in silico Molecular Understanding of Opioid Receptor Activation



Naushad AL Velgy
St Hilda's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

October 2018

Acknowledgements

First and foremost, my thanks to my fantastic supervisor, Prof. Philip Charles Biggin, for his knowledge, help, and support over these turbulent three years. Thank you for giving me the opportunity to explore this challenging project, and allowing me to learn from the best.

Secondly, my profound thanks to Dr Irfan Alibay, for the fruitful discussions and his knowledge, for reading my thesis, and for making sure I have it written on time. Profound thanks do not feel sufficient, but alas it is all I have.

I would also like to thank current and past members of the SBCB lab, particularly Dr Maria Musgaard for teaching me the ropes; Dr Laura Domicevica, for teaching me to code; William George Glass, Max Epstein, Joseph Bluck, Dr Robin Corey, Dr Afroditi Maria Zaki, Marc Damgen, Jan Domanski, Nicholas Michelarakis, and Viwan Jarerattanachat, for their constant support, especially over these last months; and many, many others (to whom I also offer apologies for not including your name).

This project would not have been possible without funding from the Engineering and Physical Sciences Research Council, or without resources from the Department of Biochemistry in Oxford.

My parents, Suzete K Velgy and Luis MRL Velgi, and my brother, Nadir FL Velgi, for their infinite love and support.

Abstract

This thesis explores the behaviour the *homo sapiens* δ opioid receptor (δ OR) in complex with a variety of morphinan (morphine derivatives) opioids, aiming to answer whether molecular dynamics (MD) tools can be used to determine the molecular underpinning of opioid structure-activity relationship.

To do so, the results of conventional MD simulations are reported and analysed to shed light on the local changes to the binding pocket, and a novel method, N-Body Information Theory, is adapted to study concerted motions within the receptor as a result of morphinan binding. The reported simulations reveal an interesting connection between Tyr130$^{3.30}$ and Trp173$^{4.50}$ (superscripts correspond to the Ballesteros-Weinstein numbering scheme), the latter of which is critical in stabilising inactive states of the receptor.

Due to sampling limitations of medium timescale (100 ns) conventional MD is not capable of exploring large conformational changes (necessary for receptor activation, e.g. outer movement of TM6), a novel collective variable (CV) based on unique contacts between two receptor states was developed. Using the μ OR, this new CV (termed FNC) was designed to explore the transition path from inactive to active states (and vice-versa) of the receptor. Umbrella sampling simulations showed that, despite the high free energy values, the presence of the agonist DIPP-NH_2 lowered the energetic barrier for activation, agreeing with previously reported experimental results.

Encouraged by this, FNC were generated for the δ OR. Enhanced sampling simulations of the δ OR bound to four ligands (the antagonists naltrindole and naltrexone, and the agonists morphine and oxycodone), however, failed to elucidate the differences between agonists and antagonists, likely due to poor phase space sampling borne from umbrella sampling. Improvements to the work performed are suggested, particularly with respect to the sampling of the FNC CV and orthogonal states.

Contents

1 Introduction **1**

1.1 G Protein Coupled Receptors 1

1.1.1 G Protein Coupled Receptors and their importance in life . . . 1

1.1.2 GPCR superfamily 1

1.1.2.1 Structure of GPCRs 2

1.1.3 Opioid Receptors 8

1.1.3.1 ORs: a compressed history 8

1.1.3.2 How opioid receptors affect nociception 9

1.1.3.3 Opioid Receptor structure 10

1.1.3.4 Opioids and derivatives 17

1.1.3.5 Morphine analogues 18

1.1.3.6 Mixed ligands 20

1.1.4 Project aims 20

2 Methods **23**

2.1 The basics of MD 23

2.1.1 Theory underlying MD 23

2.1.2 Force fields 24

2.1.2.1 Bonded Interactions 24

2.1.2.2 Nonbonded Interactions 26

2.1.2.3 Periodic Boundary Conditions 27

2.1.3 Propagation of information 29

2.1.3.1 Configurational entropy 29

2.1.3.2 Mutual Information 30

2.1.3.3 Paths from networks 31

2.1.3.4 System preparations 31

2.1.3.5 Analysis 34

2.2 Enhanced Sampling 34

2.2.1 Choice of CV 34
2.2.2 Contact Maps derived from native contacts 35
2.2.3 ALPHARMSD 38
2.2.4 Steered MD 38
2.2.5 Umbrella Sampling 44
2.2.6 Reweighting the Windows 45

3 Interactions between opioids and an opioid receptor, and resulting information pathways 46

3.1 Introduction 46
3.1.1 Revisiting the classical opioids: morphine derivatives 48
3.1.1.1 Analgesia from other opioid receptors 52
3.1.1.2 δ OR crystal structure 54
3.1.1.3 Investigating interactions between δ OR and certain ligands 56
3.1.1.4 Investigating allosteric communication between ligands and the G protein binding site 58
3.1.2 Results 59
3.1.2.1 Residency time during trajectories 59
3.1.3 Bridging the orthosteric and allosteric binding sites 65
3.1.3.1 Changes in dihedral angles of the key residues Asp128$^{3.32}$ and Asn131$^{3.35}$ 65
3.1.3.2 Connecting the two critical binding sites 67
3.1.3.3 Connecting all receptor sites 69
3.2 Discussion 74
3.2.1 Ligand contact analysis during conventional MD 74
3.2.1.1 Opioid ligands and their relationship with Tyr$^{7.43}$ . . 74
3.2.1.2 Contact with "address" residues 75
3.2.1.3 Connecting the orthosteric and allosteric binding sites 75
3.2.1.4 The allosteric sodium does not discriminate between ligands in short timescale MD 76
3.2.2 More data from trajectories: mutual information 76
3.3 Limitations and further work 81
3.3.1 Paths between residues 81
3.3.1.1 Wandering around proteins 81
3.3.1.2 Paths may not converge 83

3.3.1.3 Paths in active states . . . 83
3.3.1.4 Alternatives to G protein complexing . . . 83
3.3.2 Outlook . . . 85

4 Development and implementation of the Fraction of Native Contacts Collective Variable 87
4.1 Introduction . . . 87
4.1.1 Exploration of GPCR activation landscape . . . 87
4.1.1.1 Landmark conventional MD studies . . . 87
4.1.1.2 Limitations of conventional MD . . . 89
4.1.2 Landmark enhanced MD studies . . . 90
4.1.3 Motivation . . . 91
4.2 Results . . . 92
4.2.1 FNC changes in conventional MD . . . 92
4.2.2 Activation of μOR from FNC biasing . . . 93
4.2.3 Deactivation of μOR from FNC biasing . . . 100
4.2.3.1 Structural changes in "deactivating" simulations . . . 100
4.2.4 Combination of FNC and ALPHARMSD . . . 103
4.2.4.1 Timeseries of supported activation simulations . . . 103
4.2.4.2 Timeseries of supported deactivation simulations . . . 105
4.2.5 Umbrella Sampling of FNC CV . . . 106
4.2.5.1 Free energy landscape of μ OR activation based on FNC umbrellas . . . 107
4.3 Discussion . . . 109
4.3.1 On FNC degradation in conventional MD . . . 109
4.3.2 Helicity loss as a result of steered MD simulations . . . 109
4.3.2.1 Tweaking parameters for future steered MD simulations 110
4.3.2.2 Erratic motion of ICL3 during simulations . . . 111
4.3.3 On the resulting PMFs . . . 115
4.4 Limitations and further work . . . 116
4.4.1 Limitations of ALPHARMSD . . . 117
4.4.2 Better helix parametrisation: a potential solution? . . . 119
4.4.3 Limitations of umbrella sampling . . . 119
4.4.4 Magnitude of umbrella sampling derived energetic landscape . . . 120
4.5 Outlook . . . 120

5 Using FNCs to study opioid structure-activity relationship 121
5.1 Introduction . . . 121
5.1.1 Investigating SAR of opioids . . . 123
5.1.1.1 Preparation of FNCs CV . . . 124
5.2 Results . . . 124
5.2.1 FNC and structural changes during steered MD . . . 124
5.2.2 Supported "activation" simulations results . . . 126
5.2.3 Steered MD simulations in the presence of ligands . . . 128
5.2.4 Umbrella sampling results . . . 128
5.2.5 Ligand contact analysis during FNC exploration . . . 129
5.3 Discussion . . . 135
5.3.1 Addressing the free energy values . . . 136
5.3.2 A system issue or a methodological limitation? . . . 136
5.3.2.1 Sodium escape during sampling . . . 136
5.3.2.2 G protein binding cavity sampling . . . 140
5.3.2.3 Generation of FNCs . . . 142
5.3.2.4 FNC convergence and exploration of orthogonal states 142
5.4 Outlook . . . 143

6 Discussion and Conclusions 144
6.1 Overview of results presented . . . 144

A Code and scripts used in Chapter 3. 148
A.1 Mutual Information . . . 148

B Code and scripts used in Chapter 4. 151
B.1 FNC generation . . . 151

C Further details on FNC changes 156
C.1 Additional forward transition test systems . . . 156
C.2 Convergence of PMFs . . . 156

D Supplemental data for Chapter 5 160

List of Figures

1.1 Fold of inactive state GPCRs. 3
1.2 Fold of active state GPCRs. 4
1.3 Lifecycle of GPCRs. 11
1.4 Summary of the DRY motif in GPCRs. 12
1.5 Summary of the NPxxY motif in GPCRs. 13
1.6 Summary of the CWxP motif in GPCRs. 14
1.7 Details of the sodium allosteric site. 15
1.8 Underlying structure of morphine derivatives. 18
1.9 "Message-address" concept applied to morphine (left) and naltrindole (right). 19
1.10 Structure and fold of DIPP-NH_2 21
1.11 MP1104 binding compared to naltrindole. 22

2.1 Components of a typical force field. 25
2.2 Representation of periodic boundary conditions. 28
2.3 Example of a system used in this thesis. 32
2.4 Derivation of FNCs. 36
2.5 Example of finding nearest non-symmetrical heavy atom. 38

3.1 Structures of morphine precursors (A) neopinone and (B) thebaine. . 46
3.2 Selected morphine derivatives. 49
3.3 Effect of aromaticity in selected compounds. 50
3.4 Selected oxymophone derivatives. 51
3.5 Selective δOR agonists (A) SNC-80, (B) ARM390, (C) and UFP-512. 53
3.6 Explaining ligand selectivity in the sibling ORs. 55
3.7 Details of the naltrindole binding site. 57
3.8 Structure of the opioids simulated in this Chapter. *Antagonists. . . . 58
3.9 Interactions with $Tyr129^{3.33}$. 63

3.10 Waters bridging classical, (A) naltrindole and (B) oxycodone, and non-classical opioid (C) TAN-67, and $His^{6.52}$. 64
3.11 Summary of $Asp^{3.32}$ and $Asn^{3.35}$ dihedral exploration. 66
3.12 Alternate conformations of $Asp128^{3.32}$. 67
3.13 Effect of ligands on the allosteric sodium. 68
3.14 Residues further away within the protein structure share lower information. 69
3.15 Examples of high MI exchange hotspots in unliganded simulations. . 72
3.16 χ_1 dihedral changes in $Tyr130^{3.34}$ and $Trp173^{4.50}$ over 3 x 100 ns simulations (5 x 100 ns for apo simulations, left). *Antagonists. 73
3.17 Path between Naltrindole and the allosteric sodium. 78
3.18 Conformational flexibility of $Tyr130^{3.34}$ and $Trp173^{4.50}$. 79
3.19 Information travelling within the receptor. 82
3.20 Path generated from naltrindole to the allosteric sodium. 84

4.1 FNC variation under conventional MD. 93
4.2 FNC dependence on allosteric sodium. 94
4.3 "Ionic lock" distance dependence on force constant. 96
4.4 Decrease in helicity of TM6 during steered MD. 98
4.5 Sodium expulsion from the binding site. 99
4.6 FNC changes in deactivating μ OR. 101
4.7 Deformities in TM6 structure. 102
4.8 Correction of helicity using ALPHARMSD. 104
4.9 ALPHARMSD change in steered MD simulations. 105
4.10 Summary of changes in "deactivation" simulations. 106
4.11 Work done on FNC CV during steered MD. 107
4.12 Resulting PMFs. 108
4.13 Helicity comparison. 110
4.14 "Very inactive state captured during conventional MD. 112
4.15 Unraveling of TM6 in converional MD. 113
4.16 FNC dynamics from inactive state simulations. 114
4.17 Misfolding of TM6. 118

5.1 Structure of 5'-guanidinonaltrindole (top) and 6'-guanidinonaltrindole (bottom), two very similar compounds that have different *in vitro* activity profiles [147]. 122
5.2 Summary of "activating" simulations of the δ OR. 125

5.3 Summary of the results of steered MD simulations performed on the δ OR. 127
5.4 Summary of "activating" steered MD simulations performed in the presence of the ligands described in the Introduction. Each line corresponds to an average of triplicate 50 ns simulations. 129
5.5 Resulting PMFs. 130
5.6 Rotation of the Asn131$^{3.35}$ side chain during activation. 137
5.7 Sodium movement during activation. 139
5.8 Tyr233$^{5.58}$-Tyr318$^{7.53}$ distance during activation. 141

C.1 20 ns simulations were performed to attempt a forward transition (inactive state to active state) using a force constant of 1000 kJ/-mol. One transition was attempted in a sodium-bound receptor (top plot), whereas three were performed in sodium-absent receptors (bottom plot). This value was deemed too weak as it doesn't force the system to traverse the 0.5 FNC value. For reference, a system was considered active if the FNC CV adopted a value of 0.7 or above (see Figure 4.1.) 157
C.2 A force constant of 1000 kJ/mol fails to transition the μOR from an inactive-like to an active-like state based on the "ionic lock" distance. Top and bottom plots represent simulations in the presence and absence of an allosteric sodium. The dashed lines follow the same convention as Figure 4.3. 158
C.3 Convergence of PMF calculations in the simulations presented, in the absence or presence of DIPP-NH_2. These show convergence at 50 ns of simulation. 159

D.1 Sequence alignment between first half of the δ and μ OR. 161
D.2 Sequence alignment between last half of the δ and μ OR. 162
D.3 Convergence of PMF calculations for each condition reported in Chapter 5. 163

List of Tables

1.1 Conservation of certain residues in across all families [19, 20]. 5
1.2 Selected endogenous opioid peptides. 9
1.3 At present, all 3 OR subtypes have been captured in multiple conformations, the most recent of which also includes an inhibitory G protein. 16

2.1 Restraints to ensure ligand stability. Each equilibration step was performed for 10 ns. All restraints in kJ/mol/nm^2. 33
2.2 Atoms used in FNC generation. *backbone heavy atoms only. 39
2.3 Summary of steered MD performed on the FNC CV. 42
2.4 Summary of steered MD performed on the FNC/AR CV combination on the μ OR. 43
2.5 Summary of steered MD performed on the FNC/AR CV combination on the δ OR. 43

3.1 Summary of ligand/*delta* OR interactions. 61
3.2 Summary of DIPP-NH_2/δ OR interactions. 62
3.3 Frequency of water bridge formation between Asp128$^{3.32}$ and Asn131$^{3.35}$. 68
3.4 Average normalised MI. 70
3.5 Information transfer via TM4. 71

5.1 Summary of naltrindole contact analysis during FNC exploration. . . 131
5.2 Summary of naltrexone contact analysis during FNC exploration. . . 132
5.3 Summary of morphine (agonist) contact analysis during FNC exploration, further adding to the idea that antagonists differentiate from agonists by interactions with Tyr308$^{7.43}$. It follows the same syntax as Table 5.1. 133
5.4 Summary of oxycodone contact analysis during FNC exploration. . . 134

Nomenclature

^{19}F-BTFA 2-bromo-4-(trifluoromethyl)acetanilide

7TM 7 transmembrane

Å Ångström

K Kelvin

ABS Allosteric Binding Site

AMBER Assisted Model Building with Energy Refinement

BW Ballesteros-Weinstein

cAMP cyclic Adenosine Mono-Phosphate

CNS Central Nervous System

CV Collective Variable

DADLE [D-Ala^2, D-Leu^5]-Enkephalin

DEER Double Electron-Electron Resonance

DIPP-NH_2 H-Dmt-Tic-Phe-Phe-NH_2

DPDPE [D-Phe^2, D-Phe^5]-Enkephalin

FNC Fraction of Native Contacts

GABA γ-aminobutyric acid

GBS G Protein Binding Site

GDP Guanosine Di-Phosphate

GPCR G Protein-Coupled Receptors

GROMACS GROningen MAchine for Chemical Simulations

GTP Guanosine Tri-Phosphate

IA-PROXYL 3-(2-iodoacetamido)-2,2,5,5-tetramethyl-pyrroline-1-oxyl

JDTic James D Thomas tetraisoquilonine carboxylic acid

K_i Dissociation constant

kcal kilocalorie

kJ kilojoule

MD Molecular Dynamics

MI Mutual Information

NAMD Nanoscale Molecular Dynamics

NMR Nuclear Magnetic Resonance

OBS Orthosteric Binding Site

OR(s) Opioid Receptor(s)

PME Particle Mesh Ewald

PMF Potential of Mean Force

POPC 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

QM Quantum Mechanics

RMSD Root Mean Square Deviation

SAR Structure/Activity Relationship

TIPP H-Tyr-Tic-Phe-Phe-OH

TIPP[Ψ] H-Tyr-TicΨ[CH_2NH]Phe-Phe-OH

TM Transmembrane

TRV130 Oliceridine

WHAM Weighted Histogram Analysis Method

ZINC ZINC Is Not Commercial

Chapter 1

Introduction

1.1 G Protein Coupled Receptors

1.1.1 G Protein Coupled Receptors and their importance in life

Transmission of information from the outside to the inside of a cell is crucial in all aspects of life. Numerous protein families exist to facilitate this process, including ion channels, passive transporters, and critically to this thesis, G Protein-Coupled Receptors (among others).

Named for their ability to couple and activate G proteins, G Protein-Coupled Receptors (henceforth GPCRs) represent one of the largest family of receptors found in the cell, encoded by over 800 genes [1, 2], or approximately 3% of the human proteome [3]. GPCRs are capable of recognising a wide variety of extracellular molecules and other stimuli, including peptides, lipids, photons, ions, and small molecule ligands [1], the latter two being of great importance in the work presented here.

Because of this, and due to their widespread expression, it is estimated that 35% of all Food and Drug Administration (FDA) approved drugs target GPCRs or GPCR mediated processes [3, 4, 5, 6, 7], and that annual revenues from GPCR drug sales are estimated to be around 30 billion USD [8].

1.1.2 GPCR superfamily

Due to the large diversity of GPCRs within the superfamily, systems were introduced to separate and classify these receptors. Currently, the GRAFS classification system is the most widely used. Developed by Frederikson in 2003 [1], this system divides GPCRs into 5 different subfamilies: Glutamate, Rhodopsin, Adhesion,

Frizzled/Taste2, and Secretin. Owing to the large number of receptors in the Rhodopsin family, this group is further divided into 4 subgroups: α, β, γ, and δ [1].

1.1.2.1 Structure of GPCRs

Though the canonical 7 transmembrane (7TM) fold of GPCRs was first described in 1993 by Schertler and colleagues [2, 9], it was done so by analysing projection density images of the bovine rhodopsin. The full three dimensional projection of a GPCR was only determined via X-ray crystallography by Palczewski and colleagues in 2000 [10].

Following this, there has been an explosion in GPCR structures deposited in the Protein DataBank, the majority of which have been listed elsewhere in the literature (the reader is referred to Table 1 of [11] for a near comprehensive list). At the time of writing of this thesis, only two sub-2$\mathring{A}$ resolution GPCR structures exist: a structure of the human A_{2A} Adenosine Receptor [12], and one of the human δ Opioid Receptor [13], both captured at 1.8 Å resolution.

The following section will describe with more detail some of the critical structural aspects and motifs found conserved in GPCRs. Residues referenced from now on will be labelled according to their Ballesteros-Weinstein (BW) number, a system that attributes labels on residues based on their position on the transmembrane helices (rather than wild type or full length sequences). A BW number takes the form of X.YY, with X representing the helix number, and YY representing a relative positional index; the most conserved residue in each helix is attributed the positional index "50" [14, 15]. Every residue following the most conserved residue in each helix has their BW number incremented by 1, while every preceding residue has its BW number reduced by 1. As an example, a residue located 5 residues prior to the most conserved in helix 4 would be numbered 4.45. Despite the existence of other numbering systems, such as the ones developed by Oliveira [16], Baldwin [17], and Schwartz [18], the BW numbering scheme remains the most widely used, since it is not affected by the variability in transmembrane helix lengths [19].

Overall architecture As mentioned above, all GPCRs share the same 7TM fold (Figure 1.1). Despite the variability in residues forming each helix, certain key residues and motifs are conserved across multiple GPCR families.

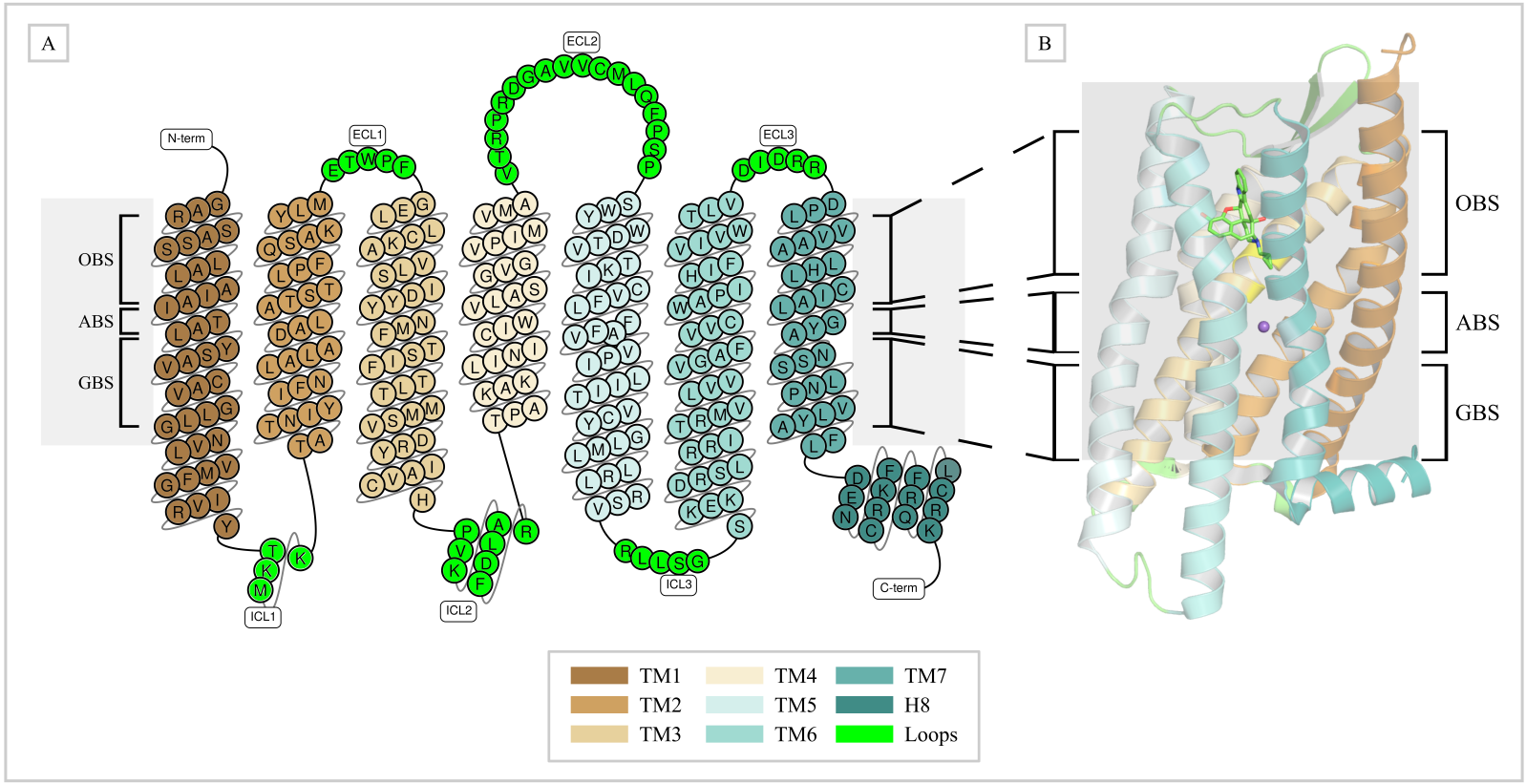


Figure 1.1: The classical 7TM architecture shared by all GPCRs. (A) Snake representation of the 7 helices of the *homo sapiens* δ opioid receptor (coloured according to the legend shown in the bottom box). All connecting loops are coloured green. Adapted from the GPCRDb [20]. (B) 3D representation of the receptor in (A), including the ligand Naltrindole and the location of the allosteric sodium (PDB accession 4N6H [13]). In both, the grey rectangles represent the position of the bilayer. OBS, Orthosteric Binding Site; ABS, Allosteric Binding Site; GBS, G protein Binding Site.

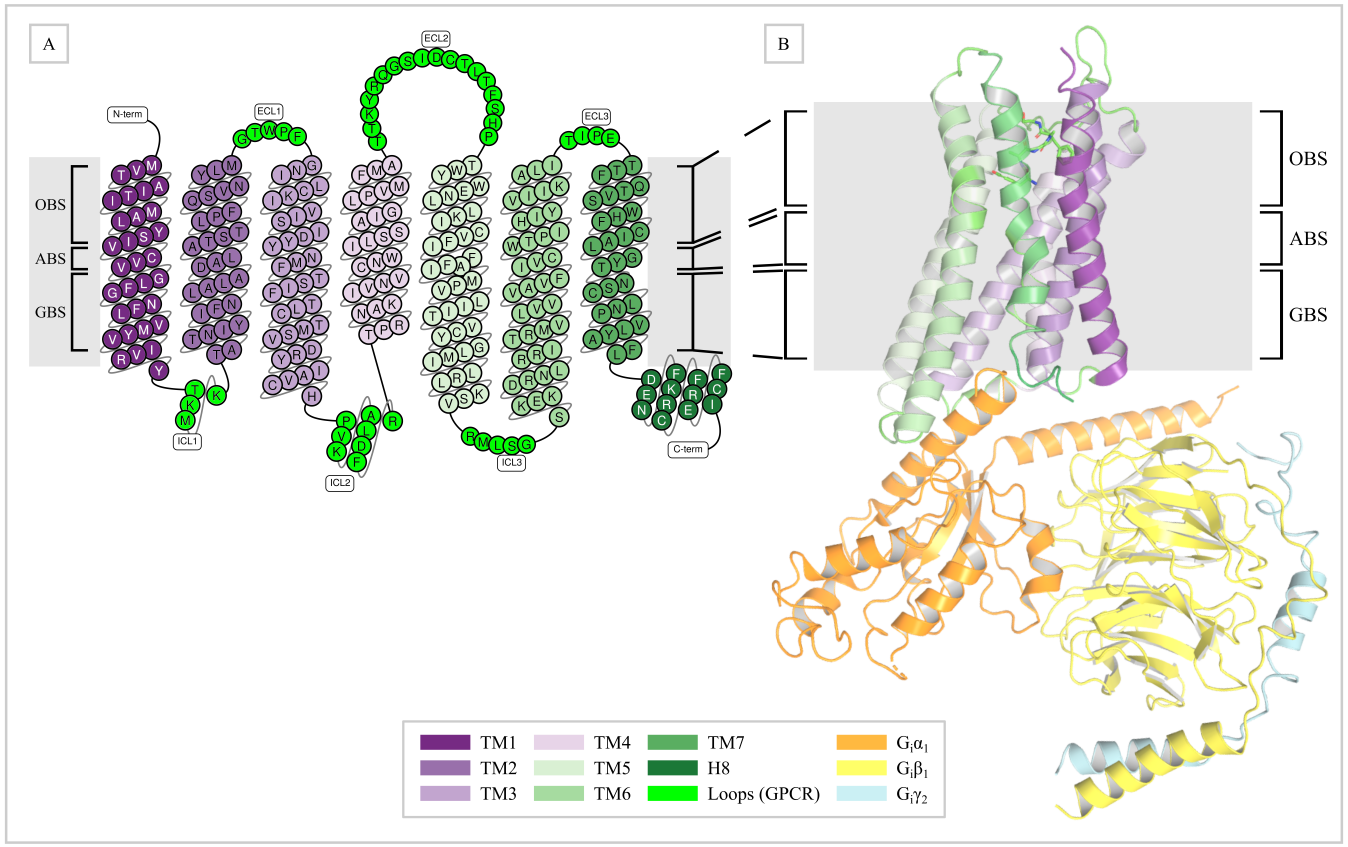


Figure 1.2: G proteins bind to the intracellular end of the GPCR. (A) Snake representation of the 7 helices of the *mus musculus* μ opioid receptor (coloured according to the legend shown in the bottom box). Helix 8 is coloured dark green. Adapted from the GPCRDb [20]. (B) 3D representation of the receptor in (A), including the ligand DAMGO and a $G_{i/o}$ protein (PDB accession 6DDF [21]). In both, the grey rectangles represent the position of the bilayer. OBS, Orthosteric Binding Site; ABS, Allosteric Binding Site; GBS, G protein Binding Site.

Table 1.1: Conservation of certain residues in across all families [19, 20].

BW Number	Residue Name	Conservation
1.50	Asparagine	98%
2.50	Aspartic acid	90%
3.50	Arginine	95%
4.50	Tryptophan	97%
5.50	Proline	78%
6.50	Proline	99%
7.50	Proline	88%

Linking these helices are 3 intracellular and 3 extracellular loops [22], commonly abbreviated to ICL1-3 and ECL1-3, respectively (see Box A of Figure 1.1). ICL1-3 link TM1-2, TM3-4, and TM5-6, respectively, with ICL3 often either completely missing in crystals or replaced with a stabilising T4 lysosyme to facilitate crystallisation [23]; the reader is referred to the aforementioned 1.8 $\mathring{A}$ resolution $A_{2A}AR$ structure (PDB entry 4EIY [12]), as well as the 2.4 $\mathring{A}$ resolution β_2 adrenoceptor structure (PDB entry 2RH1 [24]) for examples of this phenomenon. Additionally, in many molecular dynamics studies of GPCRs (detailed in later sections), this loop is often unmodelled [25, 26]. ECL1-3 link TM2-3, TM4-5, and TM6-7, respectively.

Class A GPCRs have an additional amphipathic helix adjacent to the C-terminal sequence. Extending out of TM7 and thus termed helix 8, this short peptide sequence is found perpendicular to the cell membrane. This helix contains a conserved cysteine that serves as a site for palmitoylation [27], a common post-translational modification [28]. Certain GPCRs, such as the μ Opioid Receptor, have been found to have palmitoylation sites elsewhere in their structure [29].

Ligand binding domain Facing the extracellular milieu, the orthosteric binding site (also referred to as the ligand binding site; see OBS in (A) and (B) of Figure 1.1) is the primary pocket for signalling molecules. This site is configured as a funnel, partially shielded by a highly ordered β sheet found in ECL2.

Ion binding site Located at the core of the receptor, just below the orthosteric site, sits the sodium allosteric site, one of the many allosteric sites present in GPCRs. Further details about this site, and the critical role it plays in GPCR dynamics, are given in section 1.1.3.3.

Conserved motifs and micro-switches across families GPCRs share a series of conserved motifs and micro-switches, thought to be critical in the activation process.

- The D(E)RY motif is a sequence of highly conserved residues in TM3, comprised of $Asp^{3.49}$, $Arg^{3.50}$, and $Tyr^{3.51}$. Of these, $Arg^{3.50}$ acts as a micro-switch, interchanging between a conformation in which it is interacting with $Asp^{3.49}$ (an inactive conformation, since this interaction precludes G Protein binding), and one in which it interacts with $Tyr^{5.58}$ and the G protein [15, 30]. The inactive conformation also sees $Arg^{3.50}$ interacting with $Glu^{6.30}$, known widely as the "ionic lock" [31].

- The CWxP motif is found at the centre of TM6. It is comprised of $Cys^{6.47}$, $Trp^{6.48}$, a hydrophobic residue at position 6.49, and $Pro^{6.50}$. The highly conserved $Trp^{6.48}$ is commonly referred to as the "toggle switch" [15], and was once thought to be the key to GPCR activation. In inactive conformations, the χ_1 angle of this residue is expected to fluctuate around -60°. In more active conformations, the χ_1 angle is expected to change to approximately 180°, and interact with $Phe^{5.47}$ [15].

- The NPxxY motif, sometimes extended to $NPx_2Yx_{5,6}F$ [32], is located in the intracellular end of TM7. The tyrosine residue $Tyr^{7.53}$ also acts as a micro-switch.

- The PIF motif, comprised of $Pro^{5.50}$, $Ile^{3.40}$, and $Phe^{6.44}$, is found at the core of the receptor. The latter residue faces TM3 in inactive states, but changes facing in active-like states, choosing to face TM5 instead [33].

To this date, crystallographically determined GPCR structures exist in a variety of conformational states. However, every crystallographic structure offers only static details. "Life is marked by change over time" [34], and in unearthing the basis for GPCR signalling, it is important to understand GPCR dynamics. Similarly to other proteins, GPCRs are in constant motion from thermal fluctuation [34], even in the absence of endogenous ligands and other stimuli [22]. These motions are either fast (typically the variation of chemical bond lengths, or changes in amino acid rotameric states, which occur in the range of femtoseconds to nanoseconds) or slow (helical kinks or re-orientations of helices relative to others, which occur in the range of nanoseconds to milliseconds) [34].

Experimental techniques, such as Nuclear Magnetic Resonance (NMR) and Double Electron-Electron Resonance (DEER), can help characterise some of the large scale motions that GPCRs undergo during the activation process. NMR can be used in consort with labelling of either cysteine or methionine residues.

Labelling the ϵ-methyl with $^{13}CH_3$ can be used to study large motions of proteins. The resulting resonance peak of the labelled methionine changes based on the chemical environment (global and local structure; solvent exposure), and is very sensitive to these changes [35]. Carefully mutating methionines, while retaining those located at critical locations, allows for very accurate measurements, from which distances can be inferred. In the β_2 adrenoceptor, retaining Met82$^{2.53}$, Met215$^{5.54}$, and Met279$^{6.41}$ allows measurements related to ligand binding and signal transduction. [35]. Met82$^{2.53}$ is located below the orthosteric binding site, so shifts in the NMR peak reflect changes resulting from ligand binding. The remaining two methionines are located towards the intracellular end of their respective transmembrane helices; changes to their peaks correspond to binding of a G Protein or a mimetic nanobody [35].

Cysteines located at the intracellular end of TM6 can be labelled with 2-bromo-4-(trifluoromethyl)acetanilide (or ^{19}F-BTFA) [36]. ^{19}F-NMR can then be used to detect changes to the chemical environment of the labelled cysteine. The β_2 adrenoceptor contains such a residue in position 6.27; by using a "minimal cysteine version" of this receptor (a receptor containing the Cys77$^{2.48}$Val, Cys327$^{7.54}$Ser, Cys378Ala, and Cys406Ala mutations), it is possible to selectively label certain cysteines and use NMR spectroscopy to probe the conformational flexibility of the area surrounding the labelled residue.

Results from ^{19}F-NMR can be supplemented by DEER studies to report on distance distributions between two nitroxide labelled domains. Manglik and colleagues mutated Asn$^{4.40}$ and Leu$^{6.28}$ to cysteines modified with 3-(2-iodoacetamido)-2,2,5,5-tetramethyl-pyrroline-1-oxyl (IA-PROXYL), and measured the distance distribution between TM4 and TM6 in the presence of several ligands [36].

These are some of the most ideal techniques to probe the conformational flexibility of TM6. However, large changes in proteins have been shown to be related to motions that occur in pico-, nano-, and millisecond timescale. These changes are thought to be part of hierarchical motions that enable protein function [37]. This has been demonstrated for enzyme catalysis [38, 39]. Side-chain motions, such as rotations, are thought to be critical for protein functioning. These smaller changes are much more complicated to track using experimental techniques, since current technology

allows capturing motions as slow as 3.4 μs [37]. This precludes capturing smaller, possibly concerted motions that might be critical for protein function.

Molecular dynamics (MD) methods (reviewed in Chapter 2) have proven indispensable as they allow more detailed study of protein motion. MD allows for tracking the time evolution of every receptor atom, and every atom of every molecule the receptor interacts with. MD has been used to study concerted motions within the β_2 adrenoceptor (reviewed in Chapter 4). This thesis will explore some of these motions during conventional MD of some GPCRs.

1.1.3 Opioid Receptors

Opioid receptors (OR) are a very important subclass of Class A GPCRs. Located in the Central Nervous System (CNS) and throughout peripheral tissues, these receptors are responsible for modulation of pain signalling without interfering with basic sensations [40].

Despite ample knowledge of the existence and effects of opium (specifically morphine, see section 1.1.3.5) [41], the discovery of where the drug acted on the body didn't come until 1973, when Candace Pert used radioactive morphine to determine that opioids exerted their function in the brain [42, 43]. Since then, there have been considerable efforts to various facets of opioid interactions within the human body. A compressed timeline follows, though the reader is encouraged to peruse the excellent reviews by Corbett and colleagues (2006) [41], by Trescot and colleagues (2008) [43], and by Pasternak and Pan (2013) [40].

1.1.3.1 ORs: a compressed history

In the 1960s/70s, studies by Martin and colleagues lead to the identification of μ and κ ORs [44, 45], named after their prototypical agonists (morphine and ketocyclazocine, respectively [41]). The last of the three main receptors, the δ OR, was identified in 1975 by Hughes and colleagues (specifically by studying electrically induced contraction of the *mus musculus vas deferens*), and was named after where it was discovered (the *vas deferens*) [46].

Also in 1975, Hughes and Kosterlitz reported on the first endogenous opioid peptide, termed "enkephalin" (derived from the Greek *en kephalos*, meaning "in the head" [41]). They also reported that enkephalins exists in two forms, Met-enkephalin and Leu-enkephalin, the difference between the two arising from either a Methionine

Table 1.2: Selected endogenous opioid peptides.

Name	Structure
Leu-enkephalin	Tyr-Gly-Gly-Phe-Leu
Met-enkephalin	Tyr-Gly-Gly-Phe-Leu
Dynorphin A	Tyr-Gly-Gly-Phe-Leu-Arg-Arg-Ile -Arg–Pro-Lys–Leu-Lys-Trp-Asp-Asn-Gln
Dynorphin B	Tyr-Gly-Gly-Phe-Leu-Arg-Arg-Gln-Phe-Lys-Val-Val-Thr
Endomorphin-1	Tyr-Pro-Trp-Phe-NH_2
Endomorphin-2	Tyr-Pro-Phe-Phe-NH_2

or a Leucine in their sequence (Tyr-Gly-Gly-Phe-Met for Met-enkephalin, and Tyr-Gly-Gly-Phe-Leu for Leu-enkephalin; see Table 1.2) [47]. These were shown to be moderately selective for the δ OR [48, 49].

Shortly thereafter, β-endorphin (a μ OR selective endogenous agonist) was discovered by Li and Chung [50], and dynorphin A (a κ OR selective endogenous agonist) was discovered by Goldstein and colleagues [51].

Of the three main ORs (which will be referred to as "sibling"), the δ subtype was the first to be cloned. In 1992, two groups reported on δ OR cloning [47, 52, 53].

The κ OR was the first to be captured by X-ray crystallography (PDB accession 4DJH [54]), in complex with the selective antagonist James D Thomas tetraisoquilonine carboxylic acid (JDTic). The authors reported on a structure of the inactive state, marking the first time an OR was captured in any conformation. Later efforts culminated in the 1.8 Å resolution crystallographic structure of the *homo sapiens* δ OR (PDB accession 4N6H [13]), the first sub-2 Å OR structure; in the crystallisation of the *mus musculus* μ OR in complex with a mimetic nanobody (PDB accession 5C1M [55]), the first OR captured in an active conformation; and the cryo-EM structure of the same receptor in complex with a G protein (PDB accession 6DDE/6DDF), the first reported structure of an OR in complex with a G Protein. Other structures have been solved using crystallography, and are detailed in Table 1.3.

1.1.3.2 How opioid receptors affect nociception

Activation of ORs leads to coupling to $G_{i/o}$ proteins [56] (see Box B of Figure 1.2, page 4, and $AgoR_aGProt$ of Figure 1.3, page 11). This coupling allows for the α subunit of the G protein to dissociate from subunits β and γ ('G Protein activation' in Figure 1.3, page 11), where a guanosine diphosphate (GDP) molecule is exchanged by a guanosine triphosphate (GTP) molecule [57], and subunit α is then allowed to inhibit adenylate cyclase activity [58] (thus reducing cyclic adenosine monophosphate,

or cAMP, levels in the cell). This results in activation of protein kinases, and gene transcription (the latter of which is a more long-term effect). The dissociation of the G protein also affects calcium channels present in the cell, blocking the release of pain-signalling neurotransmitters, such as glutamate and substance P, resulting in analgesia [43].

In neurons, activation of ORs inhibits the release of γ-aminobutyric acid (GABA), leading to extra dopamine being released. It is also possible that opioids antagonise N-methyl-D-aspartate (NMDA) receptors, "activating the descending serotonin and noradrenaline pain pathways from the brain stem" [43].

1.1.3.3 Opioid Receptor structure

ORs share the same structural architecture of all Class A GPCRs. The first crystallographically determined structure of an OR was that of the κ OR in the inactive state [54], and since then, structures for the δ and μ ORs have been published (see Table 1.3 for more detail of published OR strutures).

The transmembrane domain shares conserved amino acid residues with other Class A GPCRs. $Asn^{1.50}$ and $Asp^{2.50}$ are part of a polar network involved in stabilising the receptor. This network is complemented by $Asn^{3.35}$, $Ser^{3.39}$, $Asn^{7.45}$, $Ser^{7.46}$, and $Asn^{7.49}$, all shown to be involved in coordinating the allosteric sodium ion [13]. Other conserved residues are part of structural motifs, including:

- the DRY motif (Figure 1.4). No mutation studies exist on the human δ OR, but mutations of the $Asp^{3.49}$ on the μ OR alters G protein coupling to the receptor;
- and the NPx_2Yx_6F motif, bridging TM7 and H8 ($Asn^{7.49}$, $Pro^{7.50}$, $x^{7.51}$, $x^{7.52}$, $Tyr^{7.53}$, x_6, and $Phe^{8.50}$; see Figure 1.5). This motif rearranges during receptor activation, from a state that coordinates the allosteric sodium (in the inactive state), to one that coordinates with other residues and the G protein (in active-like states);
- the CWxP motif located in the central portion of TM6 (Figure 1.6), comprised of $Cys^{6.47}$, $Trp^{6.48}$, $x^{6.49}$, and $Pro^{6.50}$. Perhaps the most well documented of the residues in this motif is the $Trp^{6.48}$, known as the "tryptophan switch" (or sometimes "aromatic lock") [15, 59, 60], and $Pro^{6.50}$ is essential in allowing TM6 to hinge away from TM3 during G protein binding [48].

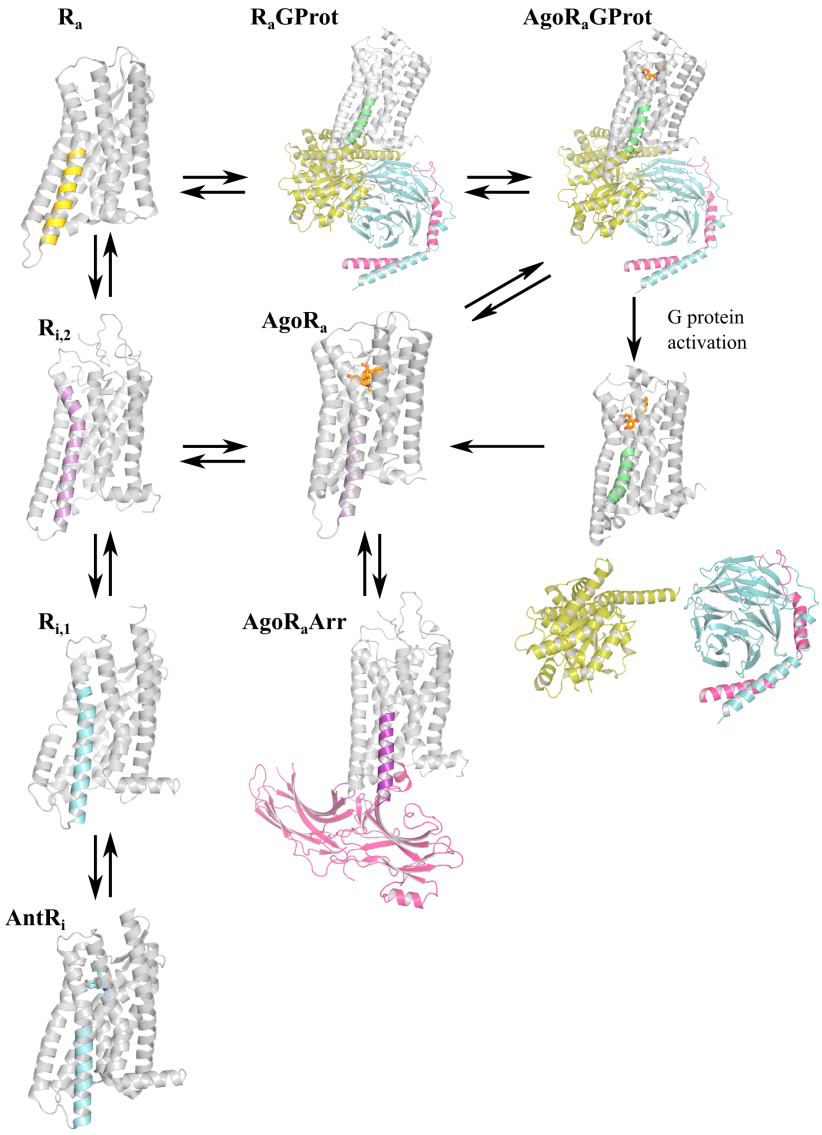


Figure 1.3: Basic schematic of the lifecycle of GPCRs. In the absence of extracellular binding partners, GPCRs exhibit a variety of conformational states, ranging from states incapable of binding G proteins (inactive or inactive-like states; $R_{i,1}$ and $R_{i,2}$), to states capable of binding G protein (active, or active-like states; R_a). Some of these states are capable of binding antagonists, such as naloxone (opioid antagonist) or propanolol (β_2 adrenoceptor antagonist), or even inverse agonists, such as naltrexone (opioid inverse agonist/antagonist) or ICI 118,551 (β_2 adrenoceptor inverse agonist). Binding to such molecules results in GPCRs sampling inactive states (represented by AntR_i). Different unbound states may be capable of either binding agonists (AgoR_a) or binding intracellular binding partners (R_aGProt). Binding to both agonists and G proteins (AgoR_aGProt) leads to irreversible dissociation between the α subunit (yellow cartoons) and the β/γ complex (cyan and pink cartoons). GPCRs in this conformational state are more likely to be phosphorilated by G protein tyrosine kinases, increasing their affinity to arrestins (AgoR_aArr). In some cases, coupling to arrestins leads to receptor internalisation.

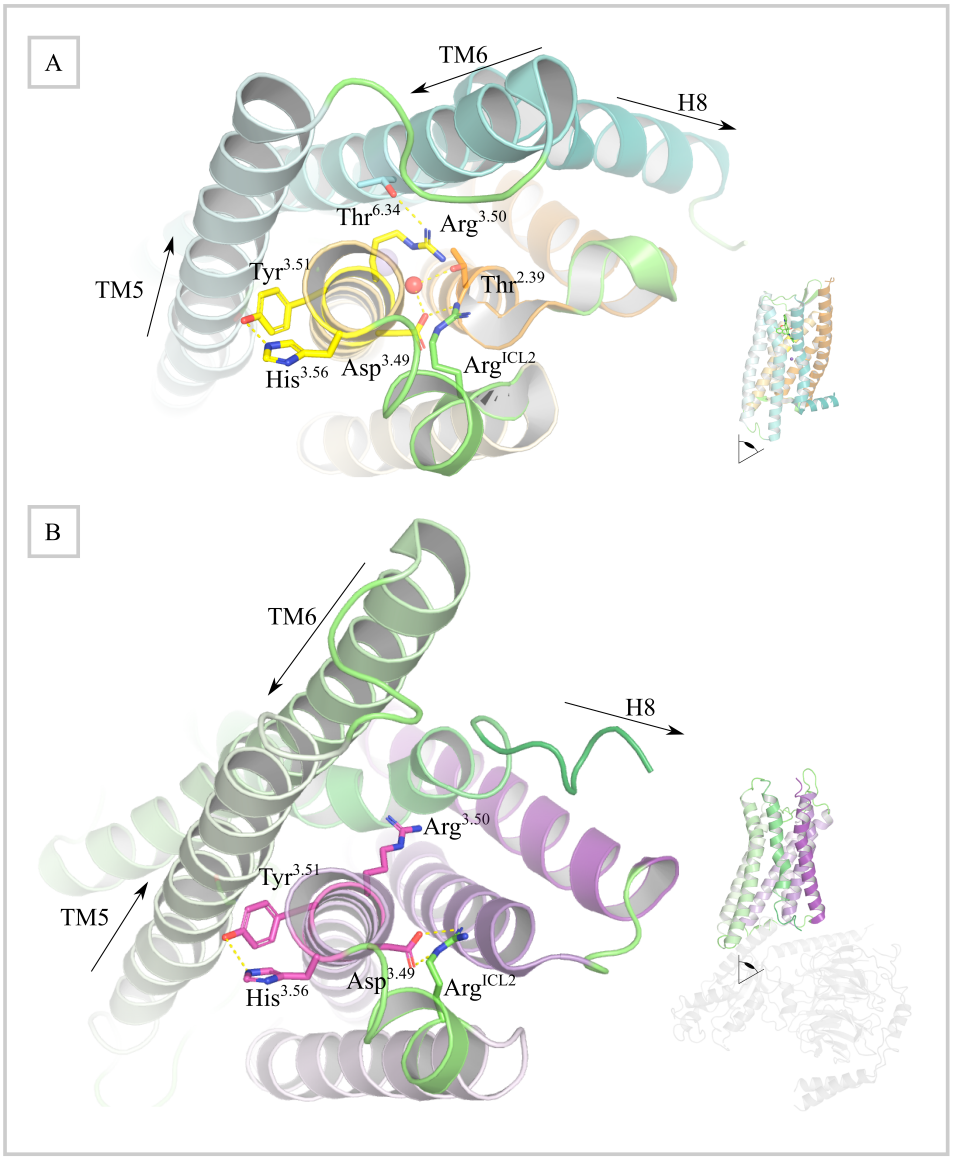


Figure 1.4: The DRY motif mediates interactions between the GPCR and the G Protein. Box A: Inactive state interactions, from the *homo sapiens* δ OR crystallographic structure (PDB accession 4N6H [13]). Box B: Active state interactions, from the *mus musculus* μ OR cryogenic electron microscopy structure (PDB accession 6DDF [21]). The accompanying G protein had been obscured. In both boxes, dashed yellow lines represent interactions between atoms at a distance <3 Å, and the schematic at the bottom right of each box shows the orientation of the image. Box A follows the colour scheme of Figure 1.1 in page 3.

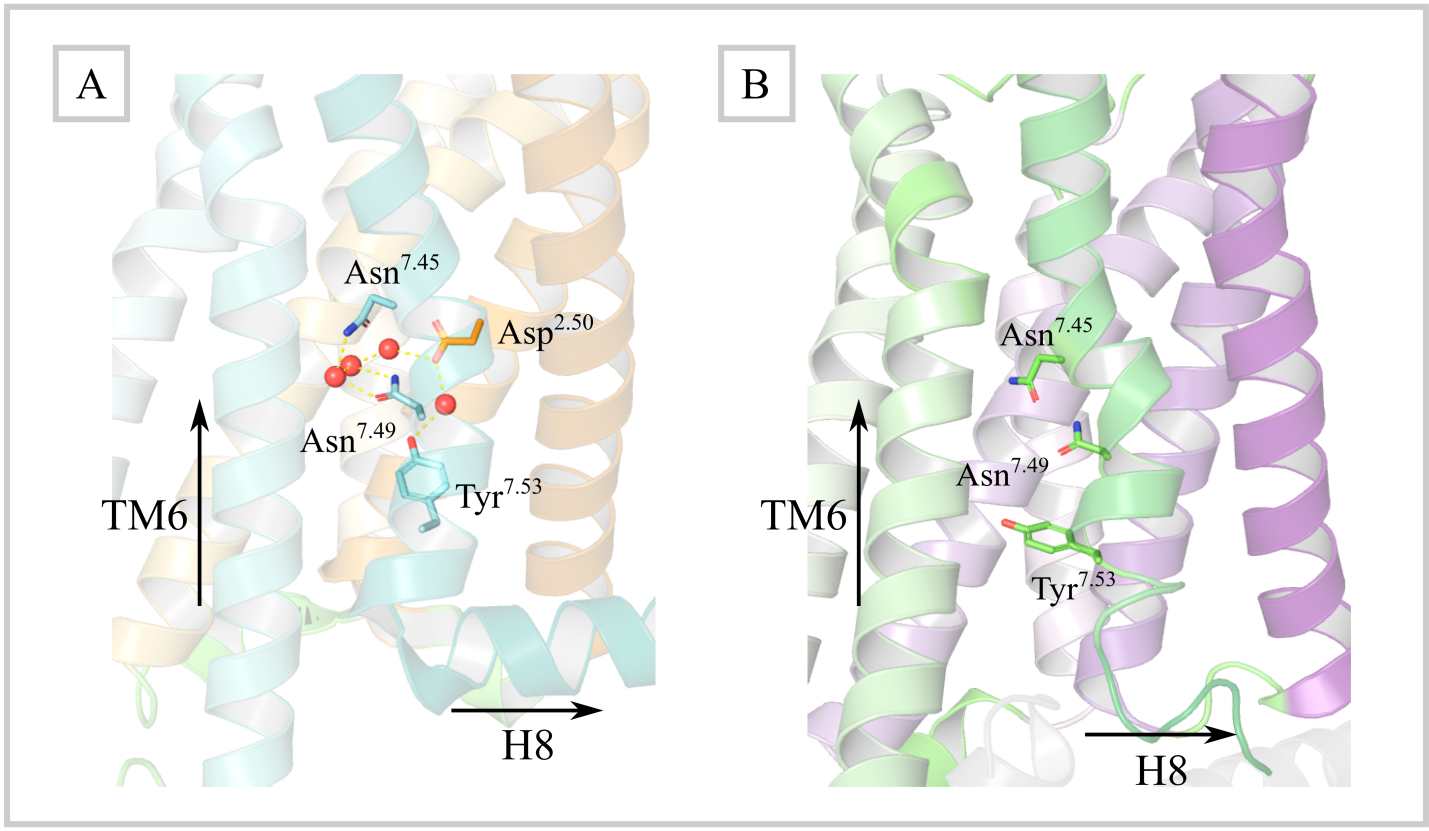


Figure 1.5: The key difference between the conformation of the NPxxY motif is the rotamer state of $Tyr^{7.53}$. Structures used to make this images were the same as Figure 1.4 (in page 12). Since H8 of *mus musculus* μ OR structure is not captured, interactions between TM7 and H8 are not shown. Colour scheme for Box A follows the same convention as Figure 1.1 in page 3.

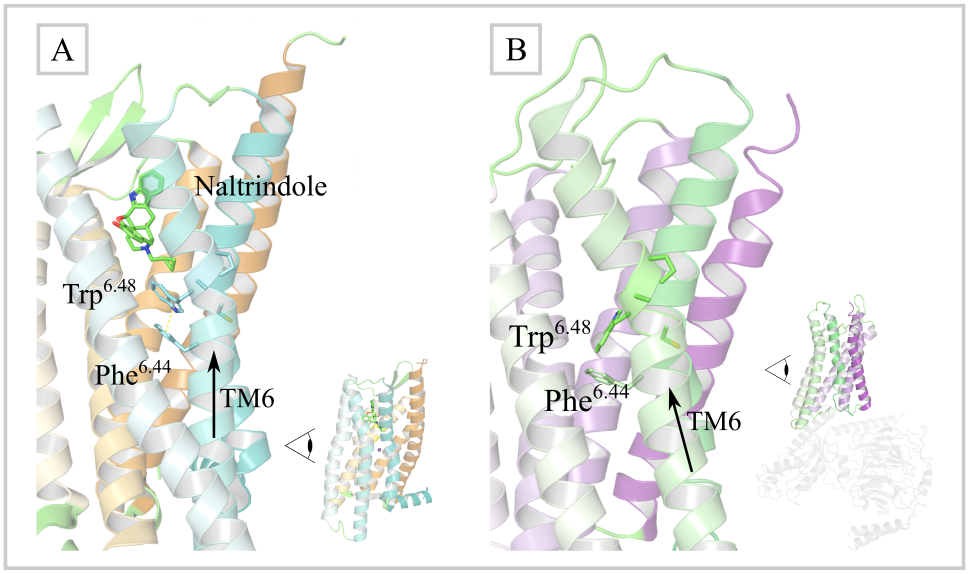


Figure 1.6: The CWxP motif contains the "toggle switch", $Trp^{6.48}$. This residue adopts multiple conformations in the activation process of GPCRs, flipping from a conformation where the side chain points towards the ligand binding domain (A), to a conformation pointing towards the G protein binding domain (B). Structures used to take make this images were the same as Figure 1.4.

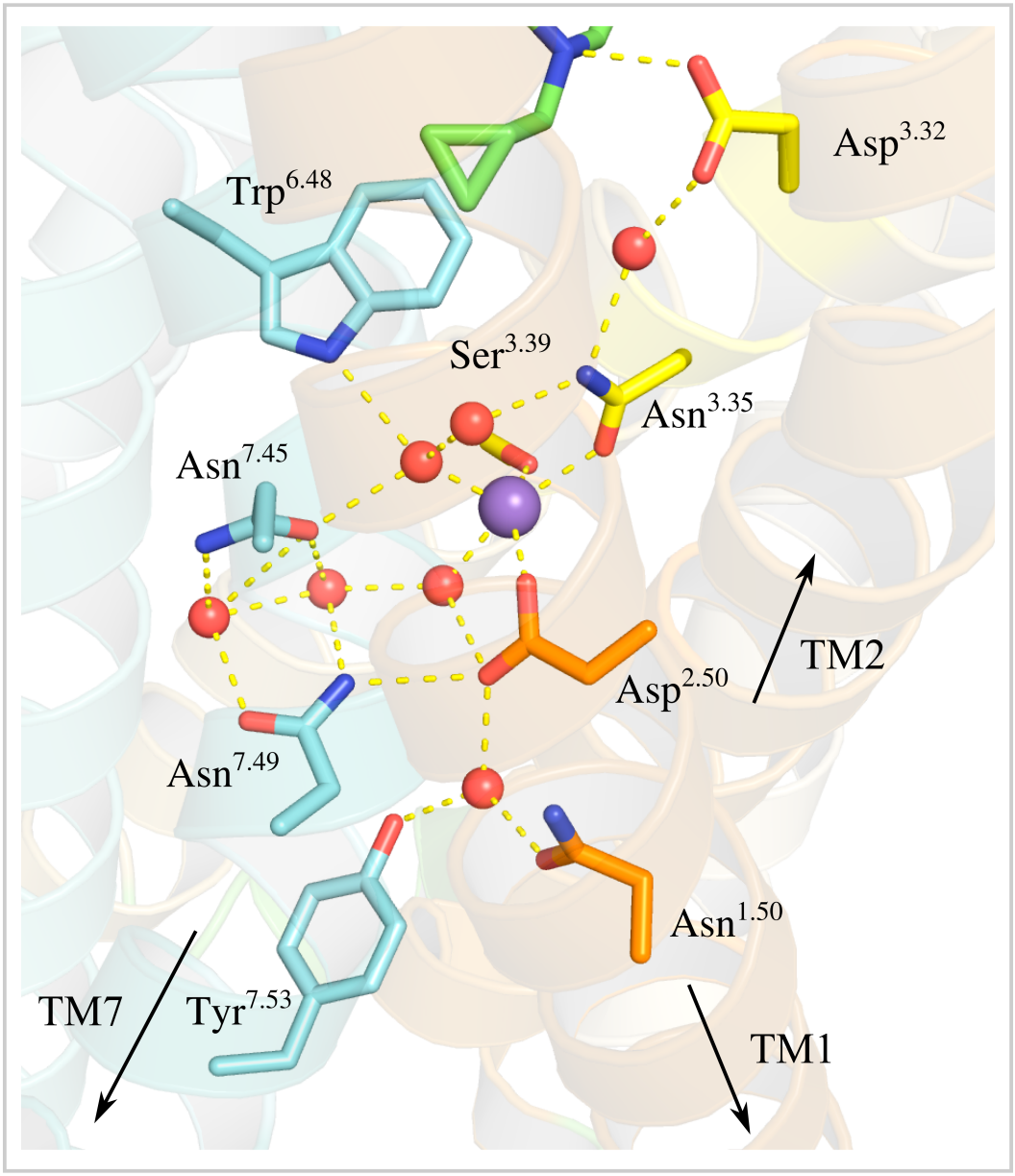


Figure 1.7: The allosteric sodium (purple sphere) resides in the core of the receptor, and has two equilibration shells. The first shell is comprised of residues within 4 Å of the allosteric sodium ($Asp^{2.50}$, $Asn^{3.35}$, $Ser^{3.39}$, and two water molecules). Generated using the *homo sapiens* δ OR structure (PDB accession 4N6H [13]).

Table 1.3: At present, all 3 OR subtypes have been captured in multiple conformations, the most recent of which also includes an inhibitory G protein.

Subtype	State	PDB ID	Reference	Resolution (Å)	Organism	Key features
μ	Inactive	4DKL	[61]	2.8	*Mus Musculus*	In complex with orthosteric ligand β-FNA and fused to a T4 Lysosyme, bridging TM5 and TM6.
	Active	5C1M	[55]	2.1	*Mus Musculus*	In complex with orthosteric ligand BU72 and Nanobody 39.
	Active	6DDF/E	[21]	3.5	*Mus Musculus*	In complex with orthosteric ligand DAMGO and a G protein. 6DDE also includes the stabilising single-chain variable fragment (scFv16).
κ	Inactive	4DJH	[54]	2.9	*Homo Sapiens*	In complex with orthosteric ligand JDTic and fused to a T4 Lysosyme, bridging TM5 and TM6.
	Active	6B73	[62]	3.1	*Homo Sapiens*	In complex with orthosteric ligand MP1104 and Nanobody 39.
δ	Inactive	4EJ4	[63]	3.4	*Mus Musculus*	In complex with high affinity antagonist Naltrindole and fused to a T4 Lysosyme, bridging TM5 and TM6.
	Inactive	4N6H	[13]	1.8	*Homo Sapiens*	In complex with high affinity antagonist Naltrindole and fused to a b562-BRIL terminal in the N-terminus.
	Inactive	4RWD/A	[64]	2.7	*Homo Sapiens*	In complex with mixed μ agonist/δ antagonist DIPP-NH_2 and fused to a b562-BRIL terminal in the N-terminus.

1.1.3.4 Opioids and derivatives

Morphine (a full agonist) is still the gold standard on treatment of acute pain [47, 48]. This compound (reviewed in section 1.1.3.5 and in Chapter 3) primarily interacts with cells via the μOR, the receptor responsible for nociception [48, 65, 66], and associated side-effects, including gastrointestinal problems, tolerance, potential for abuse, and risk of respiratory depression [66, 67].

One of the key advantages of designing drugs for the δ OR is that these show a milder side-effect profile compared to the μ OR [48, 68]. δ agonists are also capable of mediating reward behaviour, helping to mediate the effect of anxiety and mood disorders on chronic pain [69]. However, due to their increased concentration in the forebrain, activation of δ OR often lowers the threshold for seizures [70]. A more thorough review is presented in Chapter 3.

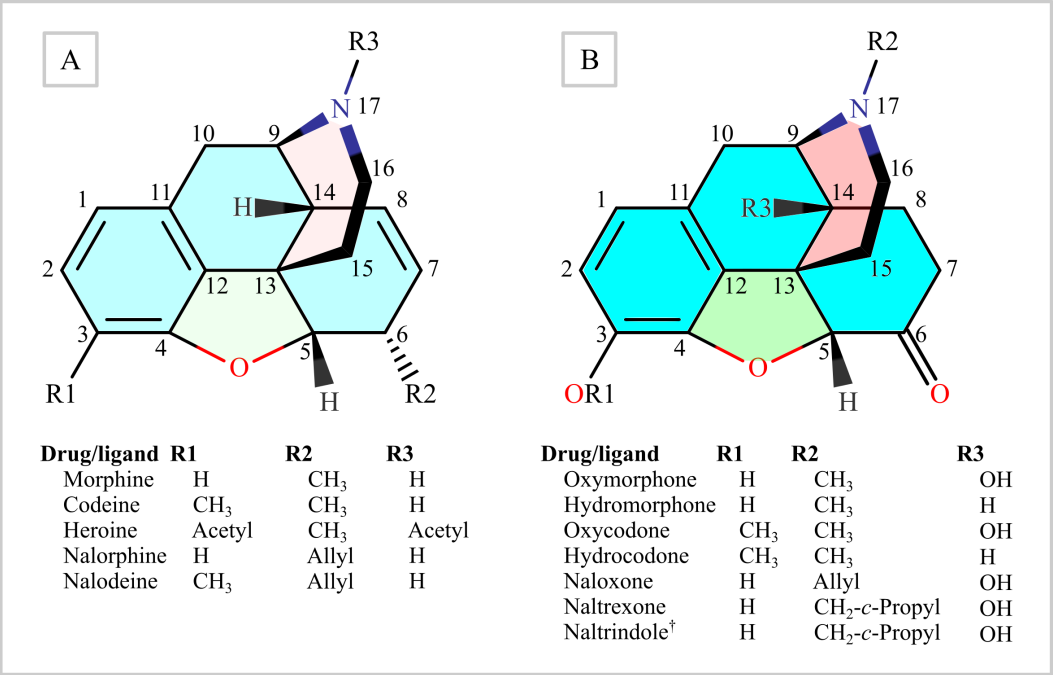
Long before the isolation of the core opioids used in clinical medicine today, Arabic physicians prescribed opium to treat dysentery [40]. The demand for opium became progressively widespread, resulting in an increase in opium trade. Eventually, this lead to the opium wars between China and England, spanning from 1839 and 1842 [40, 71].

One of the key advances in medical science, the development and widespread use of the hypodermic needle permitted direct injection of opium (particularly morphine) to the bloostream, enhancing its euphoric effects. This lead to opioid abuse, which became so problematic that international treaties were signed to limit opium trafficking [40].

Broadly speaking, the components of opium can be divided into 6 categories: morphine analogues (also referred to as 4,5α-epoxymorphinans), morphinans, benzomorphans, phenylpiperidines, acyclic analgesics, and orivapine [40]. Since this thesis focuses on morphine analogues, a more detailed description of certain members of this category follows.

The pharmacological effects of morphine have long been known, even before the drug was isolated from the *Papaver somniferum* plant (poppy) [47]. It wasn't until 1806 that German pharmacist Wilhelm Sertürner isolated the active component of the poppy [47, 72], and named it after Morpheus (the God of Dreams in Greco-Roman mythology). Morphine, alongside codeine, are present at high levels in opium [40].

Pharmacological studies have confirmed that morphine exerts its analgesic effects by activating the μ OR, and of all three subtypes, it has the highest affinity to this receptor (K_i = 1.8$\pm$0.6 nM, compared to 0.16$\pm$0.09 μM for the δ OR, and compared to 0.047$\pm$0.003 μM for the κOR [73]).



Drug/ligand	R1	R2	R3
Morphine	H	CH_3	H
Codeine	CH_3	CH_3	H
Heroine	Acetyl	CH_3	Acetyl
Nalorphine	H	Allyl	H
Nalodeine	CH_3	Allyl	H

Drug/ligand	R1	R2	R3
Oxymorphone	H	CH_3	OH
Hydromorphone	H	CH_3	H
Oxycodone	CH_3	CH_3	OH
Hydrocodone	CH_3	CH_3	H
Naloxone	H	Allyl	OH
Naltrexone	H	CH_2-*c*-Propyl	OH
Naltrindole†	H	CH_2-*c*-Propyl	OH

Figure 1.8: Changes to the core morphine results in different compounds, each with unique pharmacodynamics and pharmacokinetic properties. †To get Naltrindole, the oxygen at position 6 is removed and the single bond linking positions 6 and 7 is replaced with an indole group.

1.1.3.5 Morphine analogues

Morphine has a very defined structure (Figure 1.8), and several changes to this structure have been found to alter its properties.

Acetylation of R1 and R3 results in heroin (diacetylmorphine, or diamorphine, Box A of Figure 1.8), a substance more potent than morphine. It is thought that the presence of both acetyl group enhances blood-brain barrier penetration, thus resulting in a greater concentration of the heroin in the brain.

Changing the 6-hydroxyl group (R3 in Box A of Figure 1.8) to a ketone, in conjunction with reducing the double bond between positions 7 and 8, resulted in other powerful analgesics, including hydromorphone, hydrocodone, oxymorphone, and oxycodone, as well as antagonists naloxone and naltrexone [40] (Box B of Figure 1.8).

The final two molecules referenced, naloxone and naltrexone, also feature a modification to the *N*-methyl group. Naloxone, as well as nalorphine and nalodeine, have their *N*-methyl group replaced with an *N*-allyl group; naltrexone, as well as the highly δ OR selective neutral antagonist naltrindole, have their *N*-methyl group replaced by

a methylcyclopropyl group (Figure 1.8).

The modifications to the methyl group of morphine analogues to bulkier pharmacophores (allyl and methylcyclopropyl) led to the application of Schwyzer's "message-address" concept to opioids [74] (Figure 1.9). This resulted in discovery of selective agonists for GPCRs, including compounds developed for the δ OR by Shen and colleagues [75].

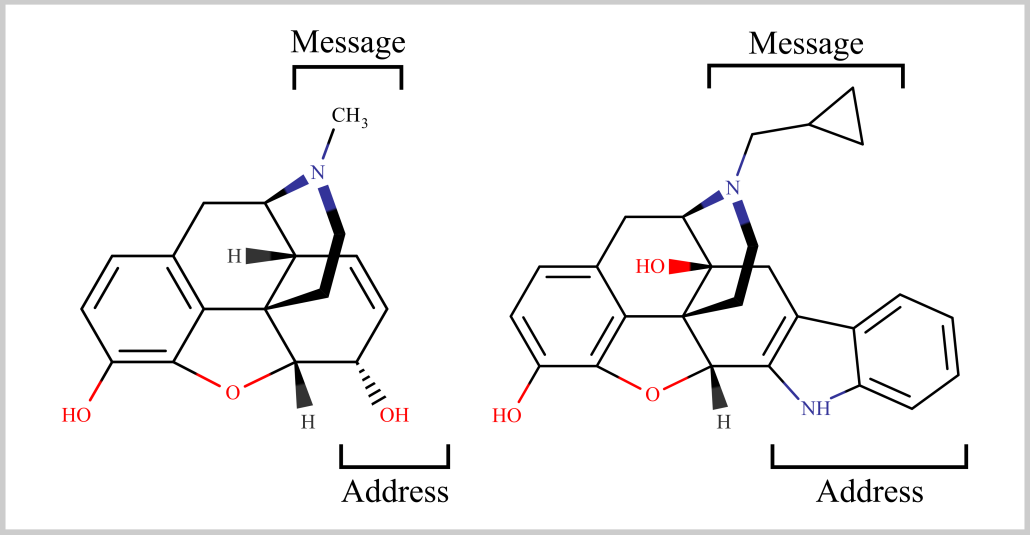


Figure 1.9: "Message-address" concept applied to morphine (left) and naltrindole (right).

In this paradigm, the group bonded to the quaternary nitrogen serves as a "message" to the opioid receptor, signalling and initiating activation. Bulky pharmacophores, such as allyl and methylcyclopropyl groups, have been shown to convert agonists to antagonists, though the precise reason for this is currently unknown. It is thought that the different pharmacophores affect the interactions between $Trp^{6.48}$ and $Tyr^{7.43}$, and this determines the "message" [76].

The group at position 6 serves as an "address" group. This group is thought to be responsible for the selectivity profile of the opioid [75]. The indole group of naltrindole, as well as the indan group of 7-spiroindanyloxymorphone [77], make these molecules selective to the δ OR. The underlying mechanism for this, explained in further detail in Chapter 3, is attributed to hydrophobic interactions between the (relatively) large indole/indan groups and small residue side-chains in position 7.35.

A more comprehensive listing of ligands can be found in Chapter 3.

1.1.3.6 Mixed ligands

Extensive research into the endogenous opioid peptides has yielded a variety of derivatives with agonist and antagonist activity.

The earliest modifications reported were performed on enkephalins, yielding, among others, the highly selective agonists H-Tyr-d-Ala-Gly-Phe-d-Leu-OH (DADLE) and H-Tyr-c[d-Pen-Gly-Phe-d-Pen] (DPDPE), the latter of which is a widely used pharmacological tool [48].

Subsequent modifications to the same endogenous opioids have yielded highly potent δ OR antagonists. Of these, members of the H-Tyr-Tic-Phe-Phe-OH (TIPP) family have been reported as the most potent and selective [48]. In particular, H-Tyr-TicΨ[CH_2NH]Phe-Phe-OH (TIPP[Ψ]) displays subnanomolar affinity and great selectivity for the δ OR (over 500 times more selective than Naltrindole) [48].

Replacement of the Tyr moiety of TIPP with 2,6-dimethyltyrosine results in DIPP-NH_2 (Figure 1.10), a highly selective and potent δ OR antagonist. Interestingly, this ligand also shows agonist activity at the μ OR, marking the first peptide-derived molecule that exhibits different activity profiles at different opioid receptors.

1.1.4 Project aims

Over the years, efforts in characterising the activity of opioids have yielded synthetic morphine analogues with higher agonist or antagonist activity profiles, increased or decreased selectivity profiles, and even peptide derivatives with mixed activity.

Since these structural changes occurred in specified regions of the morphine skeleton, the "message-address" concept was adopted to explain the dramatic activity differences that occurred *in vitro*. In this paradigm, the moiety bonded to the nitrogen at position 17 (see Figure 1.8) determines whether a molecule acts as an agonist or as an antagonist, with the prototypical *N*-methyl group being characteristic of an agonist. Bulkier moieties, such as *N*-allyl or *N*-methylcyclopropyl, confer molecules with antagonist properties.

However, a family of morphine derivatives, reviewed in Chapter 3, as well as recent crystallographic structures have called this paradigm into question.

The structure of the nanobody-stabilised κ OR was solved recently (PDB accession 6B73 [62]), in complex with the agonist MP1104. This morphine analogue contains an *N*-methylcyclopropyl group, typical of the "message" of an antagonist. This moiety even binds in a similar mode to the corresponding *N*-methylcyclopropyl in Naltrindole (see Figure 1.11). Does this invalidate the idea that this moiety determines the

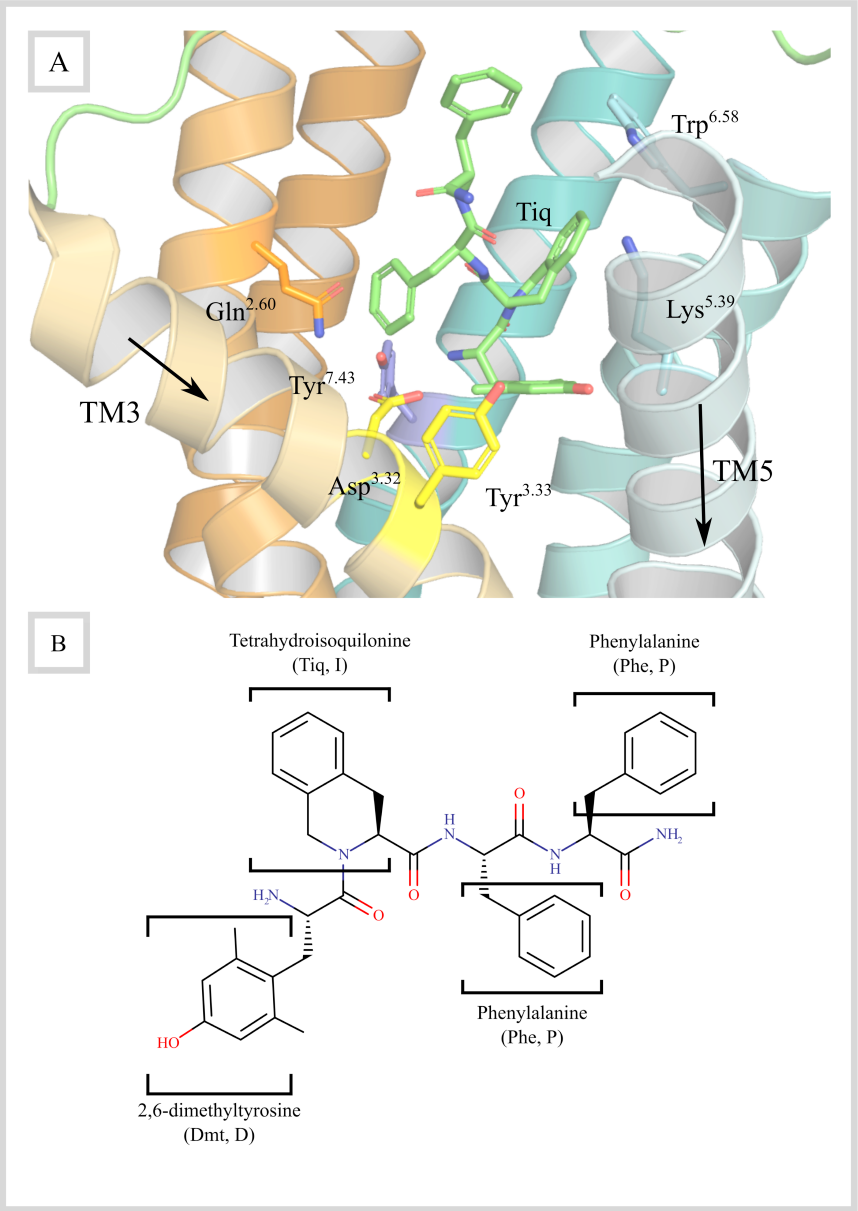


Figure 1.10: Structure of DIPP-NH_2. Box A: DIPP-NH_2 as crystallised in the *homo sapiens* δ OR (PDB accession 4RWD [64]. Each individual component of the molecule is labelled according to Box B. Box B: Chemical structure of DIPP-NH_2, and relevant functional groups.

function of the ligand? It is possible that other groups may trigger agonist activity via different pathways.

This thesis will aim to answer whether MD can be used to determine how these modifications affect the opioid receptors, and, as a consequence, how that reflects the receptor activity. In the next Chapter (starting page 23), the basic methods used

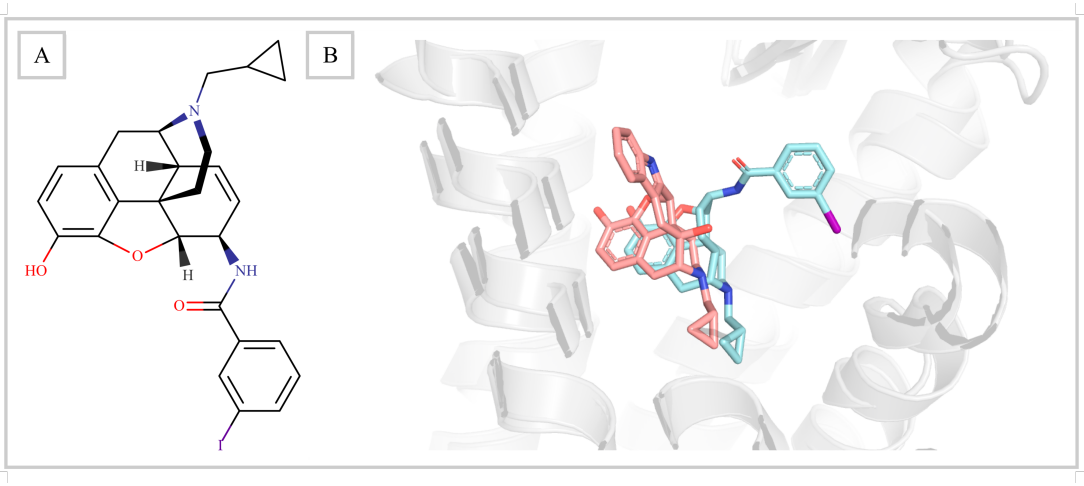


Figure 1.11: Throwing into question the "message-address" concept. Box A: Structure of the highly κ OR selective agonist, MP1104. Box B: Binding mode of MP1104 compared to Naltrindole (corresponding to PDB accessions 6B73 [62] and 4N6H [13]). Notice how the N-methylcyclopropyl moiety of both molecules bind in similar orientations.

for the remainder of this thesis will be introduced and discussed, paying particular attention to the basics of conventional and enhanced sampling MD.

These methods are then used in Chapter 3 (starting page 46), where simulations of the *homo sapiens* δ OR complexed with a variety of ligands are analysed. In order to gain an understanding of how the activation of GPCRs is affected by ligands, a novel collective variable (CV) was developed, and subsequently used to derive a free energy profile in the presence and absence of an agonist. Finally, this CV is employed to answer whether it can be used to distinguish between agonists and antagonists in the δ OR.

Chapter 2

Methods

2.1 The basics of MD

2.1.1 Theory underlying MD

Molecular dynamics (MD) simulations are an excellent tool to characterise dynamical properties of proteins that would otherwise be difficult to obtain via more established experimental methods [78, 79] (examples have been discussed in Chapter 1, section **??**). MD simulations allow for an atomistic representation of kinetics and thermodynamics of biomolecular motion, whilst allowing fine control of parameters of interest [80].

The use of MD simulations to characterise proteins dates back to 1977, when McCammon, Gelin, and Karplus published an 8.8 ps simulation of the 58 amino acid bovine pancreatic trypsin inhibitor [81]. At the time of this writing, years of theoretical progress, software [82, 83] and hardware [84] advances, allow for much longer simulations (routinely over 100 ns), of much larger systems, in more complex environments (the reader is referred to [85] for a more detailed description).

Biological phenomena and molecular systems can be accurately described by employing quantum mechanical (QM) calculations. However, since many motions of interest occur over long timescales, fully characterising a system using QM methods becomes very computationally expensive. As a result, biological systems are more commonly described using classical mechanics. In this approximation paradigm, each atom in the system is expressed as a series of three dimensional coordinates, with position $\mathbf{r}$ and mass m. This approach does not distinguish between nucleii and each electron associated with it; the two are combined into one particle in space (known as the Born-Oppenheimer approximation). These approximations allow for interactions between particles to be described using a potential energy function, from which

forces, accelerations, velocities, and positions can be calculated. This is typically done using small timesteps (one or two femtoseconds in simulations where every atom is represented, depending on the question under study).

2.1.2 Force fields

The potential energy function described above, governing all particle-particle interactions, can be expressed in two principal components:

$$U(\vec{r}) = \sum U_{bonded} + \sum U_{non-bonded} \tag{2.1}$$

where $U(\vec{r})$ is the potential energy of the system and **r** represents all positions of all particles in the system. A "forcefield" is then the functional form of Equation 2.1, describing each bonded and non-bonded interaction for all atoms in the system. It takes the form:

$$U(\vec{r}) = \sum U_{bonds} + \sum U_{angles} + \sum U_{torsions} + \sum U_{vdW} + \sum U_{electro} \tag{2.2}$$

While a visual summary of these interactions is given in Figure 2.1, each component is described below.

2.1.2.1 Bonded Interactions

Bonded interactions describe forces between two or more atoms that are covalently bonded. These are decomposed into forces that arise from bond stretching, bending, and rotations about bonds.

Bond stretching is generally described using Hooke's Law, where atomic bonds are represented as springs. With this relatively simple approximation, bond stretching can be described as

$$U_{bonds} = \frac{K_b}{2}\left(b_{ij} - b_{ij}^0\right)^2 \tag{2.3}$$

where b_{ij}^0 represents the "reference bond length" (or sometimes "natural bond length" [86]) between atoms i and j, b_{ij} is the instantaneous bond length between the same two atoms, and K_b represents the force constant (box **A** of Figure 2.1).

Following the same principle, angular variation between three atoms can be described as

$$U_{angles} = \frac{K_\theta}{2}(\theta_{ijk} - \theta_{ijk}^0)^2 \tag{2.4}$$

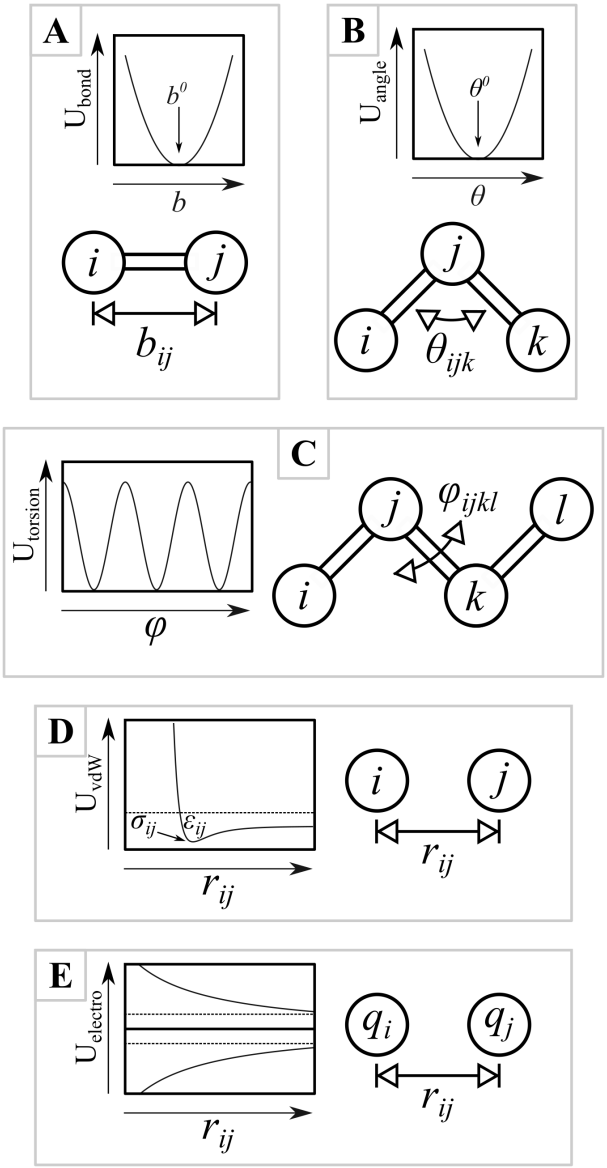


Figure 2.1: Components of a typical force field used in molecular mechanics. **A** Bond stretching (b_{ij}) between atoms i and j is modelled as a harmonic potential, with a minimum located at b^0 (the "natural bond length" [86]). **B** Variation in the angle θ_{ijk} formed between 3 atoms (i, j, and k) is also described using a harmonic potental, with a reference minimum at θ^0. **C** The torsional component describes the variation in energy when rotating about a bond. In the example provided, φ_{ijkl} describes the torsional angle between atoms i, j, k, and l, centred at the bond between j and k. Interactions between distant atoms are described using a van der Waals (vdW) component (box **D**) and an electrostatic component (box **E**). The vdW component is often modelled using a Lennard-Jones 12-6 potential, with σ_{ij} representing the separation between two atoms i and j for which the energy is 0, and ϵ_{ij} represents the well-depth. The electrostatic component between two particles with charges q_i and q_j is modelled using Coulumb's Law.

Here, θ^0_{ijk} represents a reference angle between atoms i, j, and k, θ_{ijk} represents the instantaneous angle between the three atoms, and K_θ represents the force constant (box **B** of Figure 2.1).

Torsions (or dihedral angles) are calculated as the angle between two planes made by atoms *ijkl*, where one plane is defined by atoms *ijk*, and the other defined by atoms *jkl* (box **C** of Figure 2.1).

$$U_{torsions} = \frac{K_\varphi}{2}[1 + \cos(n\varphi_{ijkl} - \delta)] \tag{2.5}$$

K_φ represents a quantitative indication of the relative barriers to rotation about the bond [86], φ_{ijkl} is the instantaneous dihedral angle, the parameter n represents the multiplicity of the dihedral (the number of minima along the energy surface as the bond rotates through 360°), and the parameter δ, known as the phase factor, indicates where the torsion passes through the minima [86].

2.1.2.2 Nonbonded Interactions

Nonbonded interactions are split into terms arising from van der Waals forces and terms from electrostatic forces.

The van der Waals forces are often modelled using a Lennard-Jones potential (box **D** of Figure 2.1):

$$U_{vdW} = 4\epsilon_{ij}\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}}\right]^{6}\right) \tag{2.6}$$

which includes a term describing attractive forces acting at long distances $\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{12}\right)$, a term describing repulsive forces at short distances $\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{6}\right)$, a term modelling the well-depth (ϵ_{ij}), and a term describing the collision diameter (σ_{ij}), all of which depend on the instantaneous separation between the two atoms ij (r_{ij}).

Finally, and in addition to the Lennard-Jones forces, two charged particles interact via electrostatic forces, modelled as a Coulombic potential (box **E** of Figure 2.1):

$$U_{electro} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.7}$$

q_i and q_j represent the (partial) charge of atoms i and j, respectively, r_{ij} represents the instantaneous separation between the two atoms, and ϵ_0 is the permitivity of free space.

Equation 2.1 is applied to every atom in the system, and finite difference methods are used integrate the equation and to determine new positions from the resulting forces. The time step parameter controls how finitely discretised Equation 2.2 is. The maximum time step has historically been limited to 1 fs, since this corresponds to the minimum of five integration steps for the slowest motion of the system [87] (typically C-H bond vibrations at 10 fs [88]). This limit increases to 2 fs when the motion of hydrogens is constrained using algorithms such as SETTLE [89] or LINCS [90]. Higher values fail to adequately sample bond vibrations, quickly leading to accuracy errors [88]. New methods are being developed to allow users to increase this limit, some of which involve modifications to hydrogen atoms in the system, and, by extension, the vibrational frequency of the C-H bond (e.g. virtual sites, hydrogen mass repartitioning [88]). A more thorough discussion of these new advances is beyond the scope of this work.

2.1.2.3 Periodic Boundary Conditions

Due to hardware limitations, the majority of atomistic MD simulations aim to study very specific phenomena over a reasonable timescale (refer to [91] for more). As a result, systems are condensed to only represent particles of interest, typically containing a protein, a set of water molecules, and a neutralising concentration of ions. Oftentimes, a lipid bilayer and a set of small molecule ligands are also included.

While this approach results in manageable systems, an important question remains; how to model the system's behaviour at it's boundaries. Throughout this thesis, periodic boundary conditions were applied, allowing particles to diffuse across the system boundaries (Figure 2.2).

As a consequence of including non-bonded interactions in forcefields (equations 2.6 and 2.7), the potential energy of every particle in the system is influenced by every other particle, including particles in periodic images. This means that, for a system with N particles, there are N^2 interactions that need calculating to accurately determine the forces acting on each particle.

Since the van der Waals interactions fade at high distances (the potential energy decays with a r^{-6} dependence), it is typical to include a cut-off for these interactions, i.e. every atoms within a certain distance is used to calculate the forces. To further facilitate computational calculations, a list of neighbours is created, which include every atom for which calculations are required, and is updated every frequently as particle positions evolve [92].

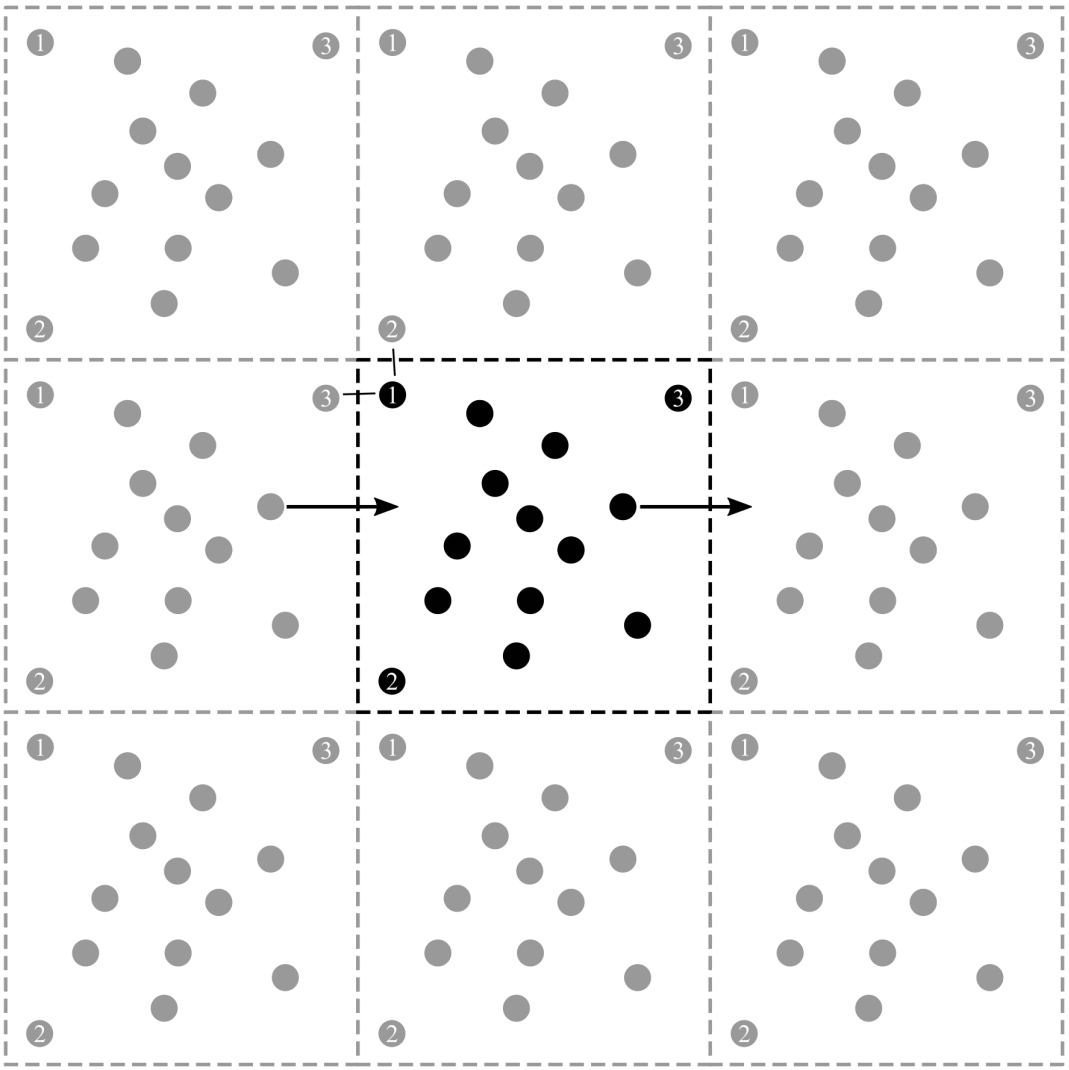


Figure 2.2: Two dimentional representation of the periodic boundary condition concept. A particle crossing the one of the boundaries (illustrated by the arrows) will "reappear" on the opposite side. Particle 1, a hypothetical particle starting at the edge of the boundary, will interact with the periodic images of particles 2 and 3.

Interactions arising from Coulombic terms are relatively stronger, decaying with an r^{-1} dependence. Consequently, these interactions cannot be managed using the same procedure as above. Instead, a Particle Mesh Ewald (PME) algorithm [93] is employed to handle these interactions.

2.1.3 Propagation of information

In recent publications, the concept of information transfer and exchange has been adapted to MD trajectories. Hinging on the concepts of configurational entropy [94] and Shannon information [95], this framework has been used to study the function of the Leucine Transporter using ≈150 ns trajectories [94], and the path of information transfer from the orthosteric site to the G protein binding site of the μ OR in complex with morphine and TRV130, using 3 x 1 µs trajectories each [96].

Following the work presented in [94], a method was developed to analyse the pairwise mutual information in all non-hydrophobic residues. The output of this was then combined with network theory to developed a set of scripts that allows one to find information pathways (see Appendix A). A brief explanation follows.

2.1.3.1 Configurational entropy

In Cartesian systems, the covariance matrix for a set of atoms can be estimated using atomic fluctuations (which arise from MD trajectories). The fluctuation of a set of atoms is correlated to the difference between the coordinates of such atoms in each frame, and the average position of all frames. The covariance between $X_{i,j}$ and $X_{k,l}$, where i and k are the atom index, and j and l are the dimension indices, is given as

$$C(X_{i,j}, X_{k,l}) =< X_{i,j} * X_{k,l} > - < X_{i,j} > * < X_{k,l} > \tag{2.8}$$

Throughout this work, covariances were calculated using a modified version of the code provided in MDAnalysis [97, 98, 99] (see Appendix A). The configurational entropy $H(X_{i,j}, X_{k,l})$ can then be obtained using the following relationship:

$$H(X_{i,j}, X_{k,l}) = \frac{1}{2} \ln |2\pi C(X_{i,j}, X_{k,l})| \tag{2.9}$$

The configurational entropy serves as an estimate of the number of states available to a set of atoms based on their position and the surrounding environment. In this paradigm, a single atom in bulk water can adopt any number of states, whereas one restricted by surrounding residues can adopt fewer; a sodium ion in bulk water has

higher configurational entropy compared to the allosteric sodium in the core of a GPCR (or in the channel of an ion conductor).

2.1.3.2 Mutual Information

Equations 2.8 and 2.9 can be applied to singular atoms, singular residues, or sets of residues. The mutual information between two sets of atoms is then

$$I_2(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \tag{2.10}$$

where X_1 and X_2 represent all atoms in sets 1 and 2, respectively. In this work, all pairwise mutual information quantities were calculated for all non-hydrophobic residues, using non-hydrogen atoms for the covariance calculation. For higher order terms, the reader is referred to [100].

Tracking atomic coordinates in space can make symmetric states appear non-symmetric [94]; this is particularly common for amino acid side-chains of unprotonated carboxylate groups of glutamic and aspartic acid, for phenylalanine, and for tyrosine. In previous works, this symmetry was accounted for by applying clustering algorithms to the side chains of the aforementioned residues (the reader is referred to [94] for more). Throughout this thesis, effects of symmetry of residues was not accounted for.

Since I_2 is a quantity in the range [0, ∞), it offers little interpretation. It is possible to make I_2 more easily interpretable by relating it to the Pearson correlation coefficient, r,

$$r(I_2) = 1 - \exp\left(\frac{-2 * I_2}{d}\right)^{0.5} \tag{2.11}$$

$r(I_2)$ is now in the range [0, 1], with 1 indicating fully correlated motion, and 0 indication fully uncorrelated motion. The parameter d is either 1 (for Gaussian distributions) or 3 (for colinear Gaussian distributions of unit variance) [101]. The value $d = 1$ is used throughout the work here presented (see also Appendix A).

2.1.3.3 Paths from networks

After calculation of pairwise mutual information, one can establish a network of interactions using residues as nodes and I_2 values as edges. This way, network properties can be obtained for each trajectory. Thoughout the work here presented, the center of mass of each non-hydrophobic residue was used to represent the node, and the python package NetworkX [102] was used to derive all graph properties.

Using this paradigm, it is possible to trace a path from one node to another distant node. A path searching algorithm was implemented; more details can be found in Appendix A.

To investigate the effect of different opioids in the δ OR binding site, several conventional MD simulations were carried and analysed. The resulting trajectories were used to study how the addition of these opioids perturb the ensemble properties, and to evaluate any potential paths between the opioids and critical residues in the δ OR architecture. All systems were prepared using the procedure described below.

2.1.3.4 System preparations

The 1.8 Å resolution crystal structure of the *homo sapiens* δ OR (PDB ascension 4N6H [13]) was used for all inactive state simulations, whilst a model based on the *mus musculus* μOR (PDB ascension 5C1M [55]) was used in all active state simulations.

To prepare the unliganded inactive state simulations, all non-receptor atoms (1-Oleoyl-R-glycerol, Oleic Acid, L(+)-Tartaric Acid, Triethylene Glycol, bRIL fusion terminal, Naltrindole) were removed from the initial PDB file. All N- and C-terminal receptor atoms (defined as all atoms before $\text{Gly39}^{1.22}$, and after $\text{Lys335}^{8.60}$) were removed. Receptor termini were not capped. Then, the receptor was embedded into a pure 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer using the MemProtMD pipeline [103]. Since this process converts the system to a coarse-grained representation (using the MARTINI forcefield [104]), it is then necessary to convert it back to an all-atom representation. In all cases, this was done using Backwards [105]. Refer to Figure 2.3 for a visual representation of the inactive simulation system.

Following embedding, the systems were neutralised with 150 nM NaCl (since waters are added during the MemProtMD stage), and energy minimisation was carried using the steepest descent algorithm (for 10000 steps or until convergence, set to 100 kJ/mol/nm). Two further equilibration steps were performed: a 100 ps equilibration in the canonical, or NVT, ensemble, using the V-rescale temperature coupling algorithm [106] and the Verlet cut-off scheme employed to allow GPU acceleration; and a 2.5 ns equilibration phase in the isothermal-isobaric, or NPT, ensemble, using a weak pressure coupling algorithm, while applying heavy restraints (1000 kcal/mol/nm^2) to all protein heavy atoms. In both cases, bonds were constrained using the LINCS algorithm [107] (using a LINCS order of 6). Additionally, PME [93] was employed to handle long-range electrostatic interactions. All simulations were performed using a temperature of 310 K.

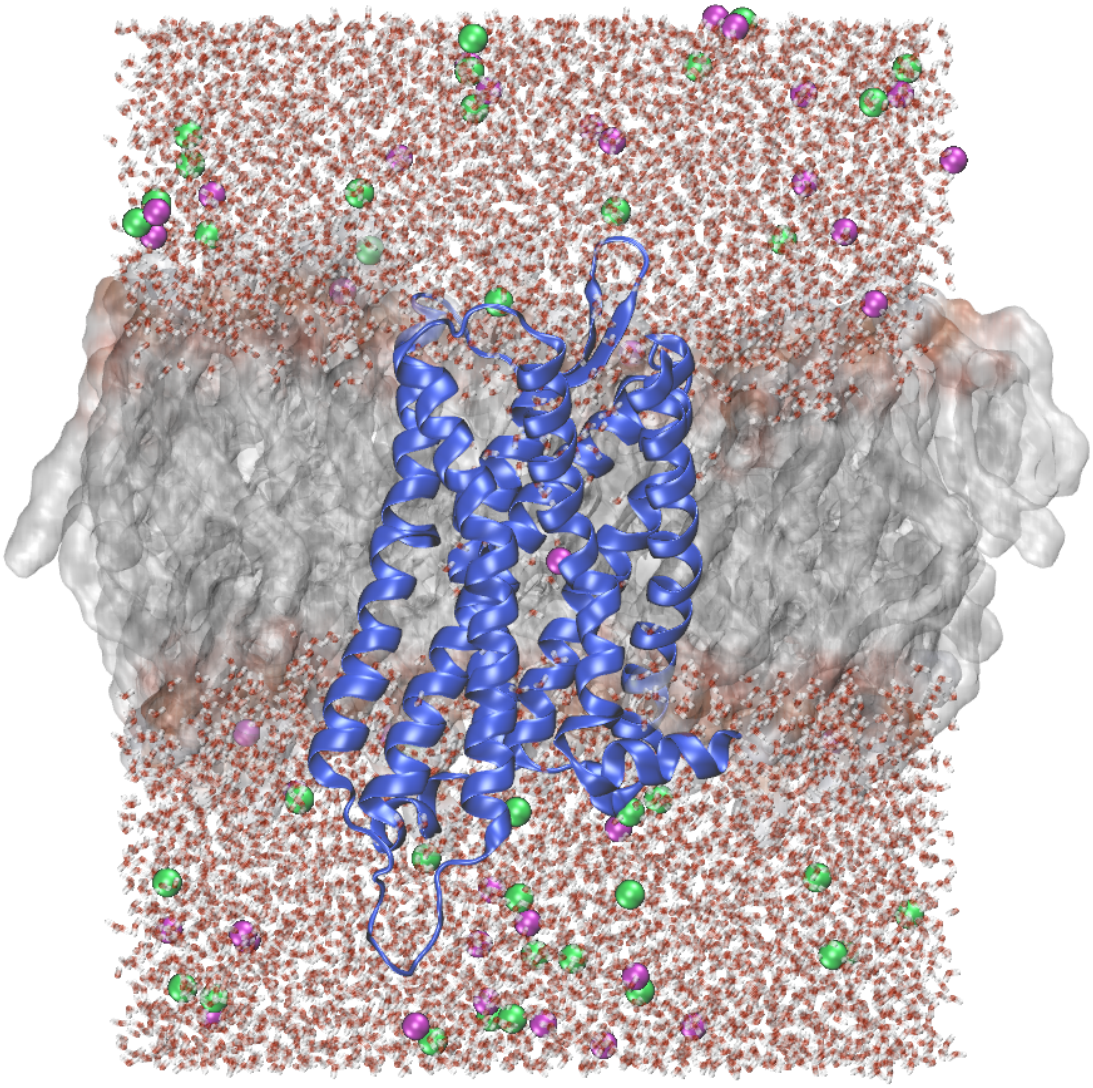


Figure 2.3: Visual representation of the apo inactive system. The receptor is represented as blue ribbons, water molecules are represented by sticks, and ions are represented as spheres (sodium ions coloured purple, chloride ions coloured green). The bilayer is composed of POPC, shown using surface representation colored grey.

Table 2.1: Restraints to ensure ligand stability. Each equilibration step was performed for 10 ns. All restraints in kJ/mol/nm^2.

Step	Protein	Ligand	Water
1	heavy, 1000	heavy, 1000	oxygen, 1000
2	heavy, 1000	heavy, 1000	oxygen, 500
3	backbone, 1000	heavy, 1000	oxygen, 500
4	backbone, 1000	none	none

For all ligand-bound inactive structures, all non-receptor atoms were removed, while preserving the coordinates of the allosteric sodium and of the critical waters (see Introduction), and all ligands were placed in the orthosteric binding pocket based on the position of naltrindole (the 4,5α-epoxymorphinan [40] skeletons were superimposed using the pair-fit function of PyMOL [108]). To allow critical waters to equilibrate in the binding pockets (ortho and allosteric), a further 4 steps of 10 ns equilibrations were carried out in the isobaric ensemble. In these last equilibration steps, restrains were applied to protein heavy atoms, critical water oxygen atoms, and ligand heavy atoms, and were subsequently reduced (see Table 2.1). This was done to ensure the ligand remained in the orthosteric binding pocket for longer MD simulations.

In all simulations, the proteins, ions, and TIP3P waters [109] were represented using the AMBER99sb-ILDN forcefield [110], while the POPC lipids were represented using the Slipids parameter set [111, 112]. All ligands were represented using the General AMBER Force Field (GAFF), version 1.7 [113], using ACPYPE [114] to generate the parameters with AMBERTools14 [115]. The inactive state systems contained ≈314 POPC molecules, ≈31000 water molecules, 126 sodium ions, 142 chloride ions, and the final box measured ≈10.7 x ≈10.7 x ≈12.05 nm^3, totalling ≈140,000 atoms. All simulations were performed using the GROMACS Molecular Dynamics engine, version 5.0.2 [116].

2.1.3.5 Analysis

Unless otherwise specified, distances and torsions were calculated using the python library MDAnalysis [98, 99], in conjunction with the NumPy library [117]. All plotting was done using the python Matplotlib package [118]. Visualisation and trajectory rendering was done in either PyMOL (version 2.2.0) [108] or VMD (version 1.9.3) [119]. Ligand analysis was performed using the python package LINTools [120].

2.2 Enhanced Sampling

Due to the time step limitations, and due to the roughness of free energy surfaces, sampling high energy states, and associated transitions to those states, can quickly become computationally expensive [121]. Critically for the work here presented, the movement of TM6 in GPCRs from inactive to active states (and vice-versa) is a process that occurs at timescales of milliseconds, making conventional MD simulations unsuitable for exploring these changes [122]. Accurate sampling of metastable states that bridge crystallographically determined structures has only been achieved using classical MD simulations on rare occasions, and often using very specialised software and hardware [26, 123].

To explore these high energy states, it is often advantageous to collapse all coordinates of the system into a defined set of coordinates of relevance, and explore the free energy landscape using this reduced ensemble. These sets of coordinates are termed collective variables (CVs). As previously mentioned, these CVs should ideally describe the evolution of the slowest processes in the system. Among the many software available to create, monitor, and force the evolution of CVs, a custom version of PLUMED v2.2.3 [124] was used throughout the work presented.

2.2.1 Choice of CV

A variety of CVs have previously been used to explore GPCR activity. A more comprehensive description is given elsewhere in the literature [121, 85]; in summary, C_α RMSD, and the rotameric (χ_1 and χ_2) angles of $Trp^{6.48}$ have been used to study how GPCR states are affected by the presence of diverse ligands. Both CVs only indirectly affect the movement of TM6 away from the core of the receptor.

2.2.2 Contact Maps derived from native contacts

In the work here presented, a novel collective variable was developed to sample the transition of GPCRs from inactive states. Throughout this thesis, it will be referred to as FNC (short for Fraction of Native Contacts).

The foundation for developing this CV is derived from the "principle of minimal frustration" (as originally applied to the folding of amino acids) [125]. A possible interpretation of this principle is that the transition free energy landscape between two distinct conformational states is correlated with the breaking or formation of contacts that are unique in either state.

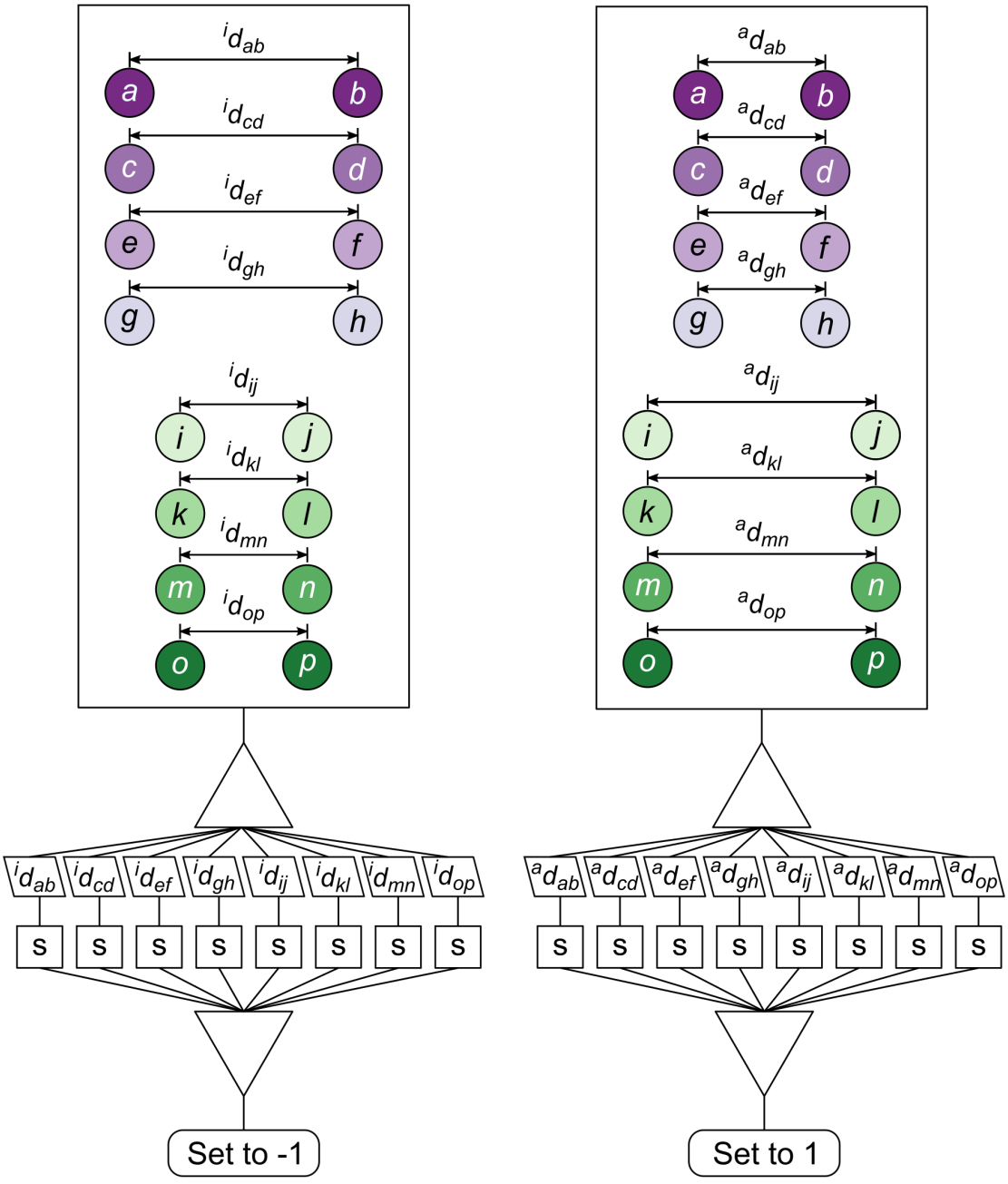


Figure 2.4: A simple example of the process of deriving native contacts from 2 distict conformational states. Here, a to p represent indices of heavy atoms, with $^{i}d_{ab}$ and $^{a}d_{ab}$ representing the Euclidean distance between atoms a and b in $i$nactive or $a$ctive conformations, respectively. A contact is then defined if the distance is below a certain cut-off. Note that, while there likely exist contacts between every atom and every other atom in the system, these are not represented for simplicity.

The derivation of native contacts is depicted in Figure 2.4. To derive these contacts, two distinct structures are required, each representing an end-point in the conformational life-cycle of the protein in question. In the work here presented, the following protocol was used to derive the FNCs (see Appendix B):

1. Protein structures were extracted from the PDB [126], or generated via a modelling software (such as Modeller [127]).

    1.1. For the initial test of the CV, the crystallographically determined inactive state of the *mus musculus* μ OR (PDB ascension 4DKL [61]) and the crystallographically determined active state of the same receptor (PDB ascension 5C1M [55]) were used. Prior to FNCs being calculated, all non-7TM atoms (lipids, ligands, ions, nanobodies) were removed. To model the missing ICL3 in the inactive state receptor, the co-crystallised T4 lysosyme was removed, and the loop was modelled using the loop refinement algorithm of Modeller 9v10 [127].

    1.2. For the δ OR case, FNCs were generated using the 1.8 Å resolution structure of the *homo sapiens* δOR (PDB ascension 4N6H [13]), and an active state model based on the active state of the *mus musculus* μ OR (PDB ascension 5C1M [55]). As per the test case, all non-7TM atoms were removed.

2. Each structure was protonated in the exact same manner, using the pdb2gmx tool of GROMACS.

3. Then, a selection was created to ensure that symmetry in amino acids was accounted for (see Table 2.2, and refer to Figure 2.5 for an example).

4. All distances were calculated between all heavy atoms using the selection generated in step 3. This included atoms in all α-helices, loops, and β-sheets. This distance is scaled by a parameter, termed "scaling factor", to allow some flexibility in the contacts. In this work, it is set to 1.8.

5. A cut-off of 6 Å was used to remove distant contacts. This returns a list of pairs of atom indices and the corresponding distance (one for each structure).

6. An element-wise comparison between the two lists was carried, and each unique contact was written to file.

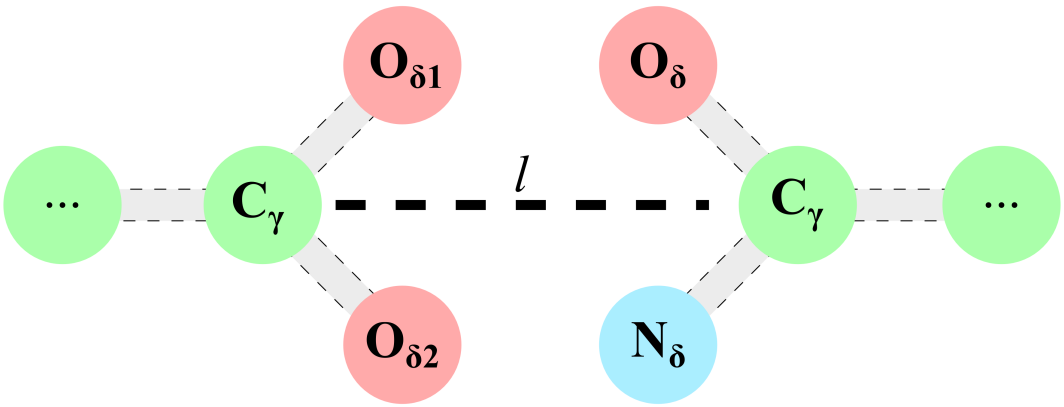


Figure 2.5: If two residues are within a certain cut-off, it is possible that they belong to a set of contacts unique to a particular state. In our selection of relevant atoms, we map each contact to the nearest non-symmetrical atom. In the example given, the distances between $O_{\delta 1}$ and O_{δ}, $O_{\delta 1}$ and N_{δ}, $O_{\delta 2}$ and O_{δ}, $O_{\delta 2}$ and N_{δ}, are irrelevant, provided the distance between the nearest non-symmetrical atom of each residue (in this case, the distance labelled l between the two C_{γ} atoms) is within a certain cut-off.

6.1. This only applied if the two atoms are separated by a minimum of 3 residues.

7. All unique contacts in each state was collapsed into a CV (using the CONTACTMAP action within PLUMED), where each contact was normalised using an exponential switching function in the form of:

$$s(r) = \exp(-\frac{r - d_0}{r_0}) \tag{2.12}$$

where d_0 is the reference distance (generated in step 6), r_0 is a smoothing parameter (set to 5 Å^{-1}, or 0.02 in the final implementation of the code), and r is the current distance between the two atoms.

8. The final step was to create a singular CV that encompasses all of the unique contacts. This was done within PLUMED (using the COMBINE function).

Table 2.2: Atoms used in FNC generation. *backbone heavy atoms only.

Amino acid	**Selected atoms***
Alanine	backbone, C_β
Arginine	backbone, C_β, C_γ, N_ϵ
Asparagine	backbone, C_γ
Aspartic acid	backbone, C_γ
Cysteine	backbone, S_γ
Glutamate	backbone, C_β, C_δ
Glutamine	backbone, C_β, C_δ
Glycine	backbone
Histidine	backbone, $C_{\epsilon 1}$, $N_{\epsilon 2}$
Isoleucine	backbone, C_β
Leucine	backbone, C_γ
Lysine	backbone, C_β, C_ζ, N_ζ
Methionine	backbone, C_β, S_δ
Serine	backbone, O_γ
Phenylalanine	backbone, C_γ, C_ζ
Proline	backbone, C_δ
Threonine	backbone, C_β
Tryptophan	backbone, $C_{\epsilon 3}$, $N_{\epsilon 1}$
Tyrosine	backbone, C_γ, C_ζ
Valine	backbone, C_β

2.2.3 ALPHARMSD

For reasons detailed in following sections and chapters, it was necessary to apply a restraint to maintain the helicity of TM6 of the GPCRs in study. PLUMED contains a CV, termed ALPHARMSD [128], used to probe the helicity of a given set of residues.

Throughout this work, a distance-RMSD (DRMSD) alignment type with a rational switching function was used for forcing helicity in a segment of amino acid residues. This function takes the form

$$s = \sum_i \frac{1 - \left(\frac{r_i - d_0}{r_0}\right)^n}{1 - \left(\frac{r_i - d_0}{r_0}\right)^m} \tag{2.13}$$

For the rational switching function, the parameter r_0 was set to 0.08 (as defined in [128]), n and m were set to 8 and 12, respectively, and d_0 was set to 0.0. This was applied to all residues between positions 6.23 and 6.44 (for the μ OR), and between positions 6.23 and 6.50 for the δ OR.

2.2.4 Steered MD

Steered MD is a method used to force the transition of a system from one state to another, based on biasing one or more CVs. A biasing potential is added to each CV depending on it's instantaneous value and a reference value:

$$V(\vec{cv}, t) = \frac{k(t)}{2}(\vec{cv} - \vec{cv}_0(t))^2 \tag{2.14}$$

In equation 2.14, *k(t)* is the force constant as a function of simulation time, and it defines how strongly the system is pulled from one state to the other [129]. Certain situations might require a gradual increase in k; in the work here presented, this parameter is constant throughout each steered MD simulation.

During a steered MD simulation, the user defines the intended change in CV. In the work presented here, each CV is varied linearly during each simulation, represented by $\vec{cv}_0(t)$ in eq. 2.14. $\vec{cv}$ is then the instantaneous value of the CV.

If two or more CVs are biased in the steered MD simulation, the total bias is the combination of the biases added to each CV:

$$V(t) = \frac{1}{2}\sum_i k_i(t)(\vec{cv}_i - \vec{cv}_{0,i}(t))^2 \tag{2.15}$$

In the work presented in this thesis, two CVs were used to explore GPCR activation: the ALPHARMSD CV is used to force helicity in a set of residues [128]; and FNCs, a novel CV based on unique contacts (see section 2.2.2).

The inclusion of the ALPHARMSD CV was deemed necessary to correct a loss of secondary structure in TM6 as a result of steered MD simulations. For the test case (using the *mus musculus* μ OR), this CV was varied from 13 to 16, corresponding to an increase in 1 turn of the helix. At the time of this writing, it is unknown whether this increase is a physiological phenomenon or an artefact of crystallisation conditions.

As mentioned above, the *mus musculus* μ OR was used as a test case for the steered MD simulations (see below).

System preparations A series of steered MD simulations were developed to test whether it was possible to transition a GPCR using the new FNC CV. The following set-ups were used:

- Transitions from the inactive to the active state in the presence of the allosteric sodium. FNCs derived from the crystallographically determined structures of

the *mus musculus* μ OR (PDB ascensions 4DKL [61] and 5C1M [55], respectively).

- Transitions from the inactive to the active state in the absence of the allosteric sodium. FNCs derived from same crystallographically determined structures mentioned above.

- Transitions from the active to the inactive state in the absence of a G protein/mimetic nanobody. No sodium was present in the sodium binding site. FNCs derived from same crystallographically determined structures mentioned above.

- Transitions from the inactive to the active state in the presence of the allosteric sodium. FNCs derived from the crystallographically determined structures of the *homo sapiens* δ OR (PDB ascensions 4N6H [13] and a model based on 5C1M [55], respectively).

The simulations above were performed using a smaller periodic box (measuring ≈8 x ≈8 x ≈10.5 nm^3), containing 165 POPC molecules, 59 sodium ions, 72 chloride ions, and ≈13,600 water molecules, to a total of ≈68,000 atoms. The box was reduced so as to accelerate sampling.

To prepare such systems, the larger boxes generated using MemProtMD (see section 2.1.3.4) were reduced, and subsequently re-solvated and neutralised with 150 mM NaCl. To allow for any water molecules to evacuate the bilayer, a longer (10 ns) NPT equilibration was performed, using the same parameters as in section 2.1.3.4, whilst applying heavy restraints to the protein heavy atoms and to the phosphorus atoms of all POPC molecules.

All production steered MD simulations were carried out in the NPT ensemble, whilst applying a constant force constant ($k(t)$ in equation 2.14). In part of the work here presented, a set of tests were developed to test whether this new CV can be used to sample the transition between two states; see Table 2.3 for more details.

In subsequent steered MD simulations, the ALPHARMSD CV [128] was applied to the range of residues between $Ser^{6.23}$ and $Phe^{6.44}$ (residue ids 268 to 289 in the PDB deposited structure). This served two functions; it prevented unravelling of TM6, and it allowed for incrementing/decrementing helicity in TM6. A second set of simulations were carried out to study the synergistic nature of the two CVs (see Tables 2.4 and 2.5).

Table 2.3: Steered MD simulations performed to test FNC CV. †I = Inactive state; I_{Na} = Inactive state, sodium bound; A = Active state. ‡-1 refers to an inactive state; 1 refers to an active state. *In kJ/mol.

Transition†	CV change‡	Force Constant*	Time (ns)	Number of Replicas
$I_{Na} \rightarrow A$	-1 $\rightarrow$ 1	1000	20	1
		2000		3
		5000		3
		10000		3
		15000		3
		20000		3
		2000	50	3
		5000		
I $\rightarrow$ A	-1 $\rightarrow$ 1	1000	20	3
		2000		
		5000		
		10000		
		15000		
		20000		
A $\rightarrow$ I	1 $\rightarrow$ -1	1000	20	3
		2000		
		5000		
		10000		
		15000		
		20000		
		2000	50	1

Table 2.4: Steered MD simulations performed to test synergy between FNC and ALPHARMSD CVs on the μ OR system. $^{\dagger}I_{Na}$ = Inactive state, sodium bound; A = Active state. ‡-1 refers to an inactive state; 1 refers to an active state. *In kJ/mol. $^{\diamond}$ALPHARMSD.

Transition†	FNC change‡	FNC Force Constant*	AR change$^{\diamond}$	AR Force Constant$^{*,\diamond}$	Time (ns)	Number of Replicas
$I_{Na} \rightarrow A$	-1 $\rightarrow$ 1	2000	13 $\rightarrow$ 16	2000	20	3
		5000		2000		
		5000		5000		
		2000		2000	50	3
		5000		2000		
		5000		5000		
$A \rightarrow I_{Na}$	1 $\rightarrow$ -1	2000	13 $\rightarrow$ 16	2000	20	3
		5000		2000		
		5000		5000		
		2000		2000	50	3
		5000		2000		
		5000		5000		

Table 2.5: Steered MD simulations performed to test synergy between FNC and ALPHARMSD CVs on the δOR system. $^{\dagger}I_{Na}$ = Inactive state, sodium bound; A = Active state. ‡-1 refers to an inactive state; 1 refers to an active state. *In kJ/mol. $^{\diamond}$ALPHARMSD.

Transition†	FNC change‡	FNC Force Constant*	AR change$^{\diamond}$	AR Force Constant$^{*,\diamond}$	Time (ns)	Number of Replicas
$I_{Na} \rightarrow A$	-1 $\rightarrow$ 1	5000	22.75 $\rightarrow$ 22.75	2000	20	3
		5000		2000	50	3

2.2.5 Umbrella Sampling

Umbrella sampling [130] is a technique used to sample specific states along a CV space. The CV of interest is discretised, and a potential is opened at each CV value (termed window). Steered MD, introduced above, can be used to generate starting coordinates for desired states along a CV path. The biasing potential is applied to ensure that the system explores states associated with the CV value of interest. For multiple CVs, this potential takes the familiar form of

$$\sum_i \frac{k_i}{2}(cv_i - cv_{0,i})^2 + m_i(cv_i - cv_{0,i})^2 \tag{2.16}$$

Contrary to the steered MD form, cv_0 does not vary over simulation time. Instead, it represents the desired value of the CV at the window of interest. In this form, either k_i or m_I are required: if k_i is present, it indicates a harmonic potential is added to the system; if m_i is present, it indicates a linear potential is added. Throughout this work, only the harmonic potential term ($\frac{k_i}{2}(cv_i - cv_{0,i})^2$) is used during umbrella sampling simulations.

Umbrella Sampling protocol Starting coordinates for umbrella windows were generated from steered MD simulations (above).

The FNC CV was discretised from -0.7 to 0.7, in steps of 0.1 (see Chapter 4) for both systems. For μ OR simulations, the ALPHARMSD CV was discretised from 13.4 to 15.6, in steps of 0.2. For the δ OR, a harmonic potential was added to maintain the helicity of TM6. A harmonic potential (k in equation 2.16) of 2000 kJ/mol was applied to each CV in each window. Each window was simulated for 10 ns, in quintuplicate.

For the μ OR systems, umbrella sampling calculations were performed for:

1. a system with sodium in the allosteric site
2. a system without an allosteric sodium
3. a system with an allosteric sodium and the agonist DIPP-NH_2 [131].

The DIPP-NH_2 was placed in the same orientation as that found in the δ OR crystallographic structure (PDB ascension 4RWD [64]).

For the δ OR systems, umbrella sampling calculations were performed for:

1. a system with sodium in the allosteric site

2. a system with an allosteric sodium and the neutral antagonist Naltrindole
3. a system with an allosteric sodium and the antagonist Naltrexone
4. a system with an allosteric sodium and the agonist Morphine
5. a system with an allosteric sodium and the agonist Oxycodone

2.2.6 Reweighting the Windows

After collecting all umbrella windows, it is then necessary to remove the influence of the harmonic potential from each replicate. Throughout this thesis, reweigthing was achieved using the Weighted Histogram Analysis Method (WHAM) [132, 133]. More recently, newer methods have been developed to reweigh biased ensembles, including UI (Umbrella Integration) [134], MBAR (Multistate Bennett Acceptance Ratio) [135], and vFEP (variational Free Energy Perturbation) [136, 137].

Chapter 3

Interactions between opioids and an opioid receptor, and resulting information pathways

3.1 Introduction

Despite the variety of pain regulating drugs available in the market, morphine remains the gold standard for moderate to severe pain [47, 138]. This small molecule is a member of the "morphinan-framed alkaloids", a family that also includes codeine (Figure 1.8A, page 18), neopinone, and thebaine (Figure 3.1). More specifically, it is part of the 4,5α-epoxymorphinan family, or more commonly, morphine analogues [40]. Some of these compounds are introduced in Chapter 1.

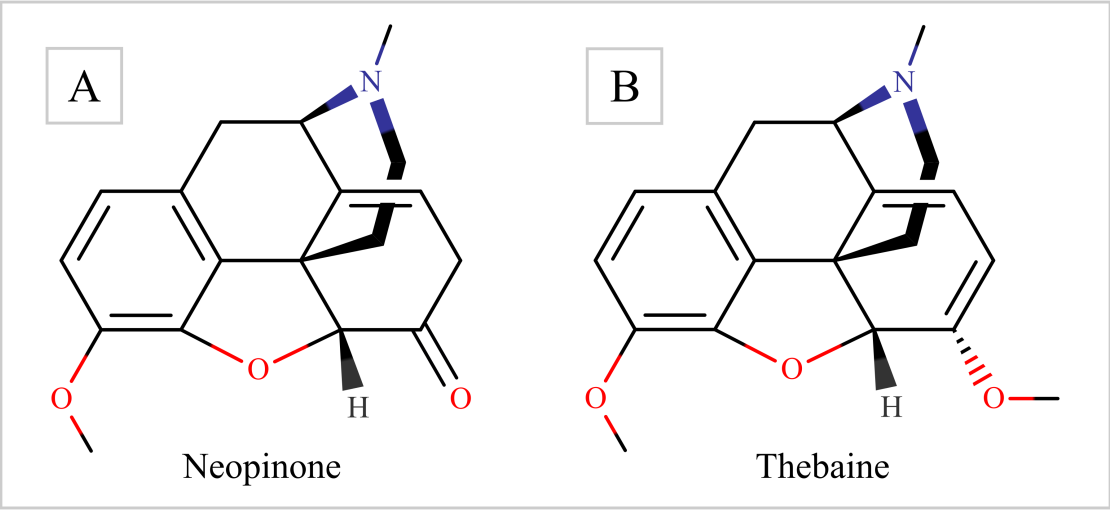


Figure 3.1: Structures of morphine precursors (A) neopinone and (B) thebaine.

Issues with morphine intake begin to emerge upon administration. The relieving

analgesia is accompanied by a variety of side-effects, including dependence, sedation, nausea, respiratory depression, seizures, constipation, and opioid bowel dysfunction [40, 138, 139, 140]. This profile of effects severely limits the clinical usefulness of, not just morphine, but also all other opioids. The combination of tolerance and dependence leads to significant costs, both financial and societal.

The search for safe analgesia has been the driving force behind the development of synthetic, side-effect free opioids. The majority of morphine derivatives and related compounds have been developed in experimental settings (reviewed below). Traditionally, this was a laborious process that involved mutagenesis studies to gain an understanding of which residues are functionally important for signalling [141], while using selective assays (radioligand binding against known high-affinity ligands, [^{35}S]GTPγS or europium-labelled GTP [8, 142], to name a few) to understand their role in the lifecycle of a GPCR [5].

Such approaches, while indispensible, are not without their drawbacks. Briefly, radioligand assays have a relatively high cost and limited potential for miniaturisation; upstream signalling assays use cell membrane preparations that typically require a filtration step to separate unbound molecules, lowering potential for automation [8]. These setbacks were some of the driving forces behind further technological development, including introduction of flourescence-based (such as time resolved fluorescence and fluorescence energy transfer), and luminescence-based (bioluminescence resonance energy transfer) read-out systems [8].

Concurrently, hardware and software advances have decreased the costs associated with *in silico* techniques. As a result, discovery of novel drugs can be aided by, for example, structure-based virtual screening. Using compounds housed in the ZINC database [143], and the then-newly released structure of the κ opioid receptor (κ OR; PDB accession 4DJH [54]), researchers docked over 4.5 million "lead-like" small molecules to the receptor's orthosteric binding site (see, for reference, Figures 1.1 and 1.2, pages 3 and 4, respectively). The best-scoring results from docking calculations were subsequently subjected to a short energy minimisation simulation, and 22 were selected for further experimental testing. Ultimately, a final compound emerged as a potential new chemotype for further study of the κ OR [144].

This study was later supplemented by Manglik and colleagues. In their study, approximately 3 million compounds from the ZINC database were docked into the inactive crystal structure of the μ OR (PDB accession 4DKL [61]) and analysed. The top 2500 scoring compounds were refined to 23. These were tested for their binding affinity against ^{3}H-diprenorphine; 7 were found to have affinities ranging from 2.3 to

14 μM, and 1, termed PZM21, was selected for further studies. PMZ21 was found to be a μ OR selective, G_i biased agonist (with antagonist activity at the κOR), with signalling bias comparable to TRV130 [145]. At the time of this writing, TRV130 is in phase III clinical trials (ClinicalTrials.gov Identifier: NCT02820324).

Despite their limitations (both studies were performed on static inactive state structures), both studies show that combining *in silico* methods with *in vivo* and *in vitro* techniques are capable of accelerating the development of reliable compounds and ligands. These also demonstrate that *in silico* methods are capable of replacing/-supplementing early stages in the drug development process, which are traditionally very labour intense [146].

3.1.1 Revisiting the classical opioids: morphine derivatives

The pursuit of safer analgesia has been a driving force long before molecular modelling techniques became cost-effective. Historically, this pursuit has been carried out by small chemical modifications to relevant molecules, performing *in vitro* and *in vivo* studies, analysing the results, and repeating the process until a compound reveals itself as a potential therapeutic agent.

Changes to the structure of that timeless opioid, morphine, have resulted in a variety of synthetic opioids. In particular, changes to the group bonded to the quaternary nitrogen, from an *N*-methyl to a bulkier group, such as *N*-allyl, seem to convert agonists to antagonists. A brief example is provided in [147], and more comprehensive examples are listed in Figure 1.8 (page 18).

Von Braun reported on the synthesis of the first opioid antagonist, *N*-allylnorcodeine (more commonly referred to as nalodeine). This compound reversed morphine-induced respiratory depression [40]. Approximately 25 years later, nalorphine (Figure 1.8A, page 18), an *N*-substituted morphine derivative, was reported [148]. This derivative reverses the analgesic effects of morphine, as well as reversing respiratory depression caused by morphine intake [140, 148]. Taken alone, nalorphine has been reported to produce antinociception and typical side effects at levels comparable to morphine [149]. Nalorphine was thus characterised as a partial agonist to the μ opioid receptor (μOR) [140].

Substituting the *N*-methyl group of morphine with *N*-phenacyl, *N*-phenoxyethyl, or *N*-benzyl (Figure 3.2), results in compounds with decimated analgesic properties compared to morphine [140]. Interestingly, the *N*-phenethyl substitution resulted in increased analgesic action (6- to 10-fold increase, studies performed in rodents), while

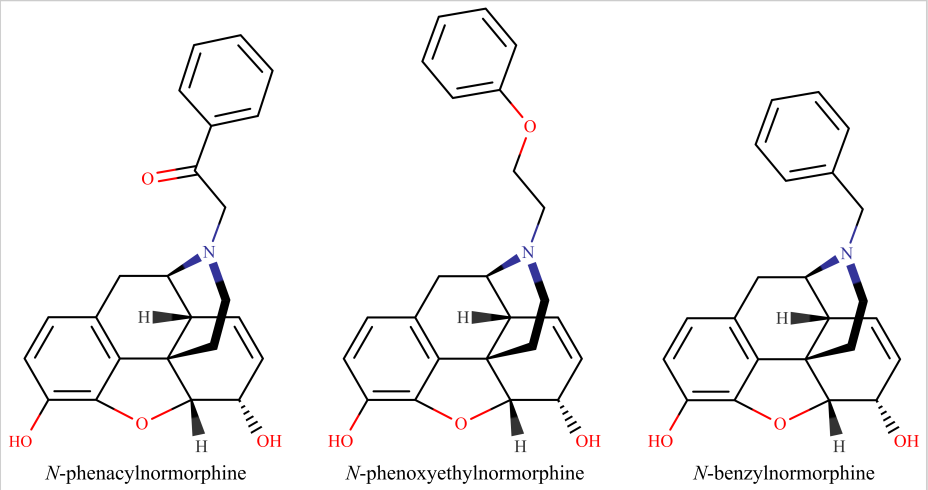


Figure 3.2: Reduced analgesia from selected morphine derivatives. These compounds have large *N*-substituted groups, but still retain some analgesic properties.

the *N*-cyclohexylethyl substitution "was only one-third as effective" [140] (Figure 3.3). An explanation for this difference is yet to be uncovered.

Another group that has received some attention, but is not considered part of the "message-address" modifications set, is the group bonded to the 14 position carbon (group R3 in Figure 1.8B, page 18). Introduction of a hydroxyl group at this position results in compounds with moderate analgesic strength [140, 150]. The accompanying reduction of the double bond between position 7 and 8, as well as conversion of the position 6 group to a carbonyl group, results in the potent analgesic oxymorphone.

This compound is used both clinically and as a scaffold for developing new μ OR ligands. Substitutions to the aforementioned quaternary nitrogen results in compounds with different activity levels. *N*-methyl, *N*-benzyl, and *N*-phenethyl substitutions (Figure 3.4, compounds A and B) result in potent agonists, the latter of which produces a 12-fold increase in antinociception [140]. *N*-dimethylallyl and *N*-cyclobutylmethyl, known as nalmexone and nalbuphone, respectively, exhibit partial agonist activity (Figure 3.4, compounds A and B). *N*-allyl and *N*-cyclopropylmethyl, known as naloxone and naltrexone (Figure 1.8B, page 18), respectively, have become the most widely used antagonists in the treatment of opioid overdose, the latter of which is also used in the management of alcohol dependence [140, 151].

Recently, the message-address concept has been called into question with the structure of the κ OR structure in complex with the agonist MP1104 (Figure 1.11, page 22). This morphine derivative contains an *N*-methyl-cyclopropyl group, a bulkier group also present in Naltrexone (Figure 1.8B, page 18). Taken together, the data

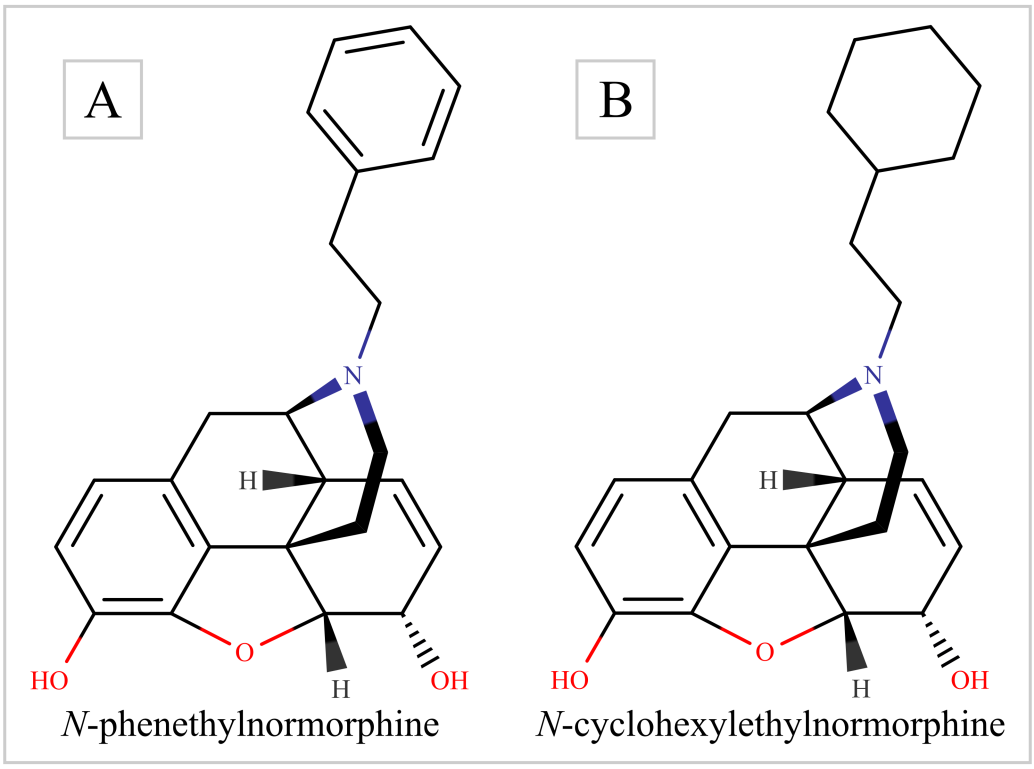


Figure 3.3: Different analgesic properties arise from the aromaticity of the *N*-substituted 6-membered ring. A saturated ring in this position (A) confers increased analgesic properties, whereas a non-aromatic ring decreases the analgesia. By which mechanisms this occurs are still elusive.

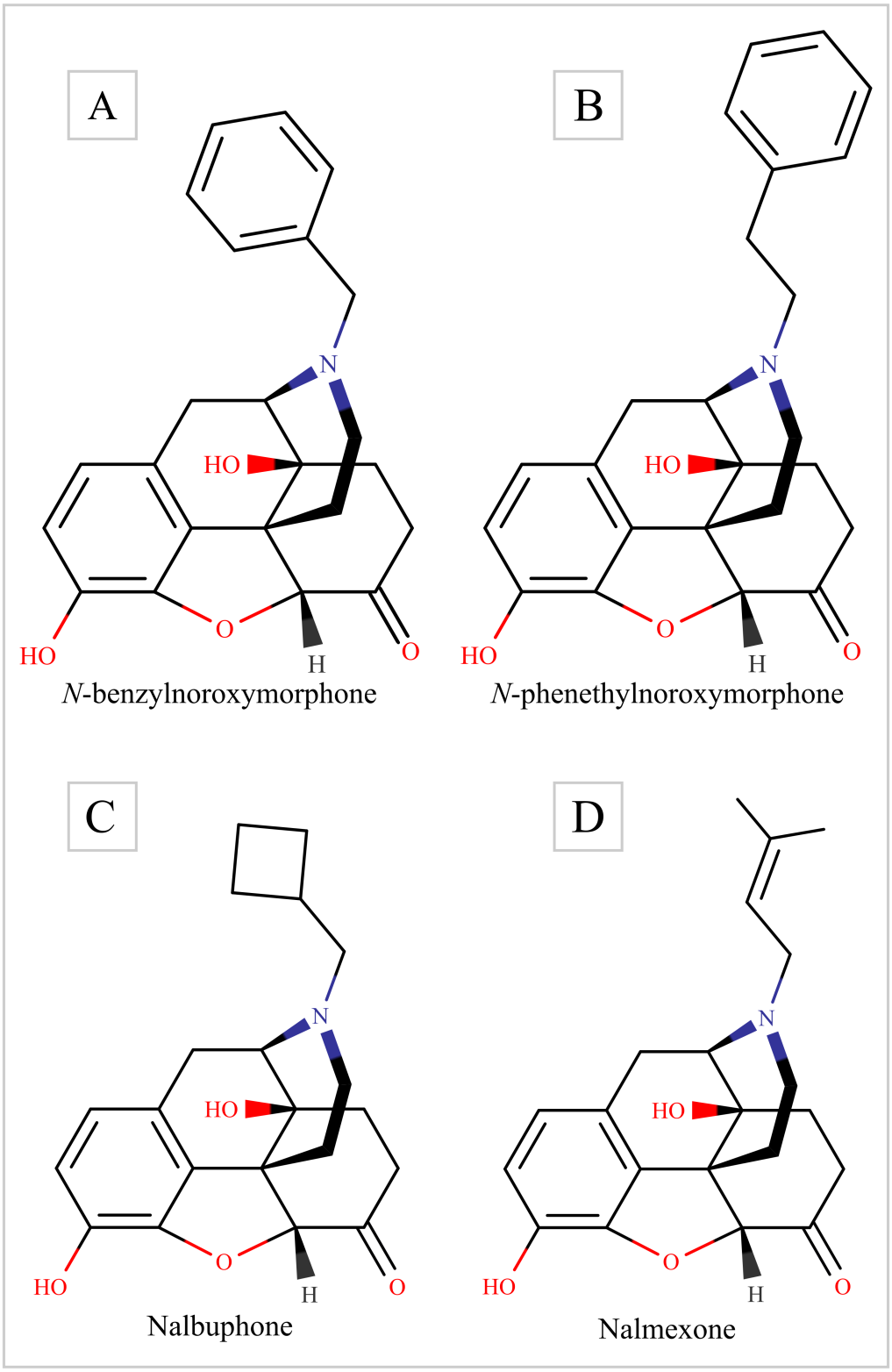


Figure 3.4: Select oxymorphone *N*-substitutes. (A) and (B) have bulky *N*-substitutes, both containing aromatic rings, are potent agonists; (C) and (D) contain bulky substitutes, but are only partial agonists.

from activity profiles of bulky N groups, and the recent crystal structure of the κ OR suggests that ligands with bulkier groups may trigger different activation pathways.

3.1.1.1 Analgesia from other opioid receptors

Later pharmacological experiments revealed that both analgesia and addictive properties are strongly mediated *via* the μ OR [70]. One of it's sibling receptors, the κ OR, also produces analgesia, though it is weaker than that induced by the μ OR. Additionally, activity of the κ OR has been linked to a different set of side-effects, including dysphoria (a profound state of unease or dissatisfaction), diuresis (increase in frequency or amount of urine), and constipation [152].

The third sibling of the classic opioid receptors, the δ OR, has also been investigated as a potential target. Found primarily in cortical regions and limbic structures (such as the hippocampus, amygdala, basal ganglia, and hypothalamus [153]), this receptor has become a model for studying GPCR trafficking, and has emerged as a potential target for to treat chronic pain and mood disorders [153]. The δ OR targets a facet of pain that is difficult to modulate: the emotional response to pain. Knockout mouse models showed that the absence of these receptors lead to higher anxiety and depressive-like behaviours, whereas the absence of κ or μ receptors didn't elicit the same responses [154]. Activation of these receptors by selective agonists such as SNC-80, ARM390, and UFP-512 (Figure 3.5), decreased anxiety-related behaviour in most classical experiments (the reader is referred to Table 1 of [70]).

One of the key advantages of using δ OR selective agonists revolves around the typical side-effects borne from opioid intake. Over the years, several research groups have found that respiratory depression is reduced in rats upon SNC-80 intake, compared to morphine intake (decreased PCO_2 in blood-gas analysis) [155]; improved gastrointestinal transit (charcoal meal test) [155, 156]; and decreased physical dependence (self-administration tests) [157]. While certain side-effects are suppressed by particular ligands, JNJ-20788560 has been shown to suppress all of the side-effect listed above [157].

In mice models of inflammatory pain, administration of SNC-80 and ARM390 revealed an interesting property of δ OR agonists. SNC80 leads to strong analgesia, robust receptor internalisation, and G protein uncoupling, resulting in "transient acute behavioural desensitisation" [158, 159]; chronic administration produces widespread receptor degradation, leading to generalised behavioural tolerance to all agonist effects [160]. Acute treatment of ARM390, a non-internalising agonist, did not produce behavioural desensitisation, but still produced analgesic tolerance. This

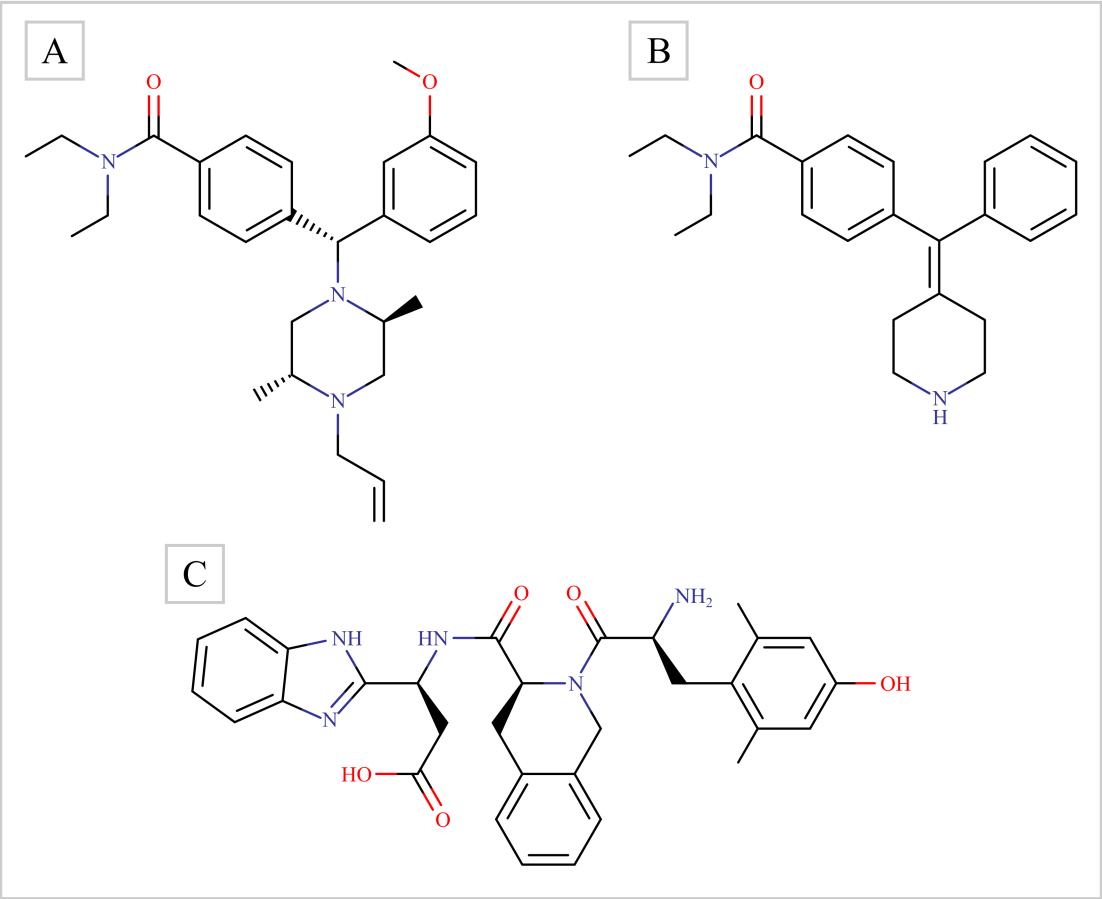


Figure 3.5: Selective δOR agonists (A) SNC-80, (B) ARM390, (C) and UFP-512.

disparity suggests different pain pathway mechanisms, thought the reason for it is still unknown [159].

All of this data suggests that δ OR agonists may provide a promising alternative to μ receptor analgesia. This analgesia, though weaker [161], is not devoid of side-effects, as the activation by some δ OR agonists has been shown to cause convulsions [48, 161].

3.1.1.2 δ OR crystal structure

Despite the seminal works by both Kieffer and colleagues, and Evans and colleagues, who independently cloned the human δOR in 1992 [52, 53], crystallisation of the receptor would only come 20 years later. First, Granier and colleagues captured the *mus musculus* δOR using X-ray crystallography (PDB accession 4EJ4 [63]). Despite it being co-crystallised with the potent and selective neutral antagonist naltrindole (see Figure 1.8B, page 18), providing great detail with respect to the binding pose of

morphine-derivatives, the relatively low resolution (3.4 Å) did not allow for capturing some more critical aspects of the δ OR structure, including the location and equilibration of the allosteric sodium binding site, and any potential interactions between naltrindole and the receptor (with the exception of the conserved interaction between Asp128$^{3.32}$ and the opioid ternary nitrogen).

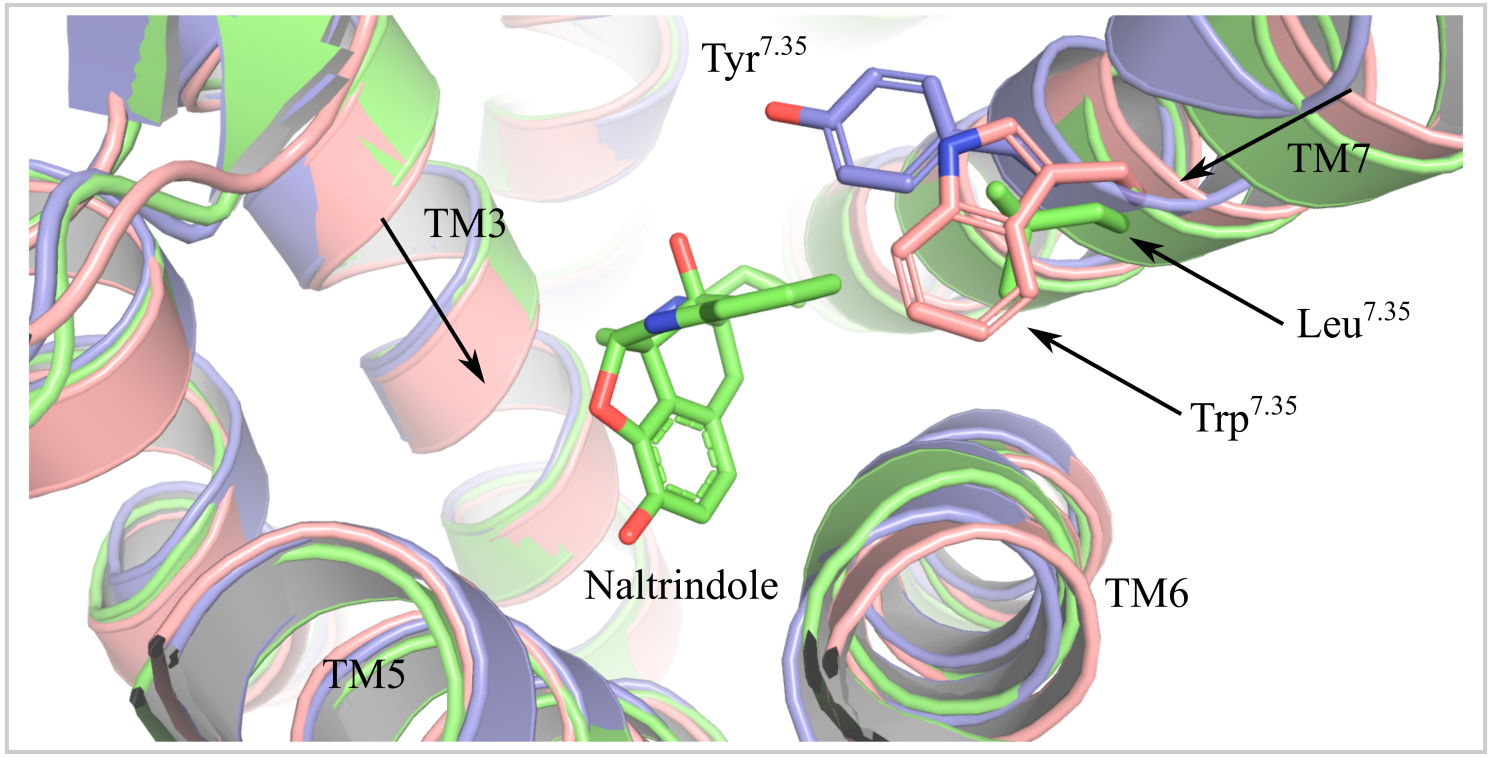


Figure 3.6: Overlay of the sibling receptors (δ, green cartoons, PDB accession 4N6H; μ, pink cartoon, PDB accession 4DKL; κ, violet cartoon, PDB accession 4DJH) showing the potential clashes between the indole moiety of naltrindole (green sticks) and the residue on position 7.35.

An interesting detail unearthed by this structure was the orientation of the indole moeity of naltrindole. The crystal structure of the *mus musculus* δ OR revealed that this group forms hydrophobic interactions with Leu300$^{7.35}$; in the sibling opioid receptors, this position is occupied by either a tryptophan (μ OR) or a tyrosine (κ OR) [162] (Figure 3.6) It is likely that bulkier side-chains in this position are responsible for the low affinity of naltrindole in the these receptors; indeed, the Trp$^{7.35}$Leu and Trp$^{7.35}$Lys mutations in the μ OR increase the affinity of δ OR selective compounds to this receptor, while also decreasing the affinity of κ OR compounds [162].

2 years later, Fenalti and colleagues resolved the *homo sapiens* δ OR [13]. This 1.8 Å resolution structure, also co-crystallised with naltrindole, revealed the position of the allosteric sodium (see Figure 1.7, page 15), and a variety of critical water molecules that equilibrate the sodium and naltrindole. An overview of the binding site can be found in Figure 3.7A.

This structure re-enforced the importance of the interaction between Asp128$^{3.32}$ and the quaternary nitrogen (Figure 3.7B); previous research had confirmed that both the carboxylate group of Asp128$^{3.32}$ and a protonated nitrogen are necessary for opioid binding [163]. The Asp128$^{3.32}$Asn and Asp128$^{3.32}$Ala mutations have been shown to increase basal signalling; both Tyr308$^{7.43}$Phe and Tyr308$^{7.43}$Ala mutations have a similar effect. These residues, Asp128$^{3.32}$ and Tyr308$^{7.43}$, connect TM3 and TM7 (Figure 3.7B), and these mutations suggested a network of hydrogen bonds that stabilise the inactive state of the receptor [164].

The structure also revealed the 2 water molecules that bridge a hydroxyphenyl moiety in opioids to His278$^{6.52}$ (Figure 3.7C); mutation of this residue to a Cys in all classical opioid receptors abolishes binding to opioids [165].

The conformations of relevant motifs can be found in Figures 1.4, 1.5, and 1.6, in Chapter 1.

3.1.1.3 Investigating interactions between δ OR and certain ligands

To investigate the dynamics of the receptor in the presence of agonists and antagonists, conventional MD simulations were performed. The δ OR (in the inactive state; coordinates taken from the Å resolution crystal structure, PDB accession 4N6H [13]) was simulated (3×100 ns) in a 100% POPC membrane, in the presence of the agonists morphine, oxycodone, TAN-67, and KNT-127, and in the presence of the antagonists naltrindole (co-crystallised), naloxone, naltrexone (all of which shown in Figure 3.8),

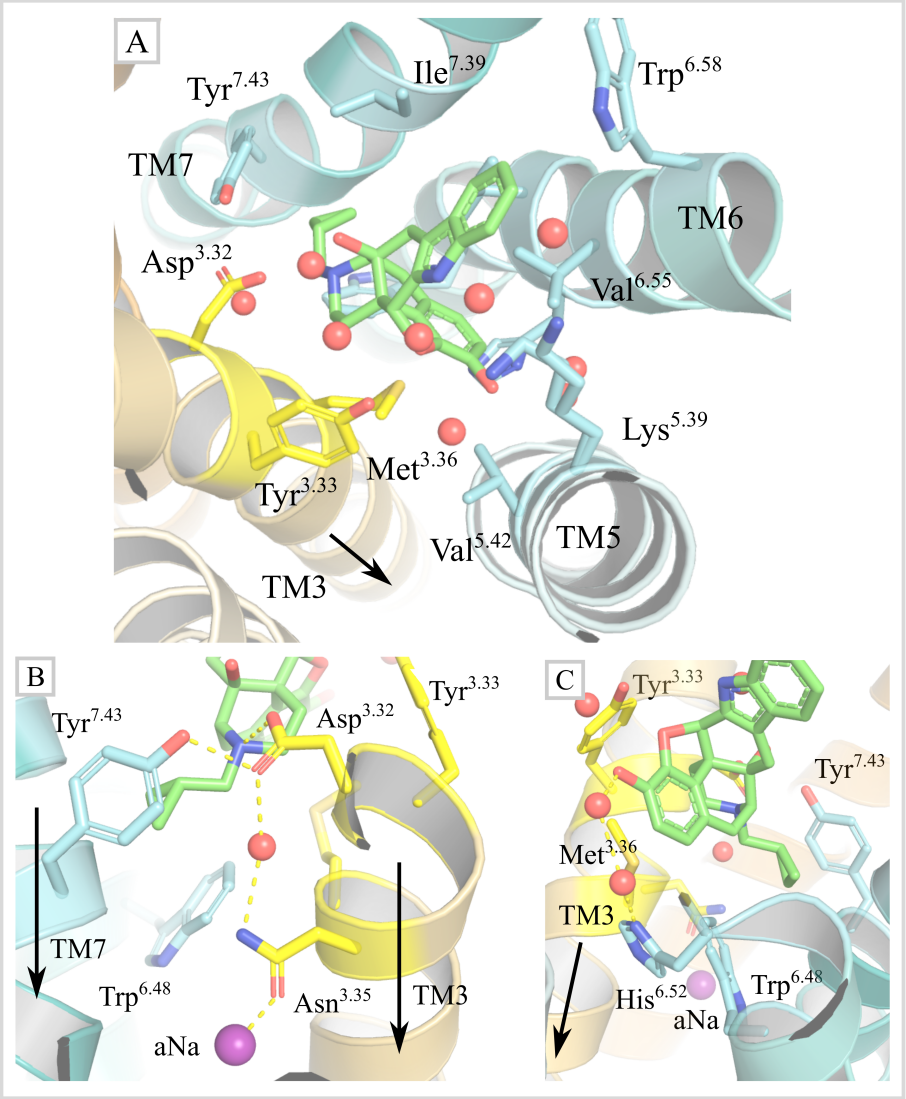


Figure 3.7: Details of the naltrindole binding site. **A** Overview of the residues surrounding the naltrindole (every residue displayed is within 4 Å of the ligand). **B** Close up of interactions between naltrindole and the highly conserved Asp128$^{3.32}$. **C** naltrindole interacts with TM6 via two water molecules. In all cases, naltrindole carbon atoms are shown as green sticks. aNa is the allosteric sodium.

and the mixed agonist/antagonist DIPP-NH_2 (a compound that demonstrates agonist activity at the μ OR, and antagonist activity at the δ OR; see Figure 1.10). All simulations were performed in the presence an allosteric sodium (also co-crystallised).

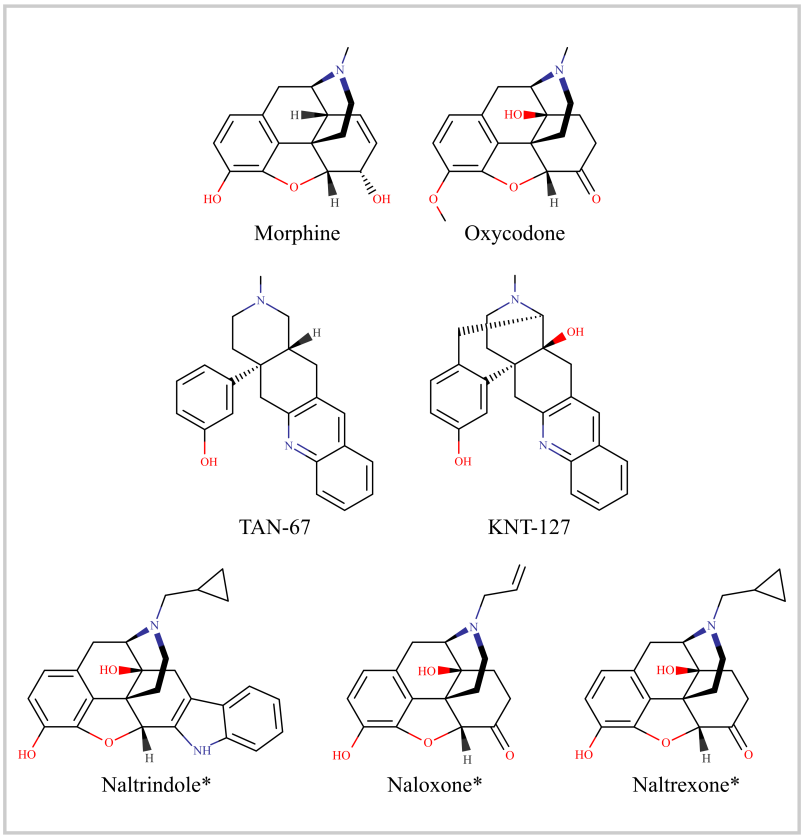


Figure 3.8: Structure of the opioids simulated in this Chapter. *Antagonists.

3.1.1.4 Investigating allosteric communication between ligands and the G protein binding site

Understanding how binding sites in remote locations within proteins communicate is critical in understanding allosteric modulation. The classical binding sites in GPCRs include the orthosteric binding site, responsible for ligand binding; the allosteric binding site, responsible for sodium binding; and the G protein binding site, responsible for binding of intracellular effectors, including G protein, mimetic nanobodies, and arrestins. While the first two sites listed are relatively adjacent in the GPCR structure, the latter is located much further away (see Figures 1.1 and 1.2).

Communication between the two earlier sites can be more directly tracked by analysing the behaviour of a small subset of residues ($Asp128^{3.32}$ and $Asn131^{3.35}$ in the case of the opioid receptors, for example). Conversely, communication between

the orthosteric binding site and the more distant G protein binding site is much more difficult to capture by tracking torsions of selected residues; instead, a more extended allosteric network might be in play that facilitates this information exchange.

Information-based methods can be used to quantify this communication, and track any information pathways between distal residues. Given that these methods require precise tracing of particle coordinates over time, conventional MD provides an ideal method to generate these coordinates. These coordinates can then be used to estimate the configurational entropy and the information exchange between residues (see section 2.1.3 of Chapter 2) [94].

The mutual information (MI) was calculated for all pairs of residues whose side-chains are capable of hydrogen-bonding (a comprehensive list of excluded residues can be found in Appendix A).

Following the calculation of the MI, each residue/residue/MI triad was placed into a network, where each residue represented a node, and the corresponding MI between residues represented the edge weight (using the Networkx Python package [102]), and a quick path-searching algorithm was developed to trace pathways between nodes.

3.1.2 Results

This section reports on results from conventional MD simulations performed on the *homo sapiens* δ OR.

3.1.2.1 Residency time during trajectories

For the remainder of this Chapter, morphine derivatives will be referred to as "classical opioids", whereas TAN-67, KNT-127, and DIPP-NH_2 will be referred to as "non-classical opioids".

From the resulting trajectories, a ligand contact analysis was performed. For this purpose, the software package LINTools [120] was used. Table 3.1 shows the contact analysis of some classical opioids.

As expected, all classical opioids are in constant contact with Asp128$^{3.32}$ and Tyr129$^{3.33}$. Met132$^{3.36}$ is also an ever-present contact in all simulations performed. The non-classical opioids TAN-67 and KNT-127 are also constantly in contact with the same residues, despite the lack of an epoxy oxygen. Instead, in the resulting simulations, the side chain of Tyr129$^{3.33}$ is maintained in its crystallographically determined state via water-meditated hydrogen bonds to the hydroxyphenyl moiety and the quinoline nitrogen atom of either molecule (see Figure 3.9A and B). DIPP-NH_2

also lacks an epoxy oxygen, but is incapable of maintaining a Tyr129$^{3.33}$ in the same conformation as the other ligands, likely due to the methyl groups in the 2,6 positions (see Figure 3.9).

In almost all simulations performed, the connection between the ligand and His278$^{6.52}$ is maintained by water-mediated hydrogen bonds. In most simulations, two waters remain bridging the ligand and the residue (Figure 3.10, A and C); these often interchange with waters from bulk. In other simulations, only one of the water molecules remains. In the oxycodone simulations, the offending waters are displaced to bulk by the methyl group in R1 (Figure 1.8B).

Table 3.1: Differences between agonists and antagonists are mostly evident in interactions between the ligand and Tyr308$^{7.43}$, as shown in the contact times for residues in Figure 3.8 (started from the inactive state; 3 x 100 ns trajectories). Highlighted in yellow are contact times different to those calculated from resulting naltrindole trajectories (as these serve as control simulations). *Antagonists.

Residue \ **Ligand**	Naltrindole*	Naltrexone*	Naloxone*	Morphine	Oxycodone	TAN-67	KNT-127
Asp128$^{3.32}$	100%	100%	100%	100%	100%	100%	100%
	100%	100%	100%	100%	100%	100%	100%
	100%	100%	100%	100%	100%	100%	100%
Tyr129$^{3.33}$	100%	99%	100%	98%	100%	100%	98%
	100%	99%	100%	100%	97%	99%	100%
	100%	99%	99%	99%	99%	100%	99%
Met132$^{3.36}$	98%	100%	100%	100%	100%	100%	100%
	99%	100%	100%	100%	99%	100%	100%
	100%	99%	100%	100%	100%	99%	100%
Lys214$^{5.39}$	49%	28%	83%	70%	98%	24%	70%
	31%	26%	64%	39%	95%	46%	39%
	24%	60%	71%	55%	98%	37%	55%
Val217$^{5.42}$	78%	85%	78%	69%	84%	71%	69%
	77%	76%	79%	71%	79%	68%	71%
	75%	81%	78%	73%	75%	65%	73%
Phe218$^{5.43}$	-	-	-	-	63%	-	-
	-	-	-	-	48%	-	-
	-	-	-	-	59%	-	-
Trp274$^{6.48}$	100%	100%	98%	94%	87%	89%	94%
	100%	100%	99%	97%	87%	97%	97%
	100%	100%	99%	83%	70%	96%	83%
Ile277$^{6.51}$	100%	100%	99%	97%	99%	98%	97%
	100%	100%	99%	98%	97%	99%	98%
	100%	100%	99%	97%	98%	96%	97%
His278$^{6.52}$	98%	96%	89%	90%	99%	96%	90%
	94%	97%	89%	95%	98%	97%	95%
	96%	96%	88%	93%	99%	94%	93%
Val281$^{6.55}$	61%	83%	73%	76%	96%	70%	76%
	76%	78%	79%	85%	92%	64%	85%
	75%	82%	76%	80%	93%	61%	80%
Trp284$^{6.58}$	87%	-	-	-	-	92%	-
	84%	-	-	-	-	93%	-
	39%	-	-	-	-	79%	-
Leu300$^{7.35}$	87%	-	-	-	-	82%	-
	84%	-	-	-	-	81%	-
	77%	-	-	-	-	84%	-
Ile304$^{7.39}$	97%	93%	93%	83%	41%	58%	83%
	71%	89%	98%	58%	73%	85%	58%
	97%	98%	98%	77%	40%	78%	77%
Gly307$^{7.42}$	47%	-	-	-	-	-	-
	53%	-	-	-	-	-	-
	53%	-	-	-	-	-	-
Tyr308$^{7.43}$	98%	100%	93%	56%	26%	-	56%
	99%	100%	98%	31%	56%	-	31%
	100%	100%	98%	14%	19%	-	14%

Table 3.2: Contact times for antagonist non-classical opioid DIPP-NH_2 (started from the inactive state; 3 x 100 ns trajectories). Notice the similarities in the contact times between the DIPP-NH_2 and Tyr308$^{7.43}$ and the same interaction time between Tyr308$^{7.43}$ and the antagonists in Table 3.1.

Gln105$^{2.60}$	Lys108$^{2.63}$	Val124$^{3.28}$	Leu125$^{3.29}$	Asp128$^{3.32}$	Tyr129$^{3.33}$	Met132$^{3.36}$
89%	86%	66%	73%	100%	93%	74%
88%	37%	84%	72%	100%	56%	83%
51%	31%	40%	75%	100%	52%	78%
Cys198^{ECL2}	Met199^{ECL2}	Leu200^{ECL2}	Lys214$^{5.39}$	Val217$^{5.42}$	Trp274$^{6.48}$	Ile277$^{6.51}$
97%	36%	1%	63%	61%	72%	97%
96%	50%	67%	69%	77%	42%	94%
73%	38%	43%	78%	64%	44%	91%
His278$^{6.52}$	Val281$^{6.55}$	Trp284$^{6.58}$	Leu300$^{7.35}$	His301$^{7.36}$	Ile304$^{7.39}$	Tyr308$^{7.43}$
77%	94%	99%	94%	43%	92%	85%
78%	96%	91%	78%	16%	78%	86%
82%	96%	96%	95%	66%	95%	87%

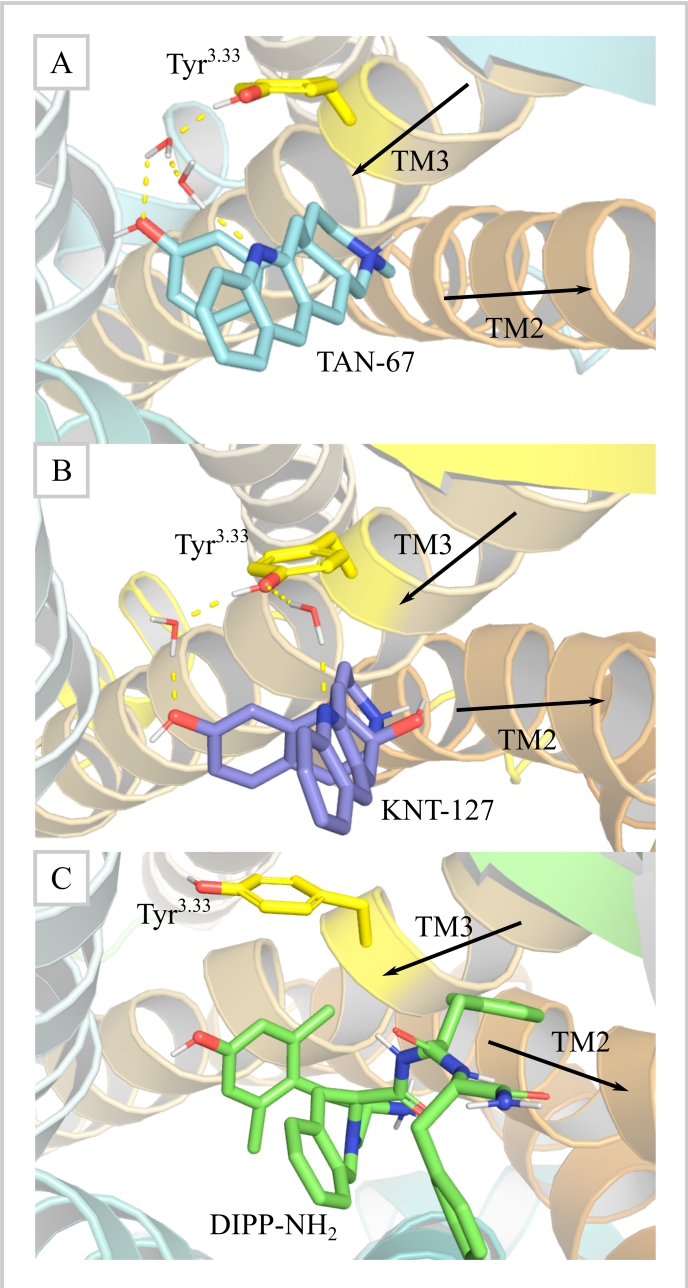


Figure 3.9: The lack of an epoxy oxygen in the structure of (A) TAN-67 and (B) KNT-127 doesn't affect their binding and their interaction with $Tyr^{3.33}$. In both cases, the interactions are mediated by two water-meditated hydrogen bonds to the hydroxyphenyl moiety of both ligands, as well as the quinoline nitrogen atom.

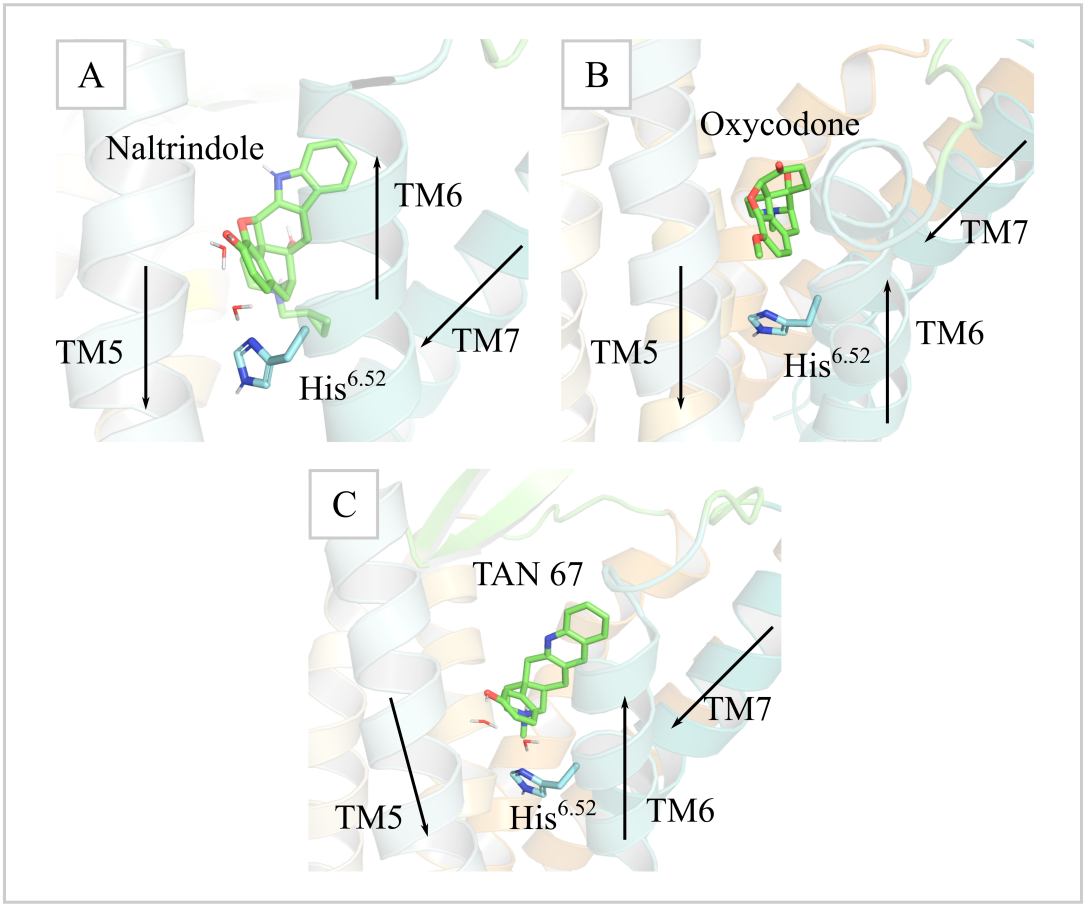


Figure 3.10: Waters bridging classical, (A) naltrindole and (B) oxycodone, and non-classical opioid (C) TAN-67, and $His^{6.52}$.

One of the more salient differences between the residues interacting with the classical opioids is interactions between the ligand and Tyr308$^{7.43}$ (see Figure 3.7B). On average, all antagonists are in near constant contact with this residue; more specifically, the *N*-substituted group is responsible for restricting the movement of the side-chain of Tyr308$^{7.43}$. Strangely, DIPP-NH_2 is also coordinated by this residue, despite the lack of a moiety that protrudes towards the sodium site. Instead, the quaternary nitrogen, critical to opioid binding, is found closer to Tyr308$^{7.43}$ in the crystallographically determined structure (PDB accession 4RWD [64]), where it remains in all simulations reported.

Naltrindole, the larger of the classical opioids, is coordinated by more residues during each simulation, particularly Trp284$^{6.58}$ and Leu300$^{7.35}$. This is also the case for TAN-67 and DIPP-NH_2. Curiously, KNT-127 is not in contact with these residues, despite sharing a similar structure to TAN-67 (Figure 3.8).

3.1.3 Bridging the orthosteric and allosteric binding sites

Changes in the binding profile of opioids may affect how the orthosteric and allosteric binding sites interact. This can be measured by tracking the behaviour of Asp128$^{3.32}$ and Asn131$^{3.35}$, the two residues that bridge the sites (see Figure 3.7B).

3.1.3.1 Changes in dihedral angles of the key residues Asp128$^{3.32}$ and Asn131$^{3.35}$

The conformational flexibility of both Asp128$^{3.32}$ and Asn131$^{3.35}$ were studied. Asp128$^{3.32}$ is involved in the binding of opioids (classical and non-classical), whereas Asn131$^{3.35}$ is heavily involved in interactions with the allosteric sodium and the "sodium effect". A summary of all replicates is shown in Figure 3.11.

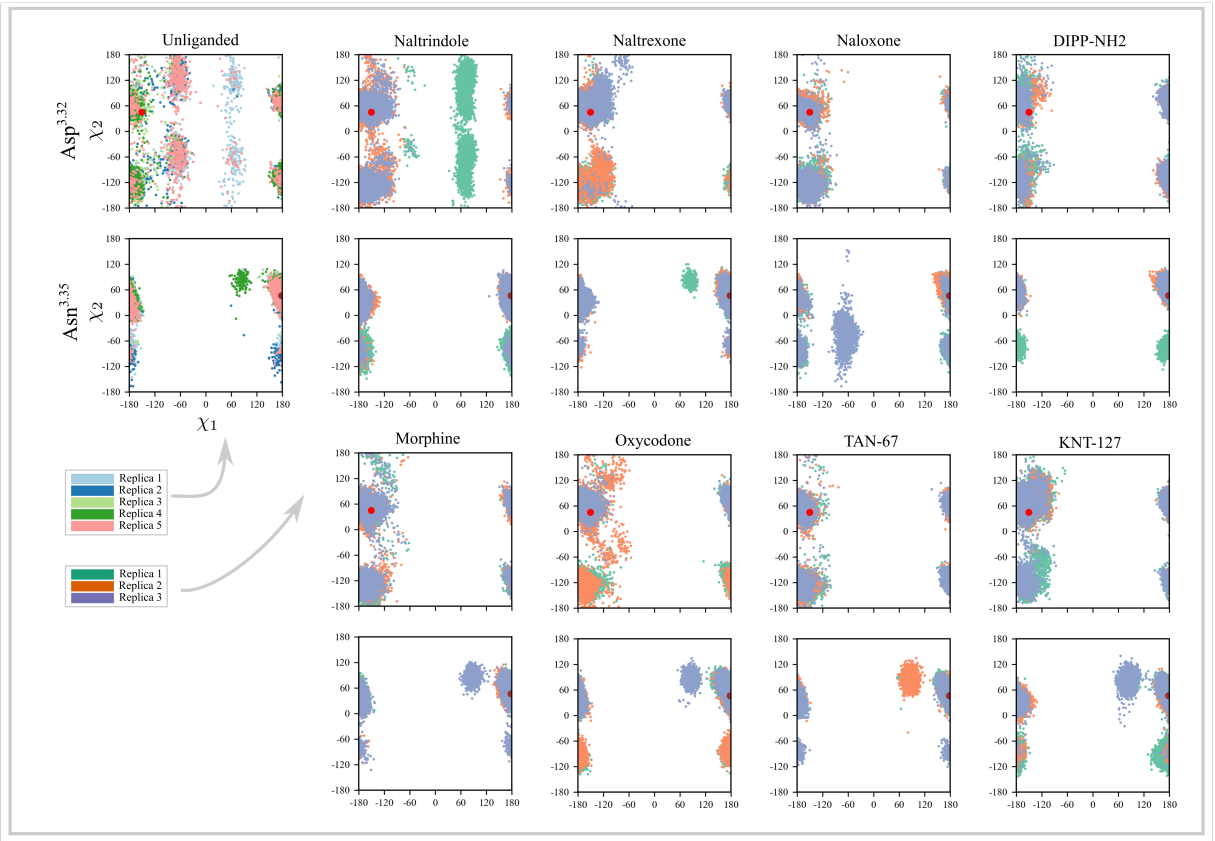


Figure 3.11: Antagonists seem to restrict the conformational freedom of both $Asp^{3.32}$ and $Asn^{3.35}$, though not significantly more so than agonists. The red and brown dots represent the values of the dihedrals in the crystallographically determined state (PDB accession 4N6H [13]). *Antagonists.

The addition of naltrindole (and other opioids) drastically restricts the conformational flexibility of Asp128$^{3.32}$. Interestingly, in one of the simulations, Asp$^{3.32}$ rotates away from the allosteric binding site, directly affecting the conformation of Asn131$^{3.35}$ (Figure 3.12). During the other 2 simulations, the Asp128$^{3.32}$ and Asn131$^{3.35}$ remain in their crystallographically determined state.

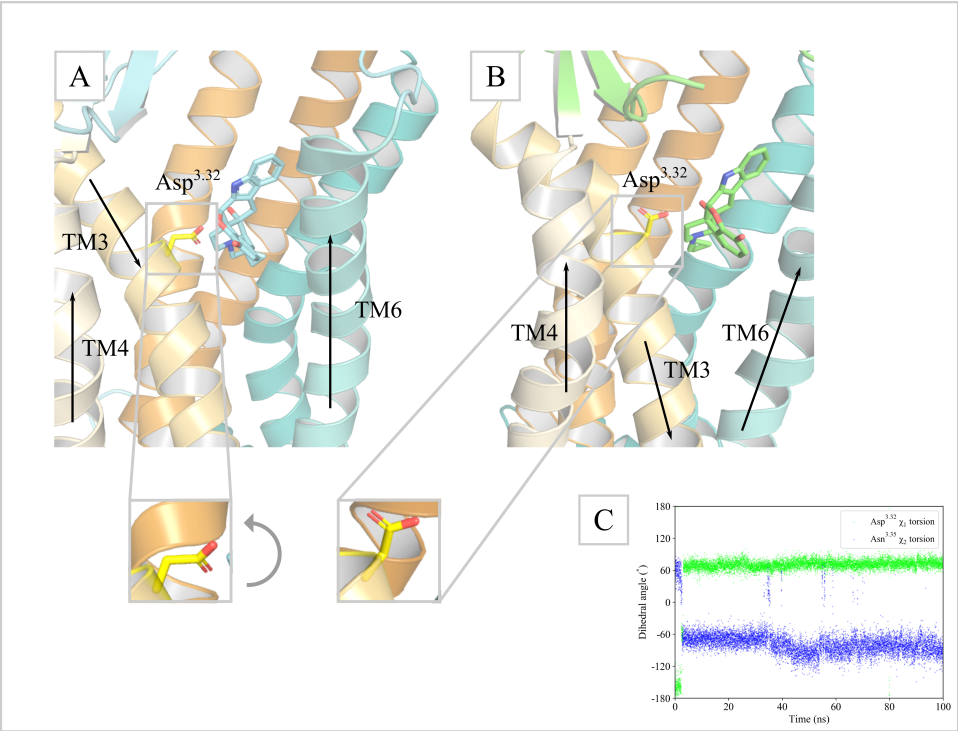


Figure 3.12: In one of the simulations, Asp128$^{3.32}$ starts in it's original conformation (A), but rotates away from the allosteric pocket (B). The timeseries plot (C) shows that this rotation only affects the χ_2 dihedral of Asn131$^{3.35}$.

3.1.3.2 Connecting the two critical binding sites

In the inactive crystal structure, a water molecule bridges the two residue, connecting the orthosteric to the allosteric binding site. The frequency of a water occupying this space in simulations was analysed using the Water Bridge analysis functionality of MDAnalysis [98]. A summary is provided in Table 3.3.

Table 3.3: Frequency of number of waters bridging Asp128$^{3.32}$ and Asn131$^{3.35}$ for all simulations. The data doesn't indicate any major differences between the different ligands, though longer simulations might. Data collected from 3 x 100 ns simulations started at the inactive state. †Data reported for two simulations. *Antagonists.

	Unliganded	Naltrindole*†	Naltrexone*	Naloxone*	DIPP-NH_2*	Morphine	Oxydocone	TAN-67	KNT-127
No waters	29 ± 9.93	10.9 ± 1.3	17.2 ± 5.23	-	-	11.1 ± 5.1	53.0 ± 5.1	14.1 ± 6.23	35.1 ± 25.1
1 water	70.3 ± 9.8	89 ± 1.3	82.7 ± 5.33	99.62 ± 0.23	99.79 ± 0.07	88.7 ± 5.1	46.7 ± 4.9	85.8 ± 6.2	64.7 ± 25.0
2 waters	0.7 ± 0.3	0.1 ± 0.03	0.13 ± 0.05	0.38 ± 0.23	0.21 ± 0.07	0.12 ± 0.07	0.27 ± 0.22	0.1 ± 0.1	0.2 ± 0.2

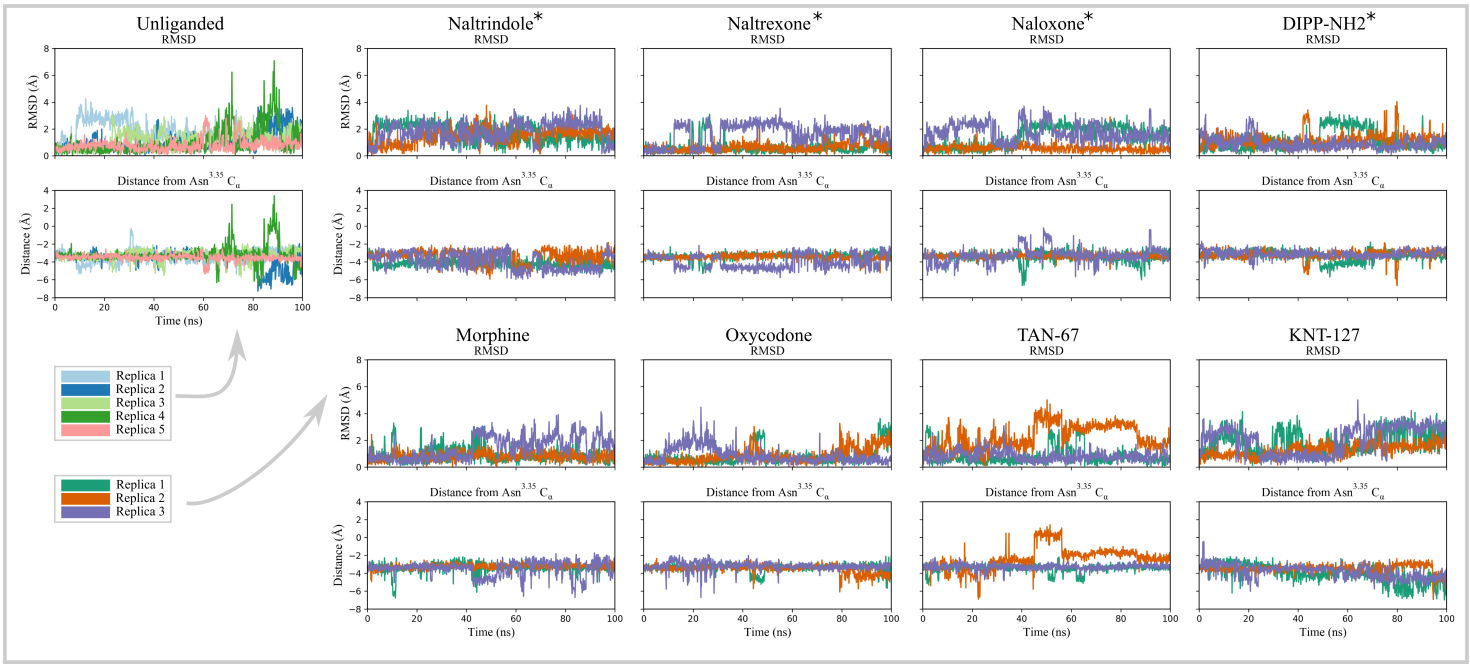


Figure 3.13: Sodium RMSD and z-distance relative to the α carbon of Asn131$^{3.35}$. Starred ligands are antagonists. From the data reported here, there does not seem to be any distinctions between antagonist and agonist-bound simulations. *Antagonists.

3.1.3.3 Connecting all receptor sites

Benchmarking MI values: correlated and uncorrelated motions during simulations As previously mentioned, it is possible to use the fluctuation of residues to calculate the MI and uncover concerted motions in the system. The calculated mutual information can reveal how correlated the motion between two residues are. This is particularly useful, since it can be used to find otherwise disregarded torsions that may be involved in allosteric signalling.

The resulting MI varies between [0, ∞) (which can be normalised to [0, 1]) such that an MI of 0 means the motion between two residues in purely random; an MI of ∞ (or normalised MT of 1) indicates correlated motion [101]. An initial observation of the top and bottom 10 residue pairs (sorted my average normalised MI) reveals a trend: motion of structurally neighbouring residues tend to be concerted, while motion of distant residues tend to be random. These are summarised in Table 3.4. This analysis extends to all residue pairs (see Figure 3.14).

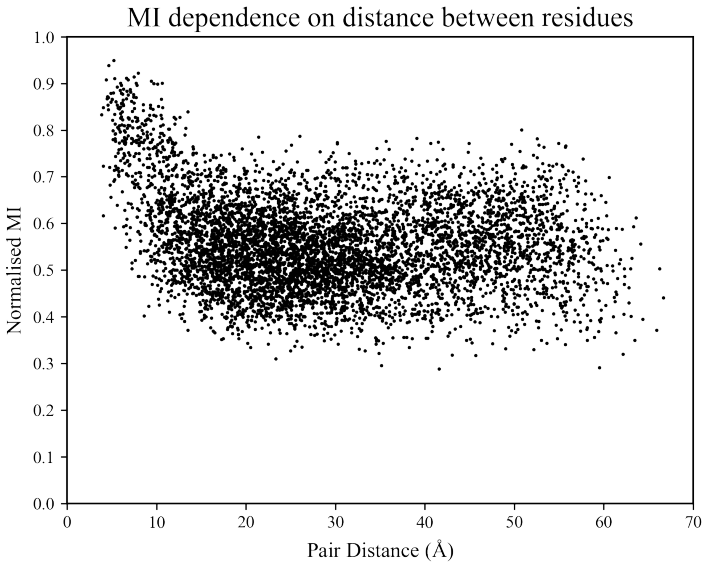


Figure 3.14: Residues further away within the protein structure share lower information.

Perhaps unsurprisingly, proximal residues capable of forming salt bridges with oppositely charged counterparts share a high MI. Asp253$^{6.27}$ and Arg257$^{6.31}$, and Glu251$^{6.25}$ and Arg254$^{6.28}$, are found on the same face of TM6, and are thus capable of forming electrostatic interactions (see Figure 3.15A). Aromatic residues may also share a high MI due to stacking interactions, directly or mediated by a third residue

Table 3.4: Average normalised MI from unliganded simulations. Left: bottom 10 pairs; Right: top 10 pairs.

Residue 1	Residue 2	MI	Residue 1	Residue 2	MI
Ser177$^{4.54}$	Ser249$^{6.23}$	0.29	Asp253$^{6.27}$	Arg257$^{6.31}$	0.95
Ser206$^{5.31}$	Ser249$^{6.23}$	0.29	Glu251$^{6.25}$	Arg254$^{6.28}$	0.94
Ser177$^{4.54}$	Ser255$^{6.29}$	0.30	Trp207$^{5.32}$	Trp209$^{5.34}$	0.92
Ser143$^{3.47}$	Ser249$^{6.23}$	0.31	Arg254$^{6.28}$	Arg257$^{6.31}$	0.91
Ser126$^{3.30}$	Ser249$^{6.23}$	0.32	Lys250$^{6.24}$	Asp253$^{6.27}$	0.91
Asp128$^{3.32}$	Ser249$^{6.23}$	0.32	Ser249$^{6.23}$	Lys252$^{6.26}$	0.91
Ser204^{ECL2}	Ser249$^{6.23}$	0.32	Trp284$^{6.58}$	Arg291^{ECL3}	0.91
Asp128$^{3.32}$	Ser255$^{6.29}$	0.32	Glu251$^{6.25}$	Asp253$^{6.27}$	0.91
Ser249$^{6.23}$	Sodium	0.33	Arg291^{ECL3}	Asp293$^{7.28}$	0.90
Ser255$^{6.29}$	Sodium	0.33	Tyr208$^{5.33}$	Thr211$^{5.36}$	0.90

(such as interactions between Trp207$^{5.32}$ and Trp209$^{5.34}$ being mediated by Tyr208$^{5.33}$; Figure 3.15B).

Information travels to downstream binding sites through TM4 In the work presented here, the information pathways spawning from the ligand to other key residues within the GPCR, particularly the allosteric sodium and key residues of the G protein binding site (Arg146$^{3.50}$, Tyr233$^{5.58}$, Tyr318$^{7.53}$) were calculated.

This was done by adapting a path searching algorithm to trace the MI from the ligand. The procedure is detailed in Chapter 2, section 2.1.3.3, and the code used is detailed in Appendix A. In essence, the residue pairs and corresponding MI are made into a network, with each residue functioning as a node; a "start" and "end" node are defined (for example, the ligand and Arg146$^{3.50}$). Then, all nodes within a cut-off distance of the "starting" node are sorted by MI; the highest MI is selected, and the corresponding residue becomes the new "starting" node. This process continues until the "end" node becomes the new "starting" node, or if no path is found.

Analysis of information pathways between the ligand and the aforementioned residues reveals that the information transfer often occurs via TM4. An interesting observation was that one the most travelled through residues in inactive state simulations was Trp173$^{4.50}$. Table 3.5 summarises these data.

Table 3.5: Number of simulations where information transfer to key residues occurred either via TM3/TM4, only via TM4, or neither. Notice how information is more likely to be transfered via TM4 in agonist simulations. *Antagonists.

	$Arg146^{3.50}$		
Ligand	$Trp173^{4.50}$	$Tyr130^{3.34}$+$Trp173^{4.50}$	Neither
Naltrindole*	1	1	1
Naltrexone*	1	-	2
Naloxone*	3	-	-
DIPP-NH_2*	2	1	-
Morphine	1	1	1
Oxycodone	1	-	2
TAN-67	-	1	2
KNT-127	1	1	1
	$Tyr233^{5.58}$		
Ligand	$Trp173^{4.50}$	$Tyr130^{3.34}$+$Trp173^{4.50}$	Neither
Naltrindole*	3	-	-
Naltrexone*	1	-	2
Naloxone*	3	-	-
DIPP-NH_2*	-	2	-
Morphine	1	1	1
Oxycodone	1	-	2
TAN-67	-	1	2
KNT-127	1	1	1
	$Tyr318^{7.53}$		
Ligand	$Trp173^{4.50}$	$Tyr130^{3.34}$+$Trp173^{4.50}$	Neither
Naltrindole*	1	1	1
Naltrexone*	1	-	2
Naloxone*	3	-	-
DIPP-NH_2*	1	2	-
Morphine	-	1	2
Oxycodone	-	-	3
TAN-67	-	1	2
KNT-127	1	-	2

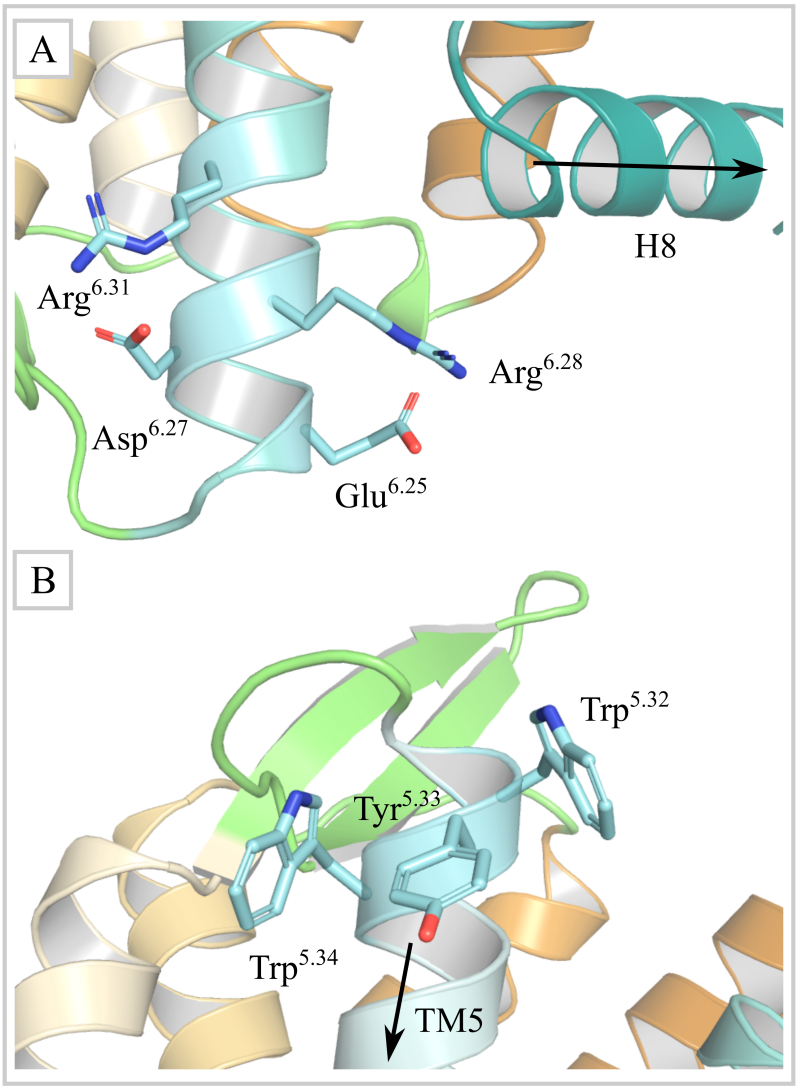


Figure 3.15: Examples of high MI exchange hotspots in unliganded simulations. (A) Proximal charged residues may share high MI due to electrostatic interactions. (B) Motion of aromatic residues may be mediated via a third residue.

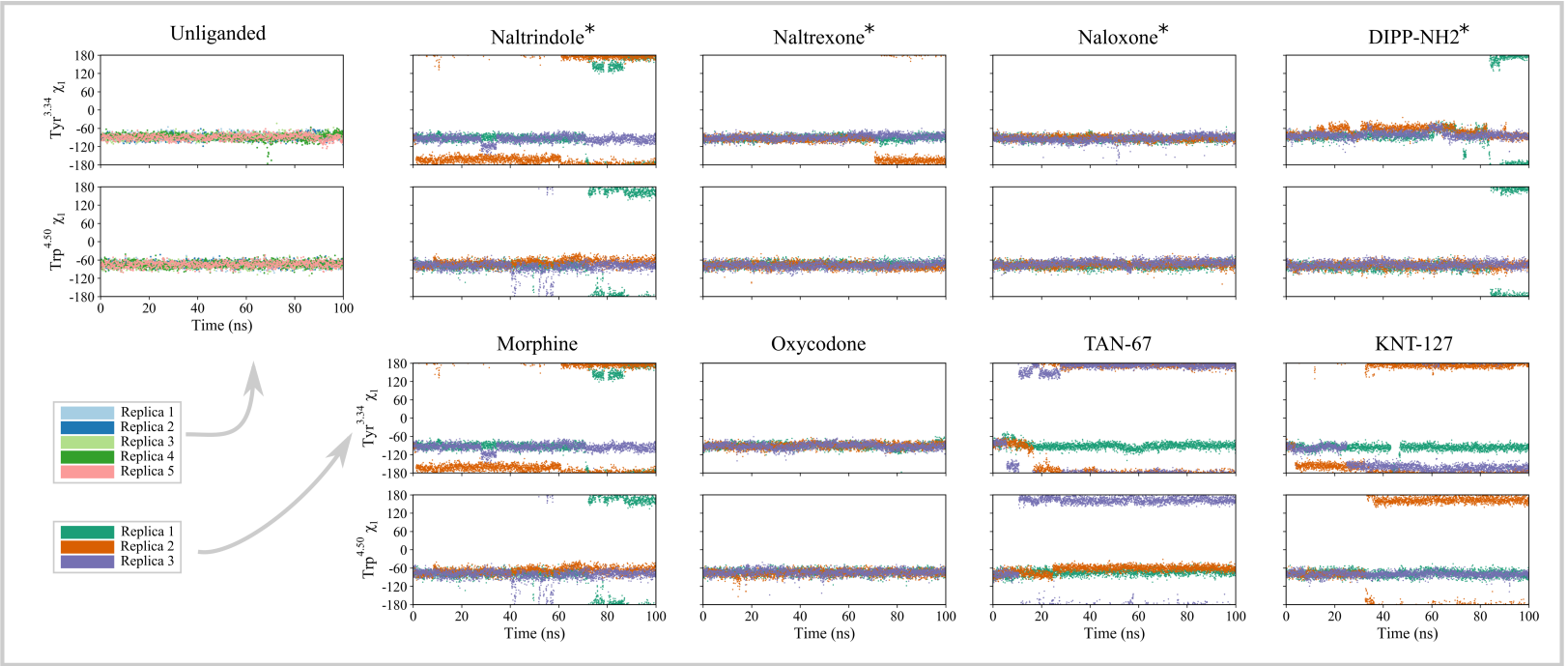


Figure 3.16: χ_1 dihedral changes in Tyr130$^{3.34}$ and Trp173$^{4.50}$ over 3 x 100 ns simulations (5 x 100 ns for apo simulations, left). *Antagonists.

As shown in the Table 3.4, communication through TM4 often occurs via Tyr130$^{3.34}$. Interestingly, this residue is located close to the orthosteric binding site, and changes to Trp173$^{4.50}$ are often accompanied by changes in Tyr130$^{3.34}$. Figure 3.16 summarises this phenomenon.

From the data summarised above, agonists have a higher chance of disturbing the rotameric state of both Tyr130$^{3.34}$ and Trp173$^{4.50}$, though strangely naltrindole exhibits similar behaviour. Simulations of antagonists are more likely to maintain the initial dihedral values throughout the course of the simulations, with oxycodone showing similar behaviour to antagonists.

3.2 Discussion

Recently, the structure of the *homo sapiens* δ OR in complex with the morphinan naltrindole and an allosteric sodium was revealed [13]. Taking advantage of this, a set of conventional MD simulations were performed. An initial set were performed on the receptor embedded in POPC, with and without the co-crystallised naltrindole, and subsequent simulations were performed by replacing naltrindole with classical agonists morphine and oxycodone, non-classical agonists TAN-67 and KNT-127, classical antagonist naltrexone and naloxone, and non-classical pseudo-peptidic antagonist DIPP-NH_2.

3.2.1 Ligand contact analysis during conventional MD

One of the most fundamental measurement one can perform on protein-ligand complexes is a contact analysis between the receptor residues and the ligand. Small changes in the ligand structure will be reflected in this.

3.2.1.1 Opioid ligands and their relationship with Tyr$^{7.43}$

Previous studies on the binding of opioid ligands highlighted Tyr$^{7.43}$ as a residue critical to the binding of opioids [166, 167] (see Figure 3.7A). Mutating this residue to a phenylalanine or to a histidine drastically changes the binding affinity of certain opioid ligands, including naltrindole. Continuing with naltrindole as the example, while the former mutation merely reduces the binding affinity [166], the latter completely abolishes binding [167] (and, interestingly, promotes constitutive activity [141]).

The resulting simulations here reported indicate that this residue is found more frequently in contact with antagonists, and more specifically, the N-substituted "message" group. Recent *in silico* studies have proposed that Trp274$^{6.48}$ and Tyr308$^{7.43}$

may act as an activation switch, and that the "message" group determines ligand activity by modulating the average distance between the two residues [76].

While this offers a partial explanation for antagonism, the recent crystallographic structure of the κ OR bound to MP1104, as well as the list of morphine derivatives containing aromatic N-substitutes presented earlier in this chapter (Figures 3.3 and 3.4A/B), has thrown this assertion into question. A more complete understanding of the relationship between the N-substituted "message" group, Trp274$^{6.48}$, and Tyr308$^{7.43}$, will require more study, using a combination of both on-site mutagenesis and conventional (medium- to long-timescale) MD of a more exhaustive list of ligands.

3.2.1.2 Contact with "address" residues

As previously mentioned, when the δ OR was first crystallised [63], Leu300$^{7.35}$ was highlighted as a residue crucial for selectivity of δ OR ligands [162]. The relatively small Leu300$^{7.35}$ allows the indole group of naltrindole to adopt the orientation reported in the crystallographic states (see Figure 3.6), where it remains throughout the reported simulations.

Comparing naltrindole to the selective agonists TAN-67 and KNT-127, both of which contain a quinoline group, seems to indicate that the "address" aromatic groups bind in similar orientations. Indeed, for TAN-67, the quinoline group also interacts with Leu300$^{7.35}$ for the majority of the simulation.

The same applies to Trp274$^{6.58}$ (Figure 3.7). Both naltrindole and TAN-67 are in almost constant contact with Trp274$^{6.58}$, a residue known for it's importance in the binding of certain δ selective ligands [168]. Mutagenesis studies on this residue (specifically Trp274$^{6.58}$Ala) showed reduced affinity for highly selective ligands (such as SNC-80).

Strangely, however, data from KNT-127 simulations do not show interactions with Leu300$^{7.35}$, or indeed with Trp274$^{6.58}$. This phenomenon might be explained by the reduced internal flexibility compared to TAN-67. It is also possible that this is a result of slight reorientation of KNT-127 within the binding pocket.

3.2.1.3 Connecting the orthosteric and allosteric binding sites

One of the key features of opioid receptor structure is the conservation of residues required to accommodate an allosteric sodium in the receptor core. This sodium is known to alter the receptor's affinity to agonists, while not affecting the affinity of antagonists; this is known as the "sodium effect", a process by which the sodium ion alters the topography of the orthosteric site.

The 1.8 Å resolution crystallographic structure of the δOR (PDB accession 4N6H) was a great step forward in understanding this effect from a structural perspective. The structure revealed more precise details on the residues coordinating the sodium ion directly (Asp95$^{2.50}$, Asn131$^{3.35}$, and Ser135$^{3.39}$), as well as residues involved in coordination of the sodium indirectly (via one or more water-bridges; Trp274$^{6.48}$, Asn310$^{7.45}$ and Asn314$^{7.49}$). A visual summary is provided in Figure 1.7.

As previously mentioned, the 1.8 Å resolution crystal structure also revealed a water molecule that bridges Asp128$^{3.32}$ and Asn131$^{3.35}$. The resulting simulation data can be used to check if this water bridge is maintained. This is summarised in Table 3.3.

The resulting simulations suggest that at least one water molecule is constantly present in all antagonist-bound δ OR simulations (compared to the unliganded simulation). Combined with the χ dihedrals distribution plots of both Asp128$^{3.32}$ and Asn131$^{3.35}$ (summarised in Figure 3.11), it would seem that the presence of antagonists reduces the conformational flexibility of both residues, whereas the presence of agonists directly affects the conformational flexibility of Asn131$^{3.35}$.

3.2.1.4 The allosteric sodium does not discriminate between ligands in short timescale MD

Given the intimate relationship between Asn131$^{3.35}$ and the allosteric sodium, the z-distance between the C_α atom and the sodium, as well as the sodium RMSD, were calculated and summarised in Figure 3.13. No pattern seems to emerge from these data, indicating that more sampling would be required to observe more interesting dynamics (such as sodium exiting the allosteric site).

3.2.2 More data from trajectories: mutual information

One critical aspect of trajectories that is often neglected is how information is propagated across proteins. In some cases, it is important to understand how the inclusion/exclusion of external stimuli (such as ligands) affects distal areas of the protein. This transfer of information, known as allostery, is thought to underpin a great number of processes in all proteins [169].

Recently, a new method to capture this allosteric communication was developed by LeVine and Weinstein. Referred to as "N-body Information Transfer", it was used to differentiate residues involved in functional sites of the bacterial LeuT passive transported. The method involves estimating the configurational entropy of residues (see equation 2.9, page 29, in Chapter 2), which can be calculated using the covariance

matrix of coordinates of each residue. In the work presented here, the covariances were calculated using a modified version of the script provided in the MDAnalysis software package [98], using only non-hydrogen atoms.

The configurational entropy can then be used to estimate the mutual information between two residues/two sets of residues using equation 2.10 (Chapter 2, page 30) [94]. For the purposes of the work here presented, an in-house script was developed to calculate the information transferred between two bodies (in this case, two residues) for all pairs of non-hydrophobic residues; the reader is requested to consult [94] for the information transfer between 3 bodies, or [100] for higher order information terms.

In this paradigm, the calculation of mutual information return a value that ranges from 0 to ∞, where high information values indicate that the motion of the two residues is highly correlated. A correction can be made to each value using equation 2.11 (Chapter 2, page 30), converting the range from [0, ∞) to [0,1], a much more intuitive range [101]. However, this is not a necessary step.

This pairwise information exchange can be used for further analysis. A simple example (used here) is considering every residue a node, and the pairwise information between two residues an edge. This way, it is possible to track a "path" between various important nodes; for example, a path can be calculated between Naltrindole and the allosteric sodium in naltrindole-bound and sodium-bound simulations (Figure 3.17).

One of the most surprising observations from the mutual information analysis was the communication between Tyr130$^{3.34}$ and Trp173$^{4.50}$.

In some simulations, Tyr130$^{3.34}$ changes it's rotameric state. In all crystallised inactive ORs, this residue points to TM4 (Figure 3.18A). Over the course of the simulations, this tyrosine "flips" downwards, forcing Trp173$^{4.50}$ to change it's conformational state as well.

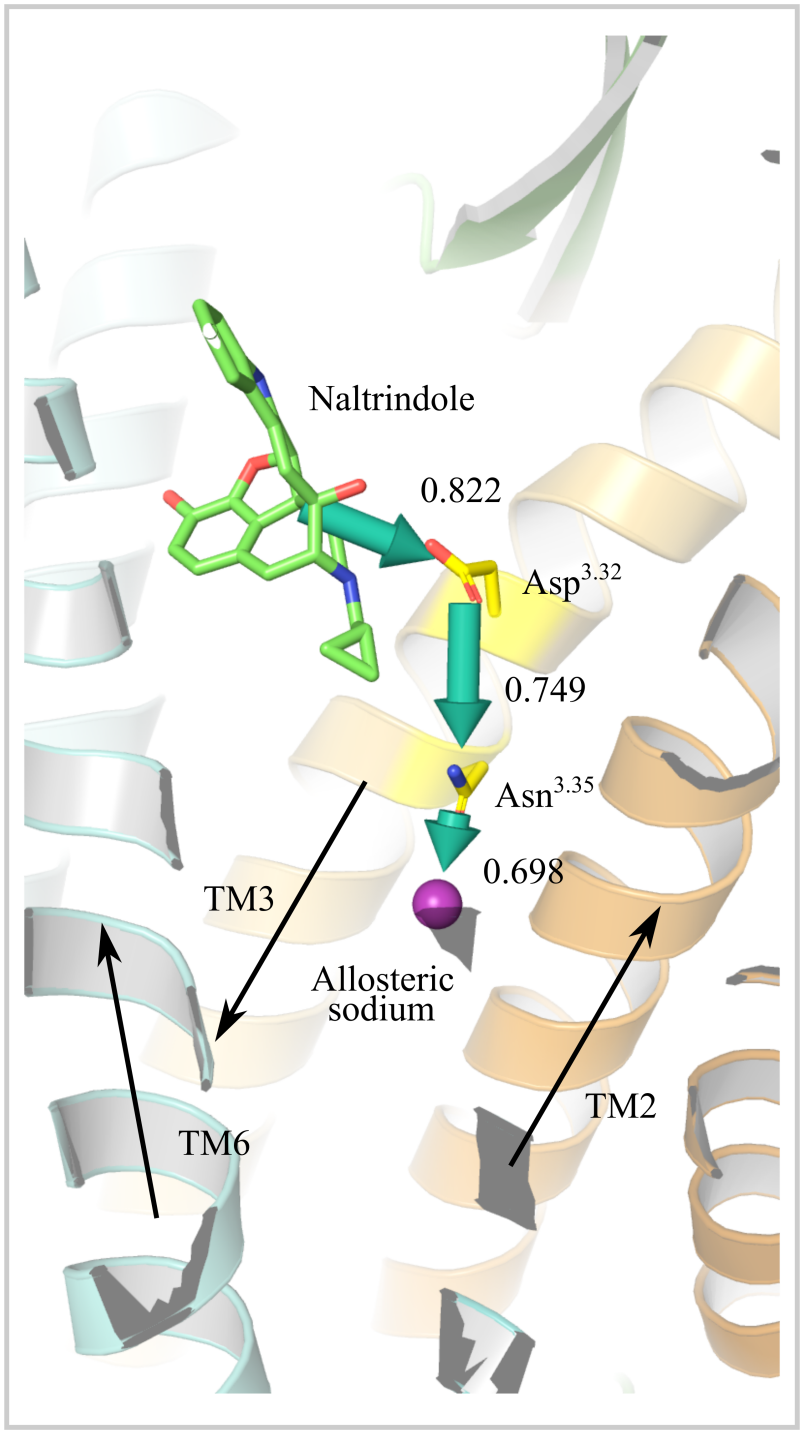


Figure 3.17: Path between Naltrindole and the allosteric sodium. The values between each residue displayed corresponds to the normalised MI between the two residues.

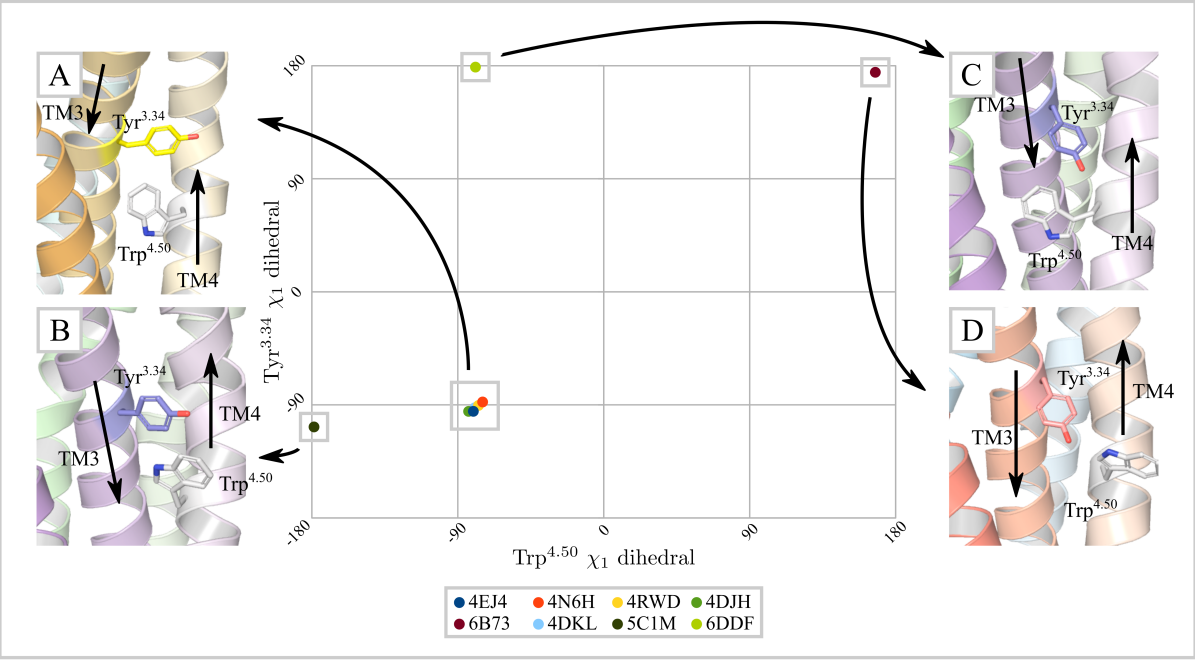


Figure 3.18: Conformational flexibility of Tyr130$^{3.34}$ and Trp173$^{4.50}$, as captured in different OR crystallographic structures. (A) Conformation of all inactive crystal states (represented here by the inactive crystallographically determined structure of the *homo sapiens* δ OR, PDB accession 4N6H [13]). (B), (C), and (D) ORs captured in the active state exhibit greater differences.

Surprisingly, there have not been many studies on either residue on the opioid receptors.

A full sequence alignment of all GPCRs (from the GPCR databank [20]) reveals that position 3.34 is only occupied by a tyrosine on the opioid receptors (including the orphanin FQ receptor; the reader is referred to the structure deposited under the PDB accession 4EA3 [170]), and on the FZD_9 and FZD_{10}, and Smoothened receptors of Class F (the latter of which has been captured via X-ray crystallography and deposited under the PDB accession 4QIN [171]). The importance of this residue in any of the aforementioned receptors is yet to be determined.

The tryptophan at position 4.50, conversely, is 97% conversed across all GPCR families [19, 20] (see Table 1.1). Despite this, not many studies have been performed on the importance of this residue.

Earlier studies on the importance of aromatic residues by Befort and colleagues have shown that Trp173$^{4.50}$ is involved in the binding of specific δ OR ligands (β-endorphin, dynorphin A, Leu- and Met-enkephalin) [166], though the precise nature of this recognition has not yet been revealed.

Decaillot and colleagues later addressed this residue in their seminal study on constitutively activating mutants (CAMs) of the δ OR. The mutations Thr134$^{3.38}$Ala, Ile170$^{4.47}$Thr, Asn169$^{4.46}$Ser, and Trp173$^{4.50}$Arg, all resulted in CAMs, with different levels of constitutive activity. The researchers proposed that these residues enveloped Trp173$^{4.50}$, which was thought to limit the mobility of TM3 by aromatic stacking with Tyr130$^{3.34}$ [141]. The mutations were then thought to perturb the stability of Trp173$^{4.50}$, which then affects the mobility of TM3.

When the aforementioned studies were published, crystallographic data was scarce. Since there were no Class A GPCR crystal structures available when the study by Befort and colleagues was published (in 1996), the researchers relied on homology modelling based on a model of the *Cricetinae* β_2 adrenoceptor, itself a model of the *homo sapiens* rhodopsin, itself a model of the bacteriorhodopsin [166]. The resulting model showed Trp173$^{4.50}$ clustering with Trp274$^{6.48}$, Phe218$^{5.43}$, and Phe222$^{5.47}$, directly involved in ligand binding (the reader is referred to Figure 1 of [166]).

However, later crystallography studies placed Trp173$^{4.50}$ parallel to the allosteric binding site, suggesting that mutations of this tryptophan disrupt the allosteric site, which then result in changes to the orthosteric site (as opposed to directly affecting the orthosteric site). It is possible that this relatively bulky residue blocks the rotation of Asn131$^{3.35}$, one of the residues responsible for coordinating the allosteric sodium.

Prior movement of Trp173$^{4.50}$ might be a required step before rotation of Asn131$^{3.35}$, though the rotation of Asn131$^{3.35}$ is not captured within 100 ns.

It may be possible to better capture the importance of this residue using long time scale conventional MD of either the Trp173$^{4.50}$Ala or Trp173$^{4.50}$Arg, and analysing the effects of the mutation on the allosteric sodium.

3.3 Limitations and further work

In this Chapter, a variety of MD simulations of were performed and analysed. The *homo sapiens* δ OR was embedded in a 100% POPC membrane, and was subjected to conventional MD in the presence of classical opioid agonists, antagonists, and non-classical agonists and antagonists.

3.3.1 Paths between residues

In the work presented here, tracking paths between residues was performed by creating networks. Each node in the network is the position of the geometric centre of non-hydrophobic residue side-chains. Each node is connected to every other node by an edge with weight equal to the mutual information between the corresponding residues.

To find a path, a start and an end node are required. The nodes around the start node are listed and sorted by the mutual information (a cut-off parameter is implemented here); the node with the highest information exchange becomes the new start node. This process continues until the end node specified is in the list of residues proximal to the current starting node.

This implementation presents certain issues, explored below.

3.3.1.1 Wandering around proteins

Depending on the cut-off value, the search algorithm will find paths that traverse a large section of the network (Figure 3.19).

This phenomenon is a consequence of how the path searching algorithm is implemented; the path is comprised of proximal residues that share high mutual information. Any concerted motion between two residues within the distance cut-off will be captured as part of the path.

It could be indicative of how GPCRs (and possibly all proteins) contain large numbers of allosteric pipelines. Some of these are more localised than others, restricted to a cluster of residues within one or two binding pockets; other pipelines might encapsulate more residues in the receptor. In the simulations presented here, the path

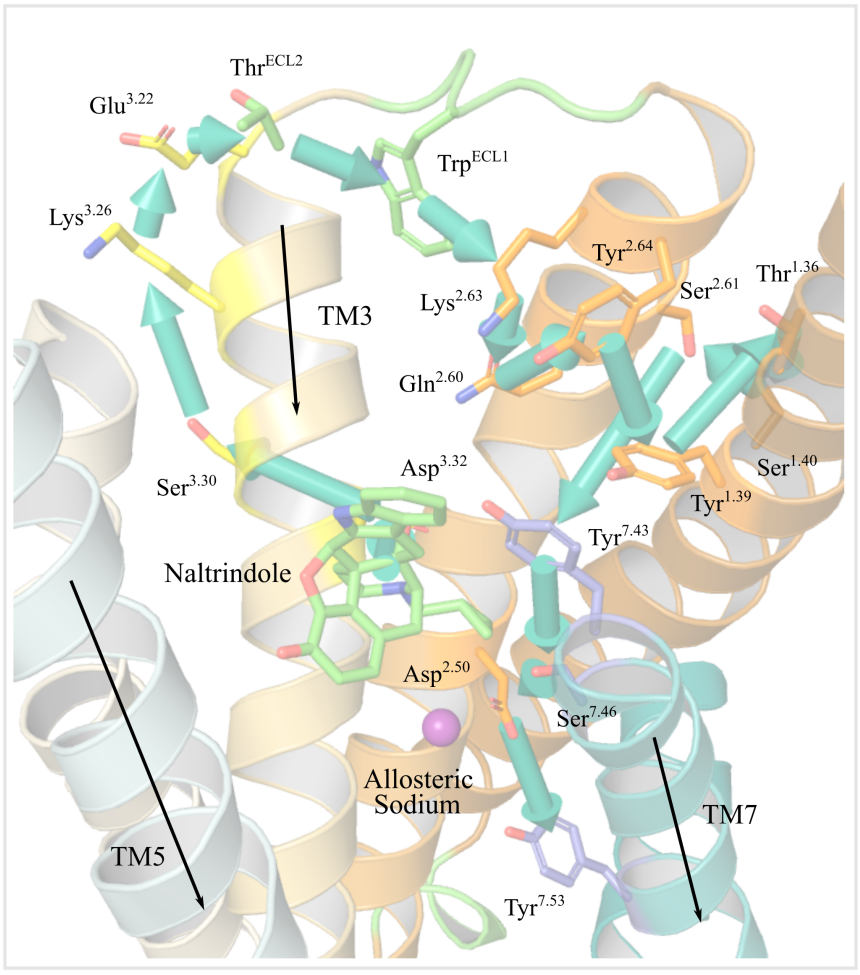


Figure 3.19: Information travelling within the receptor, from the ligand Naltrindole (green sticks) to Tyr318$^{7.53}$ (of the NPxxY motif; see also Figure 1.5, Chapter 1, page 13). Note that Asp128$^{3.32}$ is obscured behind Naltrindole.

from naltrindole to Tyr233$^{5.58}$ spans 40 residues; the path between naltrexone and Arg146$^{3.50}$/Tyr233$^{5.58}$ spans 48/50 residues, respectively.

3.3.1.2 Paths may not converge

Mutual information is not a one-way quantity. As previously explained, the mutual information between two residues approximates how concerted the motion between the two residues is. It would follow that, if a path is generated between distant residues (nodes) A and B, the path generated between B and A should be the same. However, as a result of how the path search is implemented, this isn't the case. An example is shown in Figure 3.20.

It may be possible to somehow refine this path searching by starting a search from both ends of the path (start and end nodes), and check if the paths "meet".

3.3.1.3 Paths in active states

How do these allosteric pathways change in different receptor states? At the time of this writing, there are three active state structures of opioid receptors (the reader is referred to Table 1.3 for further detail), all of which can be used to create homology models of the δ OR and investigate the behaviour of ligands in the binding site.

The resulting MD simulations could be used to develop a better understanding of the allosteric pathways within the active state receptor, and how these differ from those within the inactive state receptor. This would potentially identify regions that could targeted by allosteric modulators (which would either enhance or disrupt said communication pipelines) that could be used to stabilise metastable states otherwise difficult to capture using crystallography. Bhattacharya and Vaidehi have used this exact strategy to identify potential allosteric sites in the β_2 adrenoceptor, using data from µs long trajectories, opting to use internal coordinates for their analysis, rather than Cartesian coordinates (as was the case for the data presented here) [172].

3.3.1.4 Alternatives to G protein complexing

One of the main issues facing the field of GPCR study is capturing and simulating G protein-bound active states. While capturing GPCRs in inactive states has proven challenging (mainly because it requires removal of GPCRs from their native lipid environment), thermostabilising mutations, and truncation of highly flexible loop regions, among others, have allowed researchers to successfully crystallised a plethora of inactive state GPCRs (reviewed in [2]). The same is not always true when attempting to crystallise active-state GPCRs complexed with a native G protein.

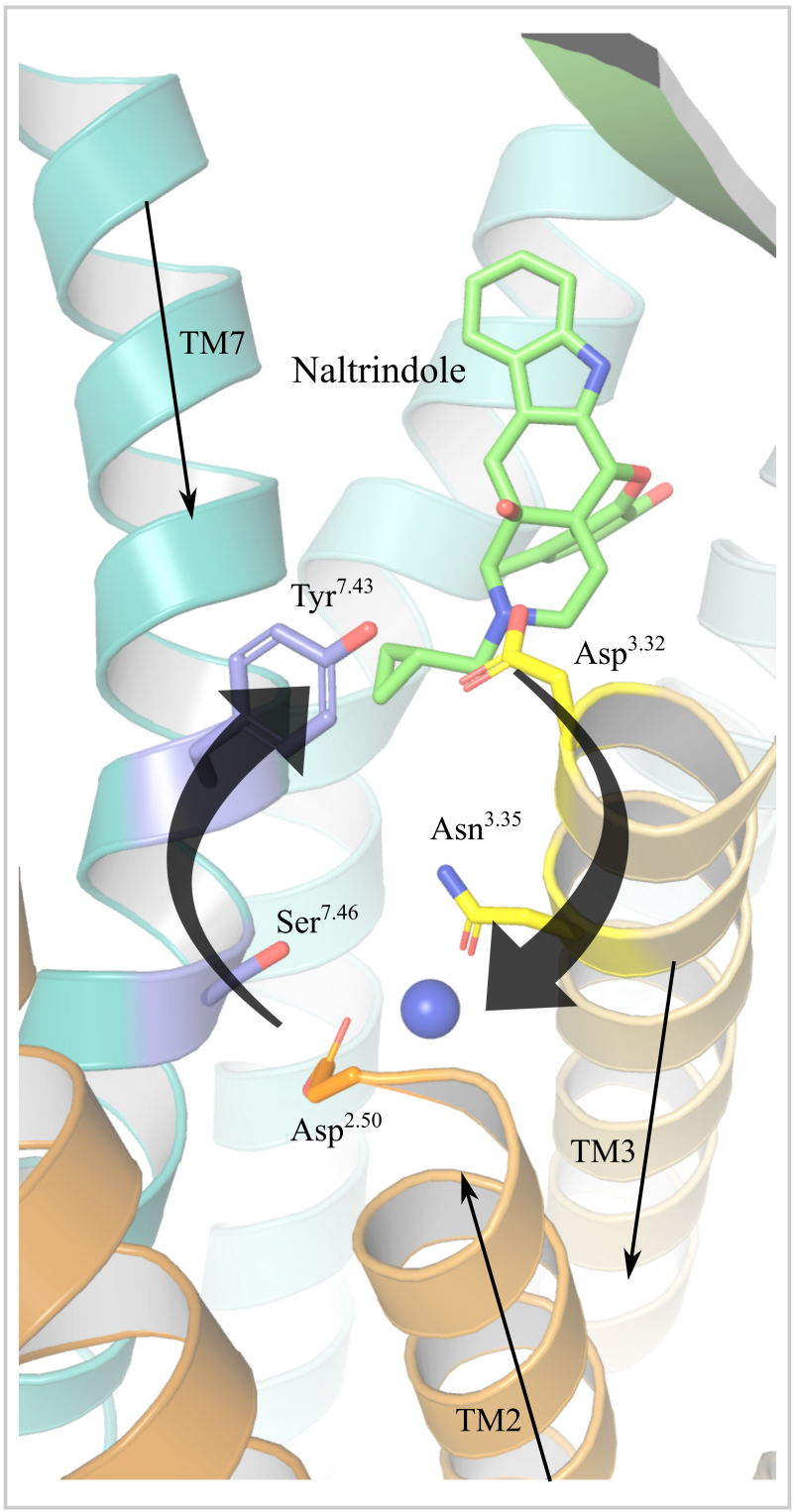


Figure 3.20: The path generated from naltrindole to the allosteric sodium (downward pointing arrow) involves Asp128$^{3.32}$ and Asn131$^{3.35}$, whereas the reverse path (upward pointing arrow) involves Ser311$^{7.46}$ and Tyr308$^{7.43}$. The allosteric sodium is represented by a blue sphere.

Inclusion of a G protein in *in silico* studies of GPCRs presents another set of issues. G proteins are large complexes, comprised of 3 subunits (α, β, and γ); their size deters long simulations of GPCR-G protein complexes, limiting how many properties researchers can study with *in silico* methods (not to mention that simulation boxes and bilayer sizes required for GPCR-G protein simulations need to be considerably bigger than boxes and bilayers when studying inactive states, which decreases performance of MD engines and increases costs associated with simulations).

Several alternatives have been introduced in the literature. The discovery that camelid nanobodies are capable of stabilising active states of GPCRs has allowed for crystallisation of several GPCRs, critically the μ OR in complex with the morphinan agonist BU-72 and nanobody 39 [55], the κ OR in complex with the morphinan agonist MP1104 (Figure 1.11) and a nanobody [62]. Re-engineering of the G protein has allowed for the crystallisation of the Adenosine A_{2A} receptor in complex with the agonist NECA and an engineered mini-G_s protein [173]. Including these in MD simulations would allow for a better understanding of the interactions between GPCRs and these constructs, while also allowing for longer simulations (since these constructs are smaller than a full heterotrimeric G protein).

In a recent study, the structure of the μ OR in complex with the agonist DAMGO and a G_i protein was simulated for ≈ 1 µs. Critically, this study reported on simulations without the co-crystallised G protein; the authors elected to, instead, prevent receptor deactivation by applying harmonic restraints to key atoms in the G protein binding cavity [21]. While this is an incredibly attractive option, particularly with respect to cost and performance, and will no doubt be more extensively used in future research, the absence of a G protein may present some limitations in the study of information cascading from the ligand binding site to critical G protein sites (such as the GDP catalyst site).

3.3.2 Outlook

While conventional MD is a tool adept at exploring the dynamics of proteins, one of its most critical limitations is difficulty in capturing high energy conformational changes. This is particularly crucial in this work, since developing an understanding of how changes in ligands affect the activity of GPCRs requires being able to capture said conformational changes.

In the case of GPCRs, the main change that is problematic to address is the outward movement of TM6 away from TM3. While research groups, particularly Kohlhoff and colleagues, have shown that conventional MD is indeed capable of briefly

capturing active-like states (reviewed in the following Chapter), it is only possible while employing long-time scale MD, alongside very dedicated software and hardware.

Alternative methods that rely on enhanced MD have been used moderate success (reviewed in the following Chapter), but still require high amounts of computational time to derive any data. In the following Chapter, a novel method is introduced that attempts to explore the main conformational changes that occur in GPCRs in more accessible timescales.

Chapter 4

Development and implementation of the Fraction of Native Contacts Collective Variable

4.1 Introduction

MD has become one of the key tools of choice for GPCR activation mechanisms studies [85]. Hardware and software advances, in particular, have allowed the development of a deeper understanding of how particular ligands modulate the energetic landscape of GPCRs (the reader is referred to [85] for a more comprehensive list of advancements).

The recent technological boom in MD hardware/software has allowed for exploration of GPCR function using conventional MD techniques in the microsecond timescale.

4.1.1 Exploration of GPCR activation landscape

4.1.1.1 Landmark conventional MD studies

One of the first reported studies of microsecond long simulations of GPCRs was published in 2011. The authors used (then) recently crystallised structures of the β_2 adrenoceptor to perform conventional MD simulations [174]. The powerful combination of the special-purpose computer, Anton [84], and the specialised software package, Desmond [83], was used to perform production runs.

In this study, the authors reported on a 30 μs simulation of the aforementioned receptor in complex with the agonist BI-167107, and a shorter 15 μs simulation of the same receptor in complex with the antagonist/partial inverse agonist carazolol. In their simulations, the authors noted that the absence of an intracellular binding partner apparently causes an active state receptor to transition to inactive-like states,

based on the C_α RMSD of TM5 and TM6. This phenomenon is present regardless of whether or not an agonist (BI-167107, in this case) is present in the orthosteric binding pocket [174].

Later in the same year, Dror and colleagues reported on an update to the study mentioned above. Again using the powerful combination of Anton and Desmond, the authors reported on 92 simulations totalling ≈656 μs of simulation time [26]. These simulations were started from active states of the β_2 adrenoceptor, and captured several intermediate states along the deactivation pathway. The results of this study shed some light into the functioning of several key residues present in conserved motifs:

- The deactivation pathway may start with the transition of $Tyr^{7.53}$ (key member of the NPxxY motif; see Chapter 1 for more information). This residue interposes TM3 and TM6 in the active state; upon switching to it's inactive conformation, the space left behind is flooded with water molecules. After this, TM6 moves closer to TM3, completing the deactivation process.

- The latter 2 residues of the PIF motif, $Ile^{3.40}$ and $Phe^{6.44}$, may be involved in loosely coupling the orthosteric binding site to the G protein binding site. In the inactive state, $Phe^{6.44}$ points towards TM3; in the active state, the same residue points away from TM3, allowing for the separation of the two helices.

The authors concluded that, from analysis of their results, the activation mechanism may begin in the G protein binding pocket, and only then are changes propagated to the orthosteric binding site, increasing it's affinity for agonists. Agonists need not stabilise a fully activated state of the receptor; it is important that they merely shift the equilibrium of orthosteric pocket states, and this shift then cascades, through the coupling region, to the G protein binding site, allowing it to adopt an intermediate state more capable of binding G proteins [26].

It is important to note that the ICL3 of the receptor was not reconstructed in either study. As a result, any TM5 and TM6 motions resulting from loop fluctuations were not studied. This may have caused either study group to overlook more intermediate conformations (druggable or otherwise) visited by the β_2 adrenoceptor during deactivation.

4.1.1.2 Limitations of conventional MD

To date, these remain some of the largest conventional MD simulations performed on protein-ligand complexes. Though they have provided invaluable insights into

the activity of GPCRs, there are significant drawbacks associated with them. Critically, there is a significant conceptual drawback, which is further compounded by a pragmatic one.

While conventional MD simulations can reliably explore low-energy events (breaking or forming salt bridges, or understanding how a ligand is equilibrated in a binding pocket), it is not the best method for exploring high energy events, or events that occur in "slow" timescales separated by several $k_B T$ (where k_B is the Boltzmann constant, and T is the absolute temperature) [34], such as the transition of the G protein binding site from an inactive to an active state [121].

The recent works from Kohlhoff and colleagues, while ground-breaking in its methods, results, and conclusions, is a perfect example of this drawback. In their seminal work, Kohlhoff and colleagues performed several simulations of the β_2 adrenoceptor using the Google Exacycle, "a cloud-computing initiative that provides an interface for running compute jobs directly on Google's production infrastructure (...)" , and combined them using Markov state models, to create a more complete picture of the β_2 adrenoceptor activation mechanism [123].

Having simulated a total of 2.15 ms, there were relatively short instances (in the μs time-scale) during the trajectories where active-like states were observed. This perfectly exemplifies the unsuitability of conventional MD as a tool for high energy, or rare, event exploration.

Moreover, this issue is only compounded further by another problem, one related to cost and availability of computational resources.

Long timescale MD simulations have proven indispensable in the study of GPCR structure and function (for more, the reader is referred to the works by Dror and colleagues, above; the works by Kohlhoff, above; the works by Neale and colleagues [30]; and the works by Schneider and colleagues [96], all of which are extensively reviewed in [85] and the latter of which is briefly reviewed in Chapter 4).

Nevertheless, running longer MD simulations accrues significant computational costs. While this problem has been somewhat diminished by integration of CUDA into MD software packages [82], which allows for calculations to be offset to GPUs, resulting in much greater performances compared to simulations run on parallel CPU environments, not all research groups have access to commodity clusters (and specialised hardware, such as Anton) on which to run their simulations.

As a consequence of both of these issues, it remains critical to find newer ways to explore high energy events.

4.1.2 Landmark enhanced MD studies

As previously mentioned, conventional MD simulations and force fields were not designed for exploration of high energy events. As a result, exploring these states requires specialised software packages and analysis methods.

Currently, the most common philosophies for enhanced MD involve either adding a biasing potential to the entire system in study, or creating a specific reaction coordinate (also referred to as a collective variable, or CV) that describes the conformational change of interest and applying an external bias to explore different values of this CV.

The former is the principle behind accelerated MD [175] and all it's variants (dual-boost MD, replica exchange accelerated MD, selective accelerated MD, rotatable dihedral accelerated MD, and Gaussian accelerated MD, among others [176]). This method allows for unconstrained exploration of free energy landscape (in that it does not require *a priori* knowledge of said landscape). This method is currently implemented, and is widely used, in the AMBER [115] and NAMD [177] simulation software packages.

Accelerated MD can be applied to a range of research questions. In 2015, Kappel and colleagues reported on the use of dual-boost accelerated MD to explore the binding of 3 ligands (tiotropium, arecoline, and acetylcholine) to the M_3 muscarinic receptor [178]. In this study, researcher reported on a method that speeds up ligand binding from bulk, compared to similar studies where bulk ligand binding in performed using conventional MD [25]. They also identified some critical residues that mediate metastable states of acetylcholine and arecoline binding (such as $Tyr^{3.33}$, $Tyr^{6.51}$, and $Tyr^{7.39}$) [178].

In 2014, Miao and colleagues applied dual-boost accelerated MD to explore the free energy landscape of the M_2 muscarinic receptor. In this study, structures of the receptor in both the apo state and antagonist-bound state were boosted and analysed with respect to the "ionic lock" distance, the 'toggle switch' χ_1 and χ_2 dihedral angles (see section 1.1.2.1 in page 6), and the $Tyr^{5.58}$-$Tyr^{7.53}$ distance. Their results suggest a strong correlation between the breaking of the "ionic lock" and formation of a hydrogen bond between the side-chains of $Tyr^{5.58}$ and $Tyr^{7.53}$ [179]. Interestingly, the formation of the latter interaction may arise as a result of differential unwinding of the intracellular segment of TM5, and subsequent interaction with $Arg^{3.50}$ are then also formed [180].

While accelerated MD methods offer the advantage of applying a boost potential to any part of the system, it is difficult to estimate equilibrium and kinetic properties. As determination of such properties depends on exponential reweighting, calculations

are prone to statistical errors. It also requires fine tuning of acceleration parameters such that transitional states are not affected by the boost potential [176].

A second strategy for enhanced sampling involved selection of one or two motions of interest of the system in study, and applying a biasing potential uniquely to those motions.

These motions are usually defined by a reaction coordinate, or a collective variable, that most critically defines the path transition between two conformational states of the system under study. In some cases, this could be a distance between two points in space; in others, the variation in side chain dihedrals for a specific residue.

One of the most common software packages for biasing these collective variables is PLUMED [124]. This software package serves two main purposes: it interfaces with other molecular dynamics software (GROMACS [116], AMBER [115], NAMD [177], to name a few) in order to bias the dynamics of the system; and it is used for post-processing of simulations to recover free energy landscapes (e.g. from metadynamics simulations) [124].

4.1.3 Motivation

As previously mentioned, a CV should represent the slowest process in the system of interest, be it a chemical reaction or a system-wide conformational change. For GPCRs, this is likely the movement of TM6 away from the receptor core, and, as a result, it became important to develop a CV that accurately described this process. It was also important to develop a CV that allowed for the exploration of the conformational landscape of GPCRs in a sensible timescale.

Introduced in section 2.2.2, a novel CV was developed to induce artificial transitions between inactive and active states of GPCRs. A series of steered MD simulations were designed to test whether this new CV, termed FNC, was sufficient to induce such transitions (see Table 2.3). Following these, a set of umbrella sampling simulations were performed, applying a biasing potential to the FNC CV (and to the ALPHARMSD CV, see section 4.2.4 for more). In this chapter, the μ OR was used as a test system; the mixed μ OR agonist/δ OR antagonist DIPP-NH_2 (Figure 1.10 in page 21) was used to test the differences in FNC energetic landscape.

The remainder of this chapter will focus on the results of the above simulations, paying particular attention to the changes in the FNC CV, and in changes to key features in the GPCR topology. The data here shows that the biasing FNC CV often results in loss of secondary structure in TM6 (and, to a lesser extent, in TM5), requiring an additional helicity CV during both steered MD and umbrella sampling

simulations. The resulting PMFs suggest that the presence of DIPP-NH_2 stabilises intermediate conformations along the FNC CV path.

4.2 Results

4.2.1 FNC changes in conventional MD

An initial set of simulations were performed to evaluate the changes in the FNC CV under conventional MD. These set out to assess the changes in FNC of the inactive state μ OR with and without an allosteric sodium, and of the active state μ OR with and without an intracellular binding partner (in this case, a mimetic nanobody). Systems were prepared using the protocol outlined in Chapter 2, section 2.2.4, simulated for 100 ns, in quintuplicate. Figure 4.1 shows the changes of the FNC CV in the 5 separate replicas.

Interestingly, the FNC CV changes drastically from the initial value (-1 for inactive states; +1 for active states) at the start of each simulation. In simulations started in the inactive state, the average FNC is ≈-0.6/-0.7 (Figure 4.1, top plots), though in certain cases the FNC value rises to ≈-0.4. There isn't a pronounced difference in the FNC variation based on the presence (Figure 4.1, top left) or absence (Figure 4.1, top right) of an allosteric sodium.

In similar fashion, in simulations started in the active state, the FNC CV instantly degrades, reaching values as low as ≈0.5 in simulations absent of any intracellular binding partner. The average FNC is ≈0.7 (4.1, bottom left).

Contrary to the inactive state simulations, the presence of an intracellular nanobody stabilises the FNC at ≈0.8 (4.1, bottom right). This agrees with previously reported studies confirming that the presence of a nanobody/G Protein is necessary to stabilise more active-like states of GPCRs.

Taken together, these results indicate that FNC values of -1 or +1 (as derived directly from crystallographic structures) are not necessarily indicative of the ensemble average of a particular conformational state, but rather a snapshot constrained by crystallography conditions. For the purposes of steered MD studies, FNC values of -0.7 and 0.7 were deemed sufficient to indicate that a structure was in an inactive-like and active-like state, respectively.

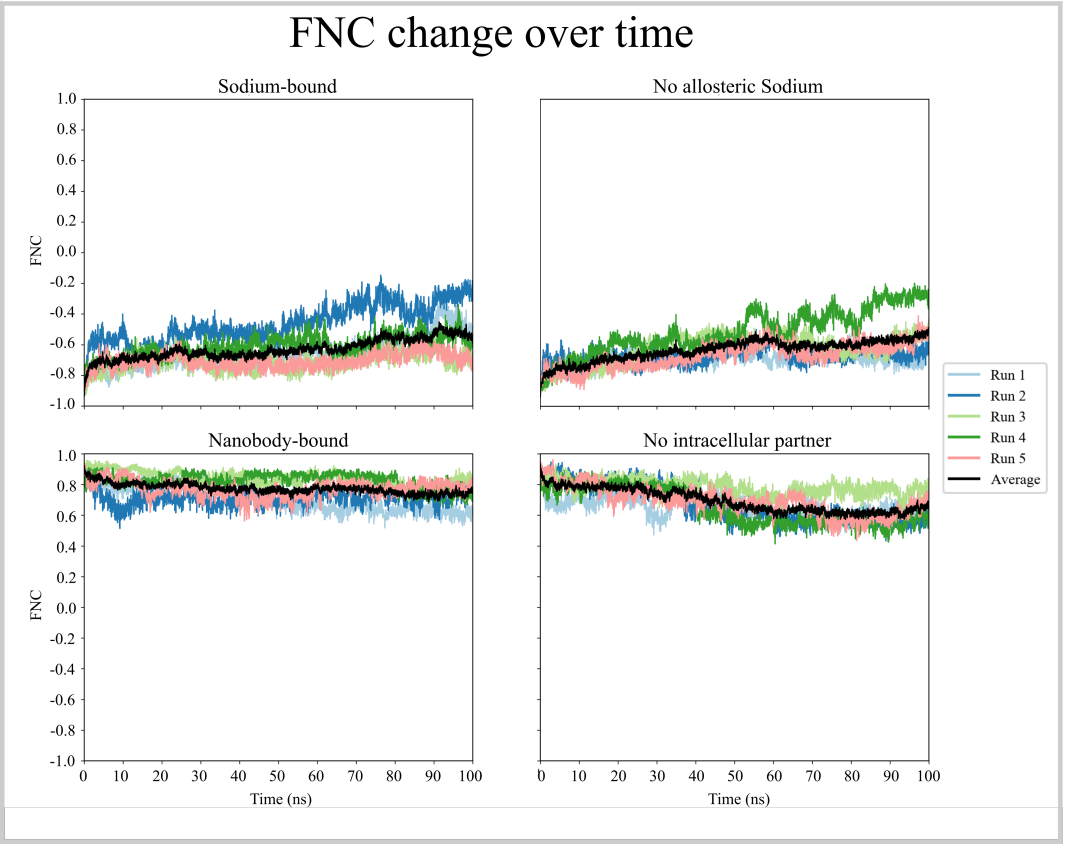


Figure 4.1: FNC variation under conventional MD. Each of the 4 plots represents data collected from quintuplicate 100 ns simulations, the average of which is shown as a black line. The simulation conditions (whether a sodium was present in the allosteric site; whether a nanobody was present in the system) are specified at the top of each plot. In all plots, FNC values of -1.0 and 1.0 corresponds to the contacts determined for the inactive and active crystallographic structures, respectively.

4.2.2 Activation of μOR from FNC biasing

A set of simulations were devised to study the effectiveness of the FNC CV to sample the transition between inactive and active states of the μ OR. In this protocol, a set of 20 ns steered MD simulations were performed, biasing the FNC CV using different force constants (see equation 2.14), which were varied as previously detailed. Figure 4.2 provides a summary of the results.

Increasing the force constant resulted in active states more closely resembling the crystallographically determined state (PDB accession 5C1M [55]), with higher force constants (over 5000 kJ/mol) resulting in higher final FNC values.

One transition was attempted using a force constant of 1000 kJ/mol, but it was deemed insufficient based on the final FNC value (≈0.4; see Figure C.1, top plot, in

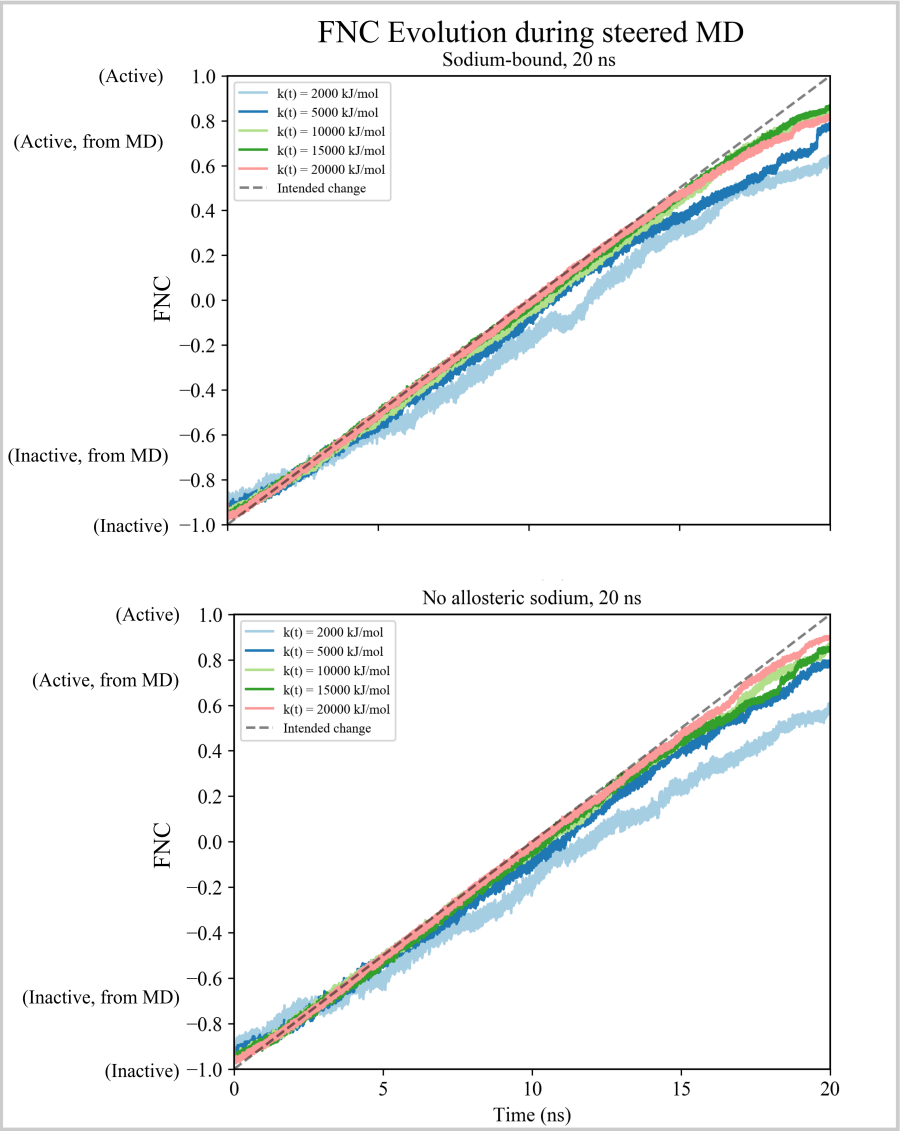


Figure 4.2: FNC changes in both sodium-bound (top) and sodium-deficient (bottom) μ OR. The data shown here is an average of 3 independent simulations (per force constant). In all simulations, the FNC CV was varied from -1 to +1.

page 157). As detailed in the previous section, the requirement was set to ≈ 0.7 (in the absence of a mimetic nanobody; see Figure 4.1).

A similar transition was attempted with the same force constant, in the absence of sodium. Despite these conditions facilitating transitions to active-like states, the transitions were deemed insufficient for similar reasons (Figure C.1, bottom plot, in page 157).

Given that the required FNC value is achieved using a force constant of 5000 kJ/mol, subsequent simulations were performed using this value for the FNC CV.

Sampling G protein binding cavity opening The use of native contacts and contact maps was introduced in this work to allow sampling of the slowest process in the activation transition, i.e. the outward movement of TM6 away from the core of the receptor. The most commonly used metric to describe this process is the distance between TM3 and TM6. In this work, the minimum distance between Arg146$^{3.50}$ and Leu256$^{6.30}$ was used for this purpose. The reader is reminded that the traditional "ionic lock" present in other GPCRs (an ionic bond between Arg$^{3.50}$ and Glu$^{6.30}$) is not present in opioid receptors. However, for simplicity, subsequent figures will still refer to this interaction as the "ionic lock" distance (see also Chapter 1, section 1.1.2.1, page 6).

The application of a force constant equal to or greater than 2000 kJ/mol is sufficient to allow for an opening of the G protein binding cavity (see Figure 4.3, felt plot). Higher force constant simulations more readily achieve an open-like state; employing force constants of 15000 kJ/mol ($\approx$3600 kcal/mol) and 20000 kJ/mol ($\approx$4800 kcal/mol) achieved an active-like "ionic lock" distance within 10 ns (on average; see Figure 4.3). Simulations performed using weaker force constants, ranging from 2000 kJ/mol ($\approx$480 kcal/mol) to 5000 kJ/mol ($\approx$1200 kcal/mol) to 10000 kJ/mol ($\approx$2400 kcal/mol), typically reached the same values within $\approx$15 ns. Force constants below 2000 kJ/mol do not elicit the same response (see Figure C.2, top plot, page 158).

Applying a force constant to the FNC CV, in the absence of an allosteric sodium, results in a similar trend (Figure 4.3, right plot). A force constant of 1000 kJ/mol is also insufficient to open the G protein binding cavity (see Figure C.2, bottom plot, page 158).

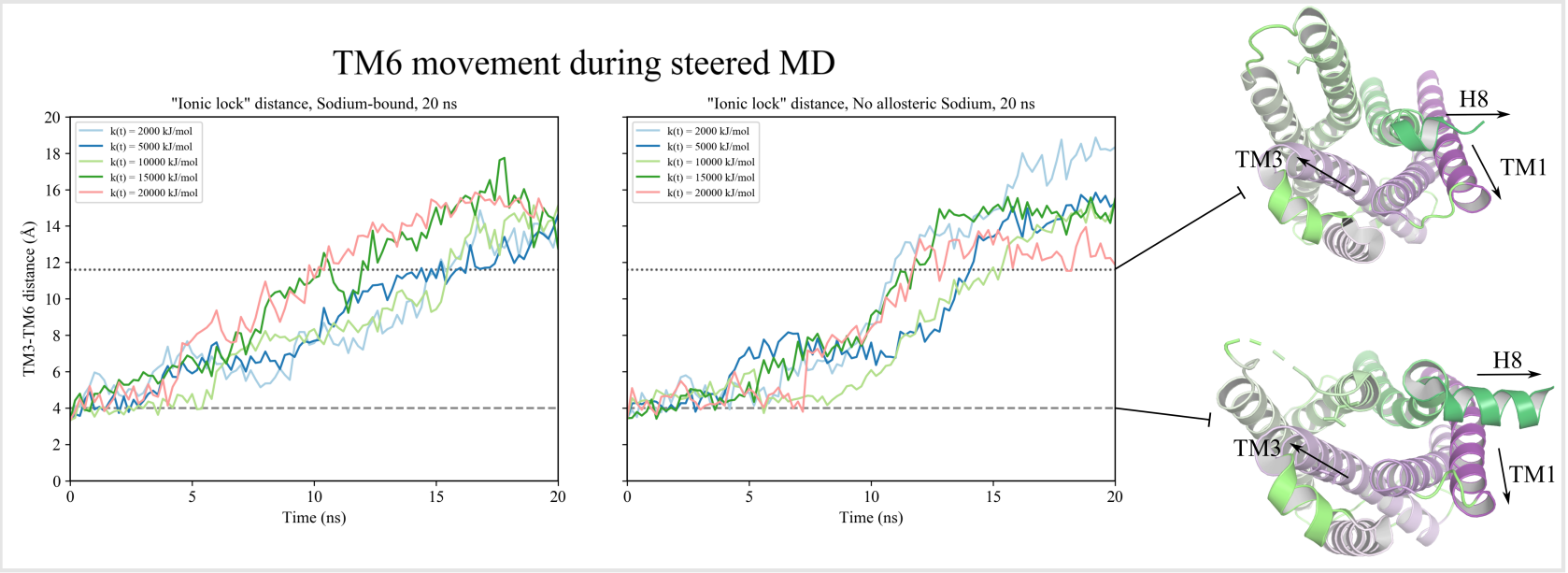


Figure 4.3: Biasing the FNC CV with force constants equal to, or above, 2000 kJ/mol results in an increase in the "ionic lock" distance (a metric for the distance between TM3 and TM6). A higher force constant more readily promotes an active-like state for the G protein binding cavity. The dashed grey line represents the inactive state crystal separation (PDB accession 4DKL [61]) and the dotted grey line represents the corresponding active state crystal distance (PDB accession 5C1M [55]). Colour scheme of the accompanying images is the same as 1.2 (page 4). Data shown here is the average of 3 independent simulations for each force constant, for both sodium-bound (left) and sodium-deficient (right) systems.

Biasing potentials cause loss of secondary structure The application of force constants, while allowing opening of the G protein binding cavity, also results in a loss of secondary structure. This loss is concentrated in the intracellular end of TM5 and TM6, the latter shown in Figure 4.4.

Sodium coordination during steered MD During steered MD simulations, the allosteric sodium exits the core of the receptor via the orthosteric binding site. As demonstrated in previously reported simulations, the sodium transitions from a state where it is equilibrated by the canonical allosteric site residues (Asp95$^{2.50}$, Asn131$^{3.35}$, Ser135$^{3.39}$; see also Figure 1.7, page 15), to a state where it is equilibrated by Asp128$^{3.32}$. Interestingly, Gln105$^{2.60}$, and to a lesser extent, Tyr308$^{7.43}$ are also involved in this equilibration (Figure 4.5).

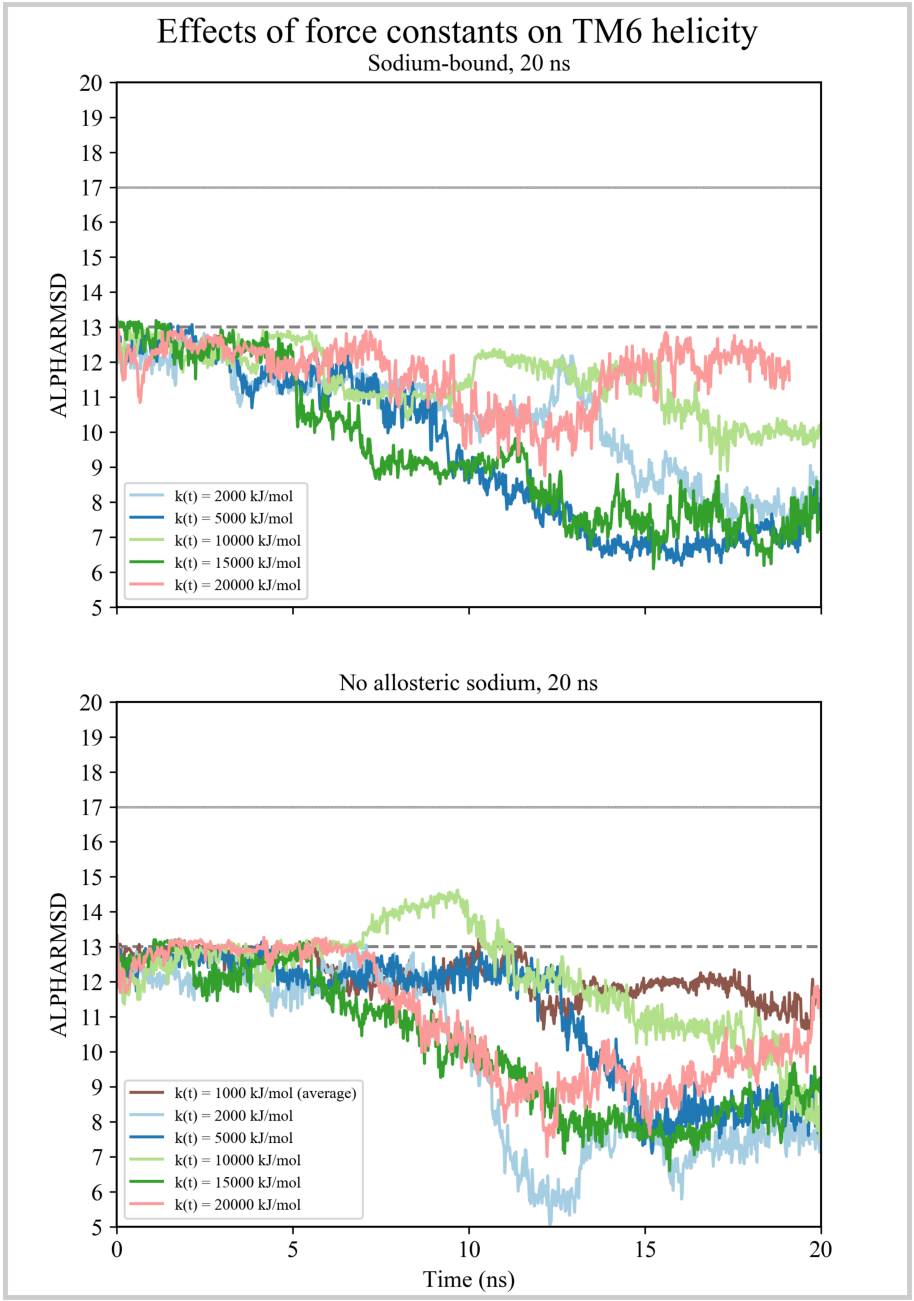


Figure 4.4: Changes in TM6 helicity as a function of the force constant applied. Independent of the force constant value, there is a marked decrease in the helicity of TM6, as measured by PLUMED, in all residues spanning positions 6.23 and 6.44 (Ser268 and Phe289). The dashed grey line corresponds to the ALPHARMSD of the inactive state receptor; the grey line corresponds to the ALPHARMSD of the active state receptor.

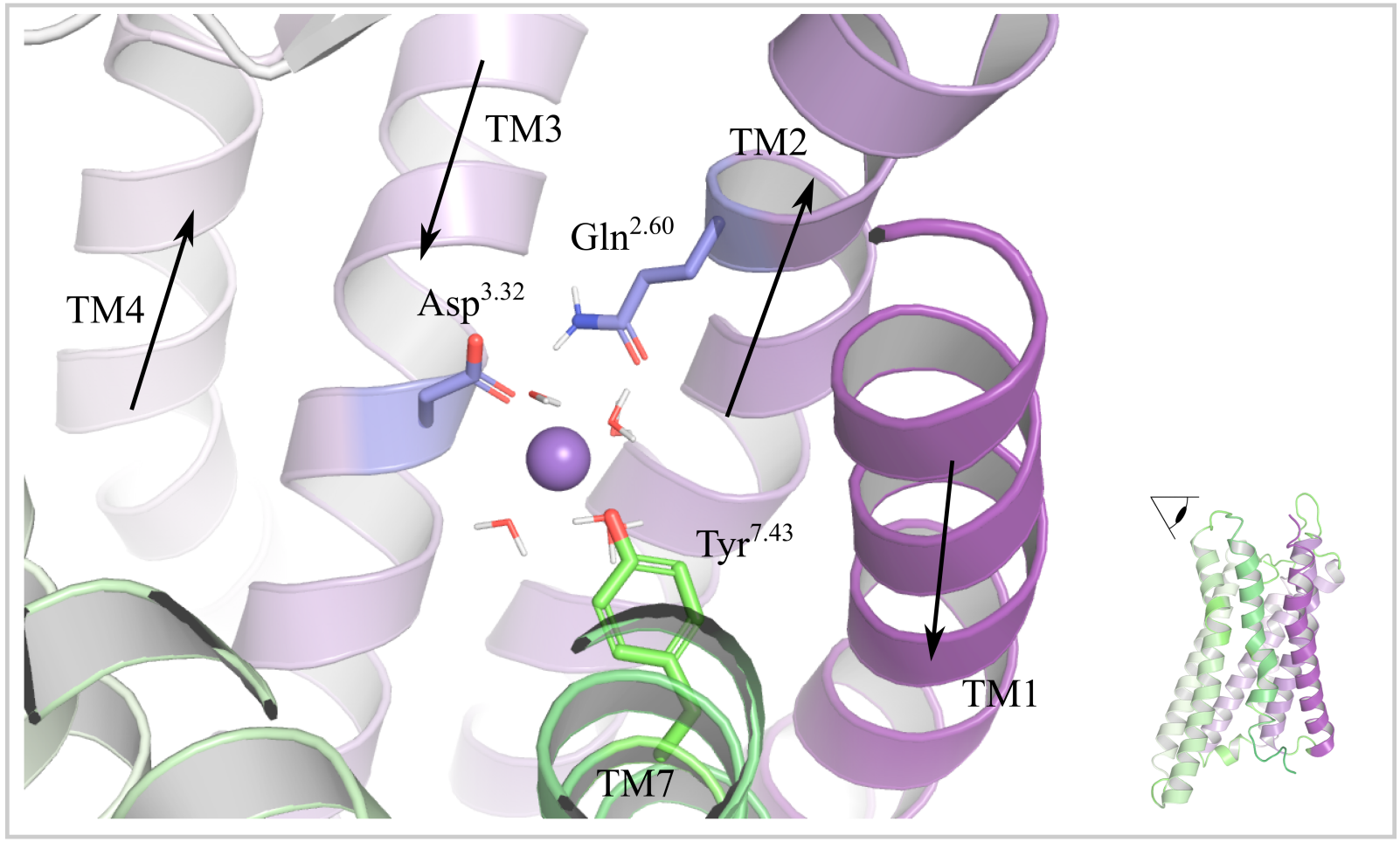


Figure 4.5: Coordination of the allosteric sodium by Asp128$^{3.32}$ has been previously observed [181]. However, Gln105$^{2.60}$ has not been shown to be involved in this equilibration. Colour scheme corresponds to that of Figure 1.2, page 4.

4.2.3 Deactivation of μOR from FNC biasing

Next, a set of simulations were performed on the active state of the μ OR. These simulations were designed to mimic the deactivation of the receptor (i.e. transitioning from active-like states, capable of binding an intracellular partner, to inactive-like states, incapable of same). Since this process typically happens in the order of µs [26, 174], a biasing potential was applied to the FNC CV to attempt deactivation in accessible timescales. Similar to the protocol described in section 4.2.2, each steered MD simulation was performed for 20 ns, employing time-independent force constants. Figure 4.6 summarises the changes in FNC (top), "ionic lock" distance (central) and ALPHARMSD (bottom) of all steered MD simulations.

4.2.3.1 Structural changes in "deactivating" simulations

The application of force constants to promote deactivation tells a contradictory story, compared to that told by the "activation" simulations.

An analysis of the FNC timeseries reveals a similar trend to it's reverse process; a higher force constant results is more inactive-like states (reflected in FNC values closer to -1; Figure 4.6, top). The two weaker force constants (1000 and 2000 kJ/mol) are not sufficient to cause transitions to inactive like states.

However, a timeseries of the "ionic lock" distance (a proxy for the G protein binding cavity; Figure 4.6, central) reveals that all force constants are capable of transitioning the GPCR to inactive-like states (since it would be incapable of binding any intracellular binding partner).

Biasing potentials cause loss of secondary structure Applying a deactivating biasing potential to the FNC CV, results in similar degradation in the secondary structure of the receptor. Figure 4.6 (bottom) shows the timeseries of this change with varying force constants. As with the activating steered MD simulations, the degradation of the secondary structure is usually concentrated in the intracellular end of TM6. Figure 4.7 shows an example of this phenomenon.

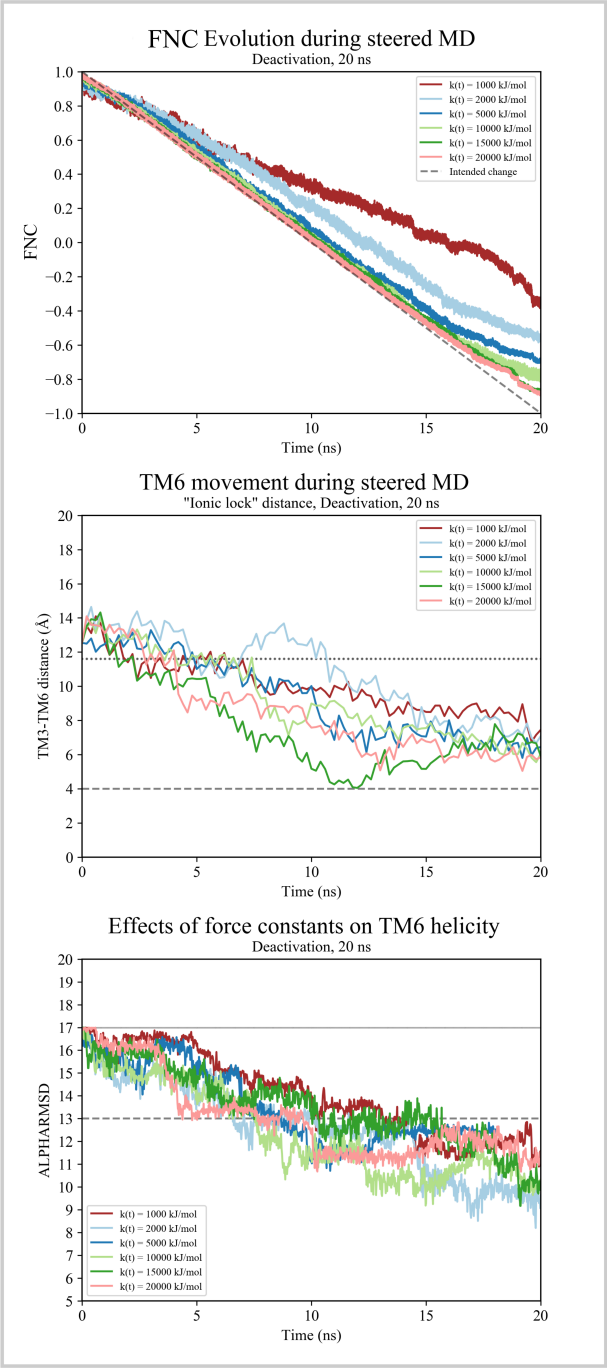


Figure 4.6: Top: FNC changes in deactivating μ OR. Middle: Decrease in G protein binding cavity distance (as measured by "ionic lock" distance). Bottom: Degradation of helicity of TM6 during steered MD simulations. As per Figure 4.4, the ALPHARMSD was calculated by PLUMED measuring all residues spanning positions 6.23 and 6.44. The decrease in helicity far surpasses the expected final value (13.0), indicating more extensive secondary structure loss in TM6. The data shown here is an average of 3 independent simulations (per force constant). In all simulations, the FNC CV was varied from +1 to -1, representing an active state and an inactive state, respectively (grey dashed line).

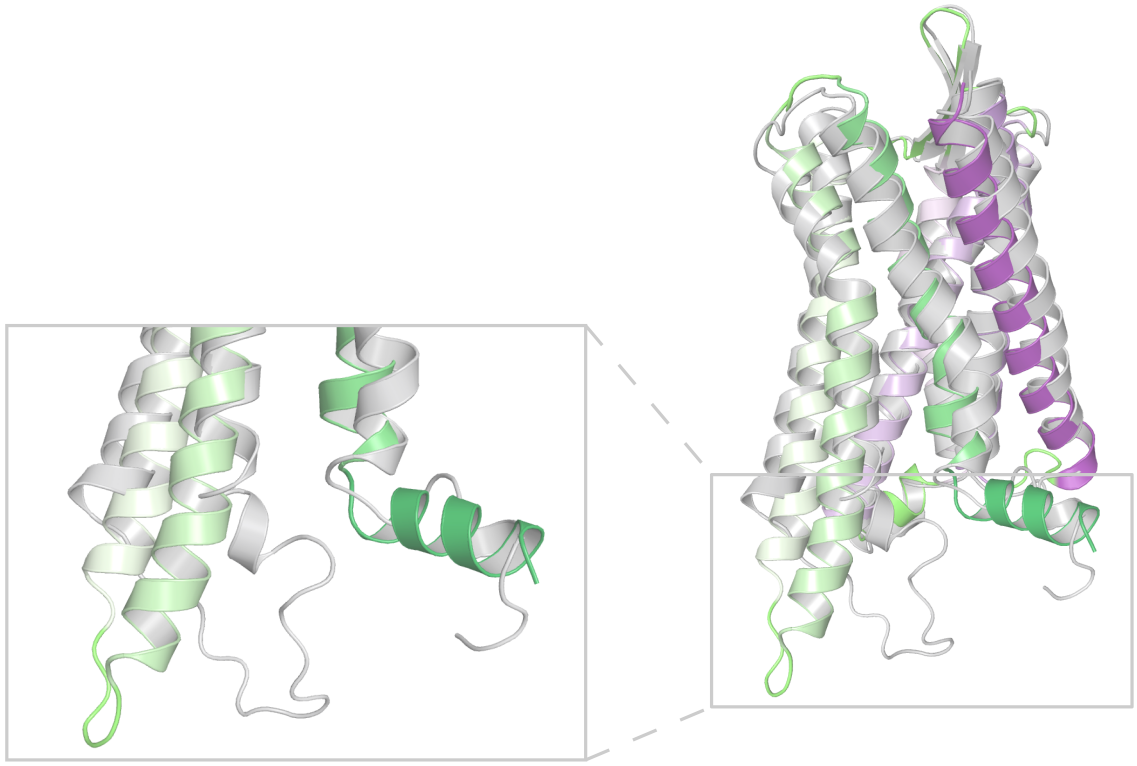


Figure 4.7: Superimposition of the final frame of a steered MD simulation (silver cartoons) and the active state of the *mus musculus* μOR (coloured cartoon; PDB accession 5C1M), showing deformities in the intracellular end of TM6. In the zoomed in box, helices other than TM5-7 and H8 have been obscured for clarity. In the ideal scenario, the loss of helicity would be restricted to one turn. The colour scheme for the active state structure follows that of Figure 1.2 on page 4.

4.2.4 Combination of FNC and ALPHARMSD

As there is no evidence of loss of helicity from structures so far, a set of simulations were performed with an additional CV that maintain the helicity of TM6. This CV is known as ALPHARMSD (please see 2.2.3). Table 2.4 details the simulations carried for this section.

4.2.4.1 Timeseries of supported activation simulations

The addition of the support CV, ALPHARMSD, allows the steered MD simulations to reach higher values of FNC (compared to similarly biased simulations without helicity correction). The correction added to the helicity of TM6 also reduces perturbations to the helicity of neighbouring TM5 (Figure 4.8).

The steered MD simulations of the AR (biased from ≈13 to ≈17) show that the changes in this CV follows the required change (contrasting the FNC changes), only occasionally falling short of the final value of 17 (see Figure 4.9).

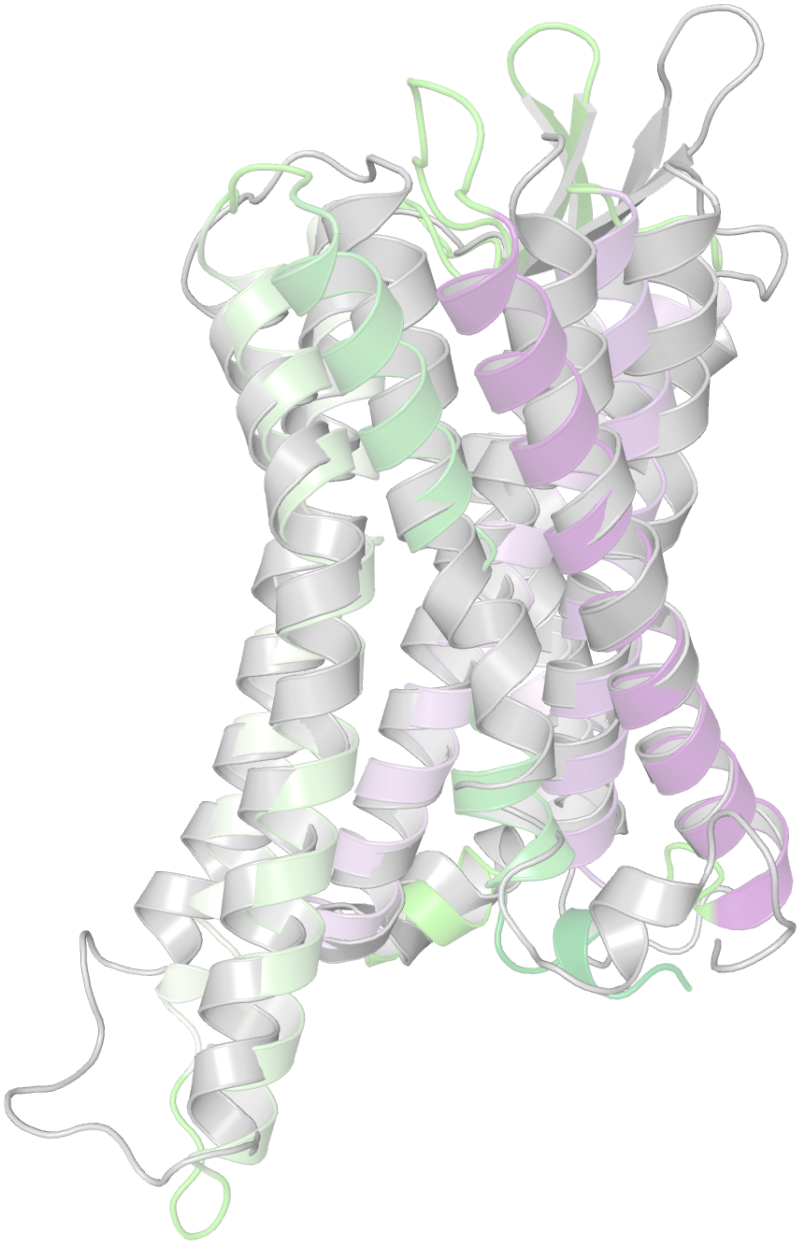


Figure 4.8: Superimposition of the final frame of a steered MD simulation (silver cartoon) and the active state of the *mus musculus* μ OR (coloured cartoon; PDB accession 5C1M) showing how the ALPHARMSD CV corrects unfolding of TM6. It also stabilises the helicity of TM5. The FNC was biased using a force constant of 5000 kJ/mol, and the ALPHARMSD was biased using a force constant of 2000 kJ/mol. The colour scheme for the active state structure follows that of Figure 1.2 on page 4.

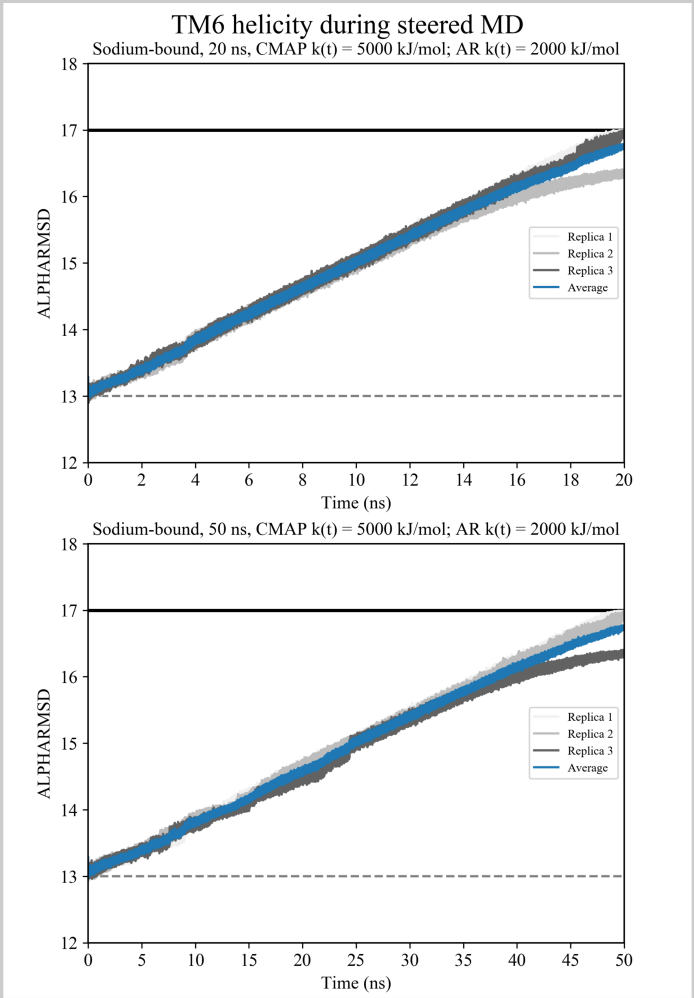


Figure 4.9: ALPHARMSD evolution in select steered MD simulations. Top plot: Change in ALPHARMSD in 3 replicates, 20 ns each. Bottom plot: Same as Box A, each replicate run for 50 ns. In both cases, the blue line corresponds to the average of the 3 replicates.

4.2.4.2 Timeseries of supported deactivation simulations

The reverse simulations (starting in the active state and forcing a transition to inactive-like states) can also be supported by the ALPHARMSD CV, by forcing the unfolding of TM6 (from 17 to ≈13, corresponding to one helical turn loss). Figure 4.10 summarises the results of certain combinations of FNC and ALPHARMSD force constants.

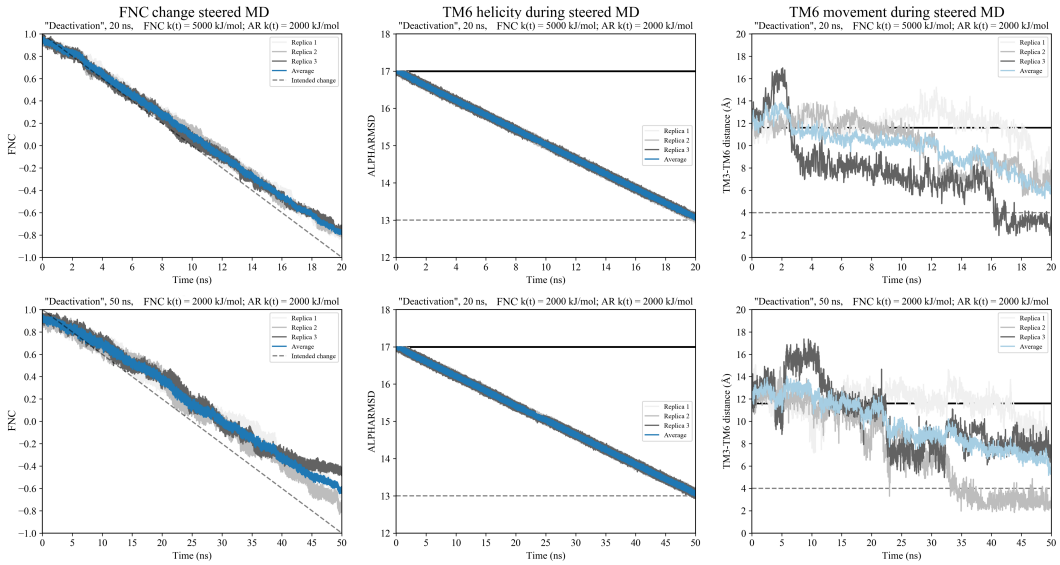


Figure 4.10: Summary of changes in "deactivation" simulations. Left to right: FNC change during simulation; ALPHARMSD change during simulation, dashed and filled line correspond to inactive and active state values, respectively; "Ionic lock" change during simulation, dashed and filled line correspond to inactive and active state values, respectively. Top plots: 20 ns steered MD simulations. Bottom plots: 50 ns simulations.

4.2.5 Umbrella Sampling of FNC CV

Steered MD simulations, similar to targeted MD and biased MD, provide an excellent tool to artificially transition proteins between multiple conformational states. However, different simulations starting from the same structure, pulling using the same force constant, result in different energetic profiles.

The energetic profiles do not converge, even in triplicate (Figure 4.11). There are two ways to overcome this barrier: one can either perform more steered MD simulations, in both the activation and deactivation pathways, until the energetic profiles converge, known as Jarzynski sampling [182]; or to perform umbrella sampling simulations (the basics of which has been described in Chapter 2).

With Jarzynski sampling, it is possible to recover the "equilibrium information from the ensemble of nonequilibrium (finite-time) measurements" [182]. Interestingly, the recovered equilibrium properties are independent of the path between the two states. However, in the interest of finding a method that can be performed in a short amount of time, Jarzynski sampling may not appropriate. One of the main drawbacks of this method is the need for a large number of samples, as equilibrium properties converge as the number of samples tends to infinity.

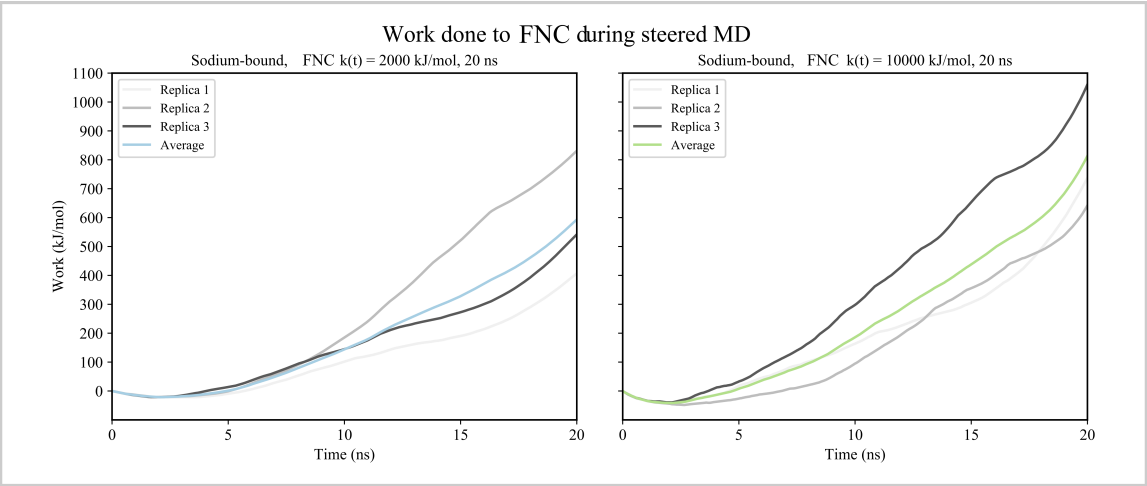


Figure 4.11: No pattern emerges in the work done to the system during steered MD simulations, other than the common increase towards higher FNC values. Left plot: simulations using a force constant of 2000 kJ/mol; right plot: simulations using a force constant of 10000 kJ/mol.

As a result, to recover the underlying free energy surface of GPCR conformational change, umbrella sampling simulations were performed, which instead discretise the CV path into a number of equally spaced windows. Using the combination of the FNC CV and the helicity enforcing ALPHARMSD, a set of windows were generated at various points along the FNC CV line. Coordinates were generated using the steered MD simulations detailed in Table 2.4 of Chapter 2.

From this point, 5 independent 10 ns simulations were performed, while applying a 2000 kJ/mol harmonic restraint to both CVs (equation 2.16). Analysis of trajectories was performed using the MDAnalysis python library, and the resulting ensembles were reweighted using the previously mentioned WHAM software [133].

4.2.5.1 Free energy landscape of μ OR activation based on FNC umbrellas

μ OR in the presence of sodium and absence of an agonist Following the protocol detailed above, a set of umbrella sampling simulations were performed on the μ OR in the presence of an allosteric sodium ion. Figure 4.12 (top plot) shows the resulting Potential of Mean Force (PMF).

Surprisingly, the energetic minima falls on a FNC value of approximately -0.45, far different from what was assumed to be the lowest energy value (-1, given that this value was obtained from crystallographically determined contacts; or -0.7/-0.6 from conventional MD simulations, Figure 4.1, top plots).

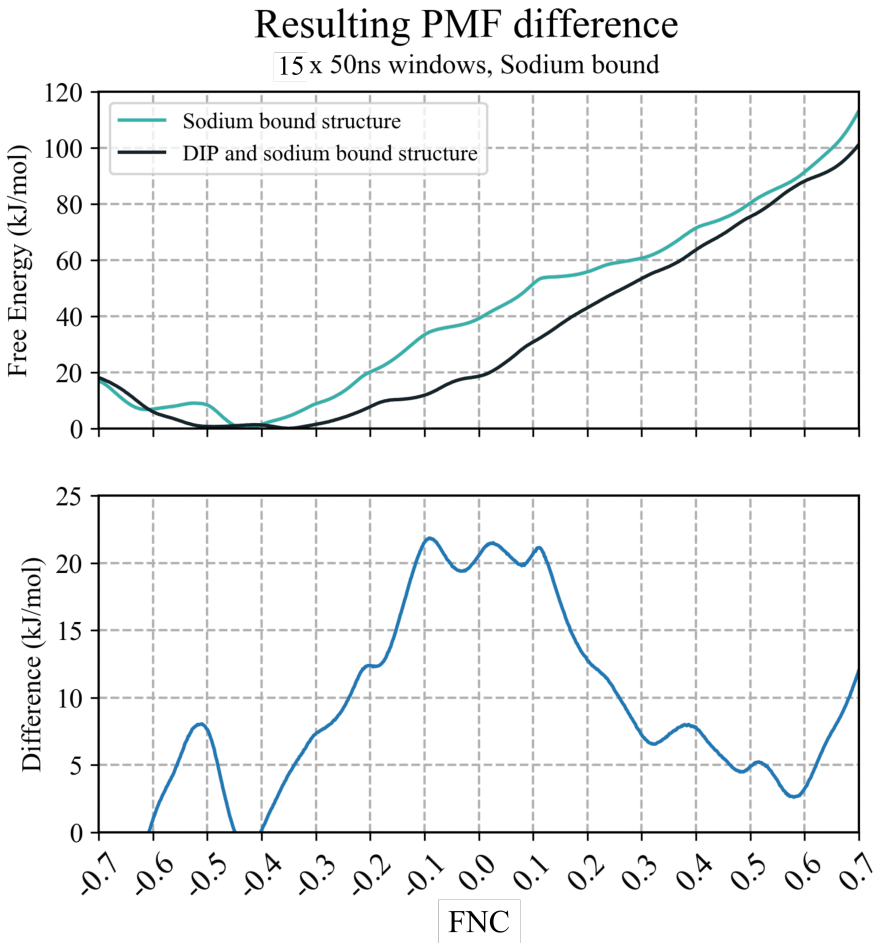


Figure 4.12: Resulting PMFs from umbrella sampling simulations. Top: combined PMFs of the receptor in the presence of an allosteric sodium (blue line) and in the presence of both sodium and DIPP-NH_2. Windows were opened from -0.7 to 0.7, at every 0.1 FNC value. For each window, an umbrella potential was opened with an biasing potential of 2000 kJ/mol. The Grossfield WHAM software was used to generate the PMFs [133]. Bottom: the differences between the free energy profiles lie in the more intermediate states, where the agonist DIPP-NH_2 reduces the transition energy.

From this value onwards, however, there is an increase in energy required to maintain high levels of FNC. There is an abrupt increase in the energetic value from 0.6 to 0.7, the range around which an intracellular binding partner is expected to stabilise. In the absence of any binding partner, the energy required to maintain these active-like states rises to ≈120 kJ/mol (or ≈28.7 kcal/mol; Figure 4.12, top plot).

μ OR in the presence of both sodium and an agonist Next, a set of umbrella sampling simulations were carried on the μ OR, this time in the presence of a sodium

ion in the allosteric binding pocket and the agonist DIPP-NH_2 in the allosteric binding pocket. The resulting PMF can be found in Figure 4.12 (top plot), which also shows the difference between the two PMFs (bottom).

4.3 Discussion

In this chapter, a novel collective variable (CV) was introduced to artificially simulate the activation procedure of GPCRs. Termed FNC, this CV was developed based on Best, Hummer, and Eaton's seminal work on native contacts [125], and the procedure for creating this collective variable is described in Chapter 2.

A series of simulations were designed to test whether transitions from inactive to active states could be performed in accessible timescales (< 100 ns), paying particular attention to the changes in the GPCR core and to changes in the key motifs. All simulations presented in this chapter use the μ OR (inactive and active state coordinated based on PDB ascensions 4DKL [61] and 5C1M [55], respectively).

4.3.1 On FNC degradation in conventional MD

Initial conventional MD simulations show that the FNC CV quickly degrades from -1 to ≈-0.6/-0.7 in the inactive state simulations (Figure 4.1, top plots). It is likely that this is a result of contacts being forced in X-ray crystallography conditions relaxing in a bilayer environment. It is also possible that different membrane compositions will allow for exploration of different FNC values.

A similar degradation is observed for the active state conventional MD simulations (from 1 to ≈0.7, Figure 4.1, bottom right), though this degradation is reduced to ≈0.8 in the presence of the co-crystallised mimetic nanobody (Figure 4.1, bottom left). This disparity is somewhat expected, since studies have shown that maintaining active-like states requires intracellular binding partners; it is possible that the presence of these binding partners forces certain contacts to remain formed.

4.3.2 Helicity loss as a result of steered MD simulations

For all simulations performed biasing only the FNC CV (Table 2.3), a significant percentage of the helicity of TM6 (and to some extent, TM5) was observed (Figure 4.4). This was observed for both “activation” and “deactivation” simulations. In some cases, the resulting helicity dropped ≈50% from the starting state (in 20 ns simulations).

This loss is not expected behaviour. From all structures of GPCRs captured with intracellular binding partners currently published, none exhibit any loss of secondary structure on TM6. In fact, comparisons between the inactive and active states of the μ OR (PDB accessions 4DKL [61], and 5C1M [55] and 6DDF [21], respectively) shows an increase of helicity in this transmembrane helix by one turn (Figure 4.13). This loss of helicity could be explained by various factors, detailed below.

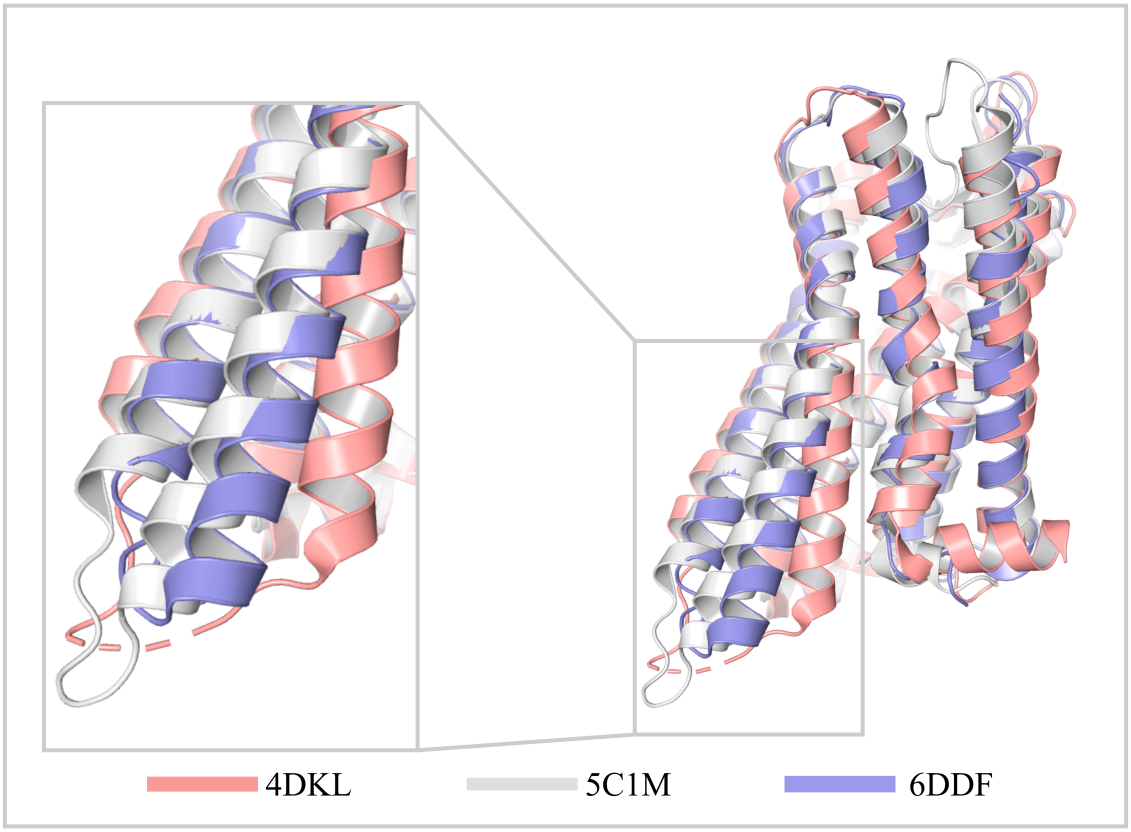


Figure 4.13: Helicity of TM6 increased in the presence of intracellular binding partners (omitted). Both active states (PDB accessions 5C1M [55] and 6DDF [21]) show an increase in helicity compared to the inactive state receptor (PDB accession 4DKL [61]).

4.3.2.1 Tweaking parameters for future steered MD simulations

As detailed in Chapter 2, some cut-offs and restrictions were applied during the creation of the FNCs (all of which can be modified in the final version of the script provided; see Appendix B). The distance cut-off between atoms was set to 6 Å, a scaling factor of 1.8 was applied, and the smoothing parameter (r_0 in equation 2.12)

was set to 0.02. The results presented in this chapter could be refined by finding a balance between these parameters.

It is worth noting that there is a final parameter required by PLUMED, WEIGHT, that defines how important that specific contact is to the overall biasing simulations. Throughout this work, the WEIGHT of each contact was set to the same value. It is, in theory, possible to modify this parameter on a per-contact basis, but this would imply that certain contacts are more important to the overall activation process than others, and modifications to this parameter are difficult to justify *a priori*.

As a result of all contacts sharing the same WEIGHT in the biasing process, the breaking and forming of contacts in steered MD becomes a stochastic process. This could lead to destruction of secondary structure; the short timescale of that these simulations explore might not prove sufficient to then reform such contacts.

As previously explained, maintaining the secondary structure of the TM6 is critical for the binding of G proteins (or equivalent intracellular binding partners). It is possible that performing steered MD simulations with an increasing force constant (starting, for example, at 500 kJ/mol, and steadily increasing the value to 20000 kJ/mol), as opposed to a fixed, high value force constant, might result in simulations without a loss of secondary structure.

4.3.2.2 Erratic motion of ICL3 during simulations

It is also possible that the disruptions in secondary structure are borne out of the erratic motion of ICL3, since this loop, too, is part of the FNCs. In all simulations presented in this chapter, the 10 residue loop was modelled into the structure (see section 2.1.3.4). In previous conventional MD studies, this loop was shown to adopt a complex set of conformations that stabilise inactive states (termed "very inactive states" by the authors) [23]; other studies simply elect to ignore this loop altogether [26, 174].

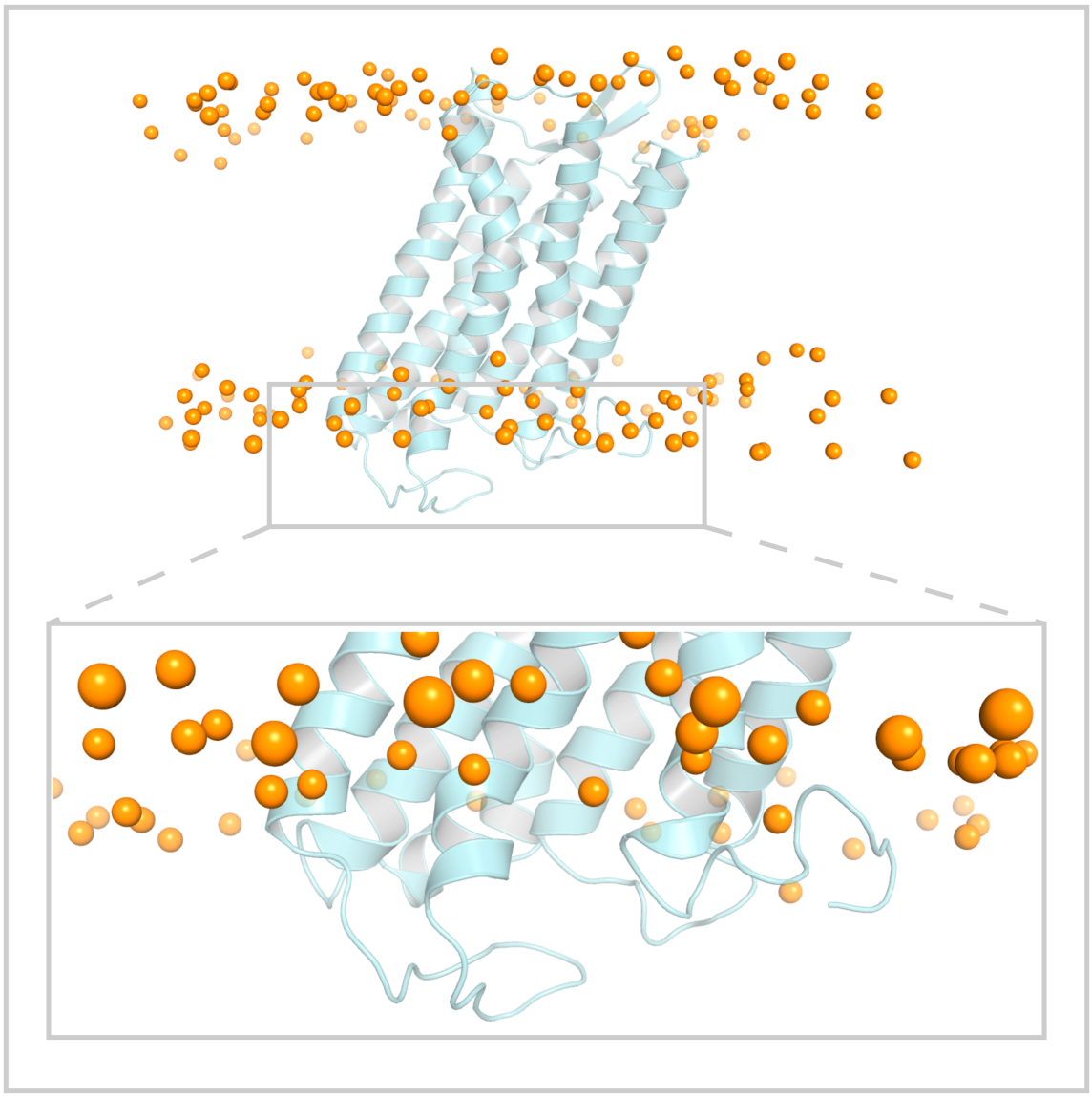


Figure 4.14: During conventional MD, simulations started from the inactive state may sometimes transition into states where ICL3 "shields" the G protein binding cavity.

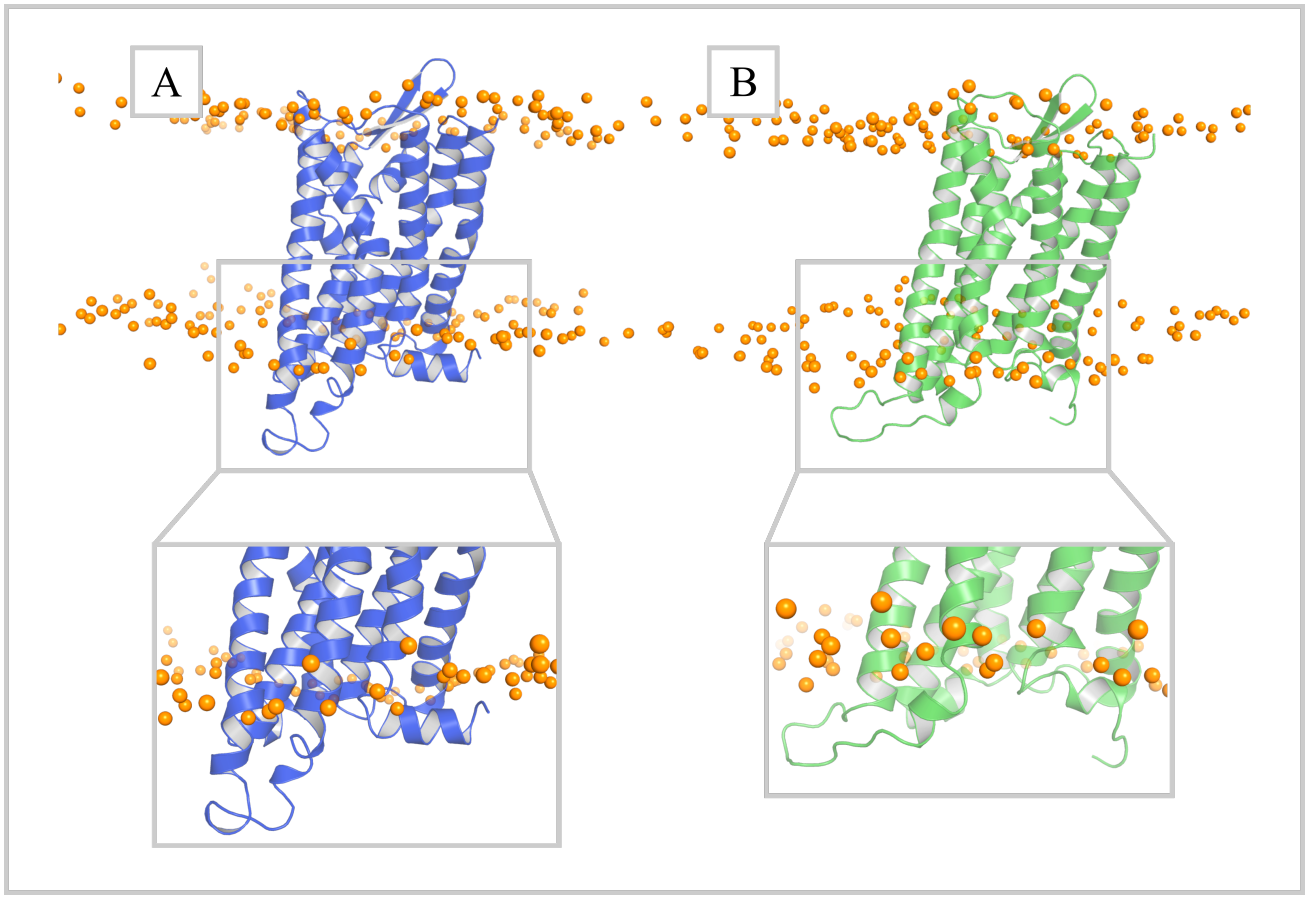


Figure 4.15: When simulations are started from the active state (in the absence of a binding partner), TM6 may begin to unravel. Boxes A and B are examples of how ICL3 may influence this unravelling.

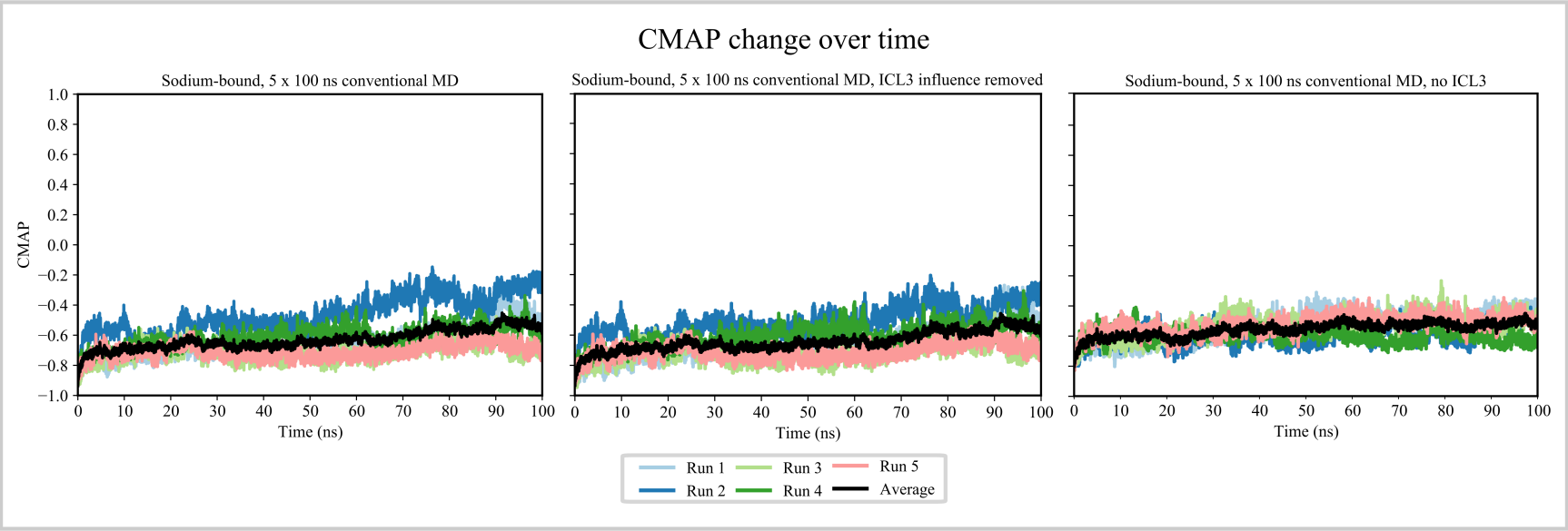


Figure 4.16: FNC dynamics from inactive state simulations. Removing the influence of ICL3 (middle) or removing ICL3 completely (right) does not appear to influence the FNC fluctuation, as compared with the fluctuation seen in a full receptor (left).

A cursory glance at some final frames of conventional MD simulations reveals the conformational flexibility of this loop, where it can adopt states resembling the "very inactive state" proposed by Ozcan and colleagues [23] (Figure 4.14), to states that may be responsible for unfolding of TM6 (Figure 4.15).

Curiously, however, removing the influence of ICL3 from the FNC file doesn't change the dynamics of the FNCs in conventional MD (Figure 4.16, middle). FNC dynamics also do not change when simulations of the inactive state are started without the ICL3 (Figure 4.16, right). Additional steered MD simulations of the transitions would shed more light into this effect.

This would, however, result in changing the dynamics of the activation/deactivation process. To prevent this, in the work here presented, a support CV was used to help maintain (and force) the helicity of TM6.

This CV, termed ALPHARMSD (or AR; see section 2.2.3), was used to force a particular set of residues to adopt an alpha helical conformation. During steered MD simulations performed biasing both CVs, the force constant for the AR was set to either 2000 or 5000 kJ/mol (see Table 2.4). In all simulations, the AR was changed from 13 to 17, corresponding to the number of turns in TM6 in the inactive and active crystal states, respectively. A similar set of simulations were performed for simulations transitioning the system from an active state to inactive-like states; in these simulations, the AR was changed from 17 to 13.

The results of the steered MD simulations can then be used to generate starting coordinates for umbrella sampling simulations at specified values of the CV. In the work presented here, umbrellas were "opened" at FNC values between -0.7 to 0.7, in steps of 0.1, using starting coordinates generated from "activation" simulations. To stabilise the helicity of TM6 (as previously discussed), the AR of this helix was varied linearly with the FNC CV (see section 2.2.5).

4.3.3 On the resulting PMFs

The finalised PMFs shown in Figure 4.12 (top plot) are the result of 50 ns simulations for each window opened (see above). Block analysis indicates that the energetic values converge in 50 ns (see Figure C.3 in Appendix C, page 159). The bottom plot on Figure 4.12 shows the difference in the energetic value between the two PMFs.

Interestingly, these results are indicative of two phenomena.

Firstly, the most stable FNC value for both apo and agonist bound states is in the -0.5 to -0.4 range, different from the expected crystallographically determined -1 state. This supplements the earlier observations that the FNC value rapidly changes from

-1 to -0.8/-0.7 under conventional MD, with 2 of the 5 certain replicate simulations reaching FNC values as high as -0.4 (see Figure 4.1, page 93, top plots).

From this value onwards, the differences between the two energetic profiles becomes more distinguished. The most pronounced difference occurs in intermediate values of the FNC CV. In particular, between the FNC values of -0.1 and 0.1, this difference is ≈20 kJ/mol, while the differences in higher FNC values (indicative of active-like states) are much less marked, reaching ≈12.5 kJ/mol.

This last observation has two consequences. The most critical consequence is that the inclusion of an agonist lowers the barrier of transition to more active-like states, which is consistent with similarly reported observations derived from experimental studies [22, 35]. The second consequence directly parallels previous *in vitro* and *in silico* studies which reported on the necessity of a G protein (or indeed any intracellular binding partner) to stabilise active-like states; the addition of agonists to the system only shifts the population of GPCRs to favour active-like states [26, 123] (the reader is referred to Figure 1 of [35] for a more concrete example of these phenomena).

In his chapter, a novel collective variable was developed to explore the transition between inactive and active states of GPCRs in a time-efficient fashion. Despite it's limitations (some of which were mentioned previously), it could provide a useful tool in exploring how different molecules affects this transition. Additionally, it could provide a useful tool in understanding how small changes in these molecules affect the activation energy landscape.

4.4 Limitations and further work

The FNC CV was developed to explore conformational changes in the system based on contacts unique to two individual states. Given that the structure is being driven from one native state to another, it is theoretically possible to access a more canonical pathway to activation.

However, GPCRs do not exist in 2 defined states. The original model for GPCR activation was developed with just two states: an inactive and an active state, depending on whether a G protein was bound to the receptor. With each new X-ray crystal and cryogenic electon microscopy structure, this model is further refined. A more accurate model for GPCR activation includes inactive states that are stabilised by antagonists (PDB ascensions 4EJ4 [63], 4N6H [13], 4DKL [61], 4RWA/D [64], 4DJH [54], among others) and agonists (PDB ascensions 4AMI [183], 2Y01 [184]); active states stabilised by agonists and arrestins (PDB ascension 4ZWJ [185]); active

states stabilised by agonists and mimetic nanobodies (PDB ascensions 5C1M [55], 6B73 [62]); active states stabilised by agonists and G proteins (PDB ascensions 3SN6 [186], 6DDF [21]) and, more recently, binding to a G protein/arrestin complexes [187] (see also Figure 1.3, page 11).

Consequently, defining FNCs based on two states might result in undersampling (or even no sampling) of potentially critical intermediate states along other pathways. It would be interesting to attempt creating various FNCs with unique contacts based on several conformational states of the same receptor, as opposed to only 2 states. This could lead to a more comprehensive understanding of the energetics at each possible state, with and without stabilising ligands, and could lead to a more profound understanding of the GPCR activity cycle.

Another drawback of this CV revolves around changes in protonation states of key residues in GPCRs.

Studies performed on GPCRs have shown that the conserved $Asp^{2.50}$ residue is in its protonated state in sodium-absent states (shown by Miao and colleagues for the M3 muscarinic receptor [188]; by Ranganathan and colleagues for the β_2 adrenoceptor [189]; and by Vickery and colleagues for the δ OR and for the M2 muscarinic receptor [190]). Dynamically capturing this could prove important in this transition.

4.4.1 Limitations of ALPHARMSD

The addition of a helicity enforcing CV proved to reliably preserve, and alter, the secondary structure of TM6, even at low force constants (2000 kJ/mol). However, the use of this CV imposes 2 limitations.

The first limitation hinges on the definition of the ALPHARMSD. From the PLUMED manual: "Any chain of six contiguous residues in a protein chain can form an alpha helix. This colvar thus generates the set of all possible six residue sections and calculates the RMSD distance between the configuration in which the residues find themselves and an [idealised] alpha helical structure." Therein lies the first limitation: TM6 may need to adopt bent conformations to accommodate binding to intracellular partners. The crystallographic structure of the β_2 adrenoceptor provides an excellent example of this phenomenon (PDB accession 3SN6 [186]). It is unclear whether ALPHARMSD accepts bent conformations as "idealised" helical structures.

The second limitation arises from the first. As exemplified in Figure 4.9, some simulations may result in lower ALPHARMSD values than the required. Upon further inspection, it transpired that this was a result of TM6 folding "in the other direction"

(i.e. the latter residues of TM6 would fold in the opposite orientation; see Figure 4.17). However, from the perspective of the ALPHARMSD algorithm, two distinct "idealised" helices exist.

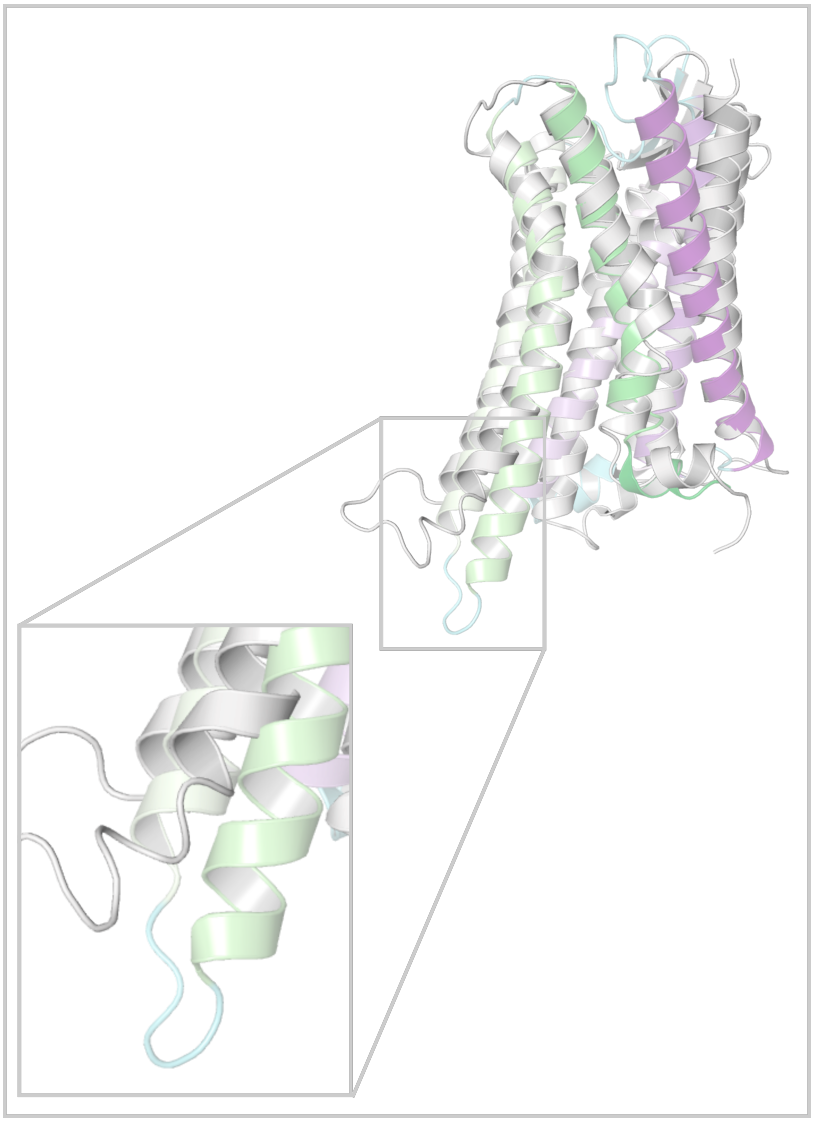


Figure 4.17: Example of TM6 folding "in the opposite direction" as a result of forcing helicity during steered MD simulations. This results in 2 "idealised" helical segments. Silver cartoons represent the final snapshot of a steered MD simulation; the coloured cartoons represent the active state μOR (PDB accession 5C1M).

4.4.2 Better helix parametrisation: a potential solution?

A potential solution to both limitations is to introduce a more rigorous set of mathematical rules defining a helix. The works by Ren and colleagues [180], if implemented

as a CV, could allow for much greater flexibility when enforcing the helicity of a particular set of residues. It would also allow for the definition of π and 3/10 helical conformations, both of which are thought to be important in transmitting information in GPCRs [180]. At the time of this writing, helical parametrisation as proposed in [180] has not yet been implemented in any major software.

4.4.3 Limitations of umbrella sampling

Umbrella sampling has proven invaluable for understanding the dynamics of many biological systems. It is an easy method to implement for relatively small systems.

One of the main assumptions made during umbrella sampling is that possible values of a particular CV are adequately sampled. For relatively simple CVs, such as variations in the ϕ dihedral angle of alanine dipeptide, this can be implemented very easily by changing the discretisation in CV space, by sampling with higher or lower biasing constants, or both. This ensures that there is a sufficient overlap between umbrella windows (a requirement for the WHAM calculations).

A secondary assumption is also present in every umbrella sampling study, is that all orthogonal states are also adequately sampled. These are all possible states that are not directly biased by the potential. Following the example of alanine dipeptide, adequate sampling is achieved when all ϕ angles are sampled, and when all possible ψ are equally adequately sampled.

This presents a significant problem for the FNCs CV. As explained, this CV is comprised of several atom pair distances, collapsed into a singular entity. This means that adequate sampling of each FNC state necessitates that all atom pairs adopt all possible distances. For the extreme conformations (inactive and active), this isn't a problem, since each atom pair is locked in a very specific conformational state. Sampling intermediate states, however, is a significant problem, since the possible distances each atom pair can visit increases, so more simulation time is required for adequate sampling. As a result, is it not always possible to guarantee "the right path" is sampled, and that it increases the likelihood of the receptor becoming locked in high energy conformations.

A possible solution to this problem would be to start multiple replicas for each FNC state, reassigning atomic velocities for each replicate prior to the start of simulation, and employing stochastic algorithms during the simulations. This allows each atom to explore more coordinate space, and as a result, allows for more efficient exploration of each FNC value. This also helps explore orthogonal states more efficiently, though some of these might require more sampling time.

4.4.4 Magnitude of umbrella sampling derived energetic landscape

In this Chapter, a novel collective variable was introduced to capture GPCR energetic landscape during transitions to active-like states in affordable timescales. The resulting PMFs (Figure 4.12) indicate that these changes are captured. However, a closer examination of the magnitude of the energy reveals that the energy required to maintain the active-like states exceeds 100 kJ/mol. This is a very high energetic value (≈24 kcal/mol), far exceeding thermal energy. This is a significant limitation, since it has been shown that, under thermal motion and in the absence of agonists, GPCRs are capable of eliciting G protein activation (known as basal signalling) [191, 192]; it is unlikely that the energy required is as high as calculated by WHAM given that spontaneous activation is possible.

4.5 Outlook

Despite the listed limitations, umbrella sampling simulations of FNCs CV (and additional supporting of this CV with ALPHARMSD) is capable of calculating the difference in the free energy landscape of a GPCR (μ OR) in the presence and absence of an agonist (DIPP-NH_2). This CV may prove useful in trying to determine how changes in ligands affect this landscape.

Chapter 5

Using FNCs to study opioid structure-activity relationship

5.1 Introduction

Investigating whether the molecular basis of structure-activity relationship (SAR) of opioids can be determined via MD serves two purposes.

If it is possible to do so for the opioid SAR, it should, in theory, also be possible to use the same methods to determine SARs for a wide variety of receptors and compounds. Just like how small changes to the opioid structure can have a drastic effect on the *in vitro* and *in vivo* activity reported, small changes in other compounds have similar effects.

While alterations to the structure of morphine have been reviewed in Chapter 3, section 3.1.1, other examples of SAR on GPCR compounds are pervasive throughout the literature. Changing the position of the guanidino group in GNTI from the 5'position (a potent κ OR antagonist) to the 6'position results in a potent κ OR agonist (Figure 5.1) [193]. Other examples are summarised in [147].

Additionally, such a discovery would have profound implications in the discovery of analgesics. In particular, this knowledge would be helpful in mitigating the current global opioid crisis, responsible for an estimated 63 thousand deaths per year (in the United States), which, according to the World Drug Report, in 2016, was the highest number on record [194]. Latest estimates by the Council of Economic Advisers place the financial burden of opioid related fatalities and non-fatalities (e.g. loss of ability to work as a result of overdoses) at $\approx$504 billion USD. These figures arise from measuring deaths related to the synthetic opioid fentanyl and its derivatives (particularly carfentanyl and sufentanyl [195]), as well as prescription and illicit opioids.

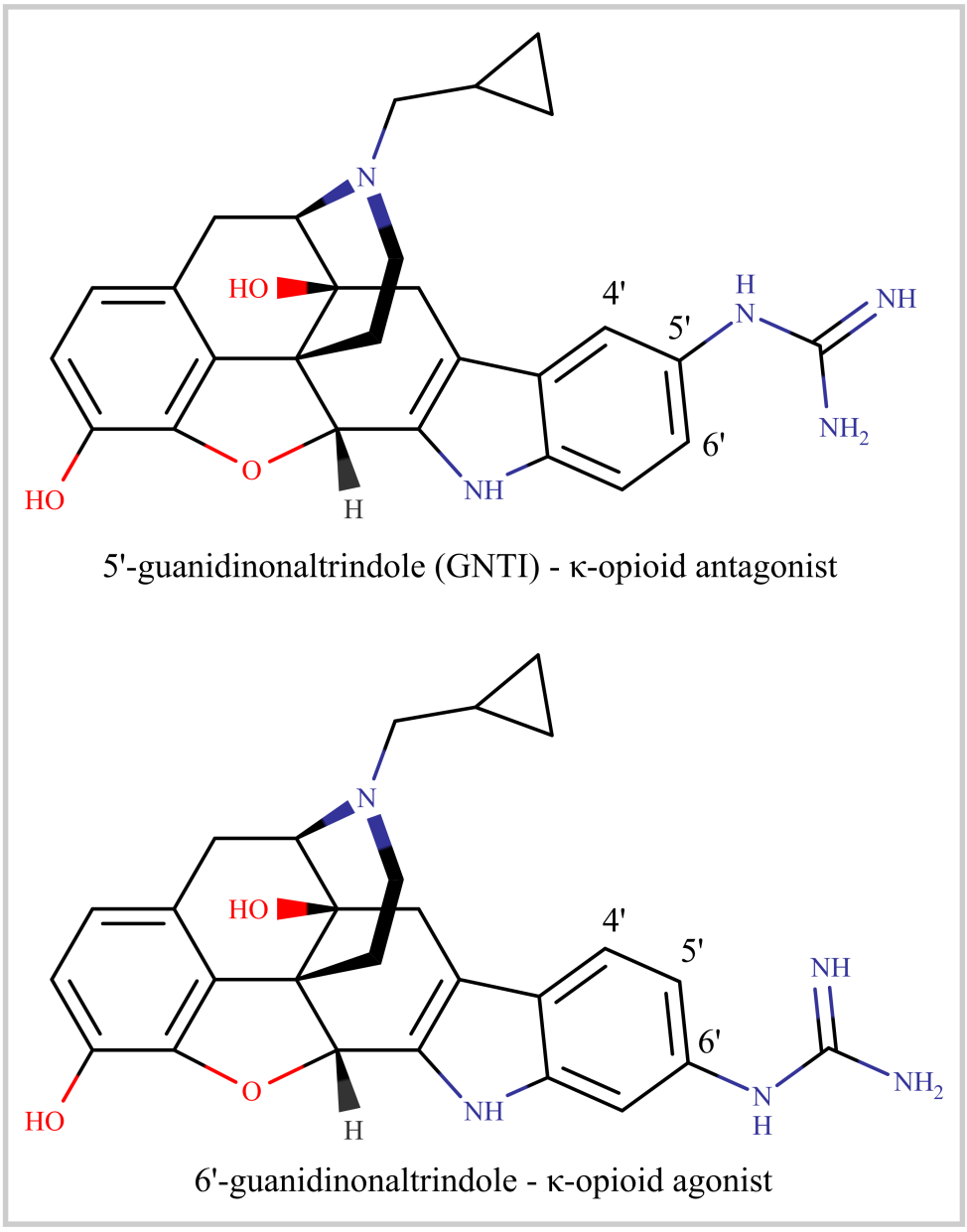


Figure 5.1: Structure of 5'-guanidinonaltrindole (top) and 6'-guanidinonaltrindole (bottom), two very similar compounds that have different *in vitro* activity profiles [147].

Gaining an understanding of the molecular basis for SAR would allow for the development of agonists with reduced side-effect profiles, mitigating some of the problems that arise from tolerance and addiction. It would also allow for the development of compounds that can be used as remedies for overdose situations. Thus, there is not only a moral imperative to find safer opioid alternatives (so as to ease societal burden), but also potential financial gain.

5.1.1 Investigating SAR of opioids

This chapter reports on the usage of FNCs on the δ OR in complex with four opioid compounds, and results of umbrella sampling simulations on the FNC CV. The four compounds were the δ OR selective naltrindole, the non-selective antagonist/inverse agonist naltrexone, and the classical agonists morphine and oxycodone.

Natrindole was the first non-peptidic δ OR selective antagonist ever synthesised [48]. It was designed by Portoghese and colleagues based on the "message-address" concept proposed by Schwyzer [74] (reviewed in Chapter 1), and on mimicking the effects of the Phe^4 residue of the endogenous enkephalins [196]. Later studies by Neilan and colleagues, measuring the effect of various opioids on sodium-deficient constitutively active receptors, classified naltrindole as a neutral antagonist [197]. Tryoen-Toth and colleagues confirmed this classification in studies of constitutively active mutants Met262$^{6.36}$Thr, Tyr308$^{7.43}$His, and Cys328$^{8.53}$Arg [167].

Naltrindole continues to be the benchmarking antagonist for δ OR studies; it is used exclusively *in vitro*, *in silico*, and *in vivo*, but not in humans. This compound was captured on two crystal structures of the δ OR (*mus musculus*, PDB accession 4EJ4 [63]; *homo sapiens*, PDB accession 4N6H [13]). Chapter 3 summarises the behaviour of this molecule in the orthosteric binding pocket of the δ OR.

Contrary to naltrindole, naltrexone is actively used in clinical settings [151]. First synthesised in 1963, naltrexone is a non-specific competitive antagonist used in the treatment of "opiate addictions and alcohol use disorders" [151]. Data suggests that naltrexone may play a role in the treatment of a wide range of disorders, including tobacco-related disorders, gambling disorders, kleptomania (the inability to refrain from theft, not necessarily for personal or financial gain), and trichotillomania (hair-pulling disorder), among others [198].

Interestingly, naltrexone doesn't always behave as a neutral antagonist; Wang and colleagues found that, after morphine pretreatment, naltrexone becomes a μ OR inverse agonist [199]. The precise mechanism behind this phenomenon is not yet known.

The agonists simulated in this study were morphine and oxycodone. The former was chosen since it is a drug of critical importance to medicine, while the latter was chosen for the interesting conventional MD results presented in Chapter 3. Some of the more interesting observations include oxycodone's ability to displace the water molecules that typically bridge opioid ligands to His278$^{6.52}$, and consequently interact with Phe218$^{5.43}$ via the methyl group in R1 (see Figure 1.8 for the structure of oxycodone and Table 3.1 for the conventional MD ligand contact analysis results).

5.1.1.1 Preparation of FNCs CV

At the time of this writing, no crystal structure of the *homo sapiens* δ OR receptor has been deposited in the Protein DataBank. As a result, to generate the unique contacts required for the FNC CV, a homology model was used, using the active state crystal structure of the *mus musculus* μ OR (PDB accession 5C1M [55]). The similarity between the two sequences is 61% over the full length sequence, 80% excluding the N- and C-termini [19, 20]. This model was built using an alignment generated from the GPCRDB [19, 20] (see Appendix D), while employing the Modeller9v10 software [127]. The 10 models with the lowest Discrete Optimized Protein Energy (DOPE) scores [200] were submitted to the Qualitative Model Energy ANalysis (QMEAN) server [201] for further scoring, from which the model with the combined lowest DOPE/highest QMEAN was selected. The FNCs were generated using the procedure described in Chapter 2, section 2.2.2.

5.2 Results

In this section, the results of steered MD simulations are reported, highlighting the changes in the FNC CV, as well as the changes to the G protein binding cavity. The changes in helicity of TM6 was also explored.

5.2.1 FNC and structural changes during steered MD

Similarly to Chapter 4, the generated FNCs were used in steered MD simulations to study transitions between inactive and active-like states, this time on the *homo sapiens* δ OR. As an initial test, and using the results from the aforementioned chapter, 3 x 20 ns steered MD simulations were performed on the FNC CV using force constants of 2000 or 5000 kJ/mol only, the results of which are summarised in Figure 5.2.

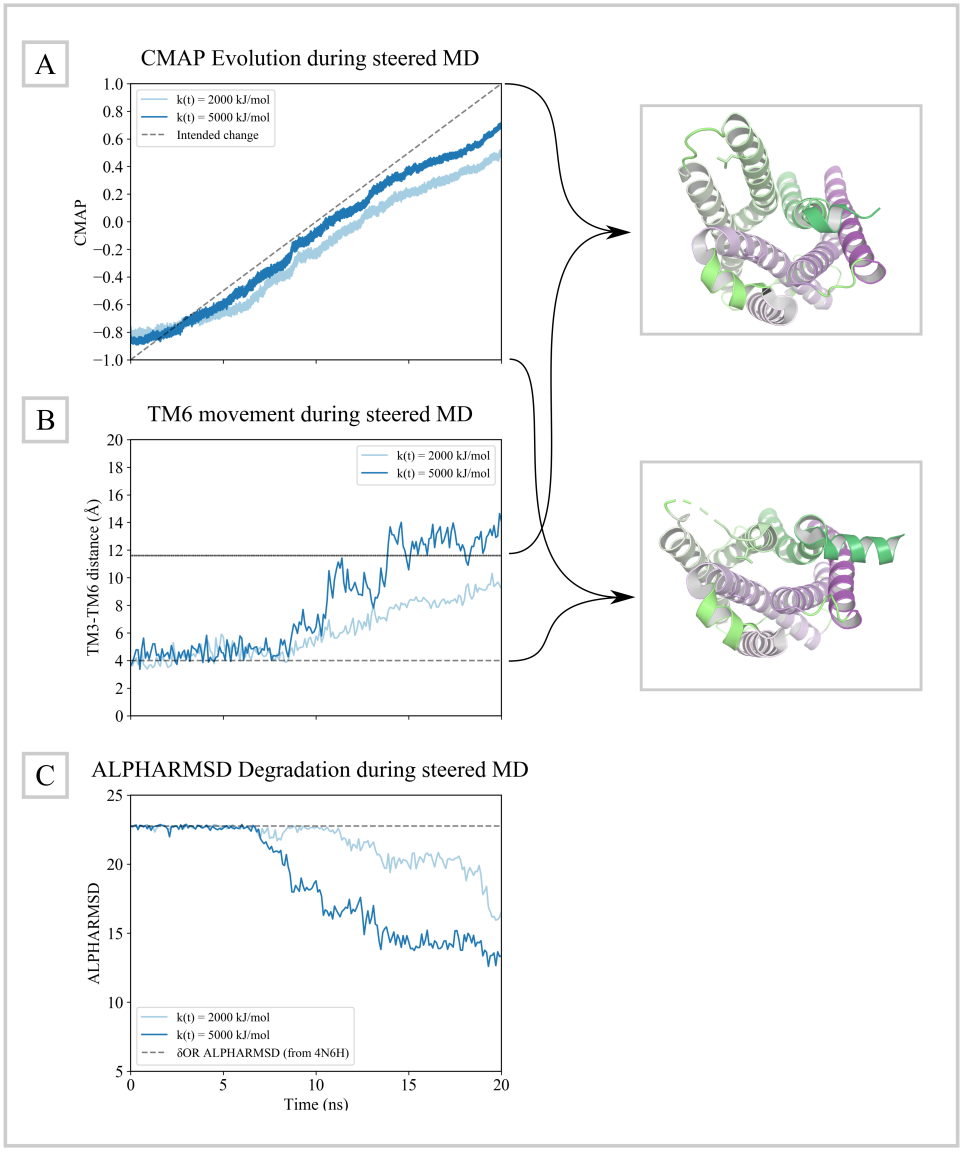


Figure 5.2: Summary of "activating" (steering the FNC CV from -1, an inactive state, to 1, an active-like state) steered MD simulations biasing the FNC CV only. (A) FNC evolution with time. (B) Opening of the G protein binding cavity (based on the "ionic lock" distance; see Chapter 1, section 1.1.2.1). The dashed and dotted lines correspond to the distance observed in the inactive (PDB accession 4DKL [61]) and active states of the μ OR (5C1M [55]), respectively. (C) Helicity change in TM6. The dashed line shows the ALPHARMSD value observed in the δ OR structure (PDB accession 4N6H [13]). All simulations were performed in triplicate, while a sodium ion was present in the allosteric binding site.

The resulting steered MD simulations performed on the *homo sapiens* δOR largely follow a similar trend to those performed on the *mus musculus* μ OR presented in Chapter 4, section 4.2.2. Despite the lack of high force constants (10000 kJ/mol, 15000 kJ/mol, 20000 kJ/mol), it is evident that lower force constants result in a lower final FNC value (Figure 5.2A).

The other aspects measured also follow similar trends to those from the simulations presented in Chapter 4.

It is noted that the G protein binding cavity opening (Figure 5.2B) adopts more active conformations with higher force constants. However, the 2000 kJ/mol simulations don't result in a completely open binding cavity. Subsequent steered MD simulations performed employed a force constance of 5000 kJ/mol when biasing the FNC CV. Furthermore, a degradation of the helicity of TM6 is also observed (Figure 5.2C).

In Chapter 4, to correct the degradation of helicity during steered MD simulations, a secondary CV was employed. While there is crystallographic data supporting the inclusion of this CV in the μ OR steered MD simulations, there is no crystal structure of the active δ OR to support the same inclusion. However, it is unlikely that an unfolding of 50% of all intracellular residues occurs during activation.

5.2.2 Supported "activation" simulations results

As a result, subsequent steered MD simulations were performed with an additional force constant employed to maintain the ALPHARMSD at a value of ≈22.8 throughout the activation process. Similarly to Chapter 4, 3 x 20 ns and 3 x 50 ns simulations were performed, using a force constant of 5000 kJ/mol for the FNC CV, and a 2000 kJ/mol for the ALPHARMSD CV. The results of these simulations are summarised in Figure 5.3.

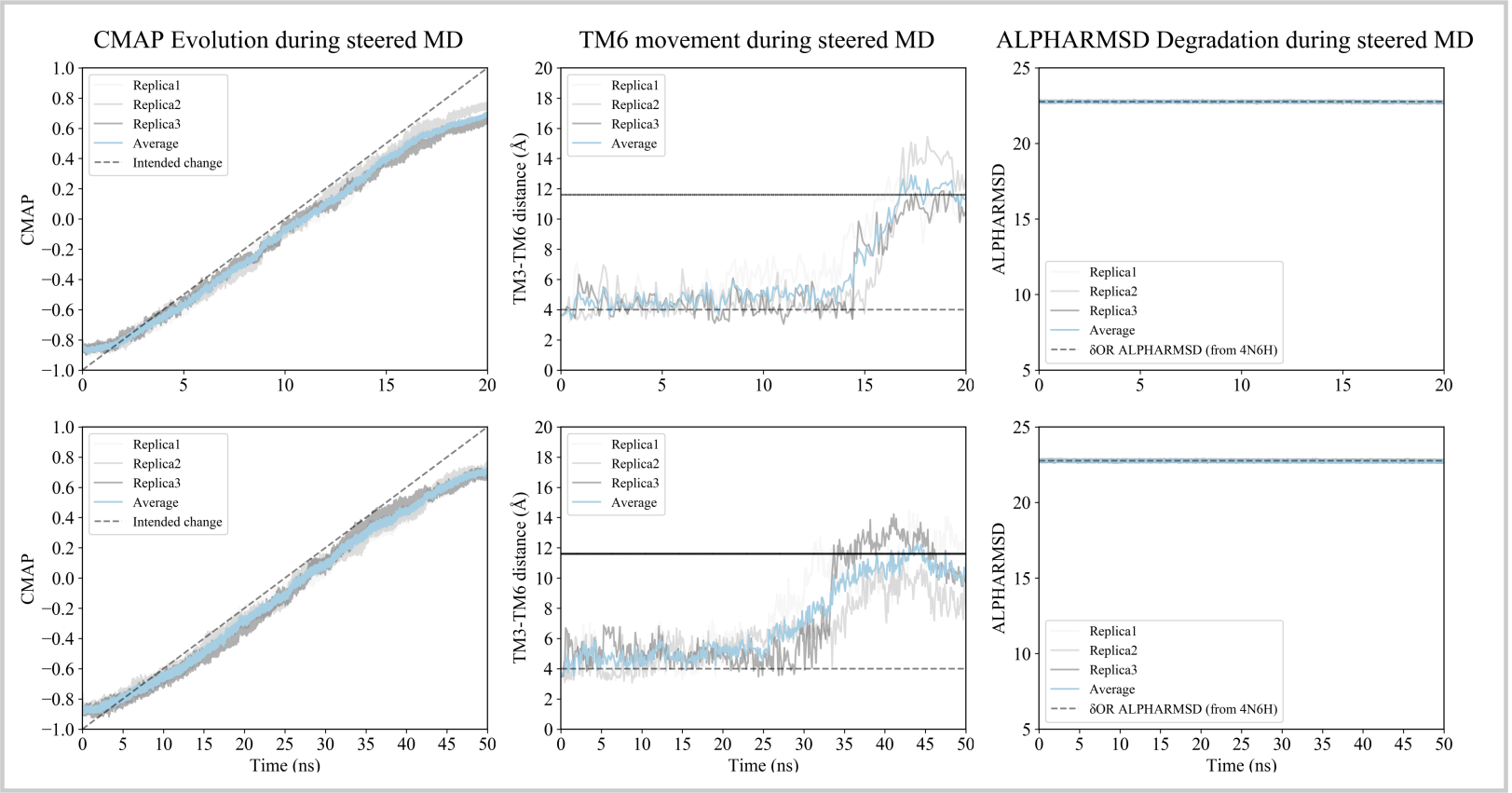


Figure 5.3: Summary of steered MD simulations performed on the δOR while biasing both FNC and ALPHARMSD with force constants 5000 kJ/mol and 2000 kJ/mol, respectively. The top row summarises the triplicate 20 ns simulations; bottom row summarises the triplicate 50 ns simulations.

In both cases, the FNC CV is pulled from -1 to +1 (dashed line in the leftmost plots of Figure 5.3, corresponding to the value at the inactive and active state, respectively), although, in the simulations performed, the CV value plateaus at a final value of ≈0.7, similar to the steered MD simulations presented in Chapter 3 (Figure 4.2). The TM3-TM6 distance (a proxy for G protein binding cavity opening) follows a similar trend to the aforementioned simulations, reaching an active-like distance at the end of the simulations (central plots of Figure 5.3).

Unsurprisingly, biasing the ALPHARMSD CV (in order to force the helicity of TM6) reports no change, as expected.

5.2.3 Steered MD simulations in the presence of ligands

To generate coordinates for umbrella sampling simulations, triplicate 50 ns simulations were performed in the presence of the 4 ligands mentioned earlier (naltrindole, naltrexone, morphine, and oxycodone). The summary of the simulations is shown in Figure 5.4.

Interestingly, the presence of ligands do not appear to alter the evolution of FNC or the change in "ionic lock" distance during the steered MD simulations. This behaviour, though expected, highlights the need for further sampling or other sampling methods that will better distinguish the ligand-bound systems, as steered MD may not prove sufficient.

5.2.4 Umbrella sampling results

The results of the umbrella sampling simulations are summarised in Figure 5.5.

Though all the PMFs show an increase in the energy required to stabilise active-like states, contrary to the PMFs presented in Chapter 4, the final energy values here reach a maximum of ≈350 kJ/mol, far higher than the final values in Figure 4.12 (page 108).

Additionally, while naltrindole and naltrexone (both antagonists) show a lower energy requirement to maintain intermediate states (compared to the unliganded profile), morphine and oxycodone (both agonists) show a higher energetic requirement, the inverse of what was expected.

The morphine simulations, in fact, show a much higher energetic profile compared to every other profile presented, requiring 25-50 kJ/mol more energy to maintain

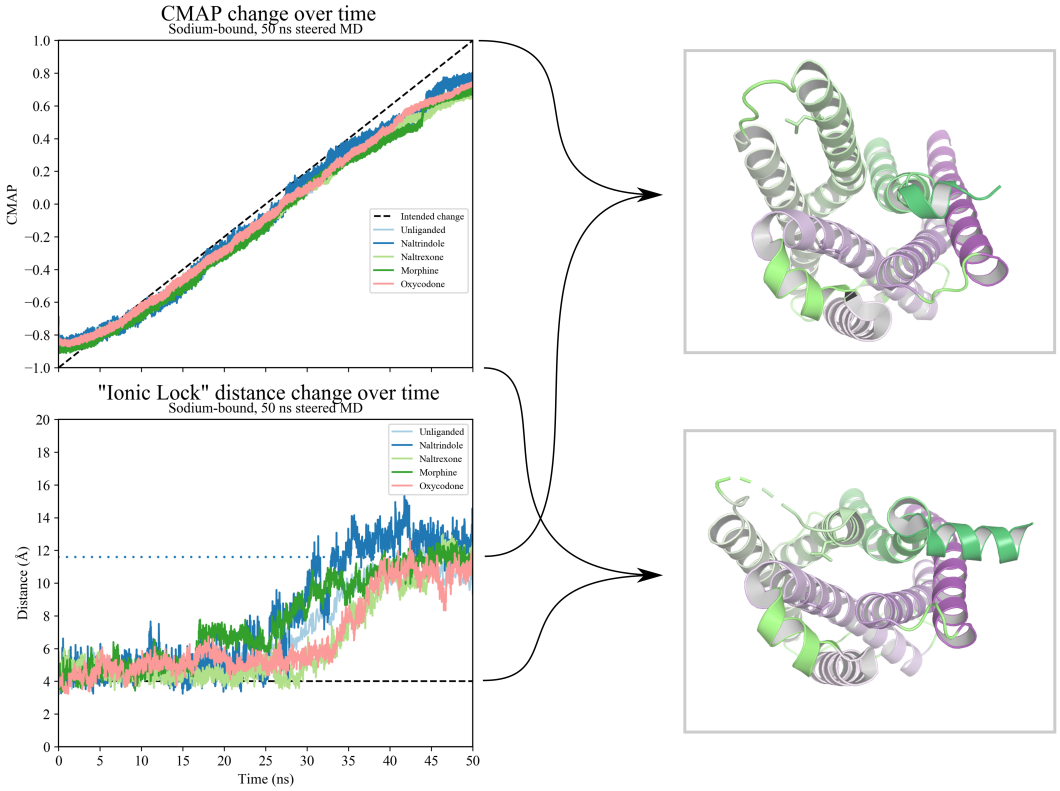


Figure 5.4: Summary of "activating" steered MD simulations performed in the presence of the ligands described in the Introduction. Each line corresponds to an average of triplicate 50 ns simulations.

active-like states compared to unliganded simulations. The same is true for the oxycodone simulations. However, there is a switch at FNC = 0.5, concluding with a final free energy value of ≈282 kJ/mol.

5.2.5 Ligand contact analysis during FNC exploration

From the resulting trajectories, a ligand contact analysis was performed. For this purpose, the software package LINTools [120] was used. A summary of these data are shown in Tables 5.1 through to 5.4, where each cell represents the average and standard deviation of quintuplicate 10 ns simulations.

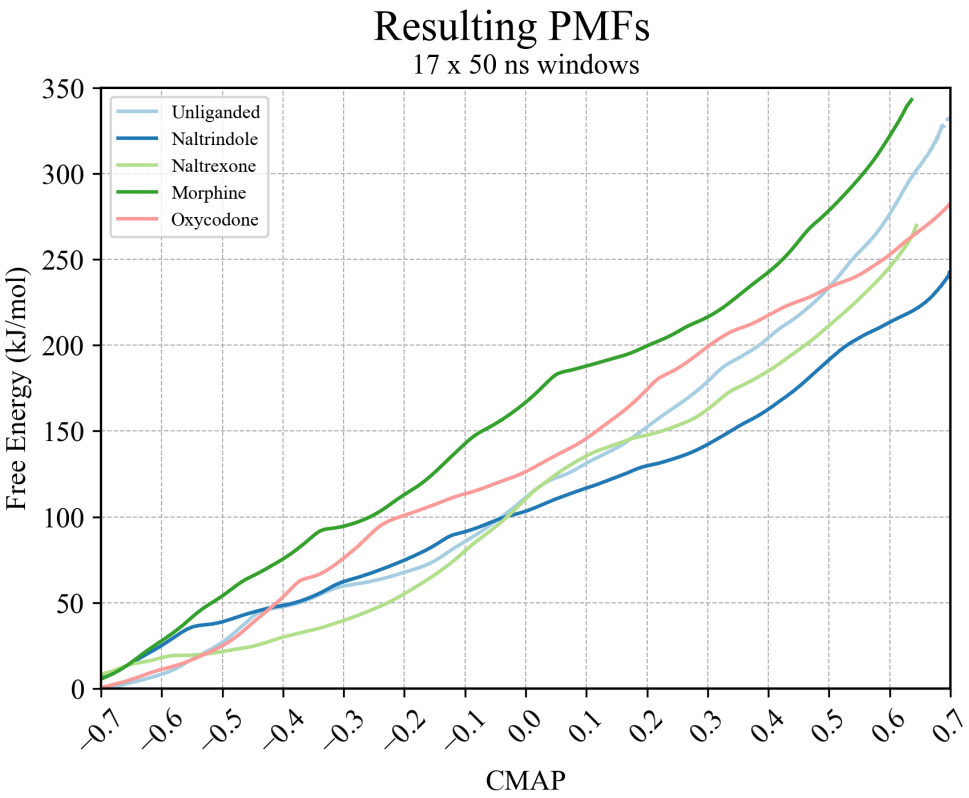


Figure 5.5: PMFs resulting from umbrella sampling simulations of the δOR in the presence of naltrindole, naltrexone, morphine, and oxycodone. These do not follow the same trend as the PMFs presented in Chapter 4. Error bars are too small to be visible.

Table 5.1: Summary of naltrindole (antagonist) contact analysis during FNC exploration. A similar trend is seen with respect to interactions with Tyr308$^{7.43}$. Highlighted in yellow are results which differ significantly from those reported in conventional MD simulations of naltrindole (Chapter 3, Table 3.1), and from other ligands in this study.

Residue \ **FNC**	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
Asp128$^{3.32}$	100.0±0.0	99.8±0.4	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
Tyr129$^{3.33}$	81.6±16.24	38.8±21.65	98.4±1.2	31.8±19.04	66.0±25.05	70.4±17.53	89.0±7.16	87.4±8.69
Met132$^{3.36}$	99.4±0.8	98.0±1.80	99.4±0.50	98.6±0.8	99.8±0.4	99.4±0.8	100.0±0.0	100.0±0.0
Lys214$^{5.39}$	-	-	-	-	-	-	-	-
Val217$^{5.43}$	80.4±3.83	84.4±4.13	71.0±5.93	90.8±2.04	88.0±5.37	88.8±4.0	81.4±1.2	83.0±1.90
Trp274$^{6.48}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.8±0.4
Ile277$^{6.51}$	99.4±0.50	99.8±0.4	99.8±0.4	99.2±0.75	100.0±0.0	100.0±0.0	99.6±0.50	99.8±0.4
His278$^{6.52}$	98.8±1.0	98.2±1.33	97.2±1.17	95.4±2.58	99.2±0.75	98.4±1.2	95.6±1.86	95.4±3.32
Phe280$^{6.54}$	-	-	-	32.6±16.31	32.4±6.02	30.4±16.74	-	-
Val281$^{6.55}$	74.0±3.10	65.4±6.10	64.2±6.68	70.6±4.36	75.8±8.50	70.2±8.13	82.8±3.66	84.8±1.60
Trp284$^{6.58}$	86.6±8.87	91.8±6.31	80.8±7.20	92.0±4.48	92.2±5.85	87.2±7.60	70.2±18.02	71.8±7.05
Leu300$^{7.35}$	74.6±3.07	72.8±6.05	72.2±5.56	67.6±21.09	77.4±3.83	70.0±10.24	75.2±5.46	66.0±3.58
Ile304$^{7.38}$	94.2±4.67	98.0±1.55	95.6±4.45	99.4±0.50	96.8±0.75	97.2±1.50	97.0±1.10	96.8±1.94
Gly307$^{7.41}$	66.8±6.27	65.8±5.53	63.4±3.26	60.8±6.94	65.4±5.61	62.8±5.42	63.2±11.46	70.8±9.11
Tyr308$^{7.42}$	99.6±0.50	99.8±0.4	100.0±0.0	100.0±0.0	99.6±0.8	100.0±0.0	100.0±0.0	100.0±0.0

Residue \ **FNC**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	
Asp128$^{3.32}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.8±0.4	100.0±0.0	100.0±0.0	
Tyr129$^{3.33}$	89.0±10.50	82.6±12.16	83.6±14.33	78.8±10.11	85.0±6.60	91.2±3.37	90.0±9.30	
Met132$^{3.36}$	99.8±0.4	99.6±0.50	100.0±0.0	99.8±0.4	99.0±0.63	99.6±0.50	99.8±0.4	
Lys214$^{5.39}$	-	-	-	-	-	41.6±25.94	-	
Val217$^{5.43}$	82.8±2.71	85.8±1.94	81.8±4.07	84.8±4.92	86.8±2.64	89.6±3.14	84.4±5.85	
Trp274$^{6.48}$	100.0±0.0	100.0±0.0	100.0±0.0	99.6±0.50	100.0±0.0	100.0±0.0	99.8±0.4	
Ile277$^{6.51}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.6±0.8	100.0±0.0	99.8±0.4	
His278$^{6.52}$	97.0±0.63	98.6±0.50	98.2±0.75	97.6±1.50	96.2±2.64	97.4±1.36	88.4±14.84	
Phe280$^{6.54}$	-	-	-	-	-	-	-	
Val281$^{6.55}$	86.4±5.75	93.4±2.73	91.8±3.50	87.6±6.71	90.2±1.167	91.8±1.72	91.4±1.50	
Trp284$^{6.58}$	79.6±10.0	55.8±26.05	73.2±11.55	81.0±11.88	89.0±4.05	86.4±5.2	88.8±6.80	
Leu300$^{7.35}$	73.4±7.55	36.0±9.21	35.0±10.0	36.8±9.20	-	34.0±13.71	55.0±5.10	
Ile304$^{7.38}$	91.0±9.10	91.2±4.75	96.2±0.75	89.0±6.60	97.0±1.67	97.2±1.94	96.4±1.36	
Gly307$^{7.41}$	71.8±5.19	75.6±5.61	78.4±4.45	63.4±6.47	68.8±5.15	62.4±5.31	64.8±4.92	
Tyr308$^{7.42}$	99.8±0.4	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.4±1.2	

Table 5.2: Summary of naltrexone (antagonist) contact analysis during FNC exploration. It is again evident that strong interactions exist between antagonists and Tyr308$^{7.43}$. It follows the same syntax as Table 5.1.

Residue \ **FNC**	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
Asp128$^{3.32}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
Tyr129$^{3.33}$	99.6±0.50	99.8±0.4	100.0±0.0	99.8±0.4	100.0±0.0	99.8±0.4	100.0±0.0	99.8±0.4
Asn131$^{3.35}$	-	-	-	-	-	-	34.8±10.4	-
Met132$^{3.36}$	99.4±0.50	99.4±0.8	99.0±0.0	99.2±0.4	99.6±0.50	99.0±0.0	99.8±0.4	100.0±0.0
Lys214$^{5.39}$	-	-	-	-	-	-	-	39.2±10.36
Val217$^{5.43}$	73.0±4.24	77.2±3.54	69.0±9.32	75.0±6.48	67.6±6.95	66.2±5.46	77.6±2.8	80.6±2.58
Trp274$^{6.48}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
Ile277$^{6.51}$	99.8±0.4	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.8±0.4
His278$^{6.52}$	97.4±0.50	98.6±1.36	98.4±0.50	98.6±0.50	98.0±0.63	98.8±0.4	98.8±0.4	96.8±1.17
Val281$^{6.55}$	93.6±4.72	90.8±4.26	94.4±1.02	93.6±2.58	93.8±1.60	92.8±3.71	88.6±5.31	82.4±2.50
Trp284$^{6.58}$	-	-	-	-	-	-	-	38.0±15.47
Ile304$^{7.38}$	94.2±2.71	94.4±3.07	93.0±5.37	93.6±3.83	94.8±3.20	94.8±0.75	92.4±1.02	97.0±1.41
Gly307$^{7.41}$	41.0±9.86	40.2±4.0	44.2±7.78	45.4±4.27	-	34.8±5.81	39.4±3.72	49.4±4.60
Tyr308$^{7.42}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.6±0.8	100.0±0.0	99.8±0.4	99.2±0.75

Residue \ **FNC**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	
Asp128$^{3.32}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	
Tyr129$^{3.33}$	99.2±0.4	100.0±0.0	100.0±0.0	99.8±0.4	99.8±0.4	99.8±0.4	99.8±0.4	
Met132$^{3.36}$	100.0±0.0	99.8±0.4	99.6±0.50	100.0±0.0	99.8±0.4	99.6±0.50	100.0±0.0	
Lys214$^{5.39}$	36.8±7.73	-	-	-	-	-	-	
Val217$^{5.43}$	79.8±2.64	76.6±5.31	80.4±3.44	81.6±1.63	83.4±2.33	83.4±1.50	82.6±2.06	
Trp274$^{6.48}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	
Ile277$^{6.51}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	
His278$^{6.52}$	96.4±0.8	98.0±0.90	97.2±0.75	98.6±0.8	97.8±1.94	97.8±1.50	98.6±0.8	
Val281$^{6.55}$	84.8±2.80	84.0±2.53	85.6±2.25	86.0±3.70	80.8±1.00	84.2±1.72	82.4±6.60	
Trp284$^{6.58}$	-	-	-	-	-	-	-	
Ile304$^{7.38}$	98.8±0.4	95.2±2.64	93.2±3.50	97.4±1.36	97.0±1.27	96.6±1.02	91.0±5.76	
Gly307$^{7.41}$	50.6±5.68	31.0±9.61	30.6±9.0	-	39.4±6.22	38.8±6.71	38.6±4.32	
Tyr308$^{7.42}$	98.8±1.94	99.6±0.50	99.4±0.8	100.0±0.0	100.0±0.0	99.8±0.4	95.6±8.31	

Table 5.3: Summary of morphine (agonist) contact analysis during FNC exploration, further adding to the idea that antagonists differentiate from agonists by interactions with Tyr308$^{7.43}$. It follows the same syntax as Table 5.1.

Residue \ **FNC**	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
Asp128$^{3.32}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
Tyr129$^{3.33}$	96.4±3.93	95.6±2.87	91.0±12.28	99.4±0.49	99.0±0.63	100.0±0.0	100.0±0.0	100.0±0.0
Met132$^{3.36}$	99.6±0.49	100.0±0.0	99.8±0.4	99.8±0.4	100.0±0.0	100.0±0.0	99.8±0.4	100.0±0.0
Lys214$^{5.39}$	-	33.4±10.05	49.2±10.17	38.4±13.41	-	-	49.8±23.42	39.0±9.96
Val217$^{5.43}$	65.6±6.95	64.2±6.2	66.0±7.75	66.6±6.34	69.6±6.74	70.8±7.47	60.0±8.67	70.4±2.8
Trp274$^{6.48}$	92.2±3.25	91.8±3.66	80.8±13.36	90.6±2.65	96.0±1.90	-	-	45.6±10.87
Ile277$^{6.51}$	90.8±4.31	84.4±5.61	83.6±13.11	91.0±2.10	92.6±3.50	79.2±6.77	77.4±10.67	84.8±5.71
His278$^{6.52}$	59.2±19.94	57.0±11.68	43.6±25.22	61.6±13.2	84.8±5.95	89.6±6.05	92.0±4.69	88.8±8.68
Val281$^{6.55}$	71.8±10.57	71.0±15.50	65.4±9.33	63.4±12.82	68.8±14.85	88.8±4.96	86.0±8.65	81.8±4.4
Ile304$^{7.38}$	74.2±15.86	74.2±7.83	67.0±12.85	73.8±13.78	82.0±7.56	53.4±14.95	62.0±12.65	72.0±16.96
Tyr308$^{7.42}$	54.8±22.80	61.4±8.04	58.6±9.35	54.8±6.01	61.2±11.39	-	-	-

Residue \ **FNC**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	
Asp128$^{3.32}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	
Tyr129$^{3.33}$	100.0±0.0	98.6±0.8	99.4±0.49	98.6±1.02	99.6±0.49	99.6±0.49	100.0±0.0	
Met132$^{3.36}$	99.8±0.4	100.0±0.0	100.0±0.0	100.0±0.0	99.8±0.4	100.0±0.0	100.0±0.0	
Lys214$^{5.39}$	31.2±15.60	44.0±11.40	33.8±14.46	-	-	-	-	
Val217$^{5.43}$	62.8±8.42	73.4±3.50	71.4±5.95	57.6±4.36	63.6±5.90	69.6±3.44	53.6±5.46	
Trp274$^{6.48}$	63.0±11.51	94.4±2.94	75.4±10.84	67.4±4.50	73.6±11.67	66.4±10.21	54.4±7.39	
Ile277$^{6.51}$	89.2±4.0	97.4±1.36	96.4±3.61	97.2±1.33	96.6±1.50	98.8±0.4	89.2±2.56	
His278$^{6.52}$	90.4±5.46	90.0±4.19	80.4±5.85	72.2±4.79	85.0±5.76	90.4±4.92	97.6±1.02	
Val281$^{6.55}$	65.2±9.66	66.2±12.54	68.8±4.31	46.6±12.69	61.4±10.37	69.4±11.84	54.6±5.31	
Ile304$^{7.38}$	74.8±8.52	79.0±12.25	83.6±10.21	86.6±3.14	89.4±8.78	86.2±5.56	56.0±7.51	
Tyr308$^{7.42}$	-	63.0±16.21	63.8±19.65	54.8±8.86	59.0±10.55	-	-	

Table 5.4: Summary of oxycodone (agonist) contact analysis during FNC exploration. Contrary to morphine, however, this agonist shows stronger interactions with Tyr308$^{7.43}$, suggesting that different agonists may activate receptors by triggering different activation pathways. It follows the same syntax as Table 5.1.

Residue \ **FNC**	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
Asp128$^{3.32}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
Tyr129$^{3.33}$	99.2±0.4	88.0±22.50	99.6±0.50	99.0±1.10	99.6±0.50	100.0±0.0	99.6±0.50	99.8±0.4
Met132$^{3.36}$	99.4±0.8	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
Lys214$^{5.39}$	-	-	-	68.0±21.38	40.2±18.70	38.25±13.94	43.0±17.29	58.8±22.75
Val217$^{5.43}$	94.4±2.06	92.6±5.90	92.6±4.03	78.8±8.61	87.6±5.46	87.75±4.27	90.6±2.87	82.8±4.53
Phe218$^{5.43}$	42.6±6.65	31.8±9.30	43.4±7.20	42.4±10.01	42.6±13.54	37.5±11.80	40.0±5.25	36.2±12.24
Trp274$^{6.48}$	55.4±8.96	71.8±7.57	72.8±5.42	62.8±2.32	59.6±8.36	59.5±8.87	81.2±5.0	81.4±7.84
Ile277$^{6.51}$	69.2±2.80	74.8±10.10	76.2±3.6	60.2±10.94	62.8±13.93	63.5±8.76	57.4±6.34	64.0±7.93
His278$^{6.52}$	92.8±2.48	86.4±7.45	85.4±9.73	68.2±9.85	75.4±12.0	77.25±6.18	81.2±4.80	73.2±6.34
Val281$^{6.55}$	46.8±9.24	42.6±8.97	49.8±6.85	34.6±10.09	36.4±15.47	38.5±13.96	45.8±8.21	30.4±12.63
Ile304$^{7.38}$	81.0±4.15	75.8±8.64	79.6±6.02	85.2±2.32	70.8±5.34	72.75±9.81	76.2±8.42	76.2±15.30
Tyr308$^{7.42}$	64.2±12.64	91.4±5.54	86.4±12.47	96.0±3.10	93.2±5.71	98.25±1.30	98.6±0.50	93.2±3.66

Residue \ **FNC**	0.1	0.2	0.3	0.4	0.5	0.6	0.7	
Asp128$^{3.32}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	
Tyr129$^{3.33}$	99.8±0.4	99.6±0.8	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	
Met132$^{3.36}$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.4±0.50	
Lys214$^{5.39}$	-	-	-	33.4±19.23	51.0±7.54	88.4±16.82	78.8±21.54	
Val217$^{5.43}$	90.2±3.66	92.8±4.21	94.8±2.04	84.0±8.0	74.4±7.74	79.8±3.50	83.0±4.86	
Phe218$^{5.43}$	-	36.0±7.67	38.2±6.01	33.4±6.71	64.0±23.52	31.8±13.83	51.6±19.73	
Trp274$^{6.48}$	71.2±8.90	80.0±9.63	73.0±7.29	61.6±9.56	71.4±7.71	85.0±9.82	73.4±9.60	
Ile277$^{6.51}$	55.4±13.22	74.0±2.28	70.6±5.31	62.8±8.70	40.6±11.67	48.8±16.30	82.4±10.82	
His278$^{6.52}$	77.4±4.27	87.8±2.04	88.8±6.85	78.0±9.38	48.4±22.64	82.6±14.09	91.8±4.87	
Val281$^{6.55}$	39.8±7.08	53.0±3.85	58.0±6.54	34.6±18.21	-	-	55.2±15.79	
Ile304$^{7.38}$	68.2±8.23	58.4±8.78	58.6±5.24	65.8±5.49	56.4±17.00	51.8±15.42	77.4±6.60	
Tyr308$^{7.42}$	97.8±1.60	97.8±1.47	92.8±3.00	90.2±6.34	96.4±3.88	89.0±4.52	76.6±9.65	

Similarly to the data reported from conventional MD (Chapter 3), interactions between ligands and the binding site show only slight differences. In particular, antagonists and agonists appear to interact differently with Lys214$^{5.39}$; inspection of the resulting trajectories indicates that this difference is due to the flexibility of the agonist (conferred to them by the lack of a bulky *N*-methyl substituent), as opposed to motion of the Lys214$^{5.39}$ side-chain.

Another difference, also observed in the conventional MD simulations, is the difference in interactions between the ligands and Tyr308$^{7.43}$. While, on average, antagonists appear to be in almost constant contact with this residue, the morphine is not. Strangely, oxycodone behaves as the antagonists do, interacting with Tyr308$^{7.43}$ near constantly during the simulations; this is in direct contradiction to the conventional MD simulations reported in Chapter 3 (Table 3.1, page 61).

5.3 Discussion

In the present Chapter, the novel CV introduced in Chapter 4 was used to explore the energetic landscape of δ OR activation in the presence and absence of a diverse set of ligands.

It was expected that there would be a similar pattern to that found in Chapter 3, where agonists (morphine and oxycodone) lowered the free energy surface at intermediate FNC values, compared to unliganded simulations, and antagonists (naltrindole and naltrexone) would have no effect on the free energy surface. Unfortunately, the results shown in Figure 5.5 do not reflect this; in fact, the PMFs point to the opposite.

Analysis of the ligand contact analysis reveals some interesting patterns that distinguishes the two agonists and the two antagonists, particularly the difference in contact times with Lys214$^{5.39}$ (see Tables 5.1 through 5.4). On average, agonists are in contact with this residue more frequently than antagonists, which is a consequence of the increased mobility of agonists in the orthosteric pocket. This residue is implicated in covalent bonding to the μ OR antagonist β-funaltrexamine [202] (captured using x-ray crystallography in PDB accession 4DKL [61]), though it is unclear whether it can distinguish between residues with different activity levels.

However, this change is not sufficient to explain the output PMFs in Figure 5.5. The free energy levels are far higher than expected (see Figure 4.12, page 108). There are several explanations for this, some of which are detailed in subsequent sections.

5.3.1 Addressing the free energy values

One of the issues found in Chapter 4 was the abnormally high free energy values calculated by the WHAM software [133]. It was found that energies in excess of ≈100 kJ/mol (or ≈24 kcal/mol) were required to stabilise active-like states of the μ OR, which, as mentioned in section 4.4.4, far exceeds thermal energy.

In the simulations reported in this Chapter, the final free energy values calculated using the WHAM software were all above ≈240 kJ/mol (for the naltrindole profile), over twofold higher than the μOR-DIPP-NH_2 profile (Chapter 4, Figure 4.12, page 108). The highest energy value profile reported, that of morphine-bound δOR, shows a maxima at the 0.64 FNC value, with a corresponding free energy value of ≈343 kJ/mol (or ≈82 kcal/mol), over threefold higher than the aforementioned μOR-DIPP-NH_2 profile. Clearly, these values are excessive.

5.3.2 A system issue or a methodological limitation?

These high energy levels may be explained by further analysis of the system and of the PMF generation. Some of the limitations have been detailed in Chapter 4 already (the reader is referred to section 4.4 for more), and so will not be addressed here again.

GPCR activation is not a process limited to the opening of the G protein binding cavity; were that the case, a collective variable based on the "ionic lock" distance (see Chapter 1, section 1.1.2.1) would suffice to calculate the free energy landscape (this, however, presents a new set of issues, and is beyond the scope of this Chapter). There are a series of side-chain rotations and helical motions (discussed in the literature) that have been indicated as critical for the activation process.

5.3.2.1 Sodium escape during sampling

One of these concerns the changes to the allosteric binding site. Analysis of the sodium site from the available GPCR crystallographic structures suggests that this site collapses in active-like states, abolishing sodium binding altogether [181]. This collapse is accompanied by a rotation of $Asn131^{3.35}$ away from the receptor core (see Figure 5.6).

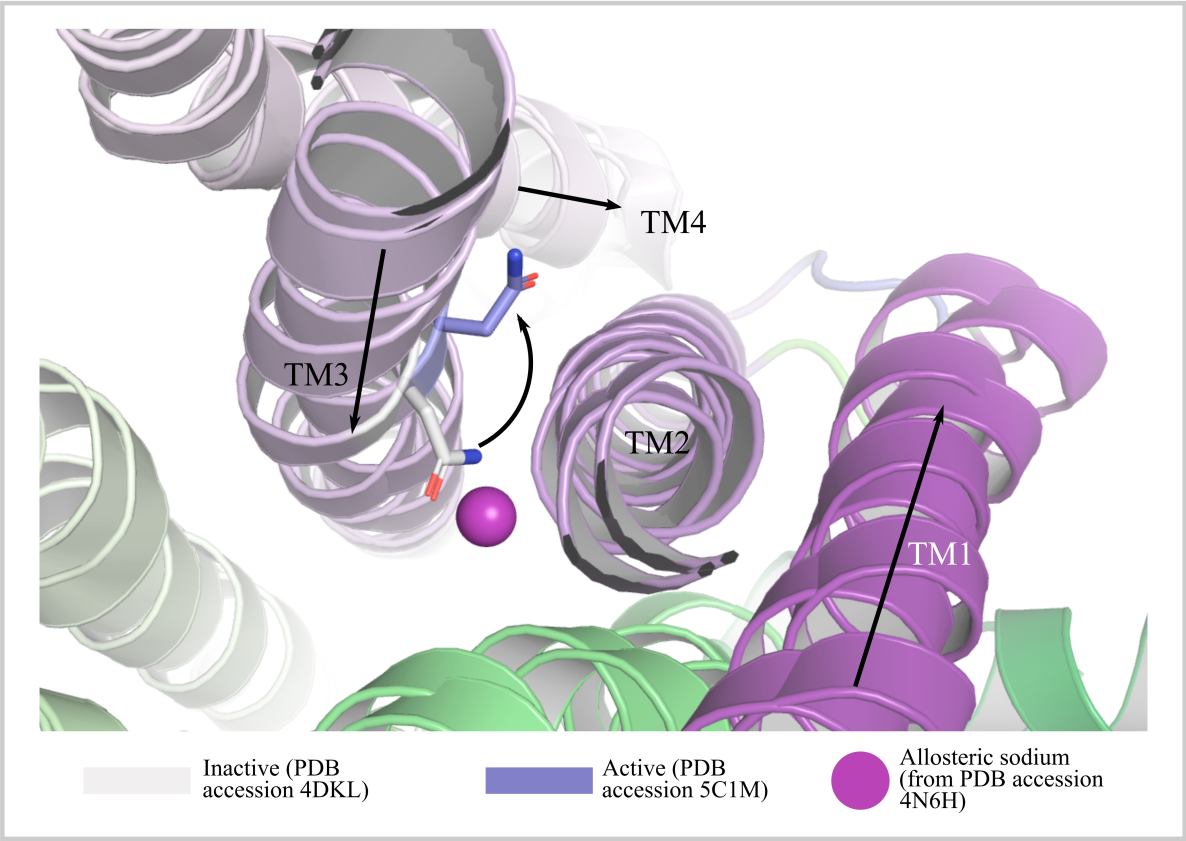


Figure 5.6: Rotation of the Asn131$^{3.35}$ side chain during activation. The sodium shown was determined in the 1.8 Å resolution structure of the δ OR (PDB accession 4N6H [13]).

The sodium is expelled from the allosteric site. Though the pathway of sodium egress is still debated, *in silico* studies by Shang and colleagues have shown that sodium is likely expelled via the orthosteric binding site, even if a ligand is still present [181]. It is important to mention, however, that the ligand used in their studies of the δ OR was SNC-80 (Chapter 3, Figure 3.5, page 53), a ligand for which a defined binding site has not yet been revealed by crystallography, and that their sodium egress pathways were calculated using random accelerated MD (or random expulsion MD), a method that relies on biasing the velocity of the sodium in random directions, in order to calculate likely pathways [181, 203]. These resulting pathways may not be physiologically relevant.

In the simulations reported in this Chapter, there is no consistent expulsion of the sodium from the allosteric site (see Figure 5.7). If the sodium remains in the allosteric site, Asn131$^{3.35}$ does not rotate away from the core of the receptor. This severely limits the conformational exploration of the δ OR (and as a result, the FNC exploration during steered MD and umbrella sampling).

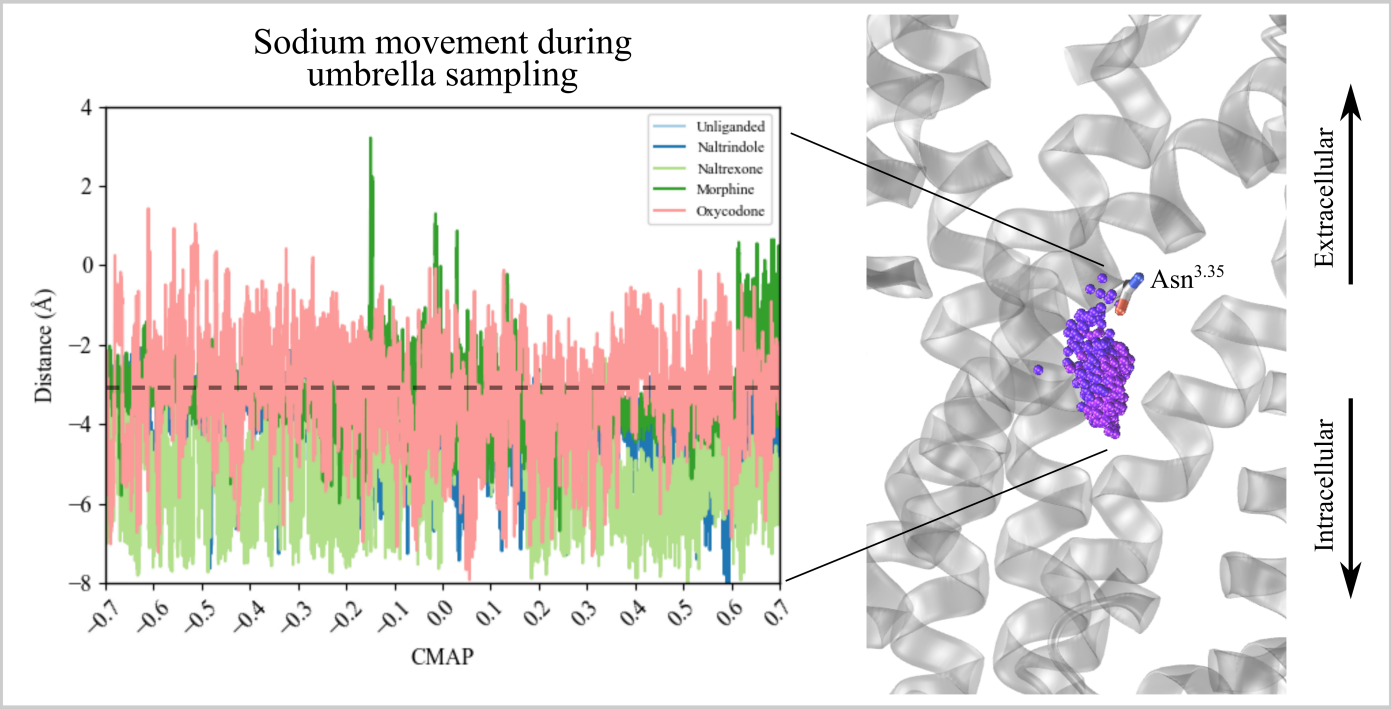


Figure 5.7: Distance between the z-coordinates of the C_α of Asn131$^{3.35}$ and the allosteric sodium, a measure for the region explored by the sodium during umbrella sampling simulations. It is clear that it does not exit the allosteric site during the sampling. The dashed line represents the distance as measured in the δOR crystal structure (PDB accession 4N6H [13]).

Mitigating the sodium problem It might be possible to mitigate this by either adding another collective variable to the steered MD to control the position of sodium, or increase simulation time to allow sodium to escape. However, either solution would impact simulation performance: adding an additional CV would not be advisable, given that there already are two CVs in each steered MD simulation; increasing simulation time would run counter to the original idea of developing a method for rapid exploration of energetic landscapes.

Alternatively, one could artificially remove the sodium from the core of the receptor at a given FNC value, so as to allow for the allosteric site to collapse during the steered MD and umbrella sampling simulations. However, it would be difficult to justify at what FNC value to remove the sodium.

5.3.2.2 G protein binding cavity sampling

As explained in Chapter 4, section 4.1.3, the motivation for developing the FNC CV was to accurately explore the movement of TM6 away from the receptor core. A number of measurements have been proposed to quantify this movement, most notably the "ionic lock" distance between the residues in position 3.50 and 6.30 (arginine and leucine, respectively, in the opioid receptors; for more information, the reader is referred to Chapter 1, section 1.1.2.1, page 6).

Another measure proposed is the distance between the tyrosines in position 5.58 (of the Px_3Ix_3Y motif [180]) and 7.53 (of the NPxxY motif; for more information, the reader is referred to Chapter 1, section 1.1.2.1, page 6), suggested to be important in active states of the rhodopsin [180]. In inactive states, the side-chain $Tyr308^{7.53}$ points towards the allosteric site of the receptor, and the two amino acids are separated by TM6. In active states, $Tyr308^{7.53}$ rotates as TM7 straightens and slightly unwinds, and, as the intracellular end of TM6 swings outwards, the intracellular end TM5 swings inwards, allowing $Tyr^{5.58}$ and $Tyr^{7.53}$ to interact via a water bridge [55].

In the simulations presented in this Chapter, FNC states where the tyrosines are proximal are poorly sampled (Figure 5.8). Similarly to the sodium expulsion issues reported (section 5.3.2.1), this may severely limit the FNC values explored during the sampling, raising the energetic values reported in the PMFs.

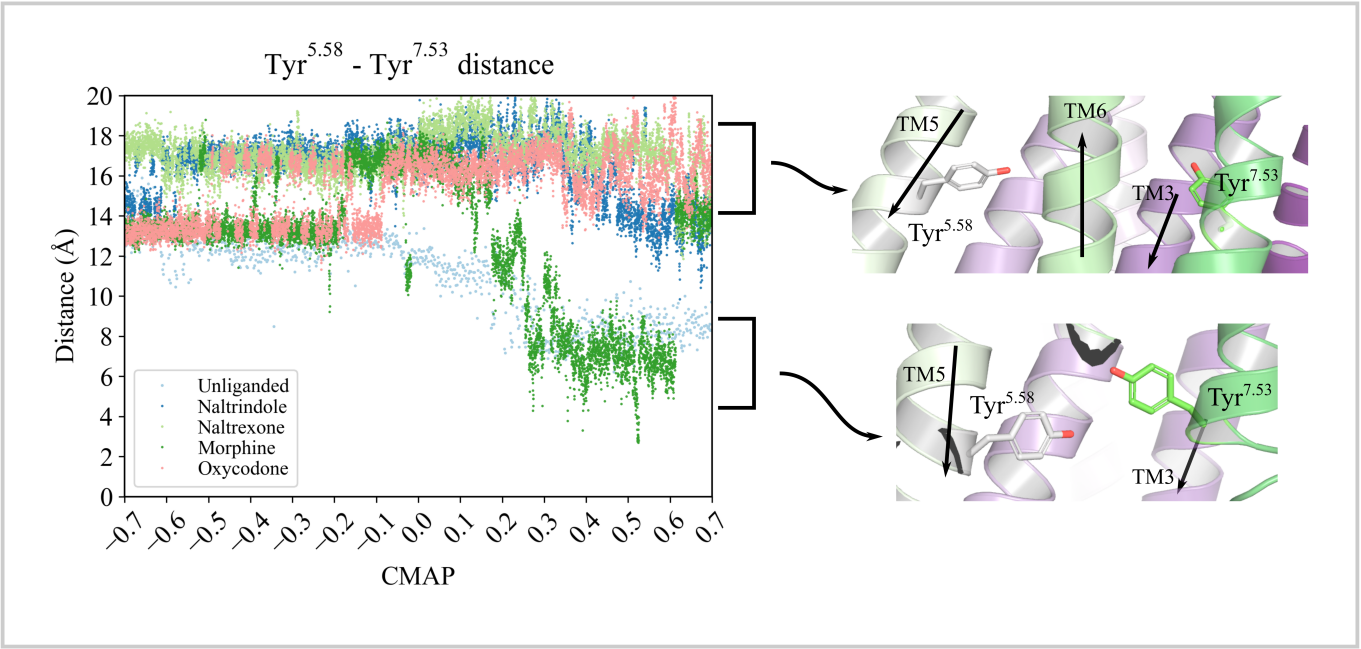


Figure 5.8: Exploration of the Tyr233$^{5.58}$-Tyr318$^{7.53}$ distance during umbrella sampling. Reference states on the right were imaged using the μ OR crystal structures (PDB accession 4DKL [61], top, and 5C1M [55], bottom).

It is worth noting that the Px_3Ix_3Y motif is not conserved within the Class A GPCR family, as the amino acid in position 5.54 is not 100% conserved. Analysis of the all Class A GPCR sequences reveals that $Ile^{5.54}$ is only 33% conversed, whereas $Pro^{5.50}$ and $Tyr^{5.58}$ are 79% and 72% conserved, respectively [19, 20].

5.3.2.3 Generation of FNCs

As explained in section 5.1.1.1, the FNC CV for the simulations presented in this Chapter were generated based on unique contacts derived from a model of the δ OR, rather than a structure determined using crystallography (as it was the case for the simulations presented in Chapter 4).

It is possible that the model used was incorrect, which would mean that the FNCs derived more readily lead to incorrect active-like conformations. It also increases the likelihood that both steered MD and umbrella sampling lead to conformationally locked states, which would result in very high energies (as the structure attempts to escape these un-physical states).

5.3.2.4 FNC convergence and exploration of orthogonal states

Similar to the PMFs reported in Chapter 4 (Figure 4.12, page 108), the energy landscapes presented here appear to show convergence after 50 ns of simulation per window (see Figure D.3). The high energetic values shown in the resulting PMFs (Figure 5.5) speaks to the main drawbacks of combining FNC exploration with umbrella sampling.

The FNC CV was designed as an all-encompassing generalised coordinate to explore the transition of receptors between distinct activity states, meant to drive the receptor to a particular state based on the unique contacts found in that state (see Chapter 2, section 2.2.2). Combined with steered MD and umbrella sampling, it was hoped that this would allow for efficient exploration of activation pathways, and that the effect of ligands would also be accounted for throughout.

At the extreme values of the FNC CV (-1 for inactive states and +1 for active states, in the work presented both here and in Chapter 4), there are very few states available for the receptor to explore. However, in states other than the extreme, this no longer holds true.

As the FNC value diverges away from either -1 or +1, the number of possible intermediate states grows exponentially, requiring much more simulation time to explore these states satisfactorily. This is compounded by the immeasurable number of states which occur orthogonal to the given FNC path. As previously reported, even in small systems, adequate sampling of these states can be difficult to achieve using the

umbrella sampling methodology [204, 205, 206]. Therefore, it is unlikely that truly converged sampling can be achieved within realistic timescales on currently available commodity hardware. It is feasible that alternative enhanced sampling methods (e.g. metadynamics, adaptive biased MD, adiabatic biased MD, etc) may partly mitigate this issue; however, due to the large nature of GPCRs, this would be difficult to evaluate.

5.4 Outlook

In this Chapter, an novel collective variable (termed FNC) was employed to explore the free energy landscape of the δ OR, in the presence and absence of two antagonists (naltrindole and naltrexone) and two agonists (morphine and oxycodone). The resulting free energy landscapes proved inconclusive, potentially owing to issues with umbrella sampling and the accessibility of microstates during transitions from inactive to active-like states of the receptor. Some of these issues are difficult to directly address and correct (such as sodium not escaping the allosteric binding site), while others may be a result of poor sampling (the distance between $Tyr^{5.58}$ and $Tyr^{7.53}$ does not reduce). In any case, future work would involve finding an alternative to umbrella sampling, as it may not be the most appropriate tool to explore the FNC landscape (owing to it's notorious problems exploring orthogonal states [204, 205, 206]), or remaking the FNC CV using different parameters (and perhaps even while excluding backbone atoms), so as to reduce the number of potential microstates available to explore in intermediate FNC values.

Chapter 6

Discussion and Conclusions

6.1 Overview of results presented

The main objective of this research was to answer whether MD is an adequate tool to investigate the molecular basis for SAR of ligands. Since, at the start of this research, the 1.8 Å resolution crystal structure of the δ opioid receptor (δ OR) had been revealed (PDB accession 4N6H [13]), it was used as the basis for conventional MD simulations in the presence (and absence) of opioid ligands.

Opioids provide an excellent platform for this research. Not only is the SAR of opioids well characterised *in vivo* and *in vitro*, their impact on society is well documented and requires multi-faceted (and urgent) solutions.

In theory, if the molecular basis for opioid SAR was possible to capture using MD, then it would follow that the same methods could be extended to other ligand/receptor combinations, and unlock the molecular basis for SAR of those compounds. An excellent review of these can be found in [147].

In attempting to find a solution to this, the first results Chapter (Chapter 3, starting page 46) reports on conventional MD simulations of the δOR in complex with a diverse set of ligands, ranging from the classical agonists morphine and oxycodone, to the selective agonists TAN-67 and KNT-127, to the antagonists naltrindole, naltrexone, naloxone, and DIPP-NH_2 (the reader is referred to Figure 3.8, Chapter 3, page 58, for the structure of these ligands). The results of these simulations, though inconclusive with respect to opioid SAR, helped highlight a connection between $Tyr130^{3.34}$ and $Trp173^{4.50}$, the latter of which has been shown to be important in the binding of certain opioid ligands [166], and also critical in stabilising inactive states of the δ OR [141].

Since Chapter 3 did not provide insights into the SAR of opioids, it was important to develop an understanding of how these ligands affected the activity profile of the δ

OR. Since this is traditionally explored using long timescale MD (the reader is referred to the excellent studies by Dror and colleagues [26], Kohlhoff and colleagues [123], and Schneider and colleagues [96]), it was necessary to develop a method to quickly explore the free energy surface of the δOR as it transitions through different activity states. Turning to the seminal work of Best, Hummer, and Eaton [125], a novel collective variable (CV) for enhanced sampling was developed and used to induce artificial transitions between inactive and active states of GPCRs. The development of this CV, termed Fraction of Native Contacts (FNCs) is detailed in Chapter 2, section 2.2.2. Since the active state of the δ OR has not yet been captured via X-ray crystallography (or cryogenic electron microscopy), this CV was tested on the μ OR, a receptor for which three different states have been captured (the reader is referred to Table 1.3 for details). The FNCs were derived from the inactive state and the nanobody-bound state, and tested for their ability to drive the receptor from inactive to active-like states (and vice-versa). Despite requiring a secondary CV to maintain the helicity of TM6, umbrella sampling simulations performed on the receptor in the presence and absence of an agonist (DIPP-NH_2, Figure 1.10) showed that agonists may reduce the energy required to maintain intermediate states of the receptor, but provide little contribution to maintaining active-like states that require stabilisation by intracellular binding partners (Figure 4.12), agreeing with previously reported experimental results [35].

Encouraged by the test results of Chapter 4, the Modeller 9v10 software [127] was used to create a homology model of the active δ OR using the nanobody-bound structure of the μ OR (PDB accession 5C1M [55]), and FNCs were created using the model and the inactive state δ OR (from PDB accession 4N6H [13]). Then, steered MD and umbrella sampling simulations were performed on the δOR in the presence of the δ OR selective neutral antagonist naltrindole [196], the non-selective antagonist naltrexone [43], and the agonists morphine and oxycodone. Unfortunately, for reasons detailed in the same chapter, the resulting free energy profiles do not follow the same pattern established in the PMFs reported in Chapter 4, and, in fact, report free energy values far exceeding thermal energy (Figure 5.5, page 130).

In future work, it would be interesting to explore system set-ups that more accurately reflect *in vivo* and *in vitro* environments. In interpreting GPCR simulations and results, the composition of the lipid bilayer has gained increasing attention in recent times.

Though a majority of reported simulations performed on GPCRs (and indeed other transmembrane proteins) involve embedding the receptor in a single lipid species

membrane (usually POPC for GPCR simulations; see [21, 25, 26, 178, 207, 208, 209] for examples), biological membranes are far more complex [210]. Recent simulations performed on GPCRs have included more complex lipids, including cholesterol. These bilayers usually include a small percentage of cholesterol (typically 10%, compared to the number of POPCs). More recent simulations have been performed with higher cholesterol content (20%, 40%) and even more native-like membranes that are composed of cholesterol, sphingomyelin, and POPE, among others [211].

In the work reported in this thesis, all receptors were embedded in a 100% POPC bilayer. To include other lipid types would require an intimate understanding of how each lipid species interacts with the receptor [210], and how this would affect the pharmacological properties of each receptor and each ligand. As an example, cholesterol has been shown to affect: the thermal stability of the β_2 adrenoceptor [212] and the oxytocin receptor [213]; the affinity of oxytocin for its native receptor [214], and 8-Hydroxy-N,N-dipropylaminotetralin (8-OH-DPAT) for the 5-HT_{1A} receptor [215]; the oligomerisation properties of the β_2 adrenoceptor [216], of the μ OR [29], of the neurotensin receptor 1 [217], and of the serotonin$_{1A}$ receptor [218]; among other properties. Other lipid species, such as sphingomyelin, monosialodihexosylganglioside, and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG), have also been implicated in *in silico* studies of GPCRs; see, for example, the study by Koldso and Sansom on the oligomerisation of sphingosine 1-phosphate receptor 1 [219] oligomerisation [220] for the former two, or the study by Neale and colleagues on the effect of POPG on stabilising certain states of the β_2 adrenoceptor [30].

Based on the importance of lipids, and their effect on GPCR, it is very likely that their inclusion would alter some *in silico* properties as well. It would be interesting to compare how the inclusion of various lipids affect the allosteric communication between the various binding sites in GPCRs (and even, other cell surface proteins), and how, if in any way, these affect the stability of any metastable states along the GPCR lifecycle.

To gain a better understanding of these properties, however, often requires µs timescale simulations, which is costly. All conventional MD simulations suffer from a common limitation: the current algorithms for controlling bond distances limit the timestep to 2 fs. As explained in Chapter 2, this value is by the slowest motion in any system, which typically is the oscillation of C-H bonds [87].

Newer studies have taken advantage of two methods that bypass this limitation. These focus on using virtual sites (VS) or using hydrogen-mass repartitioning (HMR). While a more sophisticated description of either is beyond the scope of this Thesis,

both methods allow for the removal of the C-H fast vibrations by altering hydrogen atoms (exchanging them for dummy atoms in the case of virtual sites; redistributing it from neighbouring heavy atoms in case of HMR), allowing researchers to increase the timestep to 4 fs [88], and thus perform more affordable, longer timescale conventional and enhanced MD simulations.

Both methods are becoming more common in accelerating the study of GPCRs. Vickery and colleagues used VS in their study of voltage sensing in both M2 muscarinic and δ OR [190]; Koehl and colleagues used HMR on their reported µs timescale conventional MD simulations of the μ OR in complex with the agonist DAMGO [21]. It is important to mention, however, that studies on bulk lipid properties (area per lipid, diffusivity, deuterium order parameters, among others [221]), using VS or HMR, have not yet been reported. The same is true for small molecules.

Appendix A

Code and scripts used in Chapter 3.

This Appendix explains the use of the Mutual Information code developed of Chapter 3, and lists github repositories which store this code for public use.

A.1 Mutual Information

The code used for characterising the trajectories using pairwise mutual information will be available in a public github repository in the near future; the skeleton can be found below.

It is of critical importance that the trajectory be aligned (so as to remove later protein motion) before running this code. Also, it is recommended that the structure before production MD simulations be used (in either GRO or PDB format) when running the code below, and that all waters, lipids, and non-structural ions, be removed from both structure and trajectory files, so as to speed up the analysis.

The following files can be found within this repository:

1. information_v6.py is used to generate the pairwise mutual information between all the non-hydrophobic residues. It is meant for command line usage, with the following syntax:

```
$ python information_v6.py -t structure.pdb -x trajectory.xtc
```

This generates 2 files: mut_inf_all and mut_inf_all_<date>.log, where <date> is the time and date of script running.

The former file is a tab-separated value file containing the pairwise mutual information values between all pairs of residues, while the later is a file for logging errors

and miscellaneous information (such as the version information of packages used). It also logs the differential entropy, and this is meant to be used as a checkpoint file in later versions. The filesnames of these two files can be altered upon script running using the -o/–outname and -l/–log flags, respectively.

The list of hydrophobic residues, which the code is instructed to ignore, is hard-coded:

```
hydrophobics =
    ['ALA', 'VAL', 'ILE', 'LEU', 'MET', 'PHE', 'PRO', 'GLY', 'SEC', 'CYS']
```

In future versions, analysing all residues, as opposed to just hydrophobic residues, will be made into a flag.

In this version, the reading of coordinates from input trajectories is done in memory, as opposed to in place (i.e. iteratively reading the coordinate file). In using this version, it becomes important to minimise the size of the input trajectory to avoid segmentation faults and other memory-related errors.

The non-mass weighed covariance matrix is calculated using MDAnalysis, modified from

```
coordinates = ensemble.trajectory.timeseries(ensemble.select\_atoms(selection),
                                             order='fac')
```

to

```
coordinates = AnalysisFromFunction(lambda ag: ag.positions.copy(),
                  ensemble.select\_atoms(selection)).run().results
```

where AnalysisFromFunction is a method in MDAnalysis.analysis.base. This replacement was necessary since GROMACS generated XTC files do not contain time-series information.

2. information_v6.1.py is a version of the above, extending the analysis to polar hydrogens. It is not yet recommended for use.

3. networkx_from_mutinf_v1.py is used to generate paths. This script is meant for iPython/Jupyter Notebook usage. It takes advantage of the NetworkX Python package to create networks using residues as nodes and mutual information data as edges.

4. information_v5.joblib.py uses the joblib package to distribute one of the main loops to multiple CPU cores.

5. information_v5.joblib.v2.py uses the joblib package to distribute both main loops to multiple CPU cores.

The final two modifications to the code would not have been possible without the intrumental help of Dr Irfan Alibay.

Appendix B

Code and scripts used in Chapter 4.

B.1 FNC generation

The code for generating FNCs can be found in https://github.com/ShadSBCB/python-tools/tree/thesis. Within this repository, the following files can be found:

1. plumed.py is the script used to create FNCs for two structures. It is meant for command line usage, with the following syntax:

$ python plumed.py inactive_structure.gro active_structure.gro

The code is also displayed below (generated for LaTeXrepresentation using the Pygments Python library).

```
from MDAnalysis import Universe as u
from MDAnalysis.lib import distances as cdists
import numpy as np


#Map to nearest non-symmetrical atom (often C, also N and O and S)

selection = """
protein and not name H* and (
   (resname ALA and (backbone or name CB))
or (resname ARG and (backbone or name CB or name CG or name NE))
or (resname ASN and (backbone or name CG))
or (resname ASP and (backbone or name CG))
```

```
or (resname CYS and (backbone or name SG))
or (resname GLN and (backbone or name CB or name CD))
or (resname GLU and (backbone or name CB or name CD))
or (resname GLY)
or (resname HIS and (backbone or name CE1 or name NE2))
or (resname ILE and (backbone or name CB))
or (resname LEU and (backbone or name CG))
or (resname LYS and (backbone or name CB or name CZ or name NZ))
or (resname MET and (backbone or name CB or name SD))
or (resname SER and (backbone or name OG))
or (resname PHE and (backbone or name CG or name CZ))
or (resname PRO and (backbone or name CD))
or (resname THR and (backbone or name CB))
or (resname TRP and (name CE3 or name NE1))
or (resname TYR and (backbone or name CG or name CZ))
or (resname VAL and (backbone or name CB))
)
"""

#Load receptors into universe
inactive = u('Inactive.gro')
active = u('Active.gro')

SCALING_FACTOR = 1.8

def generate_contacts(
universe,
selection,
upper_cutoff=6.0,
lower_cutoff=0.0,
resid_separation=3
):

gr = universe.select_atoms(selection)
distances = cdists.distance_array(gr.positions, gr.positions)
```

```
triu = np.triu(distances)
mask = np.argwhere((triu < upper_cutoff) & (triu > lower_cutoff))

contacts = [(atom1, atom2) for atom1, atom2 in mask \
if not abs(gr.atoms[atom1].resid -
gr.atoms[atom2].resid) <= resid_separation]
return gr, triu, set([ frozenset(pair) for pair in contacts ])

def generate_lines(gr, contacts):

return [
"ATOMS%s=%s,%s SWITCH%s={EXP R_0=0.02000 D_0=%.4f} WEIGHT%s=%.6f\n"
% (i+1, gr[idA].id+1, gr[idB].id+1, i+1,
distance * 0.1 * SCALING_FACTOR, i+1, 1.0/len(contacts)) \
for i, ((idA, idB), distance) in enumerate(contacts.items())
]



atoms_active, triu_active, contacts_active = generate_contacts(active,
selection)
atoms_inactive, triu_inactive, contacts_inactive = generate_contacts(inactive,
selection)

unique_active_contacts = contacts_active - contacts_inactive
unique_inactive_contacts = contacts_inactive - contacts_active

contacts_active = {}
for idA, idB in unique_active_contacts:
da = triu_active[idA, idB] or triu_active[idB, idA]
di = triu_inactive[idA, idB] or triu_inactive[idB, idA]
if min(da, di) * SCALING_FACTOR > max(da, di): continue
#print("active", da, di)
contacts_active[(idA, idB)] = da

contacts_inactive = {}
for idA, idB in unique_inactive_contacts:
```

```
da = triu_active[idA, idB] or triu_active[idB, idA]
di = triu_inactive[idA, idB] or triu_inactive[idB, idA]
if min(da, di) * SCALING_FACTOR > max(da, di): continue
#print("inactive", da, di)
contacts_inactive[(idA, idB)] = di

with open('plumed.dat', 'w') as f:

f.write("CONTACTMAP ...\n")
f.write("".join(generate_lines(atoms_inactive, contacts_inactive)))
f.write("LABEL=cmap_inactive\n")
f.write("SUM\n")
f.write("... CONTACTMAP\n\n")

f.write("CONTACTMAP ...\n")
f.write("".join(generate_lines(atoms_active, contacts_active)))
f.write("LABEL=cmap_active\n")
f.write("SUM\n")
f.write("... CONTACTMAP\n")

f.write("cmap: COMBINE ARG=cmap_inactive,cmap_active
COEFFICIENTS=1,-1 PERIODIC=NO\n")

f.write("PRINT ARG=* FILE=COLVAR")
```

This script generates a file called plumed.dat which contains the list of atom pairs contained within the CONTACTMAP directive. This file is ready for use in plumed. In future, this script will be re-written to allow users to change some of the currently hard-coded parameters (such as the SCALING FACTOR, or residue separation).

2. plumed_residues.py is a supplementary script used to create a pml or tcl script for visualisation of residues in either PyMOL or VMD, respectively. Usage is the same as plumed.py script.

3. readAndPlotCOLVAR_v2.py is a script developed for reading COLVAR files from PLUMED runs. Usage:

```
$ python readAndPlotCOLVAR.py -cv /path/to/COLVAR --plot \
--field1 field1 --field2 field2
```

field1 and field2 are fields in the COLVAR file. This returns a file named field1_vs_field2_COLVAR.svg (a simple plot of the data in fields 1 and 2).

```
$ python readAndPlotCOLVAR.py -cv /path/to/COLVAR --save \
--field1 field1 --field2 field2
```

as above. This returns a tab-separated file named field1_vs_field2_COLVAR.csv.

```
$ python readAndPlotCOLVAR.py -cv /path/to/COLVAR --read
```

returns all the fields available for plotting/saving.

Appendix C

Further details on FNC changes

This Appendix pertains to the data presented in Chapter 4.

C.1 Additional forward transition test systems

C.2 Convergence of PMFs

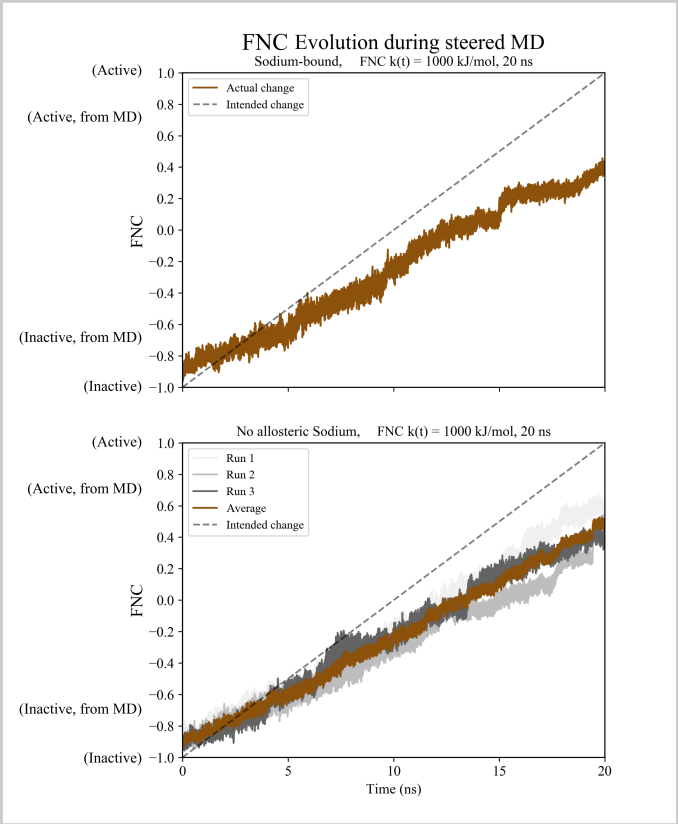


Figure C.1: 20 ns simulations were performed to attempt a forward transition (inactive state to active state) using a force constant of 1000 kJ/mol. One transition was attempted in a sodium-bound receptor (top plot), whereas three were performed in sodium-absent receptors (bottom plot). This value was deemed too weak as it doesn't force the system to traverse the 0.5 FNC value. For reference, a system was considered active if the FNC CV adopted a value of 0.7 or above (see Figure 4.1.)

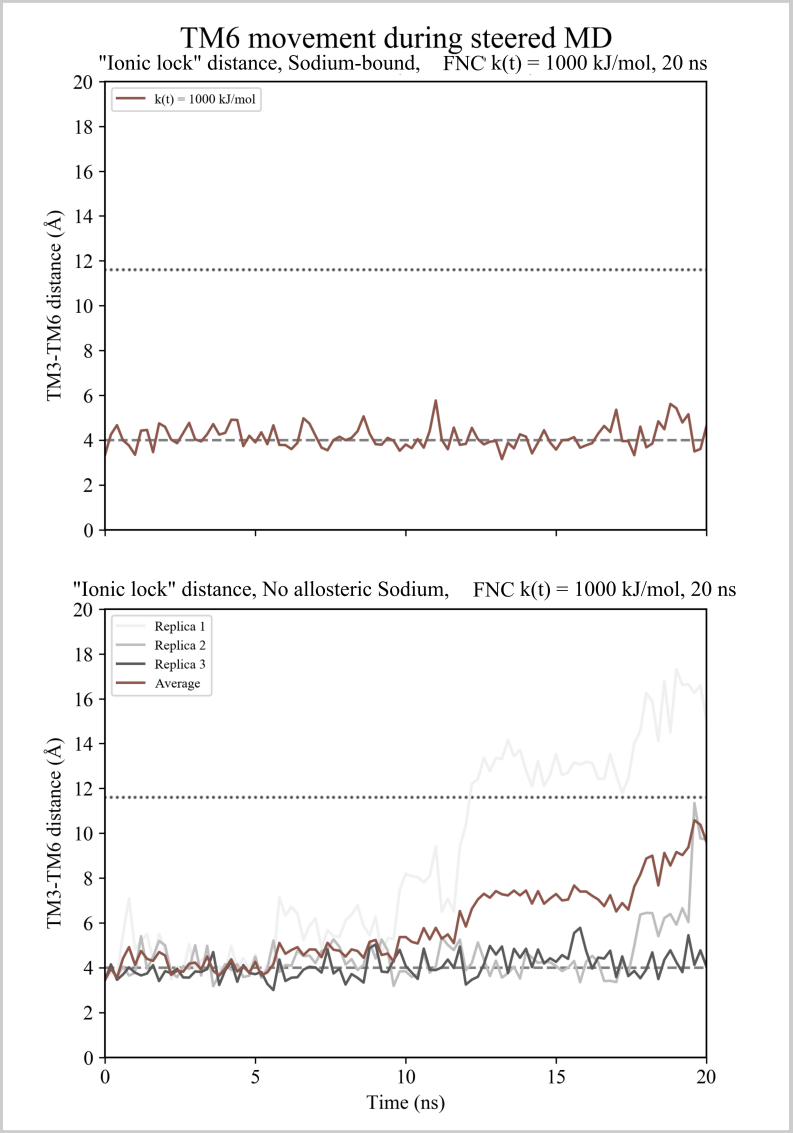


Figure C.2: A force constant of 1000 kJ/mol fails to transition the μOR from an inactive-like to an active-like state based on the "ionic lock" distance. Top and bottom plots represent simulations in the presence and absence of an allosteric sodium. The dashed lines follow the same convention as Figure 4.3.

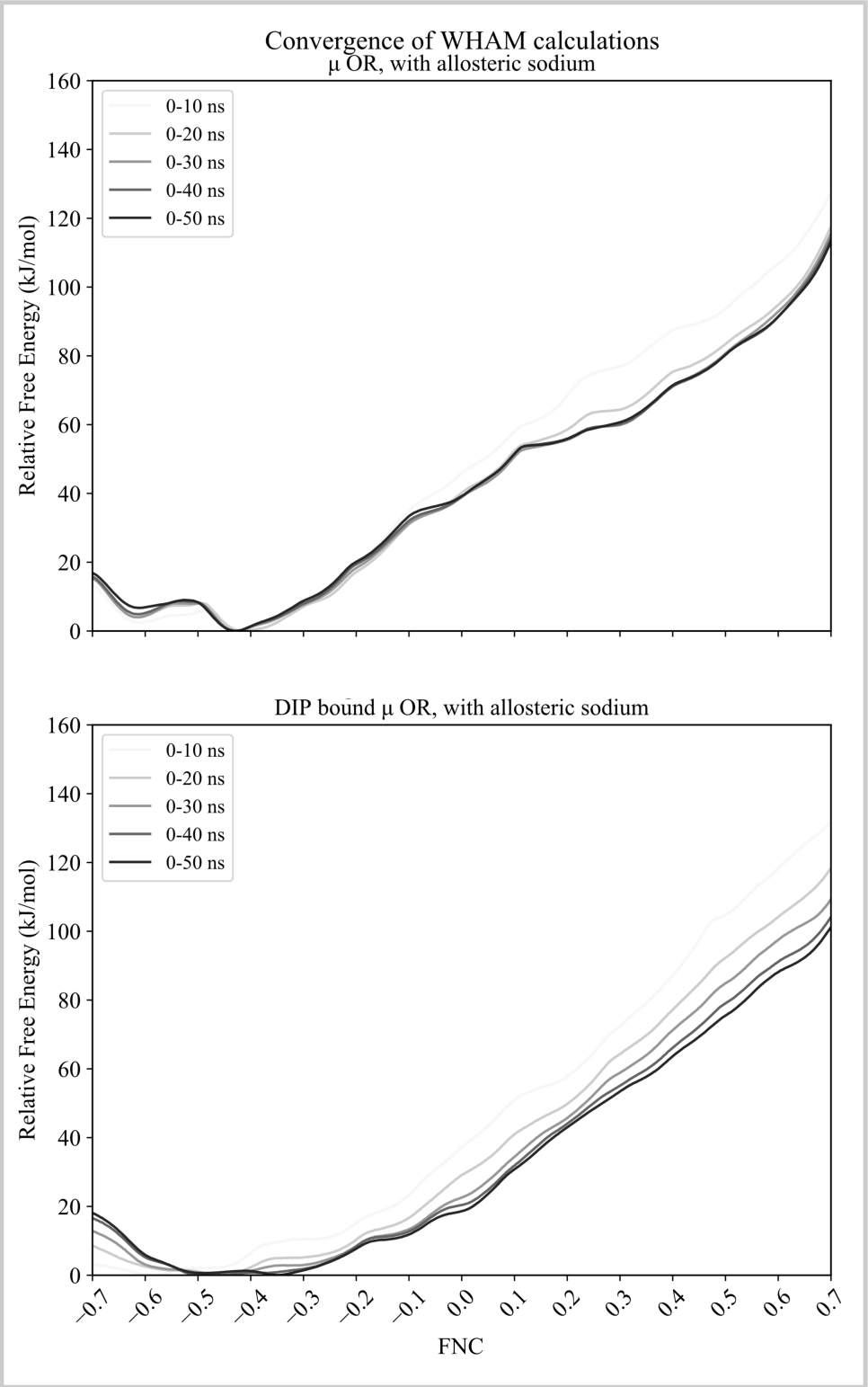


Figure C.3: Convergence of PMF calculations in the simulations presented, in the absence or presence of DIPP-NH_2. These show convergence at 50 ns of simulation.

Appendix D

Supplemental data for Chapter 5

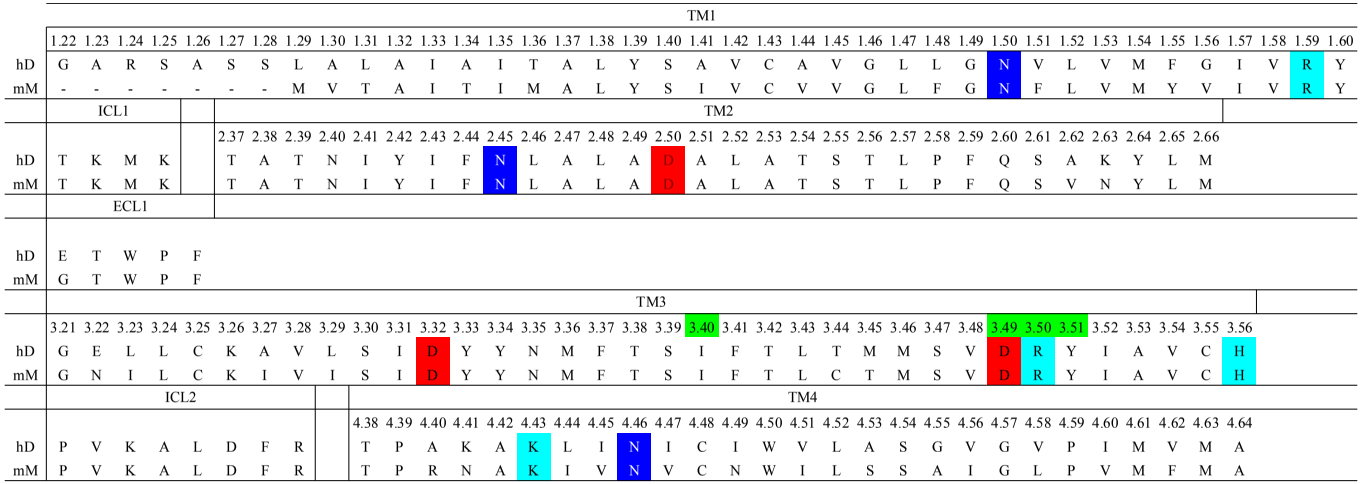


Figure D.1: Sequence alignment (from TM1 to TM4) between the *homo sapiens* δ OR (hD) and the *mus musculus* μ OR (mM), generated using the GPCRDB [19, 20]. Residues in the same location in the two receptors are highlighted in the following fashion: Arg/Lys/His, cyan; Asp/Glu, red; Asn, blue. Motifs descriped in Chapter 1 are highlighted in green.

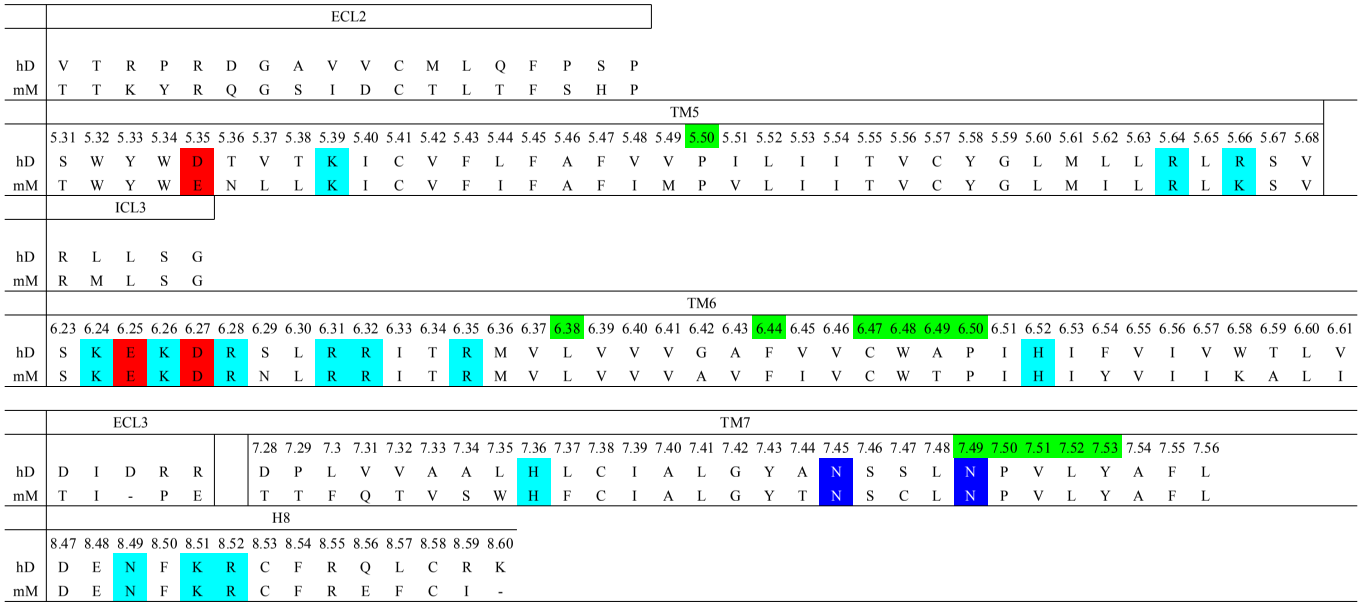


Figure D.2: Sequence alignment (from ECL2 to H8) between the *homo sapiens* δ OR (hD) and the *mus musculus* μ OR (mM), generated using the GPCRDB [19, 20]. Residue colouring follows Figure D.1.

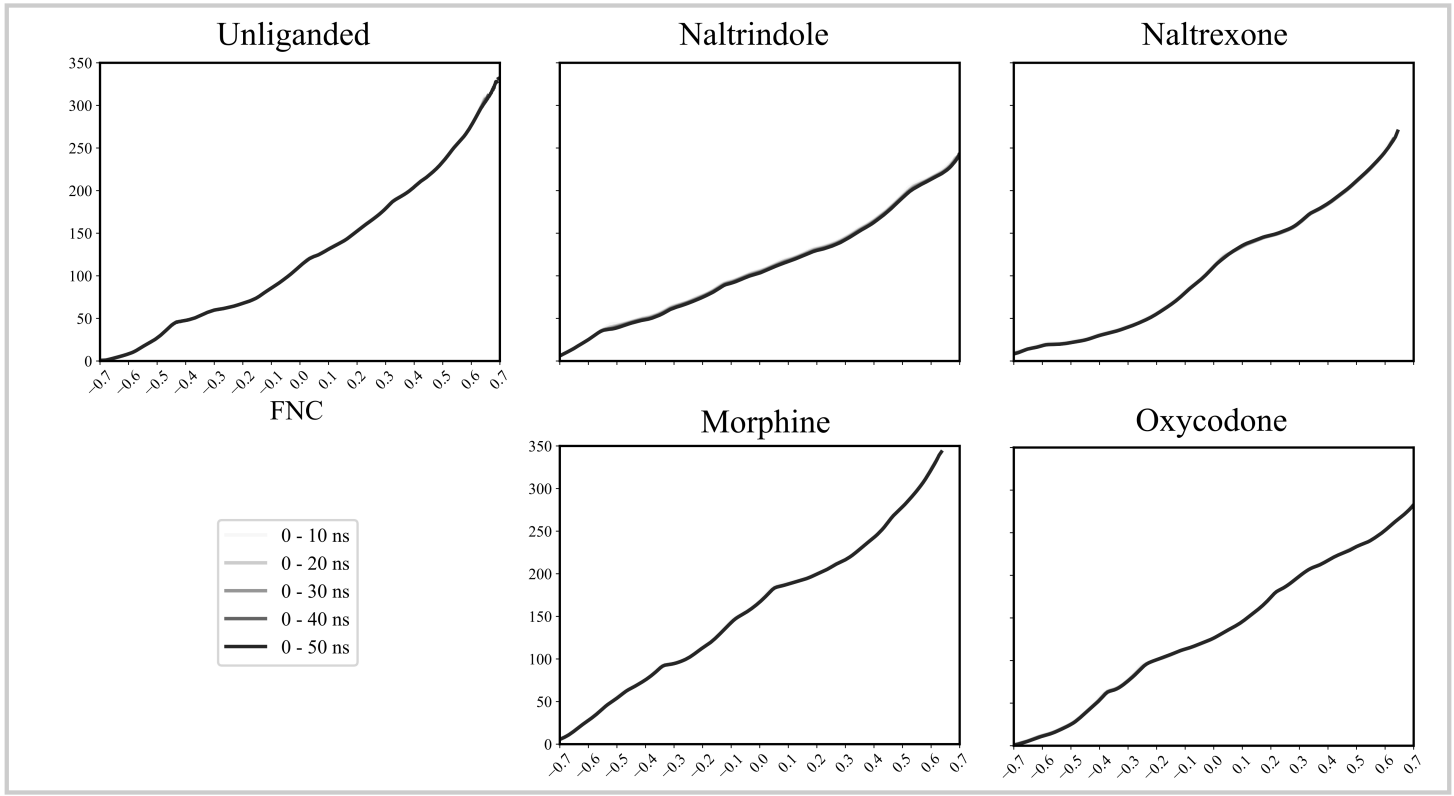


Figure D.3: Convergence of PMF calculations for each condition reported in Chapter 5.

References

1. Fredriksson, R. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* **63,** 1256–1272 (2003).
2. Shonberg, J., Kling, R. C., Gmeiner, P. & Löber, S. GPCR crystal structures: Medicinal chemistry in the pocket. *Bioorgan Med Chem* **23,** 3880–3906 (2015).
3. Cvicek, V., Goddard, W. A. & Abrol, R. Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLoS Comput Biol* **12,** 1–16 (2016).
4. Filmore, D. It's a GPCR world. *Modern D Discov* **7,** 24–27 (2004).
5. Salon, J. A., Lodowski, D. T. & Palczewski, K. The Significance of G Protein-Coupled Receptor Crystallography for Drug Discovery. *Pharmacol Rev* **63,** 901–937 (2011).
6. Hauser, A. S. *et al.* Pharmacogenomics of GPCR drug targets. *Cell* **172,** 41–54 (2018).
7. Sriram, K. & Insel, P. A. GPCRs as targets for approved drugs: How many targets and how many drugs? *Mol Pharma* **93,** 251–258 (4 2018).
8. Martins, S. M. A., Trabuco, J. R. G., Monteiro, G. A. & Prazeres, D. M. GPCR screening and drug discovery: Challenges and latest trends. *Euro Pharmac Rev* **17** (2012).
9. Schertler, G. F. X., Villa, C. & Henderson, R. Projection structure of rhodopsin. *Nature* **362,** 770 (1993).
10. Palczewski, K. *et al.* Crystal Structure of Rhodopsin : A G Protein – Coupled Receptor. *Science* **289,** 739–745 (2000).
11. Cooke, R. M., Brown, A. J. H., Marshall, F. H. & Mason, J. S. Structures of G protein-coupled receptors reveal new opportunities for drug discovery. *Drug Discov Today* **20,** 1355–1364 (2015).
12. Liu, W. *et al.* Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. *Science* **337,** 232–236 (2012).
13. Fenalti, G. *et al.* Molecular control of δ-opioid receptor signalling. *Nature* **506,** 191–196 (2014).

14. Ballesteros, J. A. & Weistein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein coupled receptors. *Methods Neurosci* **25,** 366–428 (1995).

15. Nygaard, R., Frimurer, T. M., Holst, B., Rosenkilde, M. M. & Schwartz, T. W. Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol Sci* **30,** 249–259 (2009).

16. Oliveira, L., Paiva, A. C. M. & Vriend, G. A common motif in G-protein-coupled seven transmembrane helix receptors. *J Comput Aided Mol Des* **7,** 649–658 (Dec. 1993).

17. Baldwin, J. M. The probable arrangement of the helices in G protein-coupled receptors. eng. *EMBO* **12,** 1693–1703 (Apr. 1993).

18. Schwartz, T. W. Locating ligand-binding sites in 7TM receptors by protein engineering. *Curr Opin Biotechnology* **5,** 434–444 (Aug. 1994).

19. Isberg, V. *et al.* Generic GPCR residue numbers - aligning topology maps while minding the gaps. English. *Trends Pharmacol Sci* **36,** 22–31 (Jan. 2015).

20. Isberg, V. *et al.* GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* **44,** 356–364 (2015).

21. Koehl, A. *et al.* Structure of the μ-opioid receptor–Gi protein complex. *Nature* **558,** 547–552 (2018).

22. Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR Dynamics: Structures in Motion. *Chemical Reviews,* acs.chemrev.6b00177 (2016).

23. Ozcan, O., Uyar, A., Doruker, P. & Akten, E. D. Effect of intracellular loop 3 on intrinsic dynamics of human β2-adrenergic receptor. *BMC Structural Biology* **13** (2013).

24. Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human b2-Adrenergic G Protein–Coupled Receptor. *Science* **318,** 1258–1265 (2007).

25. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *P Natl Acad Sci USA* **108,** 13188–13123 (2011).

26. Dror, R. O. *et al.* Activation mechanism of the β2-adrenergic receptor. *Proceedings of the National academy of sciences of the USA* **108,** 18684–18689 (2011).

27. Chini, B. & Parenti, M. G-protein-coupled receptors, cholesterol and palmitoylation: Facts about fats. *Journal of Molecular Endocrinology* **42,** 371–379 (2009).

28. Sensoy, O. & Weinstein, H. A mechanistic role of Helix 8 in GPCRs: Computational modeling of the dopamine D2 receptor interaction with the GIPC1-PDZ-domain. *BBA - Biomembranes* **1848,** 976–983 (2015).

29. Zheng, H. *et al.* Palmitoylation and membrane cholesterol stabilize μ-opioid receptor homodimerization and G protein coupling. *BMC Cell Biology* **13,** 1–18 (2012).

30. Neale, C., Herce, H. D., Pomès, R. & García, A. E. Can Specific Protein-Lipid Interactions Stabilize an Active State of the Beta 2 Adrenergic Receptor? *Biophysical Journal* **109,** 1652–1662 (2015).

31. Vogel, R. *et al.* Functional Role of the "Ionic Lock"—An Interhelical Hydrogen-Bond Network in Family A Heptahelical Receptors. *J Mol Biol* **380,** 648–655 (2008).

32. Fritze, O. *et al.* Role of the conserved NPxxY(x)(5,6)F motif in the rhodopsin ground state and during activation. *P Natl Acad Sci USA* **100,** 2290–2295 (Mar. 2003).

33. Wacker, D. *et al.* Structural features for functional selectivity at serotonin receptors. *Science* **340,** 615–619 (2013).

34. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).

35. Nygaard, R. *et al.* The Dynamic Process of β2-Adrenergic Receptor Activation. *Cell* **152,** 532–542 (2013).

36. Manglik, A. *et al.* Structural insights into the dynamic process of β¡inf¿2¡/inf¿-adrenergic receptor signaling. *Cell* **161,** 1101–1111 (2015).

37. Smith, C. A. *et al.* Population shuffling of protein conformations. *Angewandte Chemie - International Edition* **54,** 207–210 (2015).

38. Hammes-Schiffer, S. & Benkovic, S. J. Relating Protein Motion to Catalysis. *Annual Review of Biochemistry* **75,** 519–541 (2006).

39. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450,** 913–916 (2007).

40. Pasternak, G. W. & Pan, Y.-X. Mu opioids and their receptors: evolution of a concept. *Pharmacological reviews* **65,** 1257–317 (2013).

41. Corbett, A. D., Henderson, G., McKnight, A. T. & Paterson, S. J. 75 years of opioid research: the exciting but vain quest for the Holy Grail. *Brit J Pharmachol* **147 Suppl,** S153–62 (2006).

42. Pert, C. B., Pasternak, G. W. & Snyder, S. H. Opiate Agonists and Antagonists Discriminated by Receptor Binding in Brain. *Source: Science, New Series* **182,** 1359–1361 (1973).

43. Trescot, A. M., Datta, S., Lee, M. & Hansen, H. Opioid Pharmacology. *Pain Physician* **1975,** S133–S153 (2008).

44. Martin, W. R., Eades, C. G., Thompson, J. A., Huppler, R. E. & Gilbert, P. E. The Effects of Morphine- and Nalorphine-like Drugs in the Nondependent and Morphine-dependent Chronic Spinal Dog. *The Journal of Pharmacology and Experimental Therapeutics* **197,** 517–532 (1976).

45. Martin, W. R. History and development of mixed opioid agonists, partial agonists and antagonists. *British Journal of Clinical Pharmacology* **7,** 273S–279S (1979).

46. Hughes, J. *et al.* Identification of two related pentapeptides from the brain with potent opiate agonist activity. *Nature* **258,** 577–579 (1975).

47. Benyhe, S., Zádor, F. & Ötvös, F. Biochemistry of opioid (morphine) receptors: Binding, structure and molecular modelling. *Acta Biol Szeged* **59,** 17–37 (2015).

48. Gendron, L., Cahill, C. M., Zastrow, M. V., Schiller, P. W. & Pineyro, G. Molecular Pharmacology of δ Opioid Receptors. *Pharmacol Rev* **324876,** 631–700 (2016).

49. Lord, J. A. H., Waterfield, A. A., Hughes, J. & Kosterlitz, H. W. Endogenous opioid peptides: Multiple agonists and receptors. *Nature* **267,** 495–499 (1977).

50. Li, C. H. & Chung, D. Isolation and structure of an untriakontapeptide with opiate activity from camel pituitary glands. *Proceedings of the National Academy of Sciences* **73,** 1145–1148 (1976).

51. Goldstein, A., Fischli, W., Lowney, L. I., Hunkapiller, M. & Hood, L. Porcine pituitary dynorphin: complete amino acid sequence of the biologically active heptadecapeptide. *Proceedings of the National Academy of Sciences of the United States of America* **78,** 7219–23 (1981).

52. Evans, C. J., Keith, D. E., Morrison, H., Magendzo, K. & Edwards, R. H. Cloning of a delta opioid receptor by functional expression. *Science* **258,** 1952–1955 (1992).

53. Kieffer, B. L., Befort, K., Gaveriaux-Ruff, C. & Hirth, C. G. The delta-opioid receptor: isolation of a cDNA by expression cloning and pharmacological characterization. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 12048–52 (1992).

54. Wu, H. *et al.* Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* **485,** 327–332 (2012).

55. Huang, W. *et al.* Structural insights into µ-opioid receptor activation. *Nature* **524,** 315–321 (2015).

56. Filizola, M. & Devi, L. A. Grand opening of structure-guided design for novel opioids. *Trends in Pharmacological Sciences* **34,** 6–12 (2013).

57. Milligan, G. & Kostenis, E. Heterotrimeric G-proteins: a short history. *British Journal of Pharmacology* **147,** S46–S55 (2009).

58. Johnson, B. D., Scheuer, T. & Catterall, W. A. Voltage-dependent potentiation of L-type Ca2+ channels in skeletal muscle cells requires anchored cAMP-dependent protein kinase. *Proc Natl Acad Sci U S A* **91,** 11492–11496 (1994).

59. Surratt, C. K. & Adams, W. R. G Protein-Coupled Receptor Structural Motifs: Relevance to the Opioid Receptors. *Current Topics in Medicinal Chemistry* **5,** 315–324 (2005).

60. Martí-solano, M. *et al.* Detection of New Biased Agonists for the Serotonin 5-HT 2A Receptor : Modeling and Experimental Validation s, 740–746 (2015).

61. Manglik, A. *et al.* Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature* **485,** 321–6 (2012).

62. Che, T. *et al.* Structure of the Nanobody-Stabilized Active State of the Kappa Opioid Receptor. *Cell* **172,** 55–67.e15 (2018).

63. Granier, S. *et al.* Structure of the δ-opioid receptor bound to naltrindole. *Nature* **485,** 400–4 (2012).

64. Fenalti, G. *et al.* Structural basis for bifunctional peptide recognition at human δ-opioid receptor. *Nature Structural & Molecular Biology* **22,** 1–6 (2015).

65. Charbogne, P., Kieffer, B. L. & Befort, K. 15 years of genetic approaches in vivo for addiction research: Opioid receptor and peptide gene knockout in mouse models of drug abuse. *Neuropharmacol* **76,** 204–217 (2014).

66. Soergel, D. G. *et al.* Biased agonism of the μ-opioid receptor by TRV130 increases analgesia and reduces on-target adverse effects versus morphine: A randomized, double-blind, placebo-controlled, crossover study in healthy volunteers. *Pain* **155,** 1829–1835 (2014).

67. Morgan, M. M. & Christie, M. J. Analysis of opioid efficacy, tolerance, addiction and dependence from cell culture to human. *Brit J Pharmacol* **164,** 1322–1334 (2011).

68. Banks, M. L., Roma, P. G., Folk, J. E., Rice, K. C. & Negus, S. S. Effects of the delta-opioid agonist SNC80 on the abuse liability of methadone in rhesus monkeys: A behavioral economic analysis. *Psychopharmacol* **216,** 431–439 (2011).

69. Goldenberg, D. L. The Interface of Pain and Mood Disturbances in the Rheumatic Diseases. *Semin Arthritis Rheu* **40,** 15–31 (2010).

70. Chu Sin Chung, P. & Kieffer, B. L. Delta opioid receptors in brain function and diseases. *Pharmacology & Therapeutics* **140,** 112–120 (2013).

71. Macht, D. The history of opium and some of its preparation and alkaloids. *J Am Med Assoc* **64,** 461–477 (1915).

72. Serturner, F. Darstellung der reinen mohnsaure (opiumsaure) nebsteiner chemischen untersuchung des opiums mit vorzuglicher hinsicht auf einen darin neu entdecken stoff und die dahin gehörigen bemerkungen. *Journal der Pharmazie fur Aerzte, Apotheker und Chemisten* **14,** 47–93 (1806).

73. Mignat, C., Wille, U. & Ziegler, A. Affinity profiles of morphine, codeine, dihydrocodeine and their glucuronides at opioid receptor subtypes. *Life Sciences* **56,** 793–799 (1995).

74. Schwyzer, R. Acth: a Short Introductory Review. *Annals of the New York Academy of Sciences* **297,** 3–26 (1977).

75. Shen, Q. *et al.* Discovery of Potent and Selective Agonists of δ Opioid Receptor by Revisiting the "message-Address" Concept. *ACS Medicinal Chemistry Letters* **7,** 391–396 (2016).

76. Cheng, J. X. *et al.* Computational insights into the G-protein-biased activation and inactivation mechanisms of the μ opioid receptor. *Acta Pharmacologica Sinica* **39,** 154–164 (2018).

77. Portoghese, P. S., Moe, S. T. & Takemori, A. E. A selective $\delta 1$ opioid receptor agonist derived from oxymorphone. Evidence for separate recognition sites for $\delta 1$ opioid receptor agonists and antagonists. *Journal of medicinal chemistry* **36,** 2572–4 (1993).

78. Karplus, M. & McCammon, A. J. Molecular dynamics simulations of biomolecules. *Nature Struct Bio* **9,** 646–652 (2002).

79. Adcock, S. A. & McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem Rev* **106,** 1589–1615 (2006).

80. McRobb, F. M., Negri, A., Beuming, T. & Sherman, W. Molecular dynamics techniques for modeling G protein-coupled receptors. *Curr Opin Pharmacol* **30,** 69–75 (2016).

81. McCammon, A. J., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267,** 585–590 (1977).

82. Liu, W., Schmidt, B., Voss, G. & Müller-Wittig, W. Accelerating molecular dynamics simulations using Graphics Processing Units with CUDA. *Compu Phys Comm* **179,** 634–641 (2008).

83. Bowers, K. J. *et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters* in *Proc 2006 ACMI/IEEE Conf Supercomput* (ACM Press, 2006).

84. Shaw, D. E. *et al. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation* in *Proc 34th Annual Inter Symp Comput Architec* (2008), 91–97.

85. Velgy, N., Hedger, G. & Biggin, P. C. in *Computational Methods for GPCR Drug Discovery* (ed Heifetz, A.) 133–158 (Humana Press, 2018).

86. Leach, A. R. *Molecular Modelling: Principles and Applications* (Prentice Hall, 2001).

87. Mazur, A. K. Symplectic integration of closed chain rigid body dynamics with internal coordinate equations of motion. *J Chem Phys* **111,** 1407–1414 (1999).

88. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich Systems. *J Comput Chem* **20,** 786–798 (8 1999).

89. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry,* 952–962 (1992).

90. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18,** 1463–1472 (1997).

91. Klepeis, J. L., Lindorff-larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology* **19,** 120–127 (2009).

92. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load balanced, and scalable molecular simulations. *J. Chem. Theory and Comput.* **4,** 435–447 (2008).

93. Essmann, U. *et al.* A smooth particle mesh Ewald method Ulrich. *The Journal of Chemical Physics* **103,** 8577–8593 (1995).

94. LeVine, M. V. & Weinstein, H. NbIT - A New Information Theory-Based Analysis of Allosteric Mechanisms Reveals Residues that Underlie Function in the Leucine Transporter LeuT. *PLoS Comput Biol* **10,** e1003603 (5 2014).

95. MacKay, D. J. C. *Information Theory, Inference, and Learning Algorithms David J.C. MacKay* 1–640. arXiv: `arXiv:1011.1669v3` (2005).

96. Schneider, S., Provasi, D. & Filizola, M. How Oliceridine (TRV-130) Binds and Stabilizes a μ-Opioid Receptor Conformational State that Selectively Triggers G Protein-Signaling Pathways. *Biochem* **55,** 6456–6466 (2016).

97. Tiberti, M., Papaleo, E., Bengtsen, T., Boomsma, W. & Lindorff-Larsen, K. ENCORE: Software for Quantitative Ensemble Comparison. *PLOS Computational Biology* **11,** e1004415 (Oct. 2015).

98. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comp Chem* **32,** 2319–2327 (2011).

99. Gowers, R. J. *et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations* in *{P}roceedings of the 15th {P}ython in {S}cience {C}onference* (eds Benthall, S. & Rostrup, S.) (2016), 98–105.

100. Matsuda, H. Physical nature of higher-order mutual information: Intrinsic correlations and frustration. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* **62,** 3096–3102 (2000).

101. Lange, O. F. & Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins: Structure, Function and Genetics* **62,** 1053–1061 (2006).

102. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy 2008),* 11–15 (2008).

103. Stansfeld, P. J. *et al.* MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **23,** 1–12 (2015).

104. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & Vries, A. H. D. The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations. *Journal of Physical Chemistry B* **111,** 7812–7824 (2007).

105. Wassenaar, T. A., Pluhackova, K., Bockmann, R. A., Marrink, S. J. & Tieleman, D. P. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation* **10,** 676–690 (2014).

106. Bussi, G. *et al.* Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **014101,** 014101–1–7 (2007).

107. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation* **4,** 116–122 (2008).

108. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* Nov. 2015.

109. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of single potential functions for simulating liquid water. *J Chem Phys* **79,** 926–35 (1983).

110. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics* **78,** 1950–1958 (2010).

111. Jämbeck, J. P. M. & Lyubartsev, A. P. An extension and further validation of an all-atomistic force field for biological membranes. *Journal of Chemical Theory and Computation* **8,** 2938–2948 (2012).

112. Jämbeck, J. P. M. & Lyubartsev, A. P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *Journal of Physical Chemistry B* **116,** 3164–3179 (2012).

113. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general Amber force field. *Journal of Computational Chemistry* **25,** 1157–1174 (2004).

114. Da Silva, A. W. S. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC research notes* **5,** 367 (2012).

115. Case, D. A. *et al. Amber 14* University of California, San Francisco (2014).

116. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25 (2015).

117. Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy Array : A Structure for Efficient Numerical Computation. *Comp Sci Eng,* 22–30 (2011).

118. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9,** 90–95 (2007).

119. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14,** 33–38 (1996).

120. Domicevica, L., Paramo, T., Newport, T. & Biggin, P. C. *LINTools - dynamic representation of protein- ligand interactions* Jan. 2016.

121. Rodríguez-Espigares, I., Kaczor, A. A. & Selent, J. In Silico Exploration of the Conformational Universe of GPCRs. *Molecular Informatics* **35,** 227–237 (6 2016).

122. Pierce, L. C. T. *et al.* Routine access to millsecond timescale evenrs with accelerated molecular dynamics. *Journal of chemical theory and computation* **8,** 2997–3002 (2012).

123. Kohlhoff, K. J. *et al.* Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature chemistry* **6,** 15–21 (2014).

124. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185,** 604–613 (2014).

125. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences* (2013).

126. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28,** 235–242 (2000).

127. Sali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Bio* **234,** 779–815 (1993).

128. Pietrucci, F. & Laio, A. A collective variable for the efficient exploration of protein beta-sheet structures: Application to SH3 and GB1. **5,** 2197–2201 (2009).

129. Grubmüller, H., Heymann, B. & Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **271,** 997–999 (1996).

130. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23,** 187–199 (1977).

131. Schiller, P. W. *et al.* The opioid mu agonist/delta antagonist DIPP-NH(2)[Psi] produces a potent analgesic effect, no physical dependence, and less tolerance than morphine in rats. *Journal of medicinal chemistry* **42,** 3520–3526 (1999).

132. Souaille, M. & Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comp. Phys. Comm.* **135,** 40–57 (2001).

133. Grossfield, A. *WHAM: the weighted histogram analysis method, version 2.0.9* 2013.

134. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *The Journal of Chemical Physics* **123,** 144104 (2005).

135. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *Journal of Chemical Physics* **129.** arXiv: 0801.1426 (2008).

136. Lee, T. S., Radak, B. K., Pabis, A. & York, D. M. A new maximum likelihood approach for free energy profile construction from molecular simulations. *Journal of Chemical Theory and Computation* **9,** 153–164 (2013).

137. Lee, T. S., Radak, B. K., Huang, M., Wong, K. Y. & York, D. M. Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach. *Journal of Chemical Theory and Computation* **10,** 24–34 (2014).

138. Goodman, A. J., Le Bourdonnec, B. & Dolle, R. E. *Mu opioid receptor antagonists: Recent developments* 2007.

139. Van Rijn, R. M., Defriel, J. N. & Whistler, J. L. Pharmacological traits of delta opioid receptors: Pitfalls or opportunities? **228,** 1–18 (2013).

140. Haddou, T. B. *et al.* Pharmacological investigations of N-substituent variation in morphine and oxymorphone: Opioid receptor binding, signaling and antinociceptive activity. *PLoS ONE* **9,** 8–13 (2014).

141. Decaillot, F. M. *et al.* Opioid receptor random mutagenesis reveals a mechanism for G protein-coupled receptor activation. *Nat Struct Biol* **10,** 629–636 (2003).

142. Singh, J. K., Maniyar, C. & Shirsath, V. S. Development of time-resolved fluorescent based [E u] -GTP binding assay for selection of human Histamine 3 receptor antagonists / inverse agonist : a potential target for Alzheimer ' s treatment. *Annals of Neurosciencees* **19,** 71–75 (2012).

143. Irwin, J. J. & Shoichet, B. K. ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling* **45,** 177–182 (2005).

144. Negri, A. *et al.* Discovery of a Novel Selective Kappa-Opioid Receptor Agonist Using Crystal Structure-Based Virtual Screening. *J Chem Inf Model* **53,** 521–526 (2013).

145. Manglik, A. *et al.* Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537,** 1–6 (2016).

146. Hughes, J. P., Rees, S. S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *British Journal of Pharmacology* **162,** 1239–1249 (2011).

147. Dosa, P. I. & Amin, E. A. Tactical Approaches to Interconverting GPCR Agonists and Antagonists. **59,** 810–840 (2016).

148. Unna, K. Antagonistic Effect of N-allyl-normorphine upon Morphine. *Journal of Pharmacology and Experimental Therapeutics* **79,** 27–31 (1943).

149. Lasagna, L. & Beecher, H. K. The Analgesic Effectiveness of Nalorphine and Nalorphine-Morphine Combinations in Man. *J Pharmacol Exp Ther* **112,** 356–363 (1954).

150. Weiss, U. & Daum, S. J. Derivatives of Morphine. IV. ¡sup¿1¡/sup¿ 14-Hydroxymorphine and 14-Hydroxydihydromorphine. *Journal of Medicinal Chemistry* **8,** 123–125 (1965).

151. Mouaffak, F. *et al.* Naltrexone in the treatment of broadly defined behavioral addictions: A review and meta-analysis of randomized controlled trials. *European Addiction Research* **23,** 204–210 (2017).

152. Chavkin, C. The therapeutic potential of κ-opioids for treatment of pain and addiction. *Neuropsychopharmacology* **36,** 369–370 (2011).

153. Peng, J., Sarkar, S. & Chang, S. L. Opioid receptor expression in human brain and peripheral tissues using absolute quantitative real-time RT-PCR. *Drug and Alcohol Dependence* **124,** 223–228 (2012).

154. Filliol, D. *et al.* Mice deficient for δ- and μ-opioid receptors exhibit opposing alterations of emotional responses. *Nature Genetics* **25,** 195–200 (2000).

155. Gallantine, E. L. & Meert, T. F. A comparison of the antinociceptive and adverse effects of the mu-opioid agonist morphine and the delta-opioid agonist SNC80. *Basic & clinical pharmacology & toxicology* **97,** 39–51 (2005).

156. Feng, P. *et al.* Effects of mu, kappa or delta opioids administered by pellet or pump on oral Salmonella infection and gastrointestinal transit. *European Journal of Pharmacology* **534,** 250–257 (2006).

157. Codd, E. E. *et al.* JNJ-20788560 [9-(8-Azabicyclo[3.2.1]oct-3-ylidene)-9H-xanthene-3-carboxylic Acid Diethylamide], a Selective Delta Opioid Receptor Agonist, Is a Potent and Efficacious Antihyperalgesic Agent That Does Not Produce Respiratory Depression, Pharmacologic Tolerance, or Physical Dependence. *Journal of Pharmacology and Experimental Therapeutics* **329,** 241–251 (2009).

158. Pradhan, A. A. *et al.* In vivo delta opioid receptor internalization controls behavioral effects of agonists. *PLoS ONE* **4** (2009).

159. Pradhan, A. A., Befort, K., Nozaki, C., Gavériaux-Ruff, C. & Kieffer, B. L. *The delta opioid receptor: An evolving target for the treatment of brain disorders* 2011. arXiv: `NIHMS150003`.

160. Pradhan, A. A. A. *et al.* Ligand-Directed Trafficking of the δ-Opioid Receptor In Vivo: Two Paths Toward Analgesic Tolerance. *Journal of Neuroscience* **30,** 16459–16468 (2010).

161. Broom, D. C., Jutkiewicz, E. M., Rice, K. C., Traynor, J. R. & Woods, J. H. Behavioral Effects of δ-Opioid Receptor Agonists: Potential Antidepressants? *The Japanese Journal of Pharmacology* **90,** 1–6 (2002).

162. Lee, S. M., Booe, J. M. & Pioszak, A. A. Structural insights into ligand recognition and selectivity for classes A, B, and C GPCRs. *European Journal of Pharmacology* **763,** 196–205 (2015).

163. Li, J. G. *et al.* Asp147 in the third transmembrane helix of the rat μ opioid receptor forms ion-pairing with morphine and naltrexone. *Life Sciences* **65,** 175–185 (1999).

164. Befort, K., Zilliox, C., Filliol, D., Yue, S. & Kieffer, B. L. Constitutive activation of the δ opioid receptor by mutations in transmembrane domains III and VII. *Journal of Biological Chemistry* **274,** 18574–18581 (1999).

165. Xu, W. *et al.* Comparison of the amino acid residues in the sixth transmembrane domains accessible in the binding-site crevices of μ, δ, and κ opioid receptors. *Biochemistry* **40,** 8018–8029 (2001).

166. Befort, K., Tabbara, L., Kling, D., Maigret, B. & Kieffer, B. L. Role of Aromatic Transmembrane Residues of the d-Opioid Receptor in Ligand Recognition. *Journal of Biological Chemistry* **271,** 10161–10168 (1996).

167. Tryoen-Toth, P. *et al.* Inverse agonism and neutral antagonism at wild-type and constitutively active mutant delta opioid receptors. *J Pharmacol Exp Ther* **313,** 410–421 (2005).

168. Valiquette, M., Vu, H. K., Yue, S. Y., Wahlestedt, C. & Walker, P. Involvement of Trp-284, Val-296, and Val-297 of the human delta-opioid receptor in binding of delta-selective ligands. *The Journal of biological chemistry* **271,** 18789–18796 (1996).

169. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function and Genetics* **57,** 433–443 (2004).

170. Thompson, A. A. *et al.* Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **485,** 395–399 (2012).

171. Wang, C. *et al.* Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nature Communications* **5,** 1–11 (2014).

172. Bhattacharya, S., Lee, S., Grisshammer, R., Tate, C. G. & Vaidehi, N. Rapid computational prediction of thermostabilizing mutations for G protein-coupled receptors. *J Chem Theory Comput* **10,** 5149–5160 (2014).

173. Carpenter, B., Nehmé, R., Warne, T., Leslie, A. G. & Tate, C. G. Structure of the adenosine A2Areceptor bound to an engineered G protein. *Nature* **536,** 104–107 (2016).

174. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist-b2 adrenoceptor complex. *Nature* **469,** 236–240 (2011).

175. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys* **120,** 19919–19929 (2004).

176. Doshi, U. & Hamelberg, D. Towards fast, rigorous and efficient conformational sampling of biomolecules: Advances in accelerated molecular dynamics. *Biochimica et Biophysica Acta - General Subjects* **1850,** 878–888 (2015).

177. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26,** 1781–1802 (2005).

178. Kappel, K., Miao, Y. & Mccammon, J. A. Accelerated Molecular Dynamics Simulations of Ligand Binding to a Muscarinic G-protein Coupled Receptor. *Q Rev Biophys* **48,** 479–487 (2015).

179. Miao, Y., Nichols, S. E. & McCammon, J. A. Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. *Physical Chemistry Chemical Physics* **16,** 6398–6406 (2014).

180. Ren, Z., Ren, P. X., Balusu, R. & Yang, X. Transmembrane Helices Tilt, Bend, Slide, Torque, and Unwind between Functional States of Rhodopsin. *Scientific Reports* **6,** 1–13 (2016).

181. Chang, Y. *et al.* Mechanistic insights into the allosteric modulation of opioid receptors by sodium ions. *Biochemistry* **53,** 5140–5149 (2014).

182. Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters* **78,** 2690–2693 (1997).

183. Warne, T., Edwards, P. C., Leslie, A. G. W. & Tate, C. G. Crystal structures of a stabilized β1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* **20,** 841–849 (2012).

184. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* **469,** 241–245 (2011).

185. Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523,** 561–7 (2015).

186. Rasmussen, S. G. F. *et al.* Crystal structure of the β2 adrenergic receptor-Gs protein complex. *Nature* **477,** 549–55 (2011).

187. Thomsen, A. R. *et al.* GPCR-G Protein-β-Arrestin Super-Complex Mediates Sustained G Protein Signaling. *Cell* **166,** 907–919 (2016).

188. Miao, Y., Caliman, A. D. & McCammon, J. A. Allosteric effects of sodium ion binding on activation of the M3 muscarinic G-protein-coupled receptor. *Biophysical Journal* **108,** 1796–1806 (2015).

189. Ranganathan, A., Dror, R. O. & Carlsson, J. Insights into the Role of Asp79 2.50 in β 2 Adrenergic Receptor Activation from Molecular Dynamics Simulations. *Biochemistry* **53,** 7283–7296 (2014).

190. Vickery, O. N., Machtens, J. P., Tamburrino, G., Seeliger, D. & Zachariae, U. Structural Mechanisms of Voltage Sensing in G Protein-Coupled Receptors. *Structure* **24,** 997–1007 (2016).

191. Costa, T., Lang, J., Gless, C. & Herz, A. Spontaneous association between opioid receptors and GTP-binding regulatory proteins in native membranes: specific regulation by antagonists and sodium ions. *Molecular pharmacology* **37,** 383–94 (1990).

192. Selley, D. E., Cao, C. C., Liu, Q. & Childers, S. R. Effects of sodium on agonist efficacy for G-protein activation in mu-opioid receptor-transfected CHO cells and rat thalamus. *British journal of pharmacology* **130,** 987–996 (2000).

193. Sharma, S. K., Jones, R. M., Metzger, T. G., Ferguson, D. M. & Portoghese, P. S. Transformation of a κ-opioid receptor antagonist to a κ-agonist by transfer of a guanidinium group from the 5- to 6-position of naltrindole. *Journal of Medicinal Chemistry* **44,** 2073–2079 (2001).

194. *World Drug Report 2015* tech. rep. (2015).

195. Yeadon, M. & Kitchen, I. Comparative binding of μ and δ selective ligands in whole brain and pons/medulla homogenates from rat: Affinity profiles of fentanyl derivatives. *Neuropharmacology* **27,** 345–348 (1988).

196. Portoghese, P. S., Sultana, M. & Takemori, A. E. Naltrindole, a highly selective and potent non-peptide delta opioid receptor antagonist. *European journal of pharmacology* **146,** 185–186 (1988).

197. Neilan, C. L., Akil, H., Woods, J. H. & Traynor, J. R. Constitutive activity of the delta-opioid receptor expressed in C6 glioma cells: identification of non-peptide delta-inverse agonists. *British journal of pharmacology* **128,** 556–62 (1999).

198. Aboujaoude, E. & Salame, W. O. Naltrexone: A Pan-Addiction Treatment? *CNS Drugs* **30,** 719–733 (2016).

199. Wang, D., Raehal, K. M., Bilsky, E. J. & Sadée, W. Inverse agonists and neutral antagonists at μ opioid receptor (MOR): Possible role of basal receptor signaling in narcotic dependence. *Journal of Neurochemistry* **77,** 1590–1600 (2001).

200. Shen, M.-y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science* **15,** 2507–2524 (2006).

201. Benkert, P., Tosatto, S. C. E. & Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* **71,** 261–277 (2008).

202. Chen, C. *et al.* Determination of the Amino Acid Residue Involved in [3H]beta -Funaltrexamine Covalent Binding in the Cloned Rat Opioid Receptor. *Journal of Biological Chemistry* **271,** 21422–21429 (1996).

203. Lüdemann, S. K., Lounnas, V. & Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *Journal of Molecular Biology* **303,** 797–811 (2000).

204. Alibay, I., Burusco, K. K., Bruce, N. J. & Bryce, R. A. Identification of Rare Lewis Oligosaccharide Conformers in Aqueous Solution Using Enhanced Sampling Molecular Dynamics. *The Journal of Physical Chemistry B* **122,** 2462–2474 (2018).

205. Yang, M., Huang, J. & MacKerell, A. D. Enhanced Conformational Sampling Using Replica Exchange with Concurrent Solute Scaling and Hamiltonian Biasing Realized in One Dimension. *Journal of Chemical Theory and Computation* **11,** 2855–2867 (2015).

206. Yang, M. & Mackerell, A. D. Conformational sampling of oligosaccharides using hamiltonian replica exchange with two-dimensional dihedral biasing potentials and the weighted histogram analysis method (WHAM). *Journal of Chemical Theory and Computation* **11,** 788–799 (2015).

207. Sabbadin, D. & Moro, S. Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. *Journal of Chemical Information and Modeling* **54,** 372–376 (2014).

208. Marti-Solano, M., Sanz, F., Pastor, M. & Selent, J. A dynamic view of molecular switch behavior at serotonin receptors: Implications for Functional selectivity. *PLoS ONE* **9,** e109312 (2014).

209. Gaieb, Z., Lo, D. D. & Morikis, D. Molecular mechanism of biased ligand conformational changes in CC chemokine receptor 7. *Journal of Chemical Information and Modeling,* 1808–1822 (2016).

210. Ng, H. W., Laughton, C. A. & Doughty, S. W. Molecular dynamics simulations of the adenosine A2a receptor in POPC and POPE lipid bilayers: Effects of membrane on protein behavior. *Journal of Chemical Information and Modeling* **54,** 573–581 (2014).

211. Bartuzi, D., Kaczor, A. A. & Matosiuk, D. Activation and Allosteric Modulation of Human μ Opioid Receptor in Molecular Dynamics. *J Chem Inf Model* **55,** 2421–34 (2015).

212. Hanson, M. A. *et al.* A Specific Cholesterol Binding Site Is Established by the 2.8 Å Structure of the Human β2-Adrenergic Receptor. *Structure* **16,** 897–905 (2008).

213. Gimpl, G. & Fahrenholz, F. Cholesterol as stabilizer of the oxytocin receptor. *Biochimica et Biophysica Acta - Biomembranes* **1564,** 384–392 (2002).

214. Klein, U., Gimpl, G. & Fahrenholz, F. Alteration of the Myometrial Plasma Membrane Cholesterol Content with β-Cyclodextrin Modulates the Binding Affinity of the Oxytocin Receptor. *Biochemistry* **34,** 13784–13793 (1995).

215. Pucadyil, T. J. & Chattopadhyay, A. Cholesterol modulates ligand binding and G-protein coupling to serotonin1A receptors from bovine hippocampus. *Biochimica et Biophysica Acta - Biomembranes* **1663,** 188–200 (2004).

216. Prasanna, X., Chattopadhyay, A. & Sengupta, D. Cholesterol modulates the dimer interface of the β2-adrenergic receptor via cholesterol occupancy sites. *Biophysical Journal* **106,** 1290–1300 (2014).

217. Oates, J. *et al.* The role of cholesterol on the activity and stability of neurotensin receptor 1. *Biochimica et Biophysica Acta - Biomembranes* **1818,** 2228–2233 (2012).

218. Paila, Y. D., Kombrabail, M., Krishnamoorthy, G. & Chattopadhyay, A. Oligomerization of the serotonin1A receptor in live cells: A time-resolved fluorescence anisotropy approach. *Journal of Physical Chemistry B* **115,** 11439–11447 (2011).

219. Hanson, M. A. *et al.* Crystal Structure of a Lipid G Protein – Coupled Receptor. *Science* **335,** 851–856 (2012).

220. Koldso, H. & Sansom, M. S. P. Organization and Dynamics of Receptor Proteins in a Plasma Membrane. *Journal of the American Chemical Society* **137,** 14694–14704 (2015).

221. Kukol, A. Lipid Models for Unit-atom Molecular Dynamics Simulation of Proteins. *J of Chemical Theory and Computation* **5,** 615–626 (2009).