A novel coarse-grained molecular dynamics method for the accurate prediction of helix-helix interactions in GPCRs

Nojood Altwaijry

Thesis submitted to University College London for the degree of Doctor of Philosophy

September 2017

Structural & Molecular Biology
Division of Biosciences
University College London

Declaration

The work presented in this thesis was undertaken at University College London. I, Nojood Altwaijry, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis describes a novel computational method developed to identify and characterise points of protein-protein interaction between two G protein-coupled receptors (GPCRs). An ensemble-based coarse-grained molecular dynamics (eCG-MD) approach was applied to GPCR oligomers with experimentally-determined contact interfaces (adenosine A_{2A} receptor, rhodopsin, CXCR4 and β_1AR). Error analysis was used to determine 1) the number of replicas in an ensemble and 2) the simulation time for each replica that were needed to obtain convergence with experimental results. Error analysis also enabled identification of non-interacting regions.

This novel method yielded calculations of distance between rhodopsin, CXCR4 and β_1AR transmembrane domains reported to form contact points in homodimers that correlated well with the corresponding measurements obtained from the structural data, demonstrating an ability to predict contact interfaces

computationally. The method gave distance measurements between residues shown to be involved in oligomerisation of the fifth transmembrane domain from the adenosine A_{2A} receptor that were in very good agreement with the existing biophysical data. Further, the method provided information about the nature of the contact interface that could not be determined experimentally.

This CG-MD method was then used as a high-throughput screen to identify novel sites of interaction in the adenosine A_{2A} receptor, informing the design of future experimental work. Experimental methods to investigate interactions are also described in this thesis. These were less successful in identifying contact points, however, the present computational method will enable novel interaction points between GPCRs to be predicted and tested experimentally using assays of ligand binding and receptor signaling.

In conclusion, this work provides an accurate, reproducible and reliable method for determining the specific points of interaction between GPCR dimers. The eCG-MD method discriminates between residues in TM helices that form specific interactions and residues that are in close proximity but do not interact.

Acknowledgements

I am greatly indebted to my supervisor Professor Andrea Townsend-Nicholson and to my second supervisor Professor Peter Coveney, for their unwavering support, guidance and encouragement without which this work would not have been possible. Professor Andrea was a marvelous source of ideas and inspiration who oversaw this project with enormous patience and kindness, and guided me during the research period and the writing up of this thesis. Special thanks go to Dr. Michael Baron and Dr. David Wright for their many discussions and valuable input throughout this project.

I would not have had the opportunity to carry out my research had it not been for the generous funding provided by King Saud University, Saudi Arabia, and for that I am eternally grateful.

At a more personal level, I would like to thank my friends Jowhara Althubaiti, Suaad Alkandari, Batoul Farran and Raquel Montenegro for their continuous support and for being there for me.

I would like to express my deepest gratitude to my father for always believing in me, to my mother for her faith in me, and for her continuous encouragement. I would like to thank my brothers and sisters for all the love and support they had given me throughout the years. Special thanks go to my sisters Najwa, and to my younger brother Hani, for all their helpful input and suggestions. I would like to thank my younger sister Najd for always being there

Abbreviations

AA   All-Atom

ADP   Adenosine 5`-diphosphate

ADPβS   Adenosine 5′-O-(2-thiodiphosphate)

ATP   Adenosine 5`-triphosphate

BiFC   Biomolecular Fluorescent Complementation

bp   Basepair

BRET   Bioluminescence resonance energy transfer

FRET   Förster resonance energy transfer

dATP   Deoxyadenosine triphosphate

cAMP   Cyclic adenosine 2`,3`-monophosphate

Ca^{2+} Calcium ion

CD   Circular dichroism spectroscopy

CG   Coarse-Grained

CHO   Chinese hamster ovary cells

CXCR   Chemokine receptor

DAG   1,2-diacylglycerol

DMEM   Dulbecco's Modified Eagle's Medium

DMSO   Dimethyl Sulfoxide

DPBS   Dulbecco's Phosphate-Buffered Saline

DPPC   1,2-dipalmitoyl-sn-glycero-3-phosphocholine

DNA   Deoxyribonucleic acid

EC_{50} Half-maximum effective concentration

ECACC   European Collection for Cell Cultures

ECFP   Enhanced cyan fluorescent protein

eCG-MD   Ensemble-based Coarse Grained Molecular Dynamics

ECL   Extracellular loop

ER   Endoplasmic reticulum

EtBr   Ethidium bromide

EYFP   Yellow fluorescent protein

FP   Fluorescent protein

FRET   Fluorescence resonance energy transfer

GDP	Guanosine 5`-diphosphate
GEF	GDP-GTP exchange factor
GFP	Green fluorescent protein
GPCRs	G-protein coupled receptors
GRKs	G protein-coupled receptor kinases
GTP	Guanosine 5`-triphosphate
HA	Hemagglutin
HBSS	Hank's Balanced Salt Solution
HEK	Human embryonic kidney cells
ICL	Intracellular loop
IP_3	1,4,5-triphosphate
IC_{50}	Half maximal inhibitory concentration
kb	kilobase
kDa	kilodalton
L	Litre
LB	Luria broth
LJ	Lennard-Jones potential
MAPK	Mitogen-activated Protein Kinase
MC	Monte Carlo
MD	Molecular Dynamics
$mGlu_5$	Metabotropic glutamate type 5 receptor
MM	Molecular mechanics
NBE	New England Biolabs
PCR	Polymerase chain reaction
PDB	Protein database bank
PIP_2	Phosphotidylinositol 4,5-diphosphate
PKA	Protein kinase A
PLC	Phospholipase C
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
RE	Restriction endonuclease
RMSD	Root mean square deviation
SDM	Site-directed mutagenesis
T4L	T4-lysozyme
TAE buffer	Tris-acetate-EDTA buffer

TM	Transmembrane
UDP	Uridine 5`-diphosphate
UTP	Uridine 5`-triphosphate
WT	Wild-type
β_2AR	β2-adrenergic receptor
ml	Milliliter
μL	Microliter
μM	Micromolar
7TM	Seven transmembrane

Contents Page

List of Figures

List of Tables

Chapter 1
Introduction

1.1 G protein-coupled receptors

G protein-coupled receptors (GPCRs) are a well-studied family of membrane proteins also known as seven transmembrane (7TM) receptors. They are the largest, most diverse group of cell surface receptors located in a wide variety of tissues, and organs. They respond to various numbers of chemical signals in a highly selective way and then transduce the signal from these interactions into numerous intracellular responses. Not only are they a large and important group of signaling proteins, they are also the targets for about 40% of all therapeutic compounds in clinical use. Although over 800 human proteins are classified as GPCRs, drugs have been developed against fewer than 10% of these targets[1, 2]. GPCRs are responsible for transducing signals from the most divergent repertoire of stimuli, including light, calcium ion, nucleotides, amino acids, biogenic amines, odorants, pheromones, sweet and bitter tastes, lipids, peptides, and glycoproteins. Thus there is huge potential to expand the number of targets for which new therapies can be designed. Novel therapeutic design is also important if one of the goals of personalized medicine, to develop new drugs for patient-specific variations of GPCRs, is to be achieved.

There are five main classes of 7TM superfamily: class A (rhodopsin-like), class B (secretin receptor family), class C (metabotropic

glutamate/pheromone), class E (cAMP receptors) and class F (frizzled and adhesion). The conservation of sequence identity across the huge 7TM family members is significantly low, however, sequence alignments of 7TM receptors reveal shared traits that form the basis for classification into families[3].

1.1.1   GPCR structure

All GPCRs have a common core structure, which consists of a single peptide chain that spans the plasma membrane seven times (see Figure 1.1). Experimental visualization of this was shown with the publication of the first GPCR crystal structure in 2000, when bacteriophage rhodopsin and bovine rhodopsin were crystallized at 1.55 and 2.8Å respectively[4]. The second structure obtained was that of the β_2-adrenergic receptor by Kobilka and coworkers in 2007[5, 6].

The past 14 years, and in particular the last five years, have seen an explosion in the solving of GPCR crystal structures in multiple receptor states (for example, unliganded, in complex with agonist, in complex with antagonist, etc.). This breakthrough in the understanding of GPCR structure and function has enabled significant advances in the design of GPCR ligands using computational methods and has facilitated a better understanding of ligand-receptor interactions[2]. A list of the GPCR crystal structures published to date is given in Table 1.1.
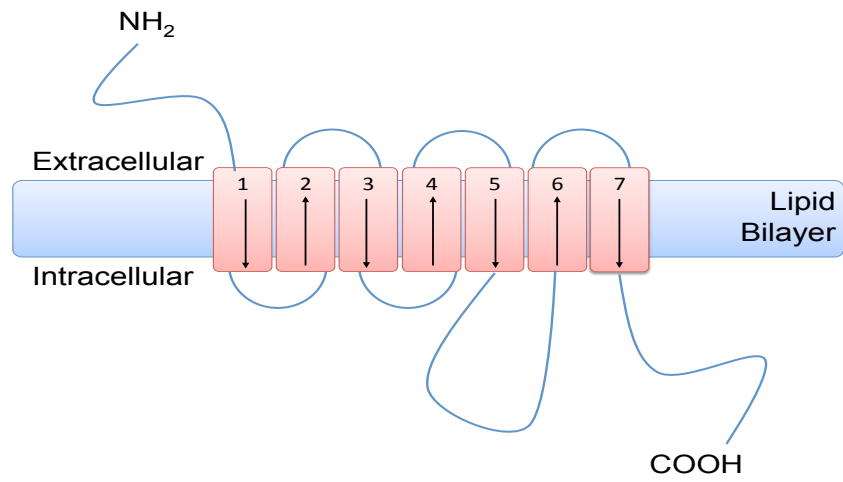
Figure 1.1 **A schematic representation of GPCR structure**

The seven TM helices embedded in the lipid bilayer of the plasma membrane (in blue) are shown (in red) with arrows indicating directionality from the amino to carboxy termini indicated at the extracellular and intracellular ends of the protein, respectively.

Class	Family	Receptor	Ligand	Resolution (Å)	Identifier
A	Adrenoreceptors	β_2-adrenoreceptor	Carazolol and Nb60	3.2	5JQH
			Carazolol	2.48	5D5A
			Carazolol	3.8	5D5B
			FAUC37	3.3	4QKX
			BI167,107	2.79	4LDE
			Hydroybenzyl isoproterenol	3.1	4LDL
			Adrenaline	3.2	4LDO
			Carazolol	3.99	4GBR
			BI167,107	3.2	3SN6
			BI167,107	3.5	3P0G
			FAUC50	3.5	3PDS
			ICI 118,551	2,84	3NY8
			Compound 2	2.84	3NY9
			Alprenolol	3.16	3NYA
			Carazolol	3.4	3KJ6
			Timolol	2.8	3D4S
			Carazolol	3.4	2R4R
			Carazolol	3.4	2R4S
			Carazolol	2.4	2RH1
		β_1-adrenoreceptor	Cyanopindolol	2.7	2VT4
			Salbutamol	3.05	2Y04
			R-Isoprenaline	2.85	2Y03
			Carmoterol	2.6	2Y02
			Dobutamine	2.5	2Y00
			Dobutamine	2.6	2Y01
			Iodocyanopindolol	3.65	2YCZ
			Cyanopindolol	3.25	2YCX
			Carazolol	3.0	2YCW
			Cyanopindolol	3.15	2YCY
			Carvedilol	2.3	4AMJ
			Bucindolol	3.2	4AMI
			Cyanopindolol	3.5	4GPO
			4-(Piperazin-1-yl)-1*H*-indole	2.7	3ZPR
			4-Methyl-2-(piperazin-1-yl)quinolone	2.8	3ZPQ
			Cyanopindolol	2.1	4BVN
	Opsin	Rhodopsin	Octylglucoside	2.7	4J4Q
			β-ionone	2.6	3OAX
			--	3.4	2J4Y
			Nonyl-glucoside	2.4	4X1H
			Chromophore	2.8	1HZX
			--	2.5	2Z73
			--	2.9	3CAP
			--	3.0	3PXO
			Arrestin	3.3	4ZWJ
			Finger-loop peptide	2.8	4PXF
			--	2.6	1L9H
			--	3.7	2ZIY
			--	2.8	1F88
			Arrestin	7.7	5DGY
			--	2.7	3C9L
			--	3.0	2PED

Purine receptors	A$_{2A}$	ZM241385	2.6	3EML
		UK-432097	2.71	3QAK
		Adenosine	3.0	2YDO
		ZM241385	3.3	3PWH
		NECA	2.6	2YDV
		XAC	3.31	3REY
		Caffeine	3.6	3RFM
		ZM241385	2.7	3VG9
		ZM241385	3.1	3VGA
		4e	3.34	3UZC
		4g	3.27	3UZA
		ZM241385	1.8	4EIY
		BRIL-ZM241385	2.5	5K2A
		BRIL-ZM241385	2.5	5K2B
		BRIL-ZM241385	1.9	5K2C
		BRIL-ZM241385	1.9	5K2D
		Mini-G$_s$	3.4	5G53
		BRIL-ZM241385	1.72	5IU4
		BRIL-compound 12c	1.9	5IU7
		BRIL-compound 12b	2.0	5IUA
		BRIL-compound 12x	2.1	5IUB
		CGS21680	2.6	4UG2
		CGS21680	2.6	4UHR
		6-(2,6-Dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine	3.27	3UZA
		4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol	3.34	3UZC
Dopamine receptors	D$_3$	Eticlopride	2.89	3PBL
Histamine receptors	H$_1$	Doxepin(*E,Z*)	3.1	3RZE
Muscarinic (acetylcholine receptors)	M$_1$	Tiotropium	2.7	5CXV
	M$_2$	Iperoxo	3.5	4MQS
		Iperoxo, LY2119620	3.7	4MQT
		QNB	3.0	3UON
	M$_3$	Tiotropium	2.8	4U15
		NMS	3.7	4U16
		Tiotropium-dsT4L	3.6	4U14
		Tiotropium	3.4	4DAJ
	M$_4$	Tiotropium-T4L	2.6	5DSG
5-Hydroxytryptamine receptors	5-HT$_{1B}$	Ergotamine	2.7	4IAR
		Dihydroergotamine	2.8	4IAQ
		Ergotamine	2.7	4IB4
		Ergotamine	2.8	4NC3
Angiotensin receptors	AT$_1$	ZD7155	2.9	4YAY
		Olmesartan	2.8	4ZUD
Cannabinoid receptor	CB$_1$	AM6538	2.8	5TGZ
Chemokine receptors	CXCR4	IT1t	2.5	3ODU
		IT1t	3.1	3OE9
		CVX15	2.9	3OE0
		IT1t	3.1	3OE8
		IT1t	3.2	3OE6
		vMIP-II	3.1	4RWS
	CCR5	Maraviroc	2.7	4MBS
Endothelin receptors	ET$_B$	--	2.5	5GLI
		ET-1	2.8	5GLH
Free fatty acid receptors	FFA1	TAK-875	2.3	4PHU
Lysophospholipid receptor	LPA$_1$	ONO-9910539	2.9	4Z35
		ONO-9780307	3.0	4Z34
		ONO-3080573	2.9	4Z36
	S1P$_1$	ML056	2.8	3V2Y
		ML056	3.4	3V2W
Nurotension receptor	NTS$_1$	NT(8-13)	2.8	4GRV
		LF-T4L	2.6	4XES
		ELF-T4L	2.9	4XEE
		NT(8-13)	3.6	4BWB
		NT(8-13)	2.8	4BUO
		NT(8-13)	3.0	3ZEV
		NT(8-13)	3.1	4BV0
Orexin receptors	OX2	Suvorexant	2.5	4S0V
	OX1	Suvorexant	2.8	4ZJ8
		Suvorexant	2.8	4ZJC
Proteinase-activated receptors	PAR1	Vorapaxar	2.2	3VW7

	P2Y receptors	P2Y$_1$	BPTU	2.2	4XNV
			MRS2500	2.7	4XNW
		P2Y$_{12}$	AZD1283	2.6	4NTJ
			2MeSADP	2.5	4PXZ
			2MeSADP	3.1	4PY0
B	Corticotropin-releasing factor	CRF$_1$	CP-376395	3.18	4Z9G
			CP-376395	2.98	4K5Y
	Glucagon receptor family	Glucagon receptor	NNC0640	3.3	4L6R
			MK-0893	2.5	5EE7
C	Metabotropic glutamate receptor	mGlu$_1$	FITM	2.8	4OR2
		mGlu$_5$	Mavoglurant	2.6	4OO9
			3-chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzonitrile-(HTL14242)	2.6	5CGD
			3-chloro-4-fluoro-5-[6-(1H-pyrazol-1-yl)pyrimidin-4-yl]benzonitrile	3.1	5CGC
F	Frizzled	SMO	LY2940680	2.5	4JKV
			Choesterol	3.2	5L7D
			Vismodegib	3.3	5L7I
			Cyclopamine	3.2	4O9R
			SAG1.5	2.6	4QIN
			SANT1	2.8	4N4W
			Anta XV	2.6	4QIM

Table 1.1   Solved GPCR crystal structures

To date, crystal structures have been obtained for four of the six classes of GPCR. The PDB database accession numbers (identifiers) and level of resolution are provided for each structure.

Analysis of the crystal structures reveals that whilst GPCRs exhibit significant structural similarities there is a rich complexity across the different receptor subtypes. The generic GPCR structure consists of a membrane spanning region of hydrophobic amino acid residues, which are linked by three extracellular (ECL) and three intracellular (ICL) hydrophilic loops found on either side of the membrane, with the N-terminus protruding extracellularly and the C-terminus extending intracellularly[7]. The sizes of N-terminal, ECLs and C-terminal tails can vary dramatically among GPCRs. Each membrane-spanning region (TM domains) is comprised of alpha helices containing approximately 21 to 28 amino acid residues [8]. A highly-conserved disulfide bridge is found between two cysteines residues in the extracellular loop II (ECL 2) in most GPCRs and the end of TM3 is key in maintaining the GPCRs structure. Viewed from within the plane of the membrane, a GPCR resembles a barrel shape where the 7TM helices form a cavity within the plasma membrane that serves a ligand-binding domain that is often covered by ECL 2. Ligands may bind elsewhere as in class C GPCR where the binding domain is in the N-terminal tail itself. Despite this structural diversity, the binding pockets of GPCRs are rigid and undergo restricted conformational rearrangements during the activation process[9, 10]. A number of interhelical bonds and hydrophobic interactions between highly conserved residues in GPCRs have been found to impart the stability of the transmembrane region[4]. In fact, the 7TM helices are also comprised of functionally important signature motifs, including the E/DRY motif in TM3, which is part of the 'ionic lock', the WXP motif in TM6 and the NPXXY motif in TM7[10]. The helices also contain a number of kinks elicited by prolein residues, which segregate the receptor into ligand binding and

receptor signalling "modules" [9]. Another striking feature of GPCRs is that the extracellular loop between TM4 and TM5 and intracellular loop between TM5 and TM6 are extended. The second extracellular loop contains a N-linked glycosylation point that functions in stabilizing the protein conformation, protection from proteases, and modulates the protein function[7]. The archetypal conserved amino acid residues in the α-helices across the whole GPCR superfamily are listed in Table 1.2.

α-Helices	Amino acid	Amino acid identifier
TM1	Asn	N 1.50
TM2	Asp	D 2.50
TM3	Arg	R 3.50
TM4	Trp	W 4.50
TM5	Pro	P 5.50
TM6	Pro	P 6.50
TM7	Pro	P 7.50

Table 1.2   Signature amino acids within GPCR TMs

Each TM contains an amino acid that is highly conserved across all members of the GPCR superfamily. These have been numbered using the Ballesteros-Weinstein method.

1.1.2 GPCR Signalling

A central feature of GPCR signalling is that the extracellular ligand that binds to the receptor does not physically traverse the plasma membrane but conveys its signal to the cellular milieu through conformational changes in the receptor elicited upon ligand binding. GPCRs exist in two principal conformational states that are in equilibrium. One, an active conformation, initiates downstream events (R), whilst the other inactive conformation (r) is unable to do so, in the presence or absence of activating ligand (agonist). Typically, the binding of an agonist to the receptor provokes a switch from its inactive to its active conformation causing an increase in the proportion of active receptors, which opens up the helical bundle and exposes the binding site of the receptor where it interacts with its cognate G-protein at the cytoplasmic side of the plasma membrane. On the other hand, the presence of an antagonist binds to the inactive conformation (r) and stops signal transfer causing the reduction of the number of active receptors. The desensitization of the GPCR is initiated by its interaction with GRKs, or G protein-coupled receptor kinases, leading to its phosphorylation, usually at its carboxy-terminal end. This in turn elicits the binding of β-arrestins, which sterically hinder additional G protein-coupling and ensure the termination of the agonist response. Partial agonists can bind to both conformational states, but they cannot achieve maximum simulation[6, 7, 11-14].

GPCRs exerts their actions through guanidine triphosphate-binding proteins (GTPases) that were discovered by Rodbell, Gilman, and co-workers. These "G-proteins" are heterotrimeric entities consisting of three subunits Gα, Gβ, and Gγ. The heterotrimeric G protein activation of the effector is mediated by the Gα-subunit. The cycle of event begins with GTP displacement of GDP

and ends with GTP-hydrolysis to GDP at the Gα-subunit, which is catalysed by a GDP-GTP exchange factor (GEF). Binding of GTP causes a conformational change that causes the G$\beta\gamma$-subunits dimer to dissociate from Gα-subunit. The dissociated G-subunits initiates input signal from membrane receptors to effector protein. The GTPase activity within the Gα-subunit causes the inactivation of the system and re-association of the heterotrimeric G-protein subunit[7, 15]. The Gα-subunit is divided into four subtypes: Gα_s, Gα_i, G$\alpha_{q/11}$, and G$\alpha_{12/13}$. Two of these subtypes transduce the membrane receptor signal through the enzyme adenylate cyclase, Gα_s stimulates it and increase the level of cAMP within cell, whereas Gα_i inhibits the enzyme and reduces the intracellular levels of cAMP. G$\alpha_{q/11}$ subtype stimulates phospholipase C$_\beta$ that mediates the release of intracellular calcium from endoplasmic reticulum stores by hydrolysing phosphotidylinositol 4,5-diphosphate (PIP$_2$) into 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP$_3$), and it includes several subgroups. The other α-subunit subtypes act on regulating the protein activity (see Figure 1.2)[7, 15].
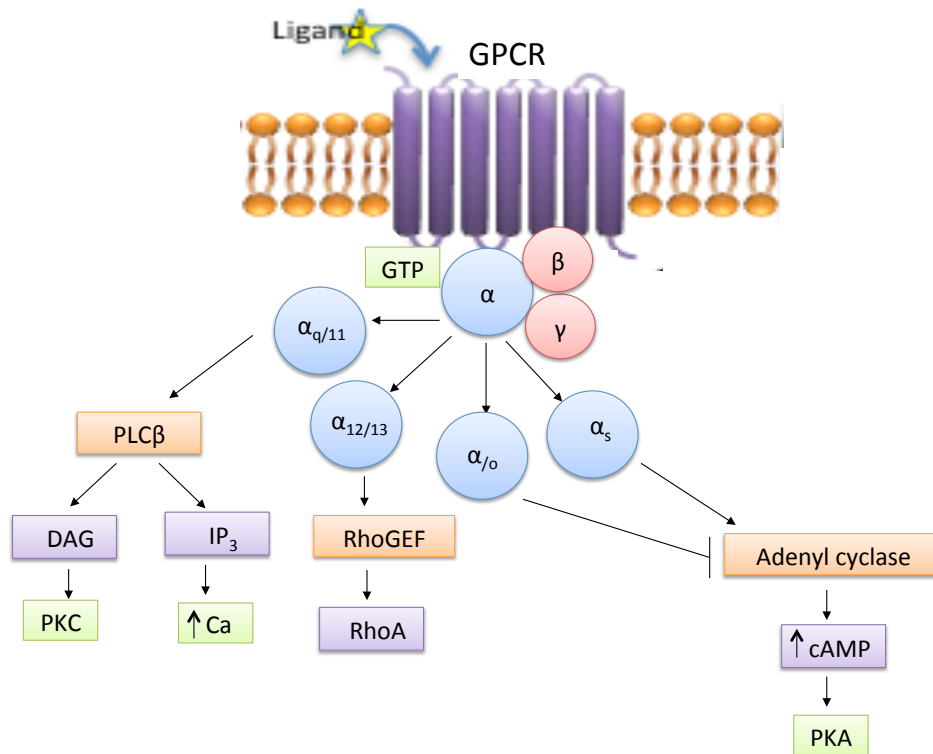
Figure 1.2 GPCRs intracellular signaling and downstream effects.

1.2 GPCR oligomerisation

For many years, the GPCRs have been viewed as monomeric entities, activating G proteins in a 1:1 ratio. This classical model of coupling is oversimplified, and many functional GPCR have been shown to exist as homodimers, heterodimers, or even as higher order oligomers[7, 16-18]. The description of GPCRs as homodimers or heterodimers refers to the physical coupling of two receptors in the cellular membrane to produce a functional centre with novel structural and functional characteristics, in comparison with the properties of each of the receptors on its own. Each receptor in the dimer is

termed a 'protomer'. Homodimers are pairs comprised of identical protomers, while heterodimers are formed from distinct receptors. An oligomer is a multimeric protein formed from a small number of subunits in a state of reversible association with each other[19]. The interaction between these two receptors is defined as the binding of a ligand to the orthosteric or allosteric sites of one receptor, causing a modification in the ligand recognition, decoding and trafficking processes of another receptor via allosteric interactions[20].

Early examples of oligomerisation include the obligate heteromeric assembly of $GABA_BR1$ and $GABA_BR2$ required to form a functional $GABA_B$ receptor[16] and the heterodimerisation of the delta and kappa opioid receptor subtypes to form an opioid receptor with the κ_2 receptor subtype pharmacology[17]. Oligomerisation can occur between different types of receptors of different classes, for example dopamine receptors have been reported to oligomerise with the adenosine A_{2A} receptor, the CB_1 cannabinoid receptor, the metabotropic glutamate type 5 receptor ($mGlu_5$) and the somatostatin-2 and -5 receptors[21-27]. The first physical proof of dimerisation, however, emerged from atomic-force microscopy analysis of the rhodopsin receptor which was shown to form structural dimers organised in paracrystalline arrays in the membrane[28] and in the model crystal structure of this GPCR (1N3M)[29].

The discovery that GPCRs are able to oligomerise has raised important questions as to where and how these complexes are formed and expressed at the cell surface. While oligomers can arise transiently in the plasma membrane in the

presence of agonists, reports have suggested that receptor dimerisation and the assembly of GPCRs with their signalling complex may occur early during biogenesis, prior to trafficking to the plasma membrane[12, 30] whilst other findings suggest that dimerisation depends on the receptor density[31]. Studies have indicated that molecular chaperones may contribute to the proper folding of receptors as well as the assembly of signalling complexes, indicating that these complexes may be delivered to the plasma membrane as "complete signalling units"[12]. The literature of receptor oligomerisation suggests that dimerisation can occur in the endoplasmic reticulum (ER) or during protein maturation in the Golgi apparatus, which might play a role in the ER quality control and exit by masking the ER-retention motifs of some receptor sequences[12, 32-34]. In support of this idea, it has been shown that the dimerisation of group I metabotropic glutamate receptor (mGluR1α) must take place in the ER[35]. Moreover, studies have established that oligomerisation has important effects on receptor function, including ligand binding, activation, desensitization, and trafficking, as well as signalling[36-38].

1.3 Experimental determination of oligomerisation

In the early 1990s, with advances in the biochemical and biophysical techniques, it became possible to detect the presence of dimeric GPCRs entities. Biological methods for studying GPCR oligomers in native cells and tissues or in recombinant mammalian expression systems include co-immunoprecipitation (CoIP), Western blot analyses, cross-linking studies, yeast two hybrid experiments, bi-molecular fluorescence complementation via GFP reconstitution

(BiFC), energy transfer-based methods (FRET and BRET), photobleaching FRET (pbFRET), sequential BRET/FRET (SRET), functional cross-talk and activation by dimeric/bivalent ligands. Table 1.3 provides examples of homo/hetero GPCR dimers entities that were identified by these biological methods. Due to diversity in both primary sequences and in the structural features of GPCRs, not all methods are applicable for the detection of a dimer under stimulated or un-stimulated conditions[39] and the results of GPCR dimerisation shown in Table 1.3 were obtained using different experimental approaches. Unfortunately, these methods do not always discriminate between proteins in close proximity and those that physically interact, something that has frequently allowed for alternative interpretations of the results and, therefore, has made it difficult to obtain unequivocal answers about the occurrence of multimerization between candidate pairs of GPCRs. In addition, contrasting results for the same GPCR pairs have been obtained using different experimental techniques. For example, the chemokine receptor CCR5 was found to form an agonist-dependent dimer using the biochemical cross-linking method[40] whilst using the BRET method on the same receptor it was shown to form a constitutive dimer during biosynthesis with no change in the intensity of BRET was observed upon agonist binding[41]. In some studies it has been reported that the cytoplasmic tail of a GPCR is essential for dimerisation of the receptor. For instance, a mutated δ-opioid receptor with a 15-residue C-terminal deletion does not show interactions as it fails to exhibit CoIP with two differently tagged receptors[42] but in other studies, the C-terminal portion of a GPCR has been shown to have no effect receptor dimerisation. As an example, the dimerisation of the H_2 histamine receptor does not involve its C-terminal portion[43]. It is a

further complication that these techniques do not yield specific details of the interface(s) between interacting GPCRs, something that would provide a capacity for further analysis of identified dimerisation through site-directed mutagenesis studies. Structural methods such as x-ray crystallography and atomic force microscopy could provide this information, however, when this project commenced, only one GPCR dimer structure had been solved and described "contact areas" rather than molecular interfaces[44].

GPCR dimers	Biological methods used to investigate receptor association	Reference
Heterodimers (a) between family subtypes		
$GABA_{B1}$:$GABA_{B2}$	TR-FRET-HTRF using Snaptag or antibodies	[45]
δ:κ	SDS-PAGE	[17]
D_1:D_2	CoIP, confocal microscopy, BiFC	[46-49]
D_1:D_3	FRET, confocal microscopy, BiFC	[50]
A_{2A}:A_1	CoIP, BRET, TR-FRET	[51]
$P2Y_1$:$P2Y_{11}$	CoIP, FRET	[52]
Heterodimers (a) between different receptors		
A_{2A}:D_2	BiFC, FRET, CoIP, Immunofluorescence microscopy	[21, 53]
D_2:SSTR5	Immunohistochemical colocalization, confocal microscopy, FRET	[27]
D_2:CB_1:A_{2A}	BiFC, BRET, sequential BRET-FRET (SRET), confocal microscopy	[54-56]
A_{2A}:CB_1	Colocalization, CoIP, BRET	[57]
A_{2A}: $mGlu_5$	CoIP	[58]
$mGlu_5$:A_{2A}:D_2	Colocalization, CoIP, BiFC, BRET and SRET	[59]
$mGlu_{2A}$:5-HT_{2A}	BRET, CoIP, colocalization	[60]
A_1:$P2Y_1$	CoIP, confocal microscopy, BiFC	[61-63]
Homodimers		
A_{2A}	CD spectroscopy and SDS-PAGE	[64]
β_2-Adrenergic	CoIP, western blot, BRET, functional complementation	[65-69]
CCR5	Cross-linking, western blot, yeast two-hybrid analysis, BRET, dominant negative	[40, 41, 70, 71]
κ	Western blot , BRET	[17, 72]
H_2	CoIP	[43]
M_3	CoIP, western blot	[73]

Table 1.3   Experimentally-identified homo/heteromeric GPCRs

Homodimer formation has been detected in at least six different receptor family GPCRs and heterodimer formation has been identified between receptor family subtypes and between different receptor families using a number of different experimental techniques.

1.3.1 Bimolecular Fluorescence Complementation

Since protein interactions regulate many cellular processes, it is of great importance when exploring the function of a protein to identify if there are other proteins with which it interacts[74, 75]. The discovery of a green fluorescent protein (GFP), composed of 238 residues, from *Aequorea Victoria* in 1962[76] has provided valuable information to our understanding of protein interactions as well as to they way in which they modulate cellular events[74]. GFP has become well-established as a marker of gene expression and protein targeting in intact cells and organisms[77].

GFP has been used in a protein complementation approach developed to directly visualize the protein-protein interaction within living cells. The bimolecular fluorescence complementation technique is based on the fact that two non-fluorescing fragments of GFP can reassociate to reconstitute the GFP protein, restoring fluorescence (see Figure 1.3)[75]. The dissection and subsequent reassembly of a protein from peptidic fragments provides a possibility for controlling its tertiary structure and hence its function[78]. The GFP is used preferentially as it is known to express, fold, and fluoresce in virtually every type of cell and subcellular structure in which it has been tested[75]. There are a large number of variants of GFP, including yellow and cyan fluorescent proteins all of which are able to reassociate and form functional fluorescent proteins when their dissected non-fluorescent fragments are brought into sufficiently close proximity[79].

Subsequently, the protein complementation approach was applied to the visualization of protein-protein interactions and to study the interaction of

31

multiple proteins in living cells. This was achieved by making fusion proteins between two target proteins of interest and different GFP variant fragments. Reassembly of the GFP fragments produces fluorescence when the complex is excited by light at the wavelength needed to visualize the variant's fluorescence. The bimolecular fluorescence complementation (BiFC) assay is a multicolored BiFC assay that uses an enhanced yellow fluorescent protein (EYFP) and an enhanced cyan fluorescent protein (ECFP)[80].
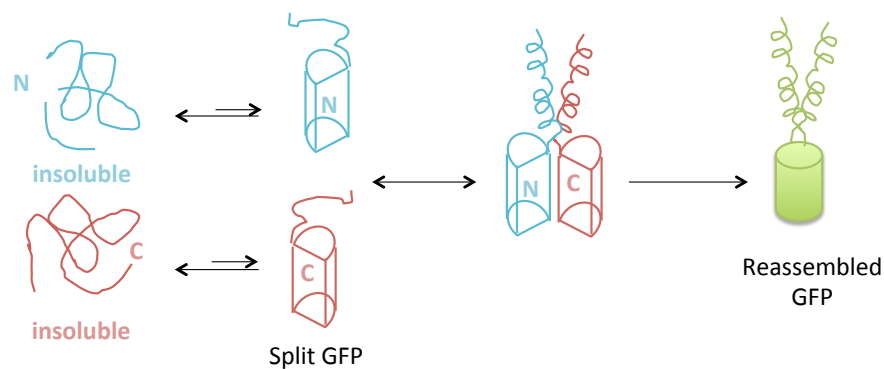
Figure 1.3 Mechanism of the GFP reassembly screen

Most of the fusions to the GFP fragments are insoluble. Interaction between the individual GFP fragments only occurs if they are fused to interacting proteins. Reassembly is essentially irreversible, which effectively pulls more of the fusions into solution.

However, EYFP has one limitation involving the chromophore maturation sensitivity to higher temperatures[77], which requires pre-incubation of cells at lower temperatures prior to visualizing the BiFC signal. Unfortunately, the decrease in temperature itself is a stressful condition[79, 81]. This difficulty prompted the discovery, by Shyu *et al.* in 2006, of several new fluorescent protein combinations able to work under physiological conditions. The new BiFC partners eliminate the need for pre-incubation at lower temperatures, show higher specificity, require lower amount of plasmids for transfection and have a shorter incubation time, leading to the increase of both the BiFC signal and its specificity[80].

The new BiFC partners consist of the N-terminal fragment of Cerulean from residues 1-172 (N173)[82] and the C-terminal fragment of Venus from residues 155-238 (C155). Cerulean is an improved ECFP mutated at (S72A/Y145A/H148D), which shows a 2.5-fold increase in fluorescence intensity compared with ECFP. Venus[83] is a new YFP mutant (F46L/F64L/M153T/V163A/S175G) engineered to be less sensitive to the environment[80].

The BiFC-fusion receptors can be studied using both stable and transient expression techniques. The advantages of the BiFC assay include, first, a high level of receptor expression is not a requirement, and therefore, it allows the detection of weak interactions between proteins. Secondly, the simple detection of the reconstituted fluorescent protein process using a fluorescent microscope because it has a strong intrinsic fluorescence, which allows direct visualization of

the protein complex under circumstances where agonist and antagonist can be applied and the results on protein-protein interactions are observed also make this technique more favorable[84]. One limitation of the BiFC approach is the time required for fluorophore maturation. This prevents real-time detection of rapid changes in interactions. In addition, bimolecular fluorescent complex formation is irreversible *in vitro*.

1.3.2 GPCR oligomerisation studied using BiFC

The occurrence of GPCR oligomerisation is now very well documented and widely accepted. Most techniques for the detection of GPCR oligomers only allow the detection of protein-protein interactions, but combining the BiFC and FRET/BRET techniques in multiple ways was established to measure ternary or quaternary higher ordered complexes[85]. The physiological and pharmacological significance of the higher-order GPCR complexes remains to be fully elucidated as these complexes may offer new drug targets.

Higher order GPCR oligomers were identified by this technique. The well documented $A_{2A}R$ receptor homodimer[86] was found to assemble into higher order oligomers by the use of a combined BiFC-FRET and BiFC-BRET technique[87, 88]. In another example, the interaction between D_2R and $A_{2A}R$[89], which was also confirmed by BiFC measurements in a neuronal cell mode[90], were also shown to form higher order oligomers[85]. Experiments using a combination of BiFC and biomolecular luminescence complementation (BiLC) have demonstrated the existence of homomeric D_2R complexes with four protomers[91].

1.4   Computational analyses of GPCR interactions

There have been many computational studies of GPCR interactions (see Table 1.4). The methodologies for modelling these have, in general, adopted one of two approaches: (i) molecular dynamics simulations using models based on homology with the nearest related GPCR for which structural data exist or (ii) docking[92, 93]. Initial GPCR MD studies were performed using CHARMM and AMBER which were subsequently integrated into NAMD[94, 95] and GROMACS[96, 97]. Although there is no established standard protocol for MD simulations of GPCRs, a number have used GROMACS with the Martini force field[98-103], which is designed specifically for lipids and membranes and allows the lipid composition most suited to the receptor in question to be incorporated into the simulation. The more recent of the GPCR dimer modeling studies have been conducted using coarse-grained simulations, which take less compute time and, therefore, provide an opportunity to perform a substantial number of replicas for each set of simulation conditions.

There is currently no standard protocol for the computational study of GPCR helix-helix interactions. By 2015, approximately 30 computational GPCR dimer models had been published (see Table 1.4). Of these, 2 are Monte Carlo-derived, 15 are based on docking and 9 have been generated using atomistic MD simulations. The rest are CG-MD models. Historically, docking was the earliest method to be employed and has been used regularly; its current use is widespread. Alternative methods of modelling began with Monte Carlo methods, moving to fully atomistic MD, with a subsequent shift to CG-MD, which is the predominant MD method currently in use. CG-MD is popular as it is cheaper,

faster and has been shown, when CG models are subsequently converted to atomistic representations, to produce similar results as models generated by atomistic MD[100, 104, 105]. CG-MD simulations have also been used to study TM helix-helix dimers for non-GPCR types of cell surface receptors such as GpA and ErbB dimers[106, 107].

Type	GPCR dimers	Method	Force field	Interface	Number of replicas	Time scale	Reference
Homodimers	Rho/Rho	Docking	CVFF	TM4,5/TM4,5		100 ps	[108]
						ND*	[109]
				TM1,2/TM1,2		ND	[110]
		MD	Gromos-87	TM4,5/TM4,5	1	45 ns	[111]
			OPLSAA		1	0.1 µs	[112]
			Amber/parm99		2	300 ns	[113]
		CG-MD	Martini	TM1,2, H8	1[a]	8 µs	[98]
				TM4,5/TM6,7	10	100 µs	[114]
	β₂-/β₂-adrenergic	CG-MD	Martini	H8/H8	2	5 µs	[103]
				TM1/TM1	4	~200 µs	[102]
				TM4,5/TM4,5	1[b]	~18 µs	[101]
				TM5/TM5			
				TM6/TM6			
				TM3/TM3			
	β₁-/β₁-adrenergic	CG-MD	Martini	TM1/TM1	3[c] (2 runs different starting point)	2 µs	[101]
				TM5/TM5			
	α₁B-/α₁B-adrenergic	Docking		TM5/TM5		ND	[115]
				TM6/TM6			
				TM7/TM7			
	5-HT₄/5-HT₄	Docking		TM2,4/TM2,4		ND	[116]
				TM4,6/TM4,6			[117]
							[118]
	5-HT₁A/5-HT₁A	Docking	CHARMM	TM4,5/TM4,5		15 ns	[119]
	CXCR4/CXCR4	Docking		TM4,5,IL2/TM4,5		ND	[110]
		MD	OPLSAA	TM3/TM4,5	1	50 ns	[120]
				TM5/TM5			
	NTSR₁/NTSR₁	Docking	CHARMM	TM1/TM4		ND	[121]
				TM4/TM4			
	δ-OR/δ-OR	CG-MD	Martini	TM2,3,4/TM2,3,4	1[d]	250 ns	[99]
				TM4/TM4 favored over TM4,5/TM4,5	2	1.5 µs	[122]
	κ-OR/κ-OR	Docking		TM1/TM2		ND	[110]
	A₂AR/A₂AR	Docking	CHARMM	TM1,2,3/TM1,2,3		ND	[123]
				TM1/TM1			
				TM1,4/TM1,4			
				TM2,3/TM2,3			
				TM6,7/TM6,7			
				H8,I3/TM6			
	A₃R/A₃R	MD	Amber7 FF99	TM4,5/TM4,5	1	500 ps	[124]
	TXA₂/TXA₂	Docking		TM1/TM1		ND	[125]
				TM1/TM2,EL2			
				H8/H8			
	D₂R/D₂R	Monte Carlo		ND		ND	[126]
	SSTR₁/SSTR₁	Monte Carlo		ND		ND	[126]
	LHR-LHR	Docking		TM4/TM6,7		ND	[125]
		MD	CHARMM	TM4/TM4	1	1 ns	[125]
				TM4/TM6			
				TM5/TM6			
				TM4/TM1,3			
Heterodimers	A₂AR/D₂R	Docking		TTM3,4/TM5,6		ND	[22]
				TM3,4,5/TM4,5			
				TM4,5/TM3,4,5			
	mGluR₂/5-HT₂A	Docking		TM4,5/TM4,5		ND	[95]
		MD	CHARMM22/27	TM4,5/TM4,5	1	40 ns	[95]
	µ-OR/δ-OR	Docking		TM6,7/TM4,5		ND	[127]
				TM1,7/TM4,7			
		MD	GROMOS87	TM1,7/TM4,5	1	5 ns	[127]
				TM4,7/TM4,5			
Homotetramer	(V₂R)₄	MD	CHARMM22/27	TM3,4/TM4,7	1	5 ns	[94]
				TM4,5/TM4,5			

Table 1.4 Computational methods for modelling GPCR dimers

Summary of the computational methods used for modeling mammalian GPCRs to date. * ND: Not defined, IL: intracellular loop, EL: extracellular loop; [a] 4 different structures; [b] 9 different structures; [c] 2 runs, different starting point; [d] umbrella sampling of 43 different starting points.

1.5   Effects of GPCR oligomerisation

Oligomerisation affects and provides a network of novel and diverse functions within cells. Heterodimerisation is able to alter the agonist affinities of individual receptors, thus generating entities with novel agonist and antagonist pharmacologies, for example, within the purinergic receptor family, the A_1R-$P2Y_1$ association provoked an A_1R with $P2Y_1R$-like agonist patterns[62]. For two other rhodopsin-like Class A GPCRs, the β-AR and AT_1R dimer provides an example of cross-inhibition, whereby the blockade of one protomer results in the uncoupling and inhibition of its interacting partner[128, 129]. It can also modify the selectivity of GPCRs for distinct G-protein subsets by changing the downstream signalling cascades; for example, the D_1-D_2 heterodimer couples to $G_{q/11}$, leading to PLC-mediated intracellular calcium increase, a pathway that is not observed for the D_1 or the D_2 receptors[46, 47]. Lastly, it might regulate desensitization and trafficking mechanisms by potentiating the internalization of receptors that cannot internalize on their own. For instance, the dimerisation of the μ- and δ-opioid receptors provokes the cross-internalization of the μ-opioid upon activation of δ-opioid, thus reducing μ-opioid receptor-induced analgesia [12].

The wide-ranging effects of GPCR oligomerisation upon cells means that there is huge potential to expand the number of targets for which new therapies can be designed. Novel therapeutic design is also important if one of the goals of personalized medicine, to develop new drugs for patient-specific variations of GPCRs, is to be achieved. Inclusion of functional GPCR homomers and heteromers in drug discovery programmes also provides a means of expanding

the range of novel targets for the development of therapeutic agents[37]. The development of therapeutic agents for purinergic GPCRs is of particular interest.

1.6 Purinergic receptors

Purine receptors, also referred to as purinergic receptors, are a subfamily of GPCRs that mediate cell signalling across the plasma membrane following activation by nucleoside and nucleotide analogues. In 1978, Burnstock *et al.* suggested that although structurally similar, these receptors should be divided into two distinct subfamilies based on their agonist selectivity: the P1 and P2 receptors. P1 receptors subtypes include A_1, A_{2A}, A_{2B} and A_3 and are activated by the nucleoside adenosine, whilst P2 receptors are activated by nucleotides, including adenosine/adenosine 5`-triphosphate (ATP), adenosine 5`-diphosphate (ADP), uridine 5`-triphosphate (UTP), and uridine 5`-diphosphate (UDP). The structures of these related agonists are shown in Figure 1.4[8]. P2 receptors can be further divided into two subtypes, P2X and P2Y, and can be further classified based on their different response profiles to ATP analogues and selective antagonism. There are eight P2Y subtypes, all of which are GPCRs: $P2Y_1$, $P2Y_2$, $P2Y_4$, $P2Y_6$, $P2Y_{11}$, $P2Y_{12}$, $P2Y_{13}$, and $P2Y_{14}$[1, 130-132].

The transmembrane domains of the P1 adenosine receptors share 11-18% sequence identity with the P2Y receptors[133]. The P2Y receptors contain between 308 to 377 amino acid residues, with positively charge amino acids in TM3, TM6, and TM7 regions that have been shown to be involved in ligand binding. Two distinct P2Y receptor subgroups have been identified. The first group includes $P2Y_{1, 2, 4, 6, 11}$ and the second group includes the $P2Y_{12, 13, 14}$. These

differ in their primary coupling to transductional G proteins. The first subgroup use the $G_{q/11}$ to activate the Phospholipase C_β, causing the formation of inositol 1,4,5-triphosphate and the release of intracellular calcium, whereas the second couples to members of the $G_{i/o}$ family of G proteins[131, 134]. The P2X receptors are structurally unrelated to the P2Y receptors, although frequently activated by the same nucleotides. P2X receptors are membrane ion channels that open in response to the binding of extracellular nucleotides. Seven genes in vertebrates encode the P2X receptor subunits, which are 40–50% identical in amino acid sequence. Each subunit has two transmembrane domains, separated by an extracellular domain about ~280 amino acids and channels form as multimers of several subunits[135].

Early studies identified that the purinergic receptors are frequently co-expressed in different proportions and different tissues. Within a receptor family, individual subtypes can form homodimers or heterodimers, something that has important implications for drug development[136].
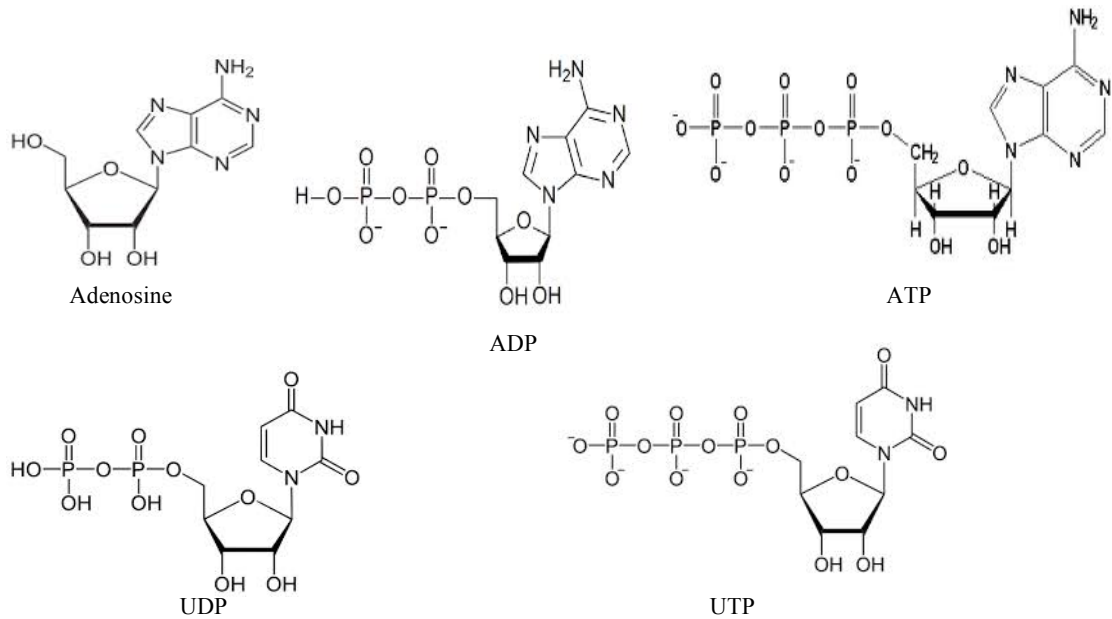
Figure 1.4   The structure of P1 and P2 receptor agonists

The nucleoside adenosine is the endogenous activator for all P1 receptors, whereas the nucleotides ADP, ATP, UDP and UTP activate P2 receptors.

1.6.1   The P1 Receptors

The P1 receptors belong to the Class A rhodopsin-like family of GPCRs. They have a short N-terminus of ~7 to 13 amino acid residues and a larger C-terminus of ~32 to 120 amino acid residues[133]. They have been subdivided into the four different subtypes according to their distinct molecular structures, tissue distributions, biomedical, and pharmacological evidence. Each of A_1, A_{2A}, A_{2B}, and A_3 contains an intron within the coding sequence that is located at the end of third transmembrane domain. These four subtypes play an essential role in responding to adenosine in the central nervous system, regulating pain, cerebral blood flow, basal ganglia functions, respiration, and sleep. It is of great

41

importance to develop more selective compounds for the P1 receptor-subtypes as this could provide therapeutics for treating numerous human diseases, such as pain, Parkinson's disease, Huntington's disease, asthma, seizures, and many other neurological disorders[137].

The transmembrane domains of the human P1 adenosine receptors share 39-61% sequence identity with each other[133]. The first cDNAs encoding the A_1 and the A_{2A} receptor subtypes were isolated in 1989 by Libert and co-workers[138]. The P1 receptors couple primarily to adenylate cyclase[131]. The A_1 and A_3 subtypes couple negatively to adenylate cyclase through $G_{i/o}$ protein alpha subunit. The A_{2A} and A_{2B} subtypes couple positively to cAMP through the G_s protein. A_{2B} has been observed to couple to through $G_{q/11}$ to regulate phospholipase C activity[133].

1.6.2   The A_{2A} Receptor

The A_{2A} receptor has the highest molecular weight of all the P1 receptors, 45 kDa compared to 36 to 37 kDa for the other adenosine receptors. Its large C-terminus domain has no effect on the coupling to G_s protein, which is mediated by the N-terminal portion of the third intracellular loop, nor does it effect the ligand binding[139].

A_{2A} has been cloned from several species including human, rat, mouse, and guinea-pig. The human gene for the A_{2A} receptor has been mapped to chromosome 22 at 22q11.2 and contains two exons and a 6-7.2 kb long intron

between the regions encoding TM3 and TM4, corresponding to the second intracellular loop of the receptor[131, 138]. The human A_{2A} receptor is polymorphic, containing two nucleotide differences, a substitution of T for C at the third base of codon 361 (nucleotide 1083) and a substitution of G for A at the first base of codon 392 (nucleotide 1174) which gives rise to a change from Arg to Gly at residue 392; the T1083C mutation is a silent mutation that is found in various populations most frequently in Asians and Caucasians[138, 140].

The A_{2A} receptor distribution is well understood due to studies performed with a number of tools including radioligand binding, *in situ* hybridisation and immunohistochemical studies. Radioligand binding and mRNA quantification show A_{2A} receptors to be concentrated in the dopamine-rich regions of the brain. They have a wide but restricted distribution with diverse biological effects in other tissues, including immune tissues, platelets, striatopallidal GABAergic neurons, olfactory neurons, central nervous system, spleen, thymus, endothelium, and vascular smooth muscle[131, 138].

The most commonly recognized signal transduction mechanism for A_{2A} receptors is through positive activation of adenylate cyclase mediated via couple to the G_s protein[131]. It has also been suggested that the A_{2A} receptor mediates dual signalling via P- and N-type Ca^{2+} channels linked to G_s / adenylate cyclase / phosphorkinase A (PKA) and cholera toxin-insensitive G-protein/PKA, respectively (see Figure 1.5)[141].
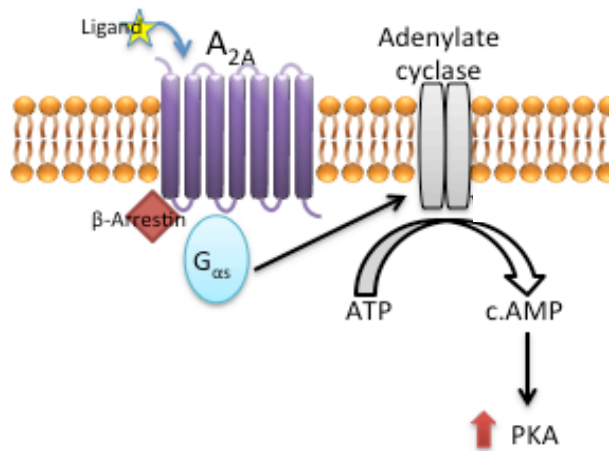
Figure 1.5 Signalling pathway for the A_{2A} receptor

Coupling of the A_{2A} receptor to the Ga_s protein leads to the accumulation of intracellular levels of cAMP produced through activation of the enzyme adenylyl cyclase.

The A_{2A} receptors are activated by the endogenous ligand, adenosine, and by adenosine analogues modified at the C2 position of the purine ring. Bulky substitutions at C2 can selectively enhance receptor affinity and have been used to characterize A_{2A} receptors. The A_{2A} receptors do not bind N^6-substituted adenosine derivatives with high affinity. The most selective A_{2A} agonist is CGS 21680; other selective agonists include: HE-NECA, CV-1808, CV-1674, ATL146e, and CVT-3146. A_{2A} receptor antagonists include: SCH 5826, ZM241385, KF 17387, CSC, SCH58261, and KW 6002, where the SCH 5826 is a selective A_{2A} antagonist, and ZM241385, KF 17387, CSC are moderately selective antagonists[142].

1.6.3   Structure of the A$_{2A}$ receptor

The first crystal structure of the A$_{2A}$ receptor in complex with antagonist ZM241385 was achieved using an A$_{2A}$-T4-lysozyme fusion (3EML)[137]. This structure revealed some unique features that were distinct from the other available GPRC structures at that time, including: 1) a different organization of extracellular loops where the surface properties of the receptor are dictated by ECL-2; 2) an extended conformation of ligand binding pocket towards the extracellular space; and 3) a subtle divergence of helical positions that redefined the receptor's binding site. This demonstrated that the ligand-binding pocket was located in a different position and orientation relative to other GPCRs and that ligand selectivity could be achieved through targeting hydrophobic residues extending from the aromatic core[2].

In the A$_{2A}$ receptor structure, the ligand-binding pocket was found to be shifted towards the extracellular side compared to that of bovine rhodopsin, β_1AR, and β_2AR[4, 6, 137, 143]. It is closer to transmembrane VI and VII where the agonist and antagonist binds in an extended conformation perpendicular to the plane of the membrane and co-linear with transmembrane helix VII, while interacting with both ECL-2 and ECL-3. The D/ERY motif of the A$_{2A}$ adenosine receptor participates in the interactions that restrain the conformation of ICL-2, as in β_1AR and β_2AR structures, and it plays a role in stabilizing the deprotonated state of the adjacent aspartate or glutamate residue that can strengthen the polar interactions between the ICL-2 and helix II with the D/ERY motif[137].

1.6.4   A$_{2A}$ receptor oligomerisation

Transmembrane (TM) domain interactions are a major determinant in the assembly and stability of the native structure of membrane proteins. It has been proposed that for proteins traversing the plasma membrane several times, the helix-helix interactions need to be controlled to give rise to the final TM protein structure[144, 145]. The existence of these specific interactions shown to give rise to an increase in helical content, especially in weakly helical TM domains, suggesting that some TM domains need a partner for complete folding in the membrane[146]. Therefore, many of the α-helices inserted into a lipid bilayer might interact to form a helix-helix dimer and, subsequently, form interactions that give rise to the formation of a higher-ordered oligomeric structure. It is these interactions affecting the packing, hydrogen bonding, aromatic interactions and salt bridges formation that will determine the sequence specific packing of the TM helices into single-span, or multi-span TM proteins[145].

The A$_{2A}$ adenosine receptor subtype has been shown to participate in the formation of heteromeric[22], homomeric[86] and even higher ordered oligomeric GPCRs[87]. The identification of heterodimeric A$_{2A}$ with several different GPCRs has been documented, including heterodimers formed with the A$_1$[51], dopamine D$_2$[22, 53], cannabinoid CB$_1$[57] and metabotropic glutamate (mGlu$_5$) receptors[58]. A$_{2A}$ receptor homodimers form through specific interactions between TM helices, leading to an increase in helical content. This has been shown to take place specifically at the weakly helical TM domains, suggesting that the helix-helix interactions, in addition to the helix-lipid interactions, facilitate homodimer formation. It has also been shown that TM

peptides interact in a similar fashion in micelles and lipid vesicles, which is not unsurprising as they exhibit thermal stability and α-helicity when inserted in SDS micelles similar to that observed when they are embedded in liposomes.

A$_{2A}$ receptor homodimerisation has been experimentally demonstrated using several different techniques[64, 86, 147]. Seven strong specific interactions have been identified between synthetic peptides corresponding to the TM domains of the human A$_{2A}$ receptor; these interactions were identified before high resolution crystal structure data became available for GPCRs and were shown to be consistent with data that had been obtained for the three-dimensional structure and stability of rhodopsin and the β$_2$ adrenergic receptor using a combination of FRET measurement and circular dichroism (CD) spectroscopy in detergent micelles[146]. In 2005, interactions between the human A$_{2A}$ receptor TM domains were studied in pairwise combinations[147] and A$_{2A}$ receptor homodimerisation was shown to be mediated by the TM5 α-helices of the two protomers[64, 147]. These data yielded a CD spectroscopy $(\Phi)_{222}/(\Phi)_{208}$ ratio > 1 for TM5-TM5 peptides, a ratio that is often associated with coiled coils or other assemblies of helical peptides. Further, it has been shown that the methionine at position 5.54 in TM5 of the A$_{2A}$R receptor (M193) influences this homodimerisation. It was also reported that none of the known motifs for TM helix dimerization (GxxxG, AxxxA, SxxSSxxT, polar clamp, serine zipper, leucine zipper) appears in TM5. However, statistical analysis of amino acid patterns in TM helices using TMSTAT reveals that the PM4 pair (PxxxM) is the most overrepresented doublet pattern from any combination of PxxxX doublet pattern[64, 148]. In addition, it has been shown that the A$_{2A}$ C-

terminal domain required for heterodimerisation[22] with the dopamine D_2 receptor is not involved in A_{2A} homodimerisation. In 2008, bimolecular fluorescence complementation (BiFC) in Cath neuronal cells confirmed the specificity of the A_{2A} - A_{2A} interaction and by combining the FRET and BiFC techniques it has been shown that at least three A_{2A} receptor protomers are able to assemble into a higher-order oligomer at the plasma membrane[87]. The major species of the A_{2A} receptor present on the cell surface is in the dimeric form, making homodimers the functionally relevant form of this GPCR[86].

1.6.5   Class A GPCR dimers

Two models of dimerisation for transmembrane proteins are described in the literature. In the first, two protomers touch each other; leading to the formation of a contact dimer. This has been established through the identification of rhodopsin dimers[28, 29], although these findings have been expressed in terms of contact areas rather than elucidating the precise molecular interactions at the dimerisation interface. In the second model, the hinge loop of the dimer opens, causing the domains to exchange and produce a domain-swapped dimer[149]. Structural methods such as x-ray crystallography and atomic force microscopy have been used to provide information about GPCR structure, but only three Class A GPCR dimer structures have been solved to date: rhodopsin, CXCR4 and β_1 adrenergic receptor (β_1AR)[44, 150, 151].

Rhodopsin has been shown to exist in a native oligomeric form[29] and an atomic model of the rhodopsin dimer has been proposed as a working model for G protein-coupled receptor dimers[28]. The rhodopsin model (1N3M) was

resolved using atomic force microscopy (AFM) and suggests that rhodopsin dimers have three contact areas. The distance between the two-rhodopsin molecules measured by luminescence resonance energy transfer was 4-5 nm. According to the model, the first contact point is between TM4, TM5 and has the largest contact area of 578Å^2; this is considered the strongest interaction between the two protomers. The second contact point between is located between TM1, TM2 and the intracellular loop connecting TM5 and TM6 (ICL-3) and covers an area of 333Å^2. The third contact point is found at the extracellular ends of TM1 with a contact area of 146Å^2; this is considered the weakest interaction in the homodimer[28, 44].

The crystal structure of the CXCR4 chemokine receptor (3ODU) bound to an antagonist small molecule IT1t has been reported and reveals a homodimer with an interface involving TM helices 5 and 6. This dimer is reported to be held together by hydrophobic interactions involving amino acid residues $L194^{5.33}$, $V197^{5.36}$, $V198^{5.37}$, $F201^{5.40}$, $M205^{5.44}$ and $L210^{5.49}$[150].

Two alternating dimer interfaces have been proposed from the crystal structure of the ligand-free basal state of the β_1AR (4GPO). The first involves TM1, TM2, extracellular loop 1 and the C-terminal H8; the second involves TM4 and TM5 and the intracellular loop 2. In the first dimer interface, the TM1-TM2-H8 dimer, the interactions identified within the TM helices are mainly through TM1 and the interacting residues include $Q38^{1.29}$, $Q39^{1.30}$, $A42^{1.33}$, $L46^{1.37}$, $L49^{1.40}$, $L50^{1.41}$, $V52^{1.43}$, $L53^{1.44}$ and $L54^{1.45}$ (from TM1) and $P96^{2.59}$, $A99^{2.62}$, $T100^{2.63}$ and $V103^{2.66}$ (from TM2). These are proposed to be mainly

hydrophobic and van dar Waals interactions. The second dimer interface includes residues from both TM4 (L171$^{4.44}$) and TM5 (R205$^{5.36}$, A206$^{5.37}$, A210$^{5.41}$, I218$^{5.49}$ and R229$^{5.60}$)[151].

1.7 Molecular modelling and computational design

High quality 3D computationally designed GPCR models are of critical importance as GPCRs are targets for drug design. Many structure-based methodologies have been developed to help identify novel pharmacological targets, by measuring the drug-ability of cavities, and also for discovering new bioactive molecules. The GPCRs, in particular, are structurally diverse especially at the ligand binding sites[152].

Computational design can be broadly divided into two classes: ligand-based design and structure-based design[153, 154]. The structure-based approach uses the three-dimensional structure of the target. By far, the largest and most reliable source of protein structures is the Protein Data Bank (PDB)[155]. Membrane proteins are particularly difficult to crystallize because they display multiple thermodynamic conformations, which greatly complicates their extraction from membranes in order to generate crystals[156]. Like most membrane proteins, the GPCRs are extremely difficult to crystallize, and only a couple of these protein receptors have been crystallized and their structures deposited into the protein bank database today[157].

In ligand-based drug design, the characteristics of known ligands are evaluated, and then virtual screening of large databases of molecules is

performed in order to find compounds that bear resemblance in structure with the desired effect. Due to the difficulty of crystallizing GPCRs, the generation of models of unknown structures from known crystal structures to assist in the study of these GPCR proteins is ideal. This is performed by a technique called homology modelling that compares unknown structures to similar known protein structures. Homology modelling coupled with molecular dynamics provides insight into the behaviour of proteins of unknown structure within the cells.

1.7.1   Molecular Dynamics

Computer simulations, in general, act as a bridge between theory and experiment. Computer simulations are usually performed to assist our understanding of the properties, structures, and microscopic interactions between the assembled molecules. They are mostly used as a complement to support or interpret the findings of well-established experiments, and to enable us to generate new findings that can be tested experimentally. Theories can be tested by conducting a simulation using an *in silico* version of the experimental model, *in silico* models can be compared with experimental results and simulations can be performed computationally that are difficult or impossible to perform in the laboratory, for example, working at extremes of temperature or pressure.

There are two main families of simulation techniques: Molecular dynamics (MD) and Monte Carlo (MC). Molecular dynamics simulations consist of the numerical, step-by-step, solution of the classical equations of Newton's law of motion (1) in computational chemistry to describe ensembles of the movement of atoms and molecules. The major use of molecular dynamics

simulations in biological modelling is to study the behaviour of macromolecules in solution or within the membranes in lipid bilayer systems[158].

$$F = m\,a \qquad (1)$$

Although it is possible and valuable to run molecular dynamics simulations using quantum-chemistry algorithms, the costs of doing so would be prohibitively expensive for the vast majority of systems of interest given the computational resources typically available. It is, therefore, crucial to use approximations, and the classical one called Molecular mechanics (MM) is the principal of these used in biological macromolecular modelling[158].

Molecular mechanics treats atoms as spheres and bonds as springs. This is clearly a very coarse approximation, as it indirectly allows unrestricted small variations in energy, it defines bonds in a fixed manner, and it ignores electronic transitions therefore making it all but impossible to model chemical reactions. Nonetheless, a large number of computational chemistry calculations are concerned with the determination of bond lengths and angles. Molecular mechanics calculations can be remarkably accurate at this, and are thus highly regarded by the scientific community as a viable alternative to the intractable calculations that would be required in quantum descriptions of the computational structural biology[158].

A molecular dynamics simulation performs its calculations using what is called a molecular mechanics force field. This is essentially a description of the

force constants, equilibrium angles and lengths, and interactions between near neighbours', including Coulombic and van der Waal's interactions. Force fields also involve the definition of "atom types" since, for example, nitrogen in an amine group is expected to behave differently from one in a nitro group. This latter point draws attention to the validity of the force field when tackling a given problem. The accuracy that a researcher can expect depends on the molecules that were used to parameterize the force field. Therefore, no force field is universal, some being more appropriate for nucleic acids, others for globular proteins or for lipids, and so on[158].

The force field equation can be summarized by six main contributions to the total potential energy of the system, namely: bond stretching, bond bending (angle), dihedral torsions, out-of-plane torsions, electrostatic/Coulombic interactions (H bond), and other non-bonded interactions (2)[158, 159].

$$F^{total} = F^{bond} + F^{angle} + F^{dihedral} + F^{torsions} + F^{H\,bond} + F^{non-bonded} \quad (2)$$

Molecular dynamic simulations are performed with two types of simulations: an all-atom (AA) and coarse-grained (CG) simulations. The all-atom molecular dynamics causes computational expense of explicitly representing every atom of the solute and solvent that limits all-atom molecular dynamic simulations to a timescale of tens of nanoseconds. To overcome this limitation, coarse grain molecular dynamic (CG-MD) force fields have been developed in which groups of atoms are represented by a single CG-MD bead. In particular, the Martini CG-MD force field that can run to a time scale of

microseconds[160]. The Martini force field equation can be summarized by four main contributions to the total potential energy of the system, namely: bond stretching, bond angle, dihedral, Lennard-Jones/Coulombic interactions (3)[161].

$$F^{total} = F^{bond} + F^{angle} + F^{dihedral} + F^{LJ} + F^{Coulombic} \quad (3)$$

The use of coarse-grained models in a variety of simulation techniques has proven to be a valuable tool to probe the time and length scales of systems beyond what is possible with traditional all-atom models[100]. Coarse-grained models, in which small group of atoms are treated as single particles, provide an approach to increasing the time scale and system dimensions of membrane simulations[162]. The Martini model is based on a four-to-one mapping, i.e., on average four heavy atoms are represented by a single interaction centre, with an exception for ring-like molecules (see Figure 1.6)[104]. For reasons of computational efficiency the mass of the CG beads is set to 72 amu (corresponding to four water molecules) for all beads, except for beads in ring structures, for which the mass is set to 45 amu[100]. Four major CG particle types are prominent: polar (P), mixed polar/aplolar (N), hydrophobic apolar (C), and charged (Q). Four subtypes for N and Q particle are found to allow more fine-tuning of the Lennard-Jones interactions reflecting the hydrogen bonding potentials. The four subtypes are: "0" for no hydrogen bonding capabilities present, "d" for groups acting as hydrogen bond donor, "a" for groups acting as hydrogen bond acceptor, and "da" for groups with both donor and acceptor options. These beads are split into four or five different levels giving a total of 20 different beads For the interactions between the beads, 10 different interaction

levels are the beads used are at normal size (4:1 mapping) for the ring structures a smaller sized mapping (2/3:1 mapping) is used. All atoms except nearest neighbours interact through a Lennard-Jones potential. The nearest neighbours are connected by a weak harmonic spring, next nearest neighbours interact through a harmonic angle potential and the charged groups also interact through a short-range electrostatic potential[106, 161].
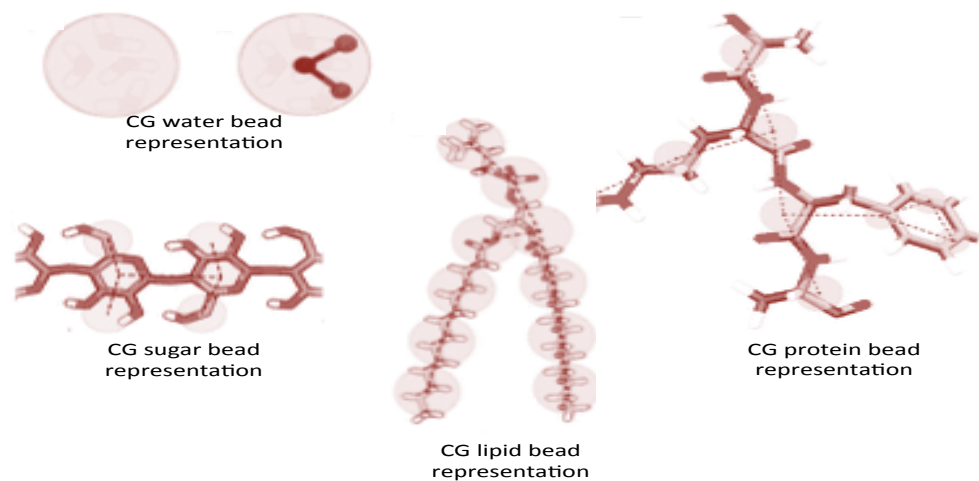
Figure 1.6 Martini force field water, sugar, protein and lipid beads representations.

1.7.2 Homology Modelling

Homology modeling is an *in silico* method, used for predicting the tertiary structure of an unknown amino acid target sequence from a close homologous crystallized template structure, in which the crystal structure has been determined experimentally, to produce an initial 3D structure of the target sequence as close to the native form of the protein[152].

In homology modelling, the two sequences are aligned together by either visual analysis or by using automatized algorithms, or a combination of the two, and then the produced alignment is fed into a specialized homology modelling software to construct the three-dimensional model of the unknown protein structure.

1.8   Aims and outline of this study

The aim of this thesis was to develop a method to investigate the helix-helix dimerisation interface of two interacting GPCR protomers. The structure of the A_{2A} homodimer is presently unknown. The variety of experimental methods with which to study oligomerisation of A_{2A}, together with the availability of crystal structure data for this receptor, make it an ideal candidate with which to integrate experimental and computational methods to better understand the precise molecular nature of GPCR oligomers.

The ability of the A_{2A} receptor to form a homodimer at $M193^{5.54}$ in TM5 formed the starting point for this study. Experimentally, a BiFC system was developed to confirm previous findings of the TM5-TM5 interaction of the A_{2A} receptor and to explore novel findings generated by the computational studies carried out in parallel. Computationally, a consistent, rapid and reproducible CG-MD methodology was developed and used to study wild type and mutated TM helices of the $A_{2A}R$ receptor within lipid bilayers using the 3EML crystal structure as a template. The results from these *in silico* analyses were compared with the available experimental biophysical data and with the structural data for

the rhodopsin (1N3M), CXCR4 chemokine receptor (3ODU) and β_1-Adrenergic receptor (4GPO) homodimers.

The work presented in this thesis defines a novel computational method of ensemble-based coarse-grained classical molecular dynamics that can be used to systematically, reproducibly and reliably determine the specific points of interaction between GPCR dimers. This method will be of great utility in further understanding GPCR function and also has broad applicability to many different types of membrane proteins.

Chapter 2
Materials & Methods

2.1 Experimental Biology Techniques

2.1.1 Cell Lines and Cell Culture Conditions

The following sections describe the basic cell culture splitting and passaging techniques employed in this project.

1321N1 (human astrocytoma) and HEK293T (human embryonic kidney) cells used were purchased from European Collection for Cell Cultures (ECACC). The cells were cultured in T_{75} tissue culture flasks using 50:50 v/v Dulbecco's Modified Eagle's Medium (DMEM/Sigma) supplemented with 50 mL of 10% fetal bovine serum (FBS/Sigma) and 5 mL of 10% L-glutamine (Sigma). The used medium was stored for a maximum length of six weeks at 4°C before replenishing the L-glutamine. Cells were grown at 5% CO_2 in a 37°C humidified tissue culture incubator (Heraeus) and were passaged, transfected and manipulated in a class II Laminar flow safety cabinet (Heraeus).

2.1.2 Cell passaging and splitting

The cells were split and passaged when they reached 100% confluence, forming a complete monolayer sheet that occupied the entire surface area of the flask. They were typically split and passed at 1:3, 1:10, or 1:20 with respect to the surface area of the flask depending on the speed of their growth cycle. The existing medium is aspirated and the cells washed with 10 mL of 1X Dulbecco's

Phosphate-Buffered Saline (DPBS) (Sigma). To detach the cells from the growth surface they were incubated in 10 mL of DPBS-EDTA buffer, containing 5 mL of 0.5 mM EDTA. The resuspended cells were then used to 1) reseed the flask or other flasks or 2) prepare ½ area plates-96 used for Lance cAMP assay.

2.1.3 Cell counting

To determine the volume of cell suspension needed to achieve the desired split, the density of resuspended cells was measured using the CASY® Cell Counter and Analyser System Model TT (INNOVATIS), a technology that performs "standardized cell counting methods, viability checks, aggregation correction, volumetric measurement with high measuring range dynamics." A tailored setup, specifying the selected measuring parameters (used capillary, X axis or range of size distribution, sample volume and number of cycles), the evaluation parameters (dilution factor, Y axis, evaluation cursor, normalization cursor, % calculation, debris, aggregation correction, and mean volume) and output parameters (interface, page feed, print mode and graphic) was created and used throughout the project for all cell count measurements. After the cells were detached from the flask, approximately 1 mL of the suspension was transferred into a clean 1.5 mL Eppendorf test tube.

In order to perform the cell count, a CASY®-cup was first filled with 10 mL of CASY®ton buffer (Roche). 100 µL of the aliquoted cell suspension was then added to the buffer, following which the lid of the cup was secured and the cup was tilted a few times to mix the sample. The vial of diluted sample was placed on the platform below the measuring capillary, allowing an external electrode to be inserted into the sample. The START button was pressed to

59

initiate the measurement, which calculates the number of cells present in a given sample based on the resistance measurement of pulse area analysis. During the measurement, 400 µL of the sample was aspirated by a precision measuring capillary and slowly passed in a measuring pore, through which a low voltage was pulsed. The cells exhibited a size or dimension-dependent resistance to the voltage, whereby an isolated intact cell generated a higher resistance than a dead lysed cell. The resistance resulted from the nucleus membrane rather than the cell surface membrane, thus allowing the discrimination between living and dead cells. The density measurement was performed in triplicate. CASY$^{\circledR}$ automatically operates the necessary steps, taking into account the dilution factor. A size distribution of the sample thus appears in the display, together with a complete numerical evaluation of the measurement data, specifying the number of viable cells in the measured sample in cells/mL and the percentage viability of the cells in the sample. The density of resuspended cells was measured for each 1) cell passage and 2) LANCE assay cell plate set-up in order to ensure that the same density of cells was plated across experiments for consistency and reproducibility purposes.

2.1.3.1 Collection and freezing of cells

Freezer cryotubes were created and labeled for the long-term storage of the newly created cell lines. The monolayer of cells in the T_{75} flask was first washed with DPBS then detached using 10 mL DPBS-EDTA buffer at room temperature for two minutes. The cell suspension was transferred to a clean 50 mL Falcon tube, centrifuged for 5 min at 1020 rcf, following which the tube was removed from the centrifuge and the supernatant was aspirated. A 10% DMSO-FBS

mixture was prepared by adding 9 mL of thawed FBS to 1 mL of DMSO (dimethyl sulfoxide). The cell pellet was resuspended in the DMSO-FBS mixture, and the obtained preparation was distributed into cryotubes (1 mL/tube). The ampoules were first stored in a -80°C freezer then transferred to liquid nitrogen.

2.1.4 Cloning Techniques

The following sections depict the cloning techniques employed in this project, as well as the protocols used to generate the stable BiFC constructs and the mutated constructs.

2.1.4.1 Primer Design

Primers were designed (Table 2.1) for: a) the A_{2A} receptor (Genbank accession number S46950) to allow the cloning of the receptor into the TOPO® TA cloning vector (LifeTechnologies) prior to the production of fusion constructs the biomolecular fluorescent protein complementation vectors were a gift from Chang-Deng Hu (pBiFC-VN173 (Addgene plasmid # 22010), pBiFC-CC155 (Addgene plasmid # 22015); b) mutated primers were designed with each containing a single point mutation to substitute the methionine residues with alanine at the M177 and M193 positions in A_{2A}R-TM5. The mutagenesis primers were designed to the specification of the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). All the primers were synthesized by eurofins genomics.

Type	Sequence 5` → 3`	GC content	Tm
A) hA$_{2A}$R Wildtype Primers			
Forward (37 bp)	ATATGAATTCGTGGCCATGCCCATCATGGGCTCCTCG	54.1%	73.9
Reverse C-terminus (31 bp)	ATATCTCGAGGGACACTCCTGCTCCATCCTG	54.8%	70.8
Reverse N-terminus (31 bp)	ATATGTCGACGGACACTCCTGCTCCATCCTG	54.8%	70.8
B) hA$_{2A}$R Mutated Primers			
1) L19M20A (M193)			
Forward (39 bp)	CTGGTGCCCCTGCTGTTCGCGCTGGGTGTCTATTTGCGG	64.1%	>75
Reverse (39 bp)	CCGCAAATAGACACCCAGCGCGAACAGCAGGGGCACCAG	64.1%	>75
2) M4AV5 (M177)			
Forward (42bp)	GTGGTCCCCATGAACTACGCGGTATACTTCAACTTCTTTGCC	50%	74.3
Reverse (42 bp)	GGCAAAGAAGTTGAAGTATACCGCGTAGTTCATGGGGACCAC	50%	74.3

Table 2.1: The different primer sequences designed to amplify the A$_{2A}$R receptor

A) The hA$_{2A}$R wildtype primer where designed to produce a wild type A$_{2A}$R receptor, B) Primers designed to produce the single point mutation in the A$_{2A}$R receptor at two different positions, M177A and M193A.

2.1.4.2 Amplification of the Plasmid

The A$_{2A}$R sequence was amplified using the proofreading enzyme Accuzyme (Bioline), and the amplicon was cloned into the TOPO© vector (Invitrogen) followed by transformation into chemically competent TOP10F' One Shot® cells (Invitrogen) and placed onto LB agar (Luria broth) plates containing 100 μg/mL ampicillin.

2.1.4.3 Polymerase Chain Reaction (PCR)

PCR reactions were carried out to amplify the hA$_{2A}$R DNA fragments. 25 μL reactions were prepared reaction mixtures were prepared in flat cap PCR tubes and contained:

        1 μL forward primers
        1 μL reverse primers
        1 μL cDNA
        12.5 μL 2X Accuzyme^{TM}
        9.5 μL of Nuclease-free water (Baxter)

The PCR reactions were conducted in the PCR Express Thermal Cycler (Thermo Hybaid) and performed in three stages. The first stage consisted of a 5 minutes denaturation at 95°C for a single cycle. This was followed by a second stage comprising of 29 cycles of 1) 15 seconds of denaturation at 95°C, 2) 30 seconds of primer annealing at 57°C and 3) 15 seconds of extension at 72°C. The final stage consisted of a 5 minutes extension at 72°C for one cycle. The finished reaction was held at 4°C. In order to check the presence of the amplified PCR products, agarose gel electrophoresis was performed.

2.1.4.4 Cloning and Transformation

The fresh amplified amplicon was transfected into TOPO© vector (Invitrogen) before ultimately transforming it into a chemically competent TOP10F' One Shot® cells (Invitrogen). The reaction mixture contained:

6 μL of the TOPO cloning reaction
4 μL of fresh PCR product
1 μL of salt solution (consisting of 200 mM NaCl and 10 mM MgCl$_2$)
1 μL of TOPO vector

The reaction was incubated for 5 minutes at room temperature then placed on ice before transformating 2 μL of cloning mixture into competent TOP10F' One Shot® *Escherichia coli* cells according to the manufacturers protocol and were plated on LB Agar plates containing Ampicillin 100 μg/mL overnight at 37°C. Only the white or light blue colonies were later selected for analysis.

2.1.4.5 Restriction Enzyme Digestion

Restriction endonuclease (RE) digestion was carried out on the plasmids produced by the mini-prep GenElute^{TM} HP Plasmid Mini-prep Kits (Sigma) and pBiFC vectors, with approximately 1 unit of enzyme(s) per 50 μg of DNA, at 37°C in water-bath (Grant) for one to one and a half hours.

For the $A_{2A}R$ PCR product digestion was prepared containing:

    3 μL plasmid DNA
    1 μL 10X Buffer H (Promega)
    0.5 μL EcoRI (10 u/μL, Promega)
    0.5 μL SalI or XhoII (10 u/μL, Promega)
    5 μL of distilled water (Baxter)

For the pBiCF vectors digestion was prepared containing:

    10 μg pBiFC-(VN173 or CC155) vector
    20 μL 10X Buffer H (Promega)
    5 μL EcoRI (10 u/μL, Promega)
    5 μL SalI or XhoII (each at 10 u/μL, Promega)
    The volume was adjusted with distilled water (Baxter)

2.1.4.6 DNA Ligation

The ligation process was prepared in 0.2 mL flat cap PCR, a molar ratio of 1:3 pBiFC vectors to DNA insertion. The maximum plasmid vector concentration used was 50 ng per reaction.

For ligation process 20 μL mixture were prepared and contained:

    1 μL of the T4 DNA Ligase (NewEnglandBiolabs-NEB)
    2 μL of 10X Ligase buffer (NEB)
    2-5 μL of DNA insert

The volume was adjusted with distilled water and the reaction was either incubated at room temperature for one hour or at 10°C overnight. Following ligation, the reaction mixtures were transformed using One Shot® TOP10F` cells, according to the manufactures instructions and were plated on LB Agar plates containing Ampicillin 100 μg/mL.

2.1.5 Gel Electrophoresis and Purification

To view the PCR products, after every amplification/restriction enzyme digestion, a 1% agarose gel in 1X TAE buffer was prepared and the gel was pre-stained with ethidium bromide (10 mg/mL) (Fluka). DNA samples were prepared by adding 2 μL of 5X Loading Buffer (Bioline) to 5 μL of PCR reaction. The gel was run in 1X TAE buffer for 30 minutes at 120 V. The Invitrogen 1 Kb Plus DNA ladder (10 μL) was used as standard marker. DNA bands were visualized under a UV-Transiluminator (Syngene).

The desired $A_{2A}R$ DNA fragment (~1100 bp) and pBiFC vector fragments (~4100-5100 bp) were excised from gel by High Pure PCR Product Purification kit (Roche) according to the manufacturers protocol.

2.1.6 Transfection of Cells with Plasmids

Generation of the transient 1321N1 cells was undertaken using the Lipofectamine^{TM} 2000 (Invitrogen) method, a patented formulation with the highest transfection efficiency for many eukaryotic cell types with circular nucleic acids.

The exponentially growing untransfected 1321N1 cells were grown in T_{75} flask until they were 90-95% confluent. The cells were then split and transferred into a clean 50 mL Falcon tube. The densities were then counted and cell suspensions of 40,000 to 80,000 cells/cm^2 cells in DMEM were each plated in either 96- or 6-well plates. The cells were transfected with 0.2 µg of the A_{2A}-pBiFC-VN173/CC155 plasmids, transfection protocol was carried-out according to the manufactures' procedure. The cells were incubated in 5% CO_2 humidified tissue culture incubator for 18-48 hours prior to testing expression.

2.1.7 Plasmid DNA Preparation

The screening for the desired recombinant plasmid constructs was conducted through the GenElute^{TM} HP Plasmid Mini-prep Kits (Sigma). A single colony was chosen and picked from the LB agar plates and inoculated in 2 mL LB broth containing 100 µg/mL Ampicillin. The culture was grown in 14 mL Falcon 2059 Polypropylene round-bottom tube (Becton Dickinson) at 37°C shaker incubator at 200 rpm overnight.

Following the mini-prep screening, large-scale transfection grade plasmid DNA of the remaining culture was isolated using QIAGEN Plasmid Maxi Kit^{TM} according to the manufactures' instructions.

2.1.8 Quantification of DNA

The extracted DNA was quantified using NanoDrop^{TM} 2000 (Thermo-scientific) measurements, which can determine the absorbance of a given sample at different wavelengths. The DNA concentration was measured in ng/µL and the $OD_{260/280}$ ratio of the analysed sample was determined. This ratio indicates the

purity of DNA, whereby a ratio of approximately 1.8 is accepted as pure for DNA.

2.1.8.1 LANCE® Ultra cAMP assay

The LANCE® *Ultra* cAMP assay (PerkinElmer) is a second generation homogeneous time-resolved fluorescence resonance energy transfer (TR-FRET) immunoassay designed to measure cAMP produced upon modulation of adenylyl cyclase activity by any G_s- or G_i-coupled GPCRs in the cellular membrane samples. The assay principle is based on the competition between the europium (Eu) chelate-labeled cAMP tracer and sample cAMP for binding sites on cAMP-specific monoclonal antibodies labeled with the U*Light*TM dye (Figure 2.1). When antibodies are bound to the Eu-labeled cAMP tracer, light pulse at 320 or 340 nm excites the Eu chelate molecule of the tracer. The energy emitted by the excited Eu chelate is transferred by FRET to U*Light* molecules on the antibodies, which in turn emit light at 665 nm. Residual energy from the Eu chelate will produce light at 615 nm. In the absence of free cAMP, maximal TR-FRET signal is achieved (Figure 2.1, left panel). Free cAMP produced by stimulated cells competes with the Eu-cAMP tracer for the binding to the U*Light*-mAb, causing a decrease in TR-FRET signal. The intensity of the signal measured at 665 nm is inversely proportional to the cAMP concentration in the sample.

This technique provides high sensitivity and signal stability, constitutes a standard approach for the pharmacological characterization and determination of relative potencies of ligands at receptors, thus allowing a quantification of agonist effects.

In the absence of free cAMP

In the presence of free cAMP
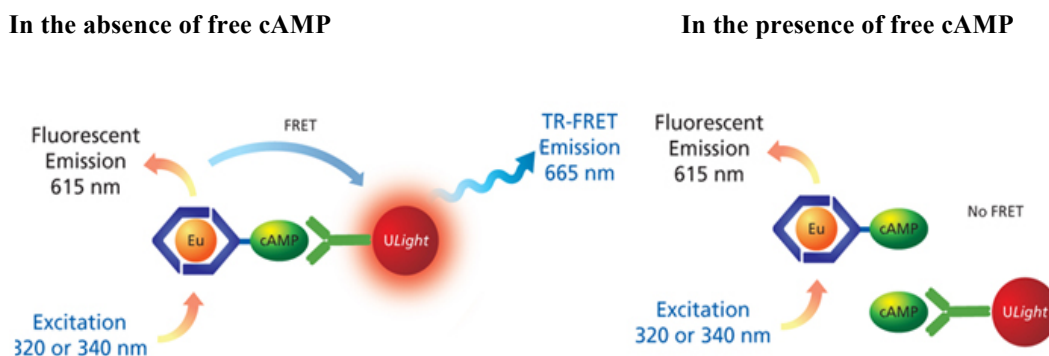
Figure 2.1 The Lance® Ultra cAMP assay[163].

2.1.8.2 The LANCE® Ultra assay procedure

The standard assay procedure is conducted in two steps. In the first step, cells or cAMP standard in suspension are stimulated for 30 min with the selected compound(s). Following stimulation, cellular cAMP is detected by the successive additions of Eu-cAMP tracer and U*Light*-anti-cAMP prepared in the cAMP Detection buffer provided with the kit. The TR-FRET signal at 665 nm is measured by SpectraMax i3 (Molecular Devices). cAMP standard curves are preformed to determine the assay sensitivity and dynamic range provided by each cAMP.

4X serial dilutions were prepared for the cAMP, Eu-cAMP tracer and U*Light*-anti-cAMP in Stimulation buffer pH 7.4. The 4X cAMP serial dilutions were prepared in half log units from 10^{-6}-10^{-11} M with a starting concentration of 50 μM cAMP standard supplied in the kit. The 4X Eu-cAMP tracer working solution was prepared by making a 1/50 dilution of the Eu-cAMP tracer stock solution in cAMP Detection buffer supplied in the kit, where 980 μL of cAMP

Detection buffer was added to 20 μL of the Eu-cAMP tracer aliquot stock solution and mixed gently. The 4X U*Light*-anti-cAMP working solution was prepared by making a 1/150 dilution of the U*Light*-anti-cAMP stock solution in cAMP Detection buffer, where 6.7 μL of the U*Light*-anti-cAMP aliquot stock solution was diluted with 998.3 μL of cAMP Detection buffer and mixed gently.

The Stimulation buffer, 15 mL pH 7.4, was prepared in 50 mL Falcon tube and contained:

> 14 mL of 1X HBSS (Invitrogen)
> 75 μL of 1M HEPES pH 7.2-7.5 (Invitrogen)
> 150 μL 100 μM RO-201724
> 200 μlL of 7.5% BSA Stabilizer (included in the kit)
> The pH was adjust with 0.1 N NaOH and the volume was
>
> completed to 15 mL with 1X HBSS

The LANCE® *Ultra* assay was prepared in ½ AreaPlate-96 white plate. A final volume of 40 μL was prepared and contained:

> 10 μL of the cells or cAMP
> 10 μL of compounds or Stimulation buffer
> 10 μL of Eu-cAMP tracer
> 10 μL U*Light*-anti-cAMP

The reaction had two incubation steps. The cells or cAMP with compounds or stimulation buffer were incubated for 30 minutes at room temperature. Subsequently, after addition of the Eu-cAMP tracer and the U*Light*-anti-cAMP the mixture was sealed with microplate seal and incubated at room temperature for one hour. The plate was then read on SpectraMax i3. White plates were used as the black plates reduce the signal.

2.1.8.3 Cell density optimization

The first step in the assay was to identify the optimal cell density by performing a forskolin and cell density cross-titration experiments. A replica of three experiments was preformed for this. Cells were grown in T_{75} flask until they reach 100% confluent, washed in DPBS and lifted in 10 mL DPBS-EDTA as previously described. The detached cells were then transferred to 50 mL Falcon tube and counted to determine the volume of cell suspension required to generate a 10×10^7 cells/mL dilution. Sequentially, 1 mL of the obtained 10×10^7 dilution was dispensed into another tube and centrifuged for 5 minutes at 1020 rcf, and the DPBS-EDTA buffer was aspirated. The cells were re-suspended in Stimulation buffer according to the cell density need for plating, approximately 10×10^5 cells/well and 10 μL of the mixture was seeded onto appropriate well of a ½ AreaPlate-96 followed by addition of the selected compounds.

2.1.8.4 Running the LANCE® Ultra Assay

The cAMP measurement assays were conducted using the SpectraMax® i3 (Microplate Plate Reader) (Molecular Devices). This technology uses an integrated optical system enabling top and bottom reads for 6 - 384 well microplates with three broad detection modes: luminescence, absorbance, and fluorescence with a possibility to upgrade to additional applications and detection modes such as cartridges for Time-Resolved Fluorescence. We have used the Cisbio HTRF cartridge for Time-Resolved Fluorescence (0200-7011).

After the one-hour incubation at room temperature the microplate seal was removed and the plate was loaded into the SpectraMax i3. First, the plate type

was selected (½ AreaPlate-96) and then an optimization run was conducted to get the recommended height read of the samples in the plate, for a 40 μL sample the recommended read height was usually between 5.38 mm to 5.56 mm. The SpectraMax i3 system was programmed to utilize a Xenon flash lamp. The excitation filter was set at 340 nm with an excitation time of 0.05 ms and 20 pules. The fluorescent signals are emitted from the Eu chelate following excitation and passed through an emission filter was a wavelength of 616 nm, the U*Light* FRET emission signal was passed through the 665 nm filter. The emission measurement delay was set to 0.03 ms after excitation. The integration time was set to 0.5 ms. These emission were passed through cooled Photomultiplier tubes (PMT) channel with two dichroic filters located on top of the cell plate, allowing fluorescent/FRET signals to be collected.

The recorded data was saved on the SoftMax Pro software for statistical analysis. They were then subjected to further analysis using GraphPad Prism.

2.1.8.5 Cisbio HTRF SpectraMax® Detection Cartridge

The HTRF technology combines standard FRET technology with time-resolved measurement of fluorescence that eliminates the short-lived background fluorescence. HTRF acceptors emit long-lived fluorescence when engaged in a FRET process this will compensate for compound interference and sample quenching and also minimizes background fluorescence detection. Therefore, the long-lived emissions signify energy transfer due to the proximity of the labeled biomolecules. The FRET technology relies on the physics of molecular proximity, when the distance between the donor and the acceptor is short

enough. A FRET system is characterized by the Förster's radius (R0) distance at which FRET efficiency is 50%. The HTRF R0 lies between 50 and 90 Å, depending on the acceptor used. The cartridge uses a Xenon flash light source, with an excitation filter of 320 or 340 nm and the emission 0 EM1 616/10 nm to EM2 665/10 nm. The HTRF cartridge used measures the emissions of the Eu-cAMP tracer and the transferred energy from Eu chelate to the U*Light*-anti-cAMP simultaneously, at the 616 nm and 665 nm respectively.

2.1.8.6 Data Analysis

The statistical and quantitative analysis of data was conducted using SoftMax Pro 6.4 and GraphPad Prism®. SoftMax Pro was used for primary calculations based on TR-FRET signal obtained from SpectraMax. Two signals were obtained at 665 nm (Ul*ight* acceptor channel) and 616 nm (Eu chelate donor channel). The TR-FRET signal at 665 nm was used directly for data analysis of: 1) cAMP standard curves, Figure 2.2 shows counts at 665 nm obtained in cAMP standard curves that allow interpolating the amount of cAMP produced in stimulated cells; 2) determination of optimal cell density; 3) for agonist and antagonist characterization. The signal at 616 nm is used to identify dispensing malfunctions and some compounds artifacts that can lead to signal decrease. To overcome this a quench correction calculation was performed using a blank-corrected normalized ratio, the following equation in SoftMax was applied:

$$F_{665,CS} = [(F_{665,S} - F_{665,BL}) \times F_{616,MAX}] / F_{616,S}$$

Where the corrected signal ($F_{665,CS}$) is produced through determining the blank value ($F_{665,BL}$) by measuring the background signal at 665 nm of buffer containing wells. The ($F_{665,S}$) and ($F_{616,S}$) signals were determined from sample wells containing the assay components and test compounds. The ($F_{616,MAX}$) was obtained from wells with no cAMP.

The Prism® software was employed for the analysis of data processed using SoftMax Pro, through curve fitting through non-linear regression. Each represented data set consisted of three or more replicates (n=3) for each concentration level.
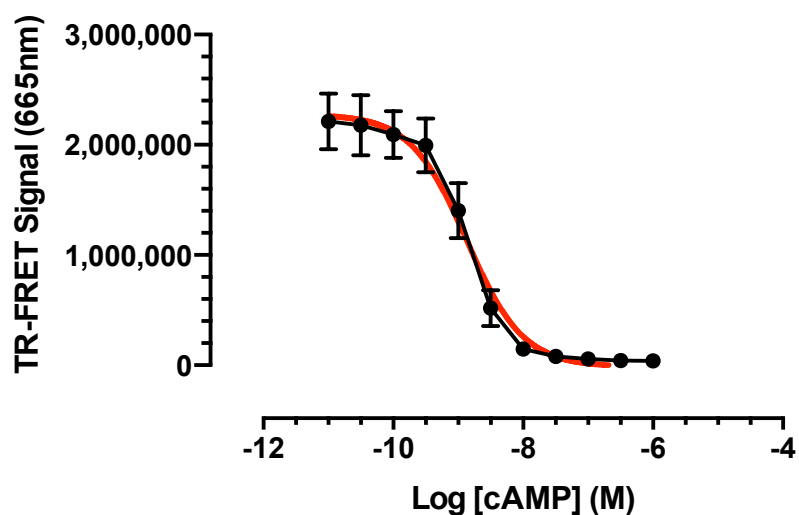
Figure 2.2 cAMP standard curve.

2.2 Computational Biology Systems

The following sections describe the modelling, simulation parameters employed in adenosine A_{2A} receptor, rhodopsin, CXCR4 and β_1AR transmembranes simulations, as well as, how the CG-MD simulations were analysed. The aim was to develop a consistent, rapid, reproducible CG-MD methodology for the study of interacting helices. This method will involve placing two GPCR transmembrane helices in a membrane and running ensemble simulations comprised of multiple replicas, each, hoping to identify interactions between the helices. In these simulations, distance will be used as a means of identifying two different types of interactions: interactions between helices and interactions between amino acid residues on each helix. For the successful interaction of both types of interaction, it will be necessary to specify the number of replicas (independent simulations identical other than for the initial velocity seeds assigned to the particles) and the run time needed to achieve converged results and see how well they reproduce experimental results. The number of replicas must be sufficient to achieve a reproducible result as evidenced by a sufficiently small error estimate.

From Table 1.4, it can be seen that the longest total simulation time for atomistic MD is 0.1 μs and for CG simulations is 200 μs. The formation of a long-lasting helix dimer was identified within a few hundred nanoseconds in CG-MD studies of Glycophorin A, a non-GPCR model for studying TM membrane protein structure[86]. The number of replicas performed in these different studies varies tremendously, but is never greater than 10. Excellent agreement has been

obtained between computed binding free energies and experimental data when ensembles of 50 replicas are used[164]. 500 ns was selected for the run time and 50 replicas as starting parameters for these studies. These calculations were run on Legion and Grace, two high-performance Research Computing cluster at UCL (for details of the machines used can be found at https://wiki.rc.ucl.ac.uk/wiki/RC_Systems#Legion_technical_specs and https://wiki.rc.ucl.ac.uk/wiki/RC_Systems#Grace_technical_specs). Our preliminary tests showed that CG-simulations (1 ensemble) run on Legion for 500 ns completed within approximately 150 hours. CG-simulations (1 ensemble) run on Grace for 500 ns completed within approximately 72 hours.

2.2.1 CG Simulations

All CG-MD simulations were performed in GROMACS (version 4.6.4) (www.gromacs.org). The temperature was equilibrated for all three groups: protein, lipid bilayer, and solvent (water) with ions to remove center of mass motion relative to the bilayer and protein. The thermalisation run was carried out for 100 ps. The simulation was then run at 310 K (the mammalian physiological temperature), which is below the phase transition temperature of pure DPPC (315 K). The system output of the temperature was evaluated to make sure that it stabilized at the required temperature (310 K) before continuing until pressure equilibration was attained. An ensemble of 50 replicas for each simulation box (see Tables 2.2 and 2.3) was performed. Each simulation was run for 500 ns. CG-atom velocities were drawn from a Maxwell-Boltzmann distribution at T=310 K, but all other variables were kept constant, therefore standard deviation was used to compare differences in mean distance outputs. Each simulation was run independently, with the initial configuration differing only by the starting

velocity; they were performed under constant temperature, pressure, and particle number using the Martini 2.2 force field[165]. The temperature of the protein and the lipid were coupled using the velocity-rescaling (modified Berendsen) thermostat at 310 K (human physiological body temperature), with a coupling constant of T_t = 1 ps. The system pressure was semi-isotropic using the Berendsen algorithm at 1 bar, with a coupling constant of T_p = 1 ps and a compressibility of 1×10^{-4} bar^{-1}. A time step integration of 30 fs was chosen and the coordinates were saved every 10000 subsequent steps for further analysis. The electrostatic interactions were shifted to zero between 0 to 1.2 nm. The Lennard-Jones (LJ) potential was shifted to zero between 0.9 to 1.2 nm to reduce the cutoff noise. The neighbour list for pairwise non-bonded interactions was determined by the Verlet cutoff scheme at 1.4 nm and updated every 10 steps.

Simulations for the $A_{2A}R$ receptor were carried out as follows: 1) wildtype TM5 helices; 2) mutated TM5 helices at M177A helices found in the other PxxxM conserved motif in all of the adenosine receptors; 3) mutated TM5 helices at M193A[64] helices; 4) mutated TM5 helices at M193 helices to isoleucine (M193I) because PxxxI is the conserved motif found in the P2Y receptors located in same position as the PxxxM motif in the P1 receptors; 5) mutated TM5 helices at Y197A and Y197F; 6) simulations of pairwise combinations of all the homo/heterologous possible pairs of the $A_{2A}R$ TM helices (TM1, TM2, TM3, TM4, TM5, TM6, and TM7); 7) mutated TM1 helix at E13A with TM2 found to be a strong interacting pair from pairwise simulation. The choice of mutating residues to alanine was motivated by standard biochemical techniques (such as alanine scanning mutagenesis) for

assessing side-chain function. The DPPC lipid bilayer used in system resembles a biological membrane designed for drug delivery[166]. Sequences used in all the simulations are found in Table 2.2.

Simulations for the other GPCRs were carried out as follows: for rhodopsin 1) TM1-TM2 helices, 2) TM4-TM5 helices; in CXCR4 1) TM5-TM6 helices, 2) TM5-TM5 helices and in β₁AR 1) TM1-TM2 helices, 2) TM4-TM5 helices and 3) TM1-TM1 helices. Sequences used in all the simulations are found in Table 2.3.

A$_{2A}$R α-Helices	Sequences
TM1	VYITVELAIAVLAILGN$^{1.50}$(24)VLVCWAV
TM1-E6A	VYITV**A**LAIAVLAILGN$^{1.50}$(24)VLVCWAV
TM2	FVVSLAAAD$^{2.50}$(52)IAVGVLAIPFAITI
TM3	L(78)FIACFVLVLTQSSIFSLLAIAI$^{3.49}$(100)
TM4	KGIIAICW$^{4.50}$(129)VLSFAIGLTPMLG
TM5-Wild-type	**MNYM**VYFNFFACVLV**P**$^{5.50}$**(189)LLLM**LGVYLRI
TM5-M177A	**MNY*A***VYFNFFACVLV**P**$^{5.50}$**(189)LLLM**LGVYLRI
TM5-M193A	**MNYM**VYFNFFACVLV**P**$^{5.50}$**(189)LLL*A***LGVYLRI
TM5-M193I	**MNYM**VYFNFFACVLV**P**$^{5.50}$**(189)LLL*I***LGVYLRI
TM5-Y197A	**MNYM**VYFNFFACVLV**P**$^{5.50}$**(189)LLLM**LGV*A*LRI
TM5-Y197F	**MNYM**VYFNFFACVLV**P**$^{5.50}$**(189)LLLM**LGV*F*LRI
TM6	LAIIVGLFALCWL**P**$^{6.50}$(248)LHIINCFTFF
TM7	WLMYLAIVLSHTNSVVN**P**$^{7.50}$(285)FIYAYRI

Table 2.2 Sequences of the A$_{2A}$R helices used in simulations

A$_{2A}$R TM5 sequence and its mutated residues suggested to play a key role in the dimerisation of the A$_{2A}$R TM helix are in bold; the mutated residues are underlined and written in italic (blue). The conserved amino acid numbered through the Ballesteros and Weinstein system of each individual transmembrane is shown in red.

Receptor	Helices	Sequences[a]
Rhodopsin	TM1	QFSMLAAYMFLLIMLGFPIN$^{1.50}$(55)FLTLYVTVQ
	TM2	NYILLNLAVAD$^{2.50}$(83)LFMVFGGFTTTLYTSLH
	TM4	ENHAIMGVAFTW$^{4.50}$(161)VMALACAAPPL
	TM5	NESFVIYMFVVHFIIP$^{5.50}$(215)LIVIFFCYGQ
CXCR4	TM5	VVVFQFQHIMVGLILP$^{5.50}$(211)GIVIL
	TM6	VILILAFFACWLP$^{6.50}$(254)YYIGISI
β_1AR	TM1	QWEAGMSLLMALVVLLIVAGN$^{1.50}$(59)VLVIAAIG
	TM2	NLFITSLACAD$^{2.50}$(87)LVMGLLVVPFGATLVV
	TM4	ARAKVIICTVW$^{4.50}$(166)AISALVSFLPIMM
	TM5	AYAIASSIISFYIP$^{5.50}$(219)LLIMIFVYLRVY

Table 2.3 Sequences of the rhodopsin, CXCR4 and β_1AR receptor helices used in ensemble simulation sets

[a] The conserved amino acid for each TM helix is shown in red and is numbered using both the Ballesteros and Weinstein nomenclature (superscript) and by residue number.

2.2.2 Construction of TM Helices and Preparation of the Simulation Box

Initial simulations were performed using TM5 of the human A_{2A} adenosine receptor, which has been shown experimentally to form a homodimer[147]. Amino acid positions have been described using amino acid number in conjunction with the Ballasteros Weinstein nomenclature[167] (in superscript). The A_{2A} TM helices sequences shown in Table 2.2, were generated using MODELLER 9.12 following the procedure detailed[168, 169], using the crystal structure of the A_{2A} receptor (PDB accession number 3EML; GI: 209447557)[137]. The atomistic helices were subsequently converted into CG models using the "martinize" Python script (see Figure 2.3)[105]. A simulation box of dimensions 8 nm x 8 nm x 8 nm was constructed containing two wild-type TM5 helices (Figure 2.4). The helices were placed 4 nm apart and aligned in a parallel orientation mimicking the natural positioning of the helix in the membrane (see Figure 2.4a), with their long axes parallel to the z-axis of the box

(see Figure 2.4b). The TM helices were separated by 4 nm at the beginning of the simulation to rule out any initial inter-helix interactions. Water and lipids were then added. Approximately ~190 molecules of the 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) lipid bilayer and additional water molecules (~2660-2690) were added in coarse-grained form in a 3-dimensional cuboid box with periodic boundary conditions using the "insane" Python script[170]. To neutralise the net charge on the protein, water molecules were replaced by counterions (either Na^+ or Cl^-, as appropriate, depending on the amino acid composition of the helices). The rhodopsin, CXCR4 and β_1AR TM helices sequences shown in Table 2.3, and were constructed as described above for the A_{2A} receptor.
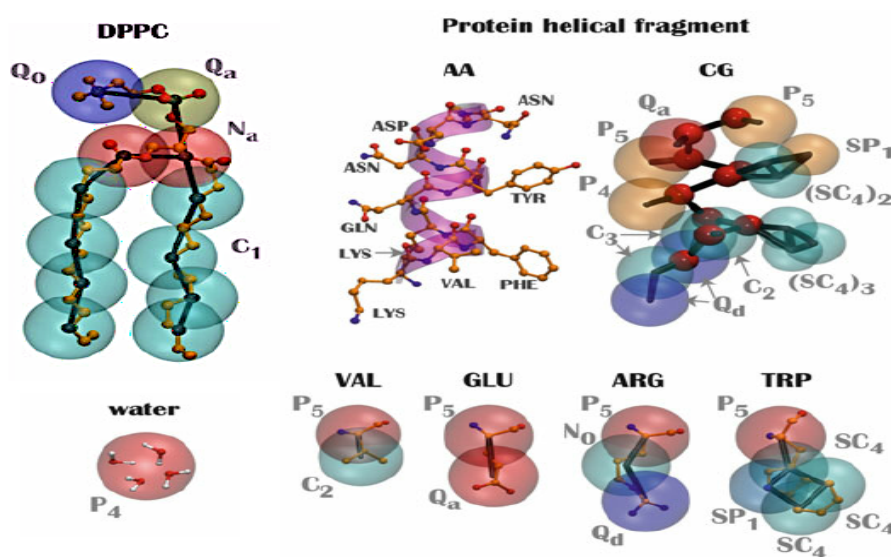
Figure 2.3 Mapping between the chemical structure at the atomistic level (AA) with the coarse-grained (CG) Martini model[100]

The CG beads are shown as transparent vdW spheres for DPPC lipid bilayer, a protein fragment, water and some amino acids.

2.2.3 Dimer Analysis

Interhelix distance matrices were calculated for the helix-helix dimer formation and contact maps used to identify specific interactions between residues were generated using the GROMACS tool g_mdmat. The individual helix-helix contacts from each replica were examined by calculating the resulting inter-helix distance matrices from the initial simulation starting distance of 4 nm (40Å) to assess the reproducibility, number of replicas and run time needed to achieve convergence through a locally written code. In those runs where dimerisation was observed, the trajectories were combined and examined in greater depth by calculating the averaged inter-helix distance matrices with three different truncation distances, 15Å, 12Å and 10Å, to determine the dimerisation properties between the helices. A cutoff distance of 10Å was then applied to identify residues involved in helix-helix interactions. To investigate the influence of the number of replica simulations on the reliability of our results, we calculated mean interaction distances for ensembles of varying size. For evaluations of run length, mean distance output for the entire ensemble was calculated at 100 ns increments.

Representative atomistic structures of the different CG dimers were generated through use of the "backward" Python script[171] and the g_cluster tool in GROMACS using the gromos algorithm at a cutoff of 2.5 nm[172]. Visualization was performed using VMD[173]. Approximate distances between the atomistic residues in interacting helices were measured using Jmol. Pairwise combinations used in the analyses were obtained from a matrix of the number of residues in helix 1 multiplied by the number of residues in helix 2. In the A_{2A}

receptor, there are 729 possible pair combinations between the two 27 residue long TM5 helices, for example, combination 552 specifies the combination of residue 23 (helix 1) with residue 24 (helix 2), representing the $V196^{5.57}$-$Y197^{5.58}$ interaction.

Given the $A_{2A}R$ seven TM helices, there are 28 possible pairs in total where 21 would be heterologous pairs and seven possibly homologous pairs. Also there are three pairs from mutated TM5 helices interactions. All of these possible self-interactions were examined *in silico*. The terms "stable dimer" and "dimerisation" to refer to interactions between helices. A 10Å truncation cutoff (backbone to backbone) has been set for dimerisation, as it has been shown experimentally that a unique FRET signal is generated when two labeled peptides are located within 10Å of each other and form an excited stated dimer[174]. The term "specific interactions" will be used to refer to interactions between amino acid side chains on the dimerised helices. Specific interactions will be identified from contact matrices (heat maps). Although a 12Å truncation cutoff had previously been used to analyse these[175], we will set our interaction cutoff to 10Å, since the existence of hydrogen bond (C_{α}-H$\cdots$O) contacts as a function of the inter-helical axial distance is between 6 and 12Å. Side-chain to side-chain distances consistent with this are used to identify specific interactions, with distances of 5-7Å reflecting stronger interactions. Not all pairs examined showed a favor to the homo- or hetero- dimer formation.
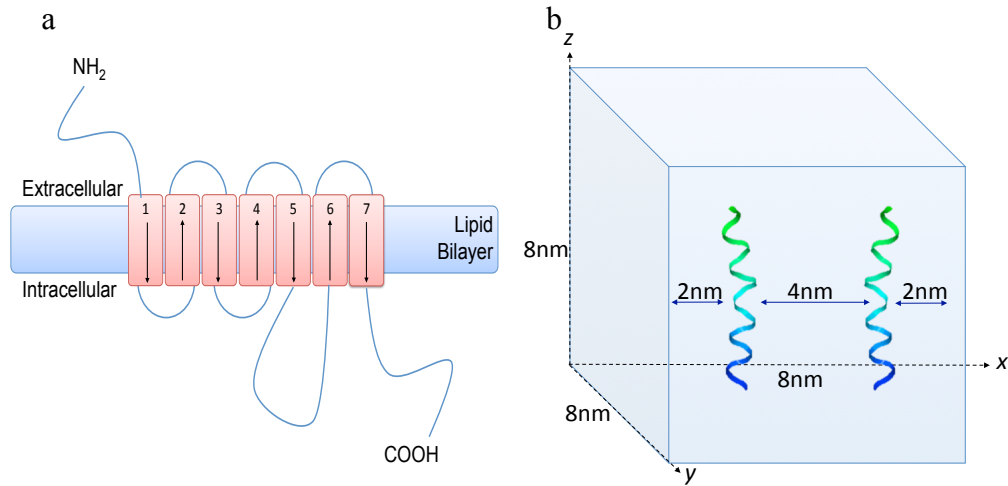
Figure 2.4 A simulation box of two wild-type TM5 helices

a) A representation of the A_{2A} receptor structure shows the direction of the TM helices in the lipid bilayer. b) A representation of the positioning of the TM in a simulation box before the addition of lipid and the water.

2.2.4 Visualisation and Data Analysis

All the simulations were extensively viewed with the graphical viewing package VMD[173] during all the steps of the simulation. The results obtained after MD runs were plotted with Xmgrace[176]. Representative structures of the interacting helices were generated using g_cluster tool in Gromacs, with the gromos algorithm[172] to cluster on RMSD with a cutoff of 2.5 nm.

Chapter 3
Development of an integrated experimental-computational approach for the characterisation of GPCR dimers

3.1 Introduction

As described in Chapter 1, GPCRs exist in various oligomeric forms that have shown to be of functional relevance. A means of accurately identifying the specific points of interaction between GPCR dimers will provide valuable information about receptor function and is essential for the design of receptor ligands against specific oligomeric receptor targets. A wide variety of different approaches have been used to characterise GPCR dimers. Biophysical methods have been used to identify protein-protein interactions, with site-directed mutagenesis used to identify key amino acid residues involved in dimer formation. These experiments are able to be replicated multiple times, producing reproducible results that identify interactions between two protein domains and provide some clue to the amino acids involved, but do not provide detailed molecular information about the interaction interface. The computational modeling of mammalian GPCR dimers, shown in Table 1.4, can provide detailed molecular information, but these have typically been single simulations and not multiple replicas performed according to a standard protocol; the reliability and

accuracy of the molecular information they provide remains to be demonstrably proven. There is, therefore, a need for a methodology that integrates both experimental and computational data in the analysis of GPCR dimerisation, in order to provide accurate information about the molecular nature of the dimer interface between two interacting GPCRs. As previously explained, the A_{2A} receptor has been shown experimentally to dimerise at M193 of TM5 and a number of A_{2A} crystal structures have been obtained, making this an ideal model with which to develop an integrated experimental-computational approach to the study of GPCR dimers.

This chapter endeavours first to establish an experimental system within which to evaluate the protein-protein interactions between two GPCR monomers at both the biochemical and the functional level. This system will be developed using the experimental data that have been obtained for the A_{2A} receptor TM5 dimerisation interface. A_{2A} dimerisation will be detected experimentally, through biochemical assays such as the BiFC technique, and through functional assays, such as the measurement of cAMP production, which is the primary signaling pathway of G_s-coupled GPCRs, including the A_{2A} receptor. This chapter subsequently aims to develop a computational method for obtaining results *in silico* that match the experimentally-obtained data for the self-association of the TM5 helices of the A_{2A} receptor. TM2 will be used as a negative control, as it was unable to form a homodimer under the same experimental conditions[64]. The computational method of choice is ensemble-based molecular dynamics simulations. This method has been chosen because neighboring trajectories diverge exponentially rapidly and only probabilistic descriptions are meaningful.

For these intrinsic reasons, collections of trajectories differing only in their initial conditions, known as ensembles, are the best means of studying the properties of such systems. Each individual system in the ensemble is referred to as a replica.

The development of an integrated experimental-computational approach to the characterization of GPCR dimers will be carried out in two sequential steps. First, the 85dimerisation of the full-length A_{2A} receptor will be structurally verified using the BiFC technique by constructing wild-type and M193A-mutated receptors and transfecting these into HEK293T cells. Signalling will be characterized by measuring cAMP levels using the Lance Ultra[®] technique in recombinant HEK293T cells transiently expressing these BiFC constructs. Second, the A_{2A} receptor 85 dimerisation interface will be investigated computationally through CG-MD simulations of wild-type and mutated TM5 in DPPC lipid bilayer. The minimum number of replicas in an ensemble and the minimum run time for each replica that is required to obtain results that converge with experimentally-obtained data will be determined.

3.2 Creation and functional testing of the A_{2A}-BiFC Constructs

The co-immunoprecipitation human A_{2A}R construct (GenBank accession number S46950) was generated for the wild-type A_{2A}R. The hA_{2A}R cDNA was then amplified using one forward and two reverse primers to accommodate the later insertion into the two BiFC plasmids minus stop codon, 1) pBiFC-CC155 and 2) pBiFC-VN173. The resulting product bearing the 5` and 3` additions was 1263 bp long. This fragment was assembled through ligation into a TOPO® TA cloning vector that possess an Ampicillin gene. The ligation reaction product was

transfected into TOP10F` One Shot® chemically competent E.coli, following which the transformation was plated on Ampicillin selective agar plates. A small-scale isolation of plasmid DNA (miniprep) was performed to screen for the ligation process products that represents the desired recombinant plasmid constructs and a restriction enzyme digestion was conducted to test for the successful insertion of $hA_{2A}R$ in the TOPO vector. The purified DNA was cut with EcoRI, SalI, and XhoI, yielding two fragments of the correct size corresponding to the TOPO vector (3900 bp) and $hA_{2A}R$ (1263 bp), respectively. Large-scale isolation of plasmid DNA was subsequently carried out using QIAGEN to $hA_{2A}R$, following which the purified DNA was digested using previously mentioned restriction enzymes.

3.3 Creation of the A_{2A}-BiFC constructs

A fluorescent $A_{2A}R$ constructs were created using the co-immunoprecipitates cDNA described previously was tagged with the two fluorescent partners, Venus and Cerulean. cDNA was cloned into Venus (pBiFC-CC155) between the EcoR1 and XholI sites and into Cerulean (pBiFC-VN173) between the EcoR1 and SalI sites (shown in Figure 3.1). The cDNAs encoding the $A_{2A}R$ were tagged at their N-terminal domain with either an HA or a FLAG tag immediately preceded by a consensus Kozak sequence to facilitate translation. The stop codon was removed from the GPCR coding sequences so that all downstream FP sequences could be fused in-frame. A linker sequence at the 3`-end of the GPCR was employed to keep the sequence in-frame. Immediately downstream from the linker, the cDNA encoding partial sequences of the relevant FP, either the EYFP N-terminal fragment comprised of amino

acids residues 155-239 (C155) of Venous, or the ECFP C-terminal fragments

comprised of residues 1-172 (N173) of Cerulean, were fused in-frame. The

resulting constructs are chimeric fusions. The cloning steps used to generate the

chimeric BiFC constructs are found in Figure 3.2.

Subsequently, two mutated $hA_{2A}R$-BiFC constructs the 1) mutated $A_{2A}R$-

M177A and the 2) mutated $A_{2A}R$-M193A were generated through amplification

of $hA_{2A}R$-BiFC constructs using a set of forward and reverse primers containing

the mutated sequences (primers length ~39-42 bp) using the QuikChange II XL

Site-Directed Mutagenesis Kit.
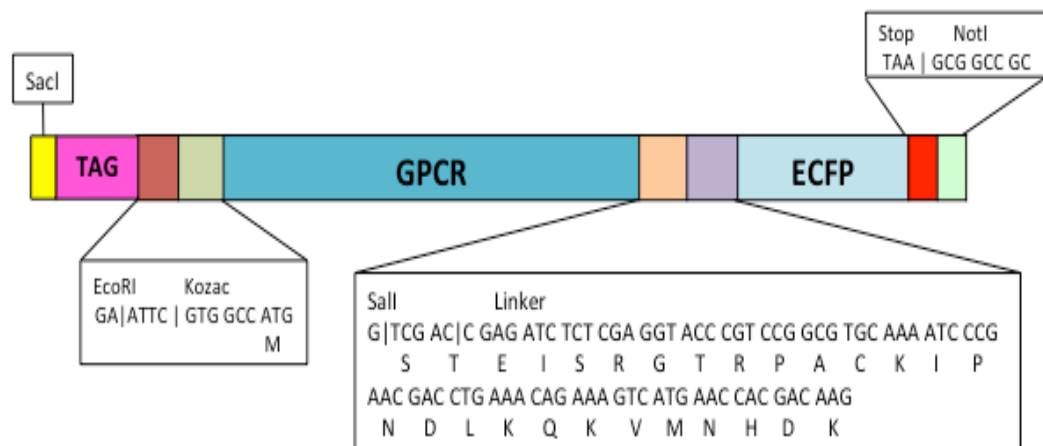
Figure 3.1 Schematic representation of the BiFC-CC155 constructed design

The GPCR were excised from TOPO-PCR vector and inserted in the pBiFC-VN173 FP vector. The GPCR sequence is cloned between EcoRI and XhoI sites and is fused to an epitope HA/FLAG at the 5` end of the receptor sequence, it lacks a stop codon. The FP sequence is cloned between KpnI and NotI sites. The linker sequence encodes a peptide consisting of 27 amino acids in length that is a part of the polylinker of the mammalian vector.
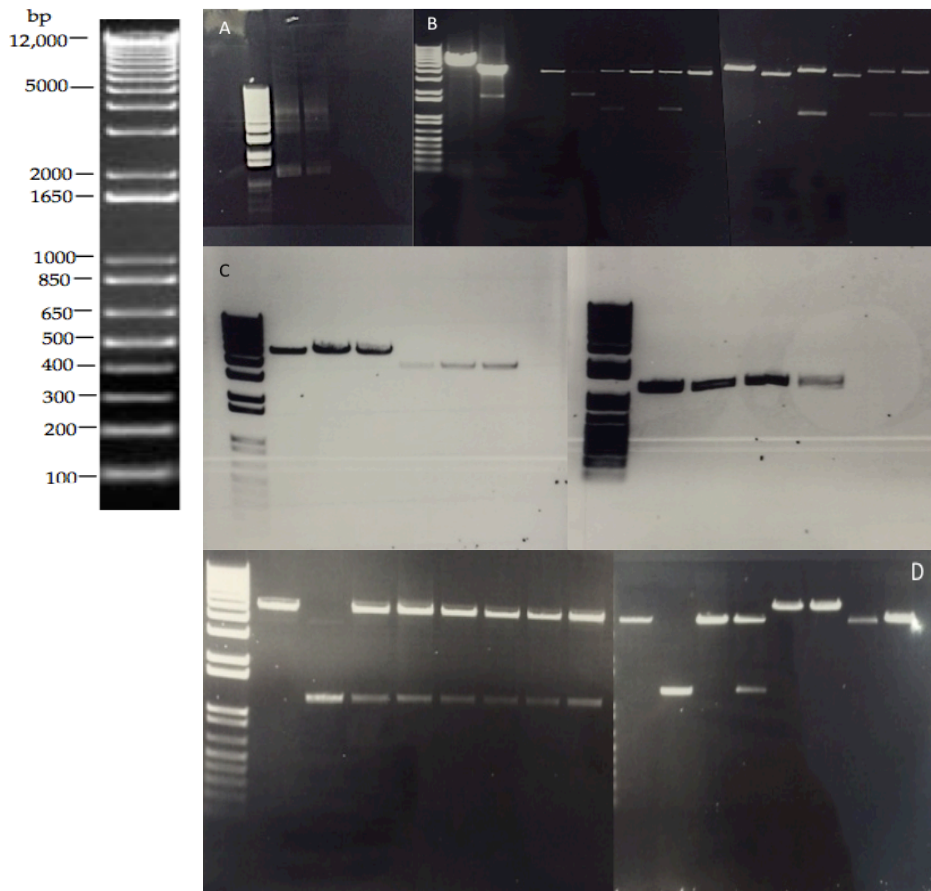
Figure 3.2 Agarose gel analysis of the cDNA constructs building steps to prepared for the BiFC assay

The size of the DNA Ladder is on the left. A) The amplified $A_{2A}R$ cDNA fragments prepared for the insertion into the TOPO10F' vector. From left-to-right: DNA Ladder; HA-$A_{2A}R$ cDNA; FLAG-$A_{2A}R$ cDNA. B) The $A_{2A}R$ cDNA fragments were excised from the TOPO TOP10F' vector for insertion into pBiFC vectors (EcoRI and SalI/XhoI). From left-to-right: DNA Ladder; pBiFC-VN173 vector; pBiFC-CC155 vector; blue control colony; HA-$A_{2A}R$; FLAG-$A_{2A}R$ fragments. C) Purified digested fragments, two different concentrations of the $A_{2A}R$ cDNA and pBiFC-vectors were placed. From left-to-right: DNA Ladder; pBiFC-VN173 vector; pBiFC-CC155 vector; HA-$A_{2A}R$; FLAG-$A_{2A}R$ fragments. D) The ligated $A_{2A}R$ cDNA into the pBiFC-vectors. From left-to-right: DNA Ladder; HA-$A_{2A}R$ with pBiFC-VN173 vector (ECFP 1-172) (EcoRI and SalI); FLAG-$A_{2A}R$ with pBiFC-CC155 vector (EYFP 155-239) (EcoRI and XhoI).

3.4 Analyses of A$_{2A}$-BiFC constructs

The generated BiFC constructs, the pBiFC-VN173-A$_{2A}$R plasmid and the pBiFC-CC155-A$_{2A}$R plasmid, were then transfected according to Lipofectamine® 2000 protocol instructions where 0.2 μg of constructs plasmid were transfected into 1321N1 cells, astrocytoma cell line that endogenously express muscarinic M$_3$ receptor, and incubated for 48 hours at 37°C, the physiological temperature of mammalian cells, to determine whether the receptor protein-BiFC was expressed and the fluorescent signal was detected to confirm the homodimerisation of the receptor. Two fusion protein controls, a positive control (pBiFC-bFos-VN173 with pBiFC-bJun-VC155) and a negative control (pBiFC-bFos-ΔzipVN173 with pBiFC-bJun-VC155), were also used to compare the number of cells that has detectable BiFC constructs with the proportion of cells that expresses the fusion proteins using fluorescence microscopy.

No fluorescence signal was detected in the constructed protein complexes and in positive control only ~5-10 cells. Troubleshooting of the method was addressed in two ways: 1) by increasing the concentration of the plasmid DNA vectors to 0.4 μg, 0.6 μg, 0.8 μg, 1.2 μg and 2.4 μg; 2) by incubating for 24 hours at different temperatures 4°C, 25°C and 30°C. No fluorescence signal was detected in any of these different conditions, which may be due to the fact that the BiFC analysis does not support interaction between the proteins under investigation interact in the cell line used. This was addressed by changing the type of cell line from 1321N1 to HEK293T cells. Unfortunately, no fluorescence was detected in HEK293T cells.

Although no fluorescence was detected, we cannot rule out that there is an interaction between the A_{2A} receptor as the interaction may still exist between the A_{2A} receptor because the creation of the fusion protein may alter the structure or interaction face of the target protein or the fluorescence fragments may be physically unable to associate. Although the A_{2A} receptor has been documented to form homodimers and higher ordered oligomers with the BiFC technique[87, 177], they did not describe the intensity of the signal nor the number of cells it was seen in. They have used different types of cells (CAD) and/or another techniques for transfection that had a higher cDNA concentration (10 μg).

3.5 LANCE Ultra assay development and optimization with the A_{2A}-BiFC constructs

Several steps needed to be established before the agonist and antagonist of A_{2A} may be tested because the required cell number and stimulation conditions can vary for each cell line and for each receptor tested. The first step of the assay development consisted of running a cAMP standard curve to determine the assay sensitivity and dynamic range of the cAMP kit. This was established by preparing serial dilutions of 50 μM standard cAMP in half log units from 1 μM to 10 pM (log values: -11 to -6). The assay shows an IC_{50} value for cAMP of 1 nM (see Figure 3.3). The next step was to identify an optimal cell density by performing a forskolin and cell density cross-titration experiment; this was achieved by forskolin dilutions from 10 mM stock solution in half log units from 40 μM to 120 nM (log values: -8 to -3). The cAMP experiments were conducted in parallel with the BiFC work and were discontinued when it became

clear that troubleshooting the BiFC was problematic. It was deemed unwise to continue the Lance studies, which were being performed using the BiFC constructs, until such time as the BiFC experiments were able to be successfully performed.
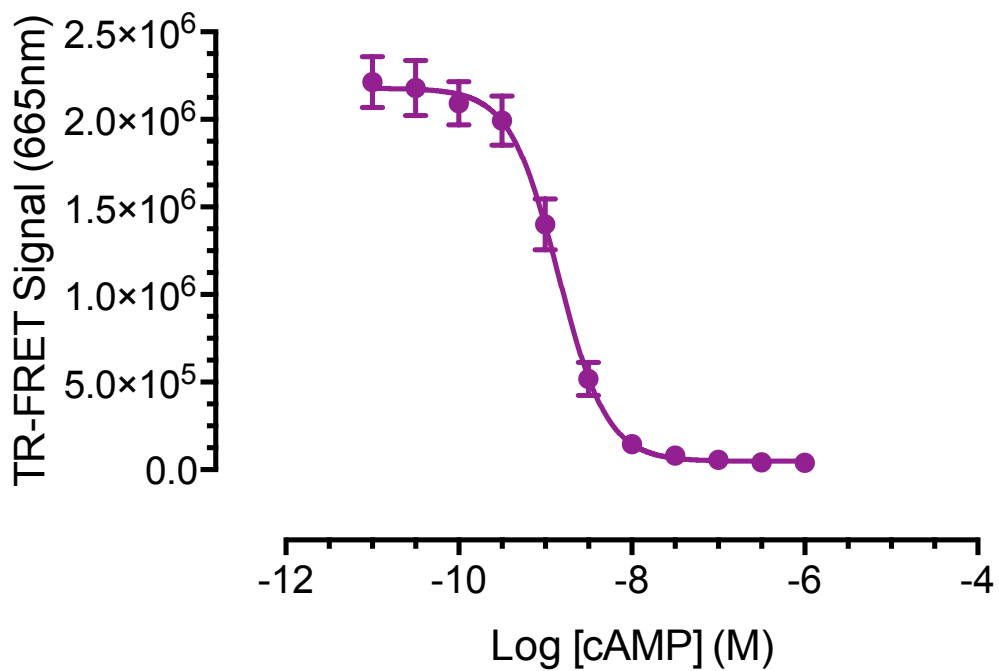
Figure 3.3 cAMP standard curve

Data are presented as mean ± s.e.m of triplicate measures and are representative of three independent experiments. Concentration-response curves were analyzed by fitting data to the four-parameter logistic equation using GraphPad Prism®.

3.6 Computational analyses of the generated TM5-TM5 ensembles

The following work aims to develop an ensemble-based coarse-grained molecular dynamics (eCG-MD) methodology for the analysis of interactions between two GPCR TM helices (see Figure 3.4). To do this, it is necessary investigate the computational parameters required to obtain converged results computationally and to identify whether these results match the experimentally-obtained data for the self-association of the TM5 helices of the A_{2A} adenosine receptor.

A

B

C
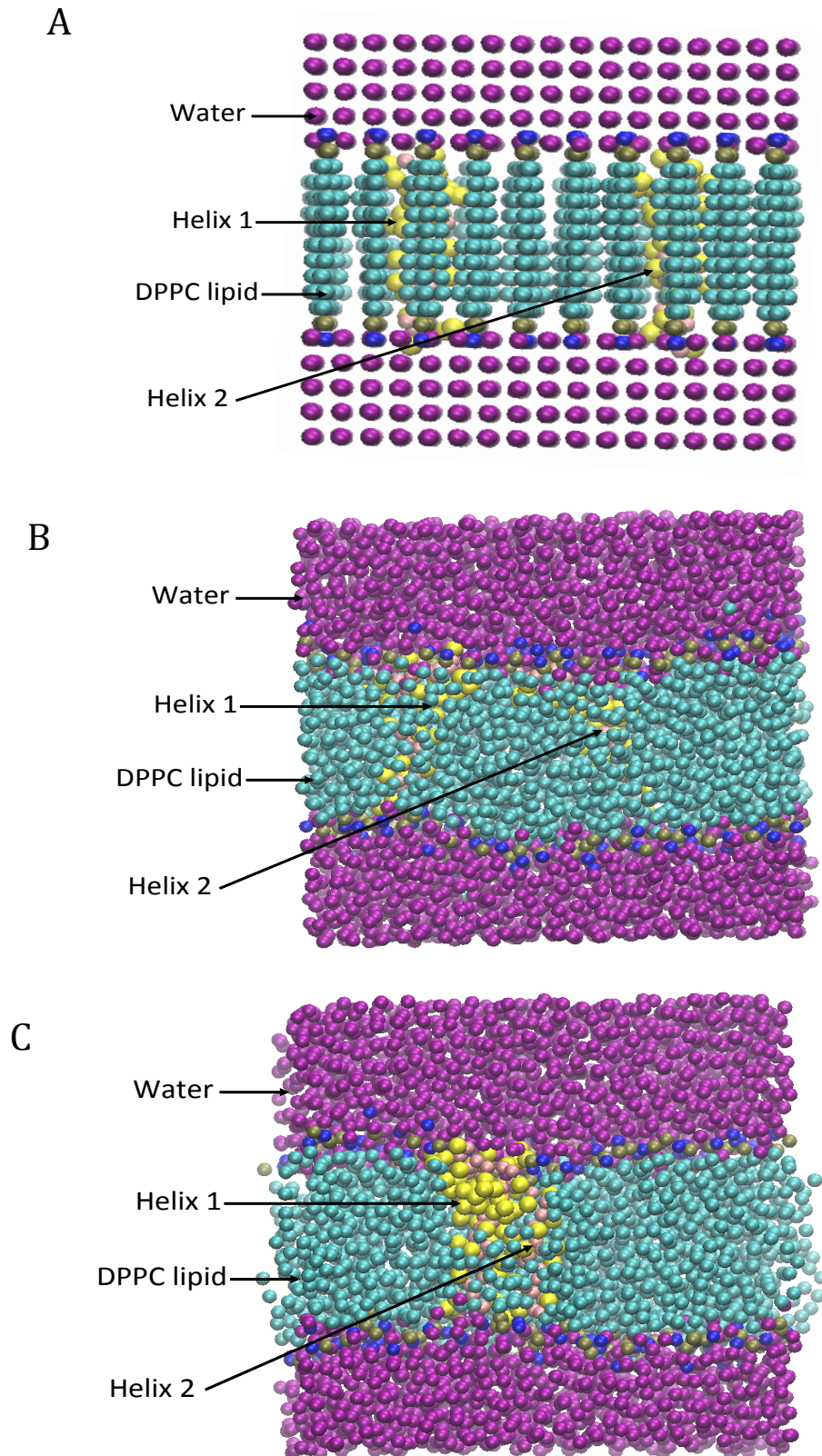
Figure 3.4 A screenshot representation of eCG-MD simulation

(a) At the start of the simulation, (b) at the middle of the simulation and (c) at an end of a simulation.

93

3.6.1 Ensembles interacting Interfaces

Seven simulation sets were tested that included: 1) both wild-type TM5 and TM2, 2) five types of mutated TM5 (M177$^{5.38}$A, M193$^{5.54}$A, M193$^{5.54}$I, Y197$^{5.58}$A and Y197$^{5.58}$F). First, the final mean distance between the two helices in an ensemble of 50 replicas in each set was used to identify the specific interactions between the A$_{2A}$ homodimers. Following the application of the 10Å cut-off, 26% of the ensemble formed stable dimers in the wild-type TM5-TM5 simulations. In the mutated TM5-M177$^{5.38}$A and in the TM5-M193$^{5.54}$I simulation sets, 16% of the ensemble formed stable dimers while in TM5-M193$^{5.54}$A, dimers were detected in 28% of the ensemble. In the mutated TM5-Y197$^{5.58}$A and in the TM5-Y197$^{5.58}$F, dimers were detected in 16% and 18% of the ensembles respectively. For all six of the TM5 simulation sets, the detected interactions took place at the same position within the helices, indicating that a defined orientation is needed to establish a specific interaction. In the negative control (the TM2-TM2 simulation set), 24% of the ensemble resulted in the formation of stable dimers but there were no specific interactions identified between residues. For all simulations, of those pairwise combinations in which dimerisation was identified after the cut-off 10Å had their trajectories combined and the results with heat maps of interactions observed at 12Å and 15Å compared (see Figure 3.5). The location of the contact interface was then mapped by comparison with the crystal structure of A$_{2A}$R (3EML) on which the helices were modelled.
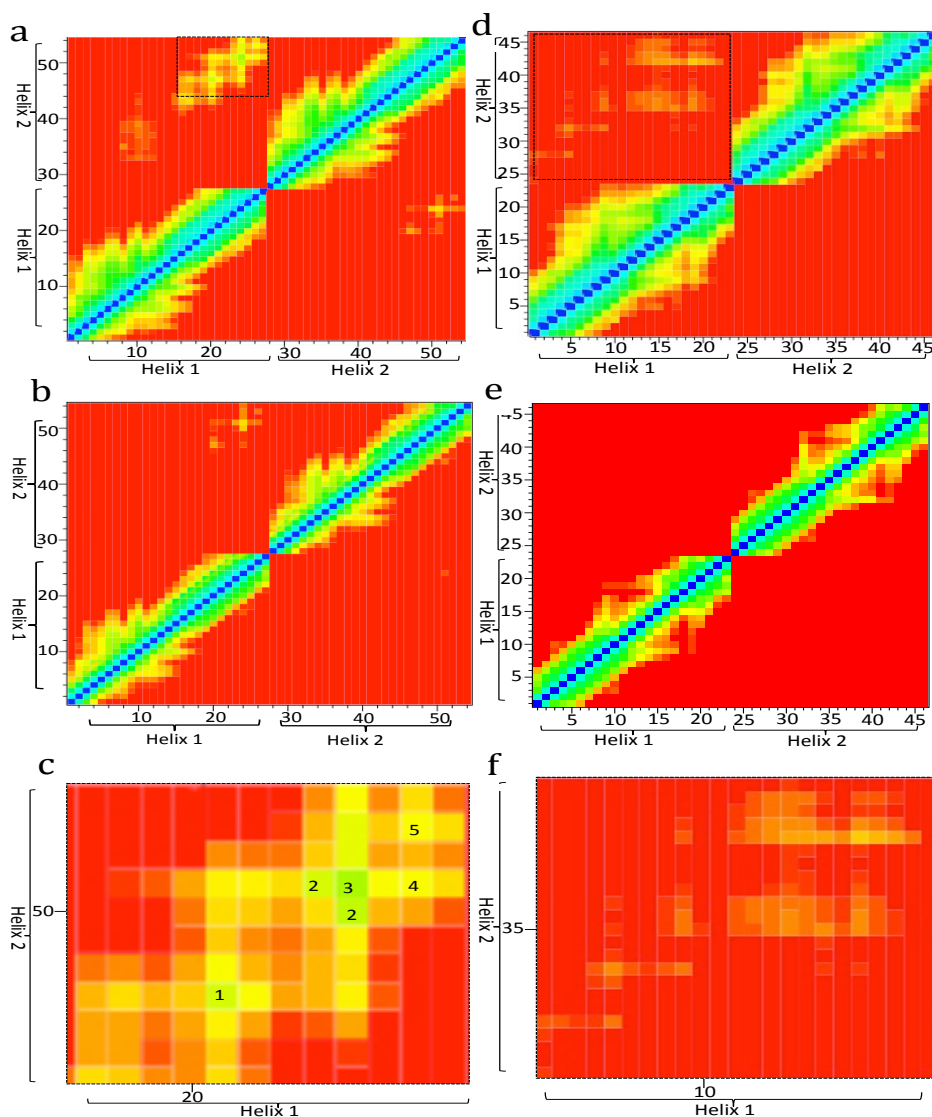
Figure 3.5 Contact matrices (heat maps) showing specific interactions between residues, as measured by distance, between two A$_{2A}$ helices ("Helix 1" and "Helix 2")

For the wild-type TM5-TM5 simulation (a-c) and for the TM2-TM2 negative control (d-f). Results shown are the average for each ensemble. Interhelical distances at the 15Å cutoff are shown in the top left quarter of panels (a) and (d). Interhelical distances at the 12Å cutoff are shown in the top left corner of (b) and (e) and in the lower right quarter of panels (a) and (d). Interhelical distances at the 10Å cutoff are shown in the lower right quarter of panels (b) and (e). The region shown in the black rectangle in (a) and (d) is magnified in (c) and (f), respectively. The five numbered interactions shown in (c) are identified in Table 3. The colour scale indicates distance between helices: blue corresponds to 0Å (superposition of the two helical backbones at all cutoffs); green corresponds to 5Å (10Å cutoff), 6Å (12Å cutoff), 7.5Å (15Å cutoff); yellow corresponds to 7Å (10Å cutoff), 8Å (12Å cutoff), 12Å (15Å cutoff); red corresponds to the cutoff distances applied (10Å, 12Å or 15Å).

3.6.2 Identification of contact interface for the wild-type TM5-TM5 homodimer

After identifying the interacting replicas their trajectories were combined in the TM5-TM5 ensemble, the contact interface were identified (the contacting residues) by measuring the average interhelical contact distance between the two wild-type TM5-TM5 helices (see Figure 3.5a-c) or between the negative control TM2-TM2 helices (see Figure 3.5d-f) was determined and results plotted at two different distances, 12Å and 15Å as heat maps of interactions. The proximity of the wild-type helices is best visualised at 15Å (see Figure 3.5a and c). The interacting residues in the wild-type TM5-TM5 simulation are found in the bottom third of the C-terminal end of TM5. From the averaged interhelix contact matrices, the specific interactions were found to be within the experimentally identified $M193^{5.54}xxVY197^{5.58}$ motif at an interhelical distance of ~8-9Å. The methionine at position $193^{5.54}$ of helix 1 interacts with the methionine at the same position on helix 2, reinforcing the suggestion[64] of its importance in the formation of the TM5 homodimer. From Figure 3.5d, it can be seen that the distance between TM2-TM2 is close enough to form potential specific interactions, however, none were detected in the combined trajectories for this negative control. Results obtained at the 15Å cutoff (see Figure 3.5f) were random and non-specific, supporting the selection of a minimum cutoff distance of 12Å. The frequency of specific interactions identified in the wild-type TM5-TM5 ensemble was determined by calculating the mean distance for each frame of every replica individually. Table 3.1 shows that the five most prominently occurring interactions were between $M193^{5.54}$-$M193^{5.54}$, $V196^{5.57}$-$Y197^{5.58}$, $Y197^{5.58}$-$R199^{5.60}$, $R199^{5.60}$-$R199^{5.60}$ and $R199^{5.60}$-$I200^{5.61}$.

These findings are consistent with the experimental results[64] identifying that the interaction between two wild-type A_{2A} TM5 peptide sequences involved amino acid residue M193[5.54]. These findings are also consistent with experimental data showing the formation of A_{2A} receptor homodimers using bioluminescence resonance energy transfer (BRET)[86] and bimolecular fluorescence complementation (BiFC)[87]. The presence of specific interaction between TM2 helices was experimentally investigated and none were detected[64, 146]. The eCG-MD simulations produced the same results as experimentally obtained findings, with the formation of wild-type TM5-TM5 dimers involving the M193[5.54] residue and no specific interaction detected between TM2-TM2 helices *in silico*.

Figure label	Interacting residues	Replica number											Mean Distance ± Standard Deviation (in Å)
		1	5	10	15	20	25	30	35	40	45	50	
1	M193-M193	0	0	1	1	1	2	2	3	5	6	6	7.59 ± 2.89
2	V196-Y197	0	1	2	2	2	4	4	6	9	12	13	9.16 ± 2.5
3	Y197-Y197	0	1	1	1	1	1	2	4	6	7	8	9.11 ± 2.85
4	Y197-R199	0	1	1	1	1	1	1	3	5	6	6	9.83 ± 3.57
5	R199-R199	0	1	1	1	1	1	1	3	4	5	5	8.06 ± 2.99

Table 3.1 The number of interactions (hits) for specific interacting residues identified in the contact matrices for the wild-type TM5-TM5 simulation at the 10Å cut-off.

3.6.3 Identification of contact interface for the mutated TM5-TM5 helices

The identification of the presence of M193[5.54] in the wild-type contact interface suggested that this residue may play a significant role in how the two TM5 helices interact. To investigate this possibility, three sets of ensemble simulations were performed that included helices that included mutations at

M177$^{5.38}$ or at M193$^{5.54}$ (M177$^{5.38}$A, M193$^{5.54}$A, and M193$^{5.54}$I). Two types of point mutations were used: (1) the substitution of methionine to alanine; and, (2) the substitution of methionine to isoleucine. Investigation of the TM5 peptide sequence revealed that two separate PxxxM motifs existed within the same helix, with a methionine residue present at M177$^{5.38}$ as well the methionine residue identified at M193$^{5.54}$. Each of these methionine residues was mutated to alanine. M193$^{5.54}$ was also mutated to isoleucine because a conserved PxxxI motif is found in the related family of P2Y receptors at the same location as the originally-identified PxxxM motif in A$_{2A}$R.

Specifically interacting residues in the TM5-M177A simulation set were identical to those identified in wild-type TM5-TM5 dimers (see Figure 3.6a) and included the M193$^{5.54}$xxxVY197$^{5.58}$ motif. M177$^{5.38}$ was not directly involved in the dimerisation between the two helices in any simulation. The specific interactions observed in the TM5-M193I$^{5.54}$ simulation set were almost identical to those found in the wild-type but included I193$^{5.54}$ in the interaction, despite the loss of the methionine at position 193 (see Figure 3.6b). In contrast, the TM5-M193A$^{5.54}$ mutation completely changed the contact interface of the helices (see Figure 3.6c) and the key interacting residues were identified at a similar distance, but contained within a newly identified V196$^{5.57}$YxR199$^{5.60}$ motif. This provides a molecular explanation for the finding that mutation of the full-length A$_{2A}$R at position M193$^{5.54}$ noticeably alters the monomer:dimer ratio, as observed with SDS-PAGE[64]. Mutation of M193A$^{5.54}$ causes a change in the way in which the two helices come together that prevents formation of TM5 homodimers,

emphasizing the importance of the M193$^{5.54}$ residue in the specificity of TM5-TM5 dimer formation *in vivo*.

The identification of the Y197$^{5.58}$ in the wild-type contact interface as the residue with most hits (highest number of occurrence) suggested that this residue may play a significant role in how the two TM5 helices interact. To investigate this possibility, two sets of ensemble simulations were performed with helices mutated at Y197$^{5.58}$ (Y197$^{5.58}$A, and Y197$^{5.58}$F). Two types of point mutations were used: (1) the substitution of tyrosine to alanine; and, (2) the substitution of tyrosine to phenlalanine, phenylalanine is identical to tyrosine, with the exception of the absence of the phosphorylatable hydroxyl group. Thus, a phenylalanine mutation is often used to make the residue non-phosphorylated. Alanine, on the other hand, lacks the aromatic ring present in tyrosine and phenylalanine, so a mutation to alanine effectively removes the tyrosine side chain.

The interacting residues in the both ensemble sets TM5-Y197A and TM5-Y197F simulation set were identical to those identified in wild-type TM5-TM5 dimers (Figure 3.7) and included the M193$^{5.54}$xxxVYxR199$^{5.60}$ motif but with more residues at longer distances as seen in the TM5-M193A mutation. Mutation of Y197$^{5.58}$A/F does not causes a change in the way in which the two helices come together and does not effect the formation of TM5 homodimers, emphasizing that the Y197$^{5.58}$ may not be a key residue in the specificity of TM5-TM5 dimer formation *in vivo* but rather only a part of the interacting motif in the wild-type homodimer.

Figure 3.6 Contact matrices (heat maps) showing specific interactions between two mutated A$_{2A}$ TM5 helices ("Helix 1" and "Helix 2") with the following residues mutated: M177A (a), M193A (b) and M193I (c)

Results shown are the average for each ensemble. Interhelical distances at the 15Å and 12Å cutoff distances are shown in the top left quarter and in the lower right quarter of panels (a-c), respectively. The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: (1) M193$^{5.54}$ with M193$^{5.54}$; (2, 3) V196$^{5.57}$ with Y197$^{5.58}$ and Y197$^{5.58}$ with Y197$^{5.58}$; (4) Y197$^{5.58}$ with I200$^{5.61}$ and R199$^{5.60}$ with R199$^{5.60}$; (5) L192$^{5.53}$ with I193$^{5.54}$, V196$^{5.57}$ with Y197$^{5.58}$ and Y197$^{5.58}$ with R199$^{5.60}$, (6) Y197$^{5.58}$ with I200$^{5.61}$ and Y197$^{5.58}$ with R199$^{5.60}$; and (7) R199$^{5.60}$ with R199$^{5.60}$.
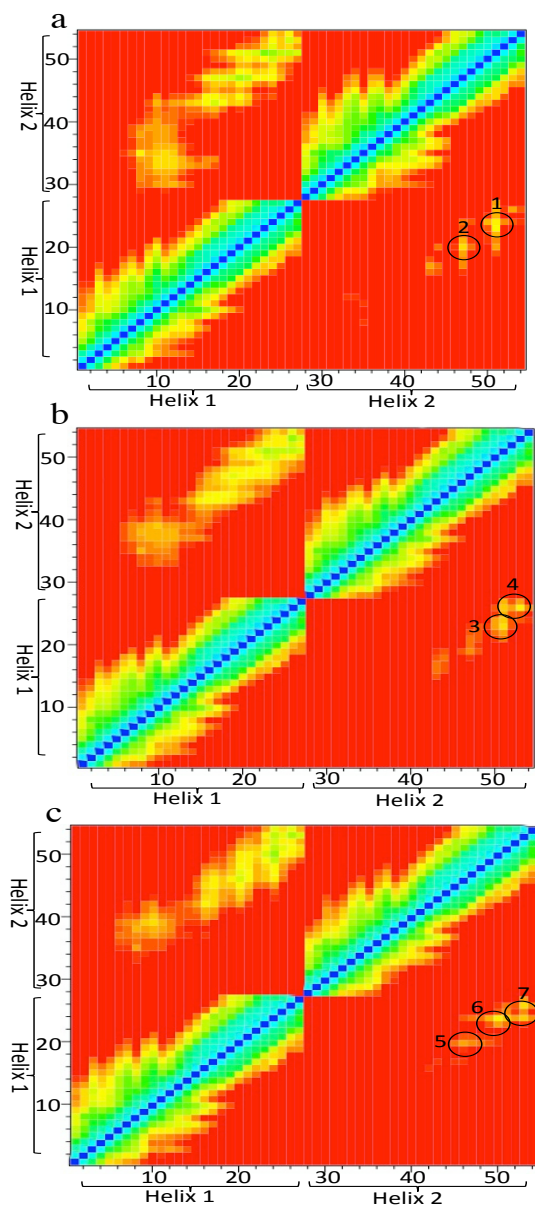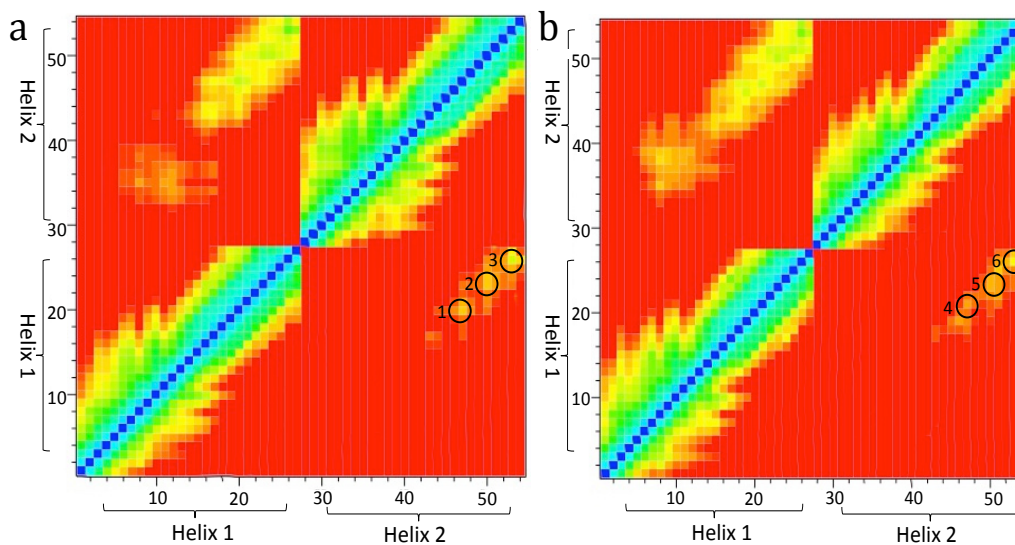
Figure 3.7 Contact matrices (heat maps) showing specific interactions between two mutated A$_{2A}$ TM5 helices ("Helix 1" and "Helix 2") with the following residues mutated: Y197A (a) and Y197F (b)

Results shown are the average for each ensemble. Interhelical distances at the 15Å and 12Å cutoff distances are shown in the top left quarter and in the lower right quarter of panels (a-c), respectively. The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: (1) M193$^{5.54}$ with M193$^{5.54}$; (2) V196$^{5.57}$ and A197$^{5.58}$ with A197$^{5.58}$ and A197$^{5.58;}$ with I200$^{5.61}$ and R199$^{5.60}$ with R199$^{5.60}$; (3) R199$^{5.60}$ with R199$^{5.60}$; (4) M193$^{5.54}$ with M193$^{5.54}$; (5) V196$^{5.57}$ and F197$^{5.58}$ with F197$^{5.58}$ and F197$^{5.58;}$ with I200$^{5.61}$ and R199$^{5.60}$ with R199$^{5.60}$; (6) R199$^{5.60}$ with R199$^{5.60}$.

3.7 Statistical analyses of the computationally generated ensemble sets

This portion of the work aims to investigate the computational parameters required to obtain converged results using error analysis to determine the ensemble size and individual replica simulation run-time required.

The five different ensembles, one for each A_{2A} receptor simulation set, were run independently in CG-MD simulations for the total run time of 500 ns. These data, which included both wild-type and mutated helix sequences, were used to investigate whether variations in the optimal replica number required would occur between different simulation sets. This information was used to identify the minimum replica number required to achieve convergence for any given simulation set. At the start of the simulation, the behaviour of the TMs was monitored and the TM helices were observed to diffuse freely in the DPPC lipid bilayer. The kernel density estimation was used to estimate the probability of the mean distance between the two wild-type TM5 helices at $t = 0$ and at increments of 100 ns up to completion of the simulation at 500 ns across the 50 replica ensemble is shown in Figure 3.8. At $t = 0$, the two helices are at their starting positions 40Å (4.0 nm) apart. At $t = 500$ ns, the mean distance between the helices has adopted a normal distribution with a mean distance of ~16Å between them. The intermediate time points show the redistribution of the distance from the starting point at $t = 0$ to the final mean distance between the helices at 500 ns. A graphical representation of the number and time of interactions observed in each replica for the wild-type TM5-TM5 simulation within the ensemble of 50 replicas is shown in Figure 3.9; these were produced by a locally written code.

102

Four of the 50 replicas showed no contact between the helices, which gives rise to the small peak at 40Å at the end of the replica run (t = 500 ns) in Figure 3.7. Three of the replicas began to show contact towards the end of the run; this corresponds to the smaller peak seen at 30Å at t = 500 ns in Figure 3.7.
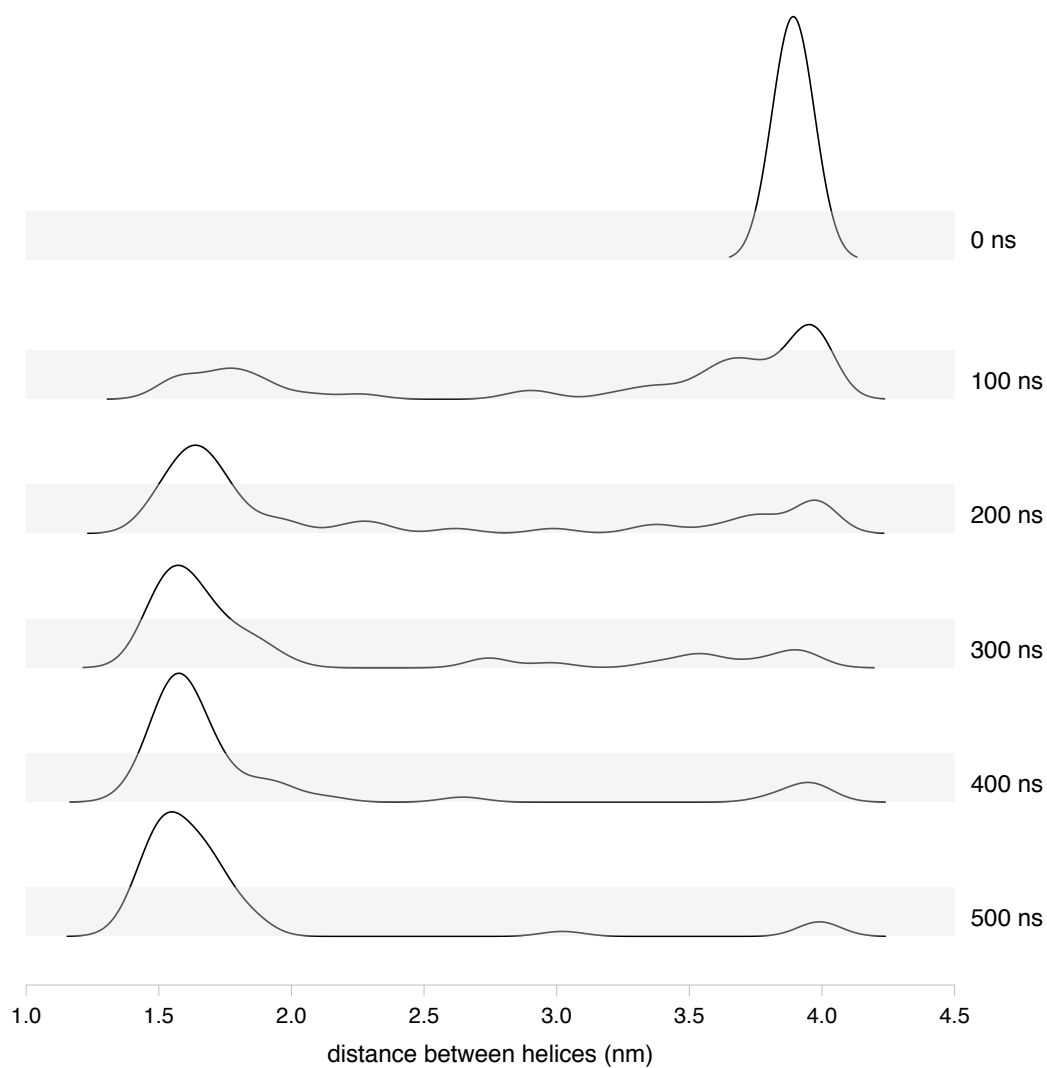
Figure 3.8 Distribution of the mean distance between the two TM5-TM5 wild type helices at 0, 100, 200, 300, 400 and 500 ns in all 50 replicas.
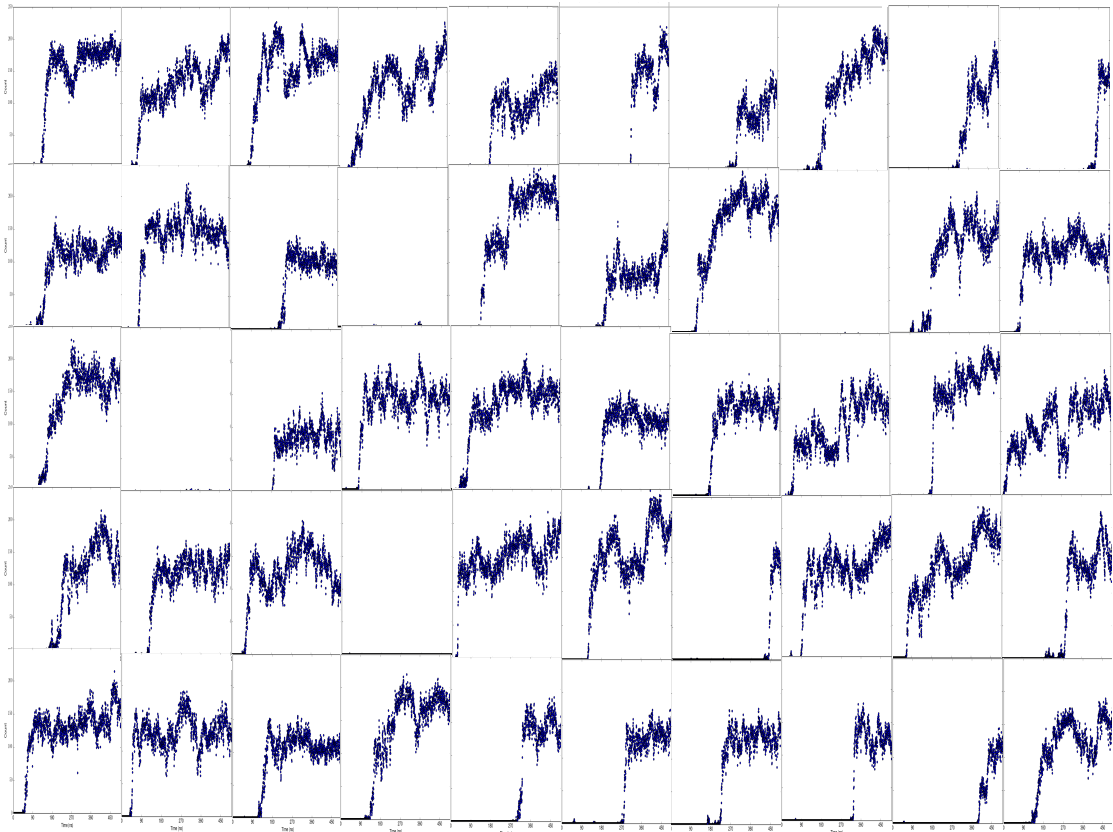
Figure 3.9 The number and timing of pairwise interactions for each of the 50 replicas within the wild-type TM5-TM5 dimer ensemble are shown
The x and y axes are linear and represent run length from 0 to 500 ns and the number of interaction events from 0 to 250 counts, respectively.

3.7.1 The optimal replica number required

The optimal replica number required for simulations was identified by plotting the mean distance between the two helices of each of the ensemble sets against the replica number (see Figure 3.10a). This revealed that there is no statistically significant difference in the mean distance as a function of replica number. However, a decrease in the error of the mean is observed with increasing ensemble size (see Figure 3.10b) and it can be seen that the rate of decrease in the error slows after approximately 15 replicas have been included in the ensemble. For each of the five sets, larger ensembles provide less variation in the error of the mean and an ensemble of 30 replicas represents a good compromise between computational effort and minimisation of the error in the mean distance calculated. From Figure 3.10, it can be seen that an ensemble of 30 replicas is sufficient to achieve convergence.

Figure 3.10 Variation in (a) the mean distance between TM helices and (b) the error (standard deviation) is shown as a function of the number of replicas performed
For the following simulation sets: (•) wild-type TM5-TM5 helices**,** (■) M177A-mutated TM5-TM5 helices, (▲) M193A-mutated TM5-TM5 helices, (▼) M193I-mutated TM5-TM5 helices, and (♦) wild-type TM2-TM2 helices.

3.7.2 The minimum run time length required

The minimum run time length required for each replica in an ensemble was estimated by calculating the mean distance and the standard deviation of the mean within the 50 replicas for simulations of varying duration (0, 100, 200, 300, 400 and 500 ns). The standard deviation of the mean provides an estimate of the error associated with the finding of an interaction between two helices.

Figure 3.11a shows a significant effect of run length on both mean distance and standard deviation and reflecting the time required for interactions to take place. For four of the five simulation sets, the standard deviation increases as a function of time, with the rate of increase slowing as the run length becomes longer. In contrast, no change in the standard deviation over time is seen in the TM2-TM2 set (Figure 3.11b). The low standard deviation between the two TM2 helices indicates that the mean distance between the helices was about the same throughout the simulation. Interestingly, the TM2-TM2 mean distance was close enough to form potential specific interactions, however, none were detected in the combined trajectories for this negative control, this finding agrees with the experimental findings where TM2 homodimers could not be detected[146]. The absence of an increase in error in the mean distance as a function of time may serve as an indicator of an absence of interaction between two helices within an ensemble. From the figure, it can be concluded that an ensemble run for a simulation time of 300 ns is sufficient to achieve convergence.
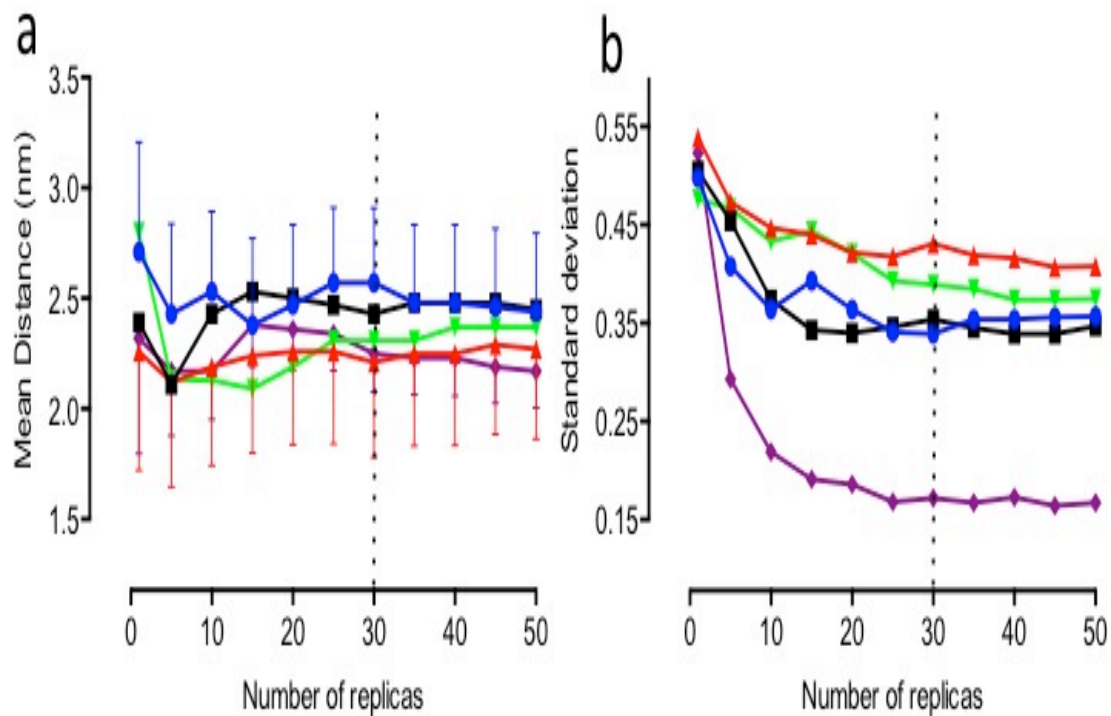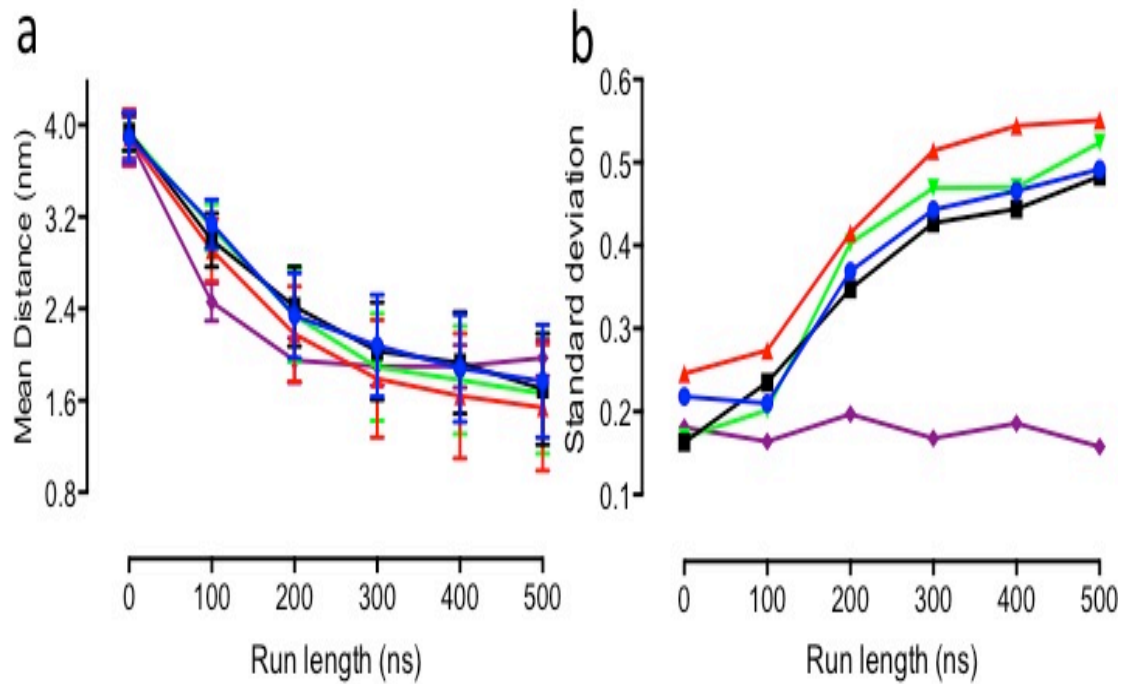
Figure 3.11 Variation in (a) the mean distance between TM helices and (b) the error (standard deviation) is shown as a function of the run length
For the following simulation sets: (•) wild-type TM5-TM5 helices, (■) M177A-mutated TM5-TM5 helices, (▲) M193A-mutated TM5-TM5 helices, (▼) M193I-mutated TM5-TM5 helices, and (♦) wild-type TM2-TM2 helices.

3.8 Summary

From the computational eCG-MD results, specific interactions involving the PxxxM motif of TM5 were identified, specifically, at the M193$^{5.54}$ residue within that motif. The eCG-MD method, therefore, accurately identified residues shown experimentally to be involved in TM5 homodimerisation. The M193$^{5.54}$ was found to interact with the other M193$^{5.54}$ on the other helix. The role of M193$^{5.54}$ was further investigated by characterizing the effect of mutation on the methionine, using the M193$^{5.54}$A mutation. From this, it was clear that the contact interface of the helices was completely changed and that the key interacting residues identified in the wild-type conformation had moved to a new position, preventing formation of TM5 homodimers. The results provided a molecular explanation for the experimental finding that the M193$^{5.54}$A mutation alters the monomer:dimer ratio at a level of detail that could not be determined biophysically and would require structural biology studies to confirm experimentally. M177$^{5.38}$ was not involved in the interaction between the two helices and by mutating it the contact interface was not affected. The M193I$^{5.54}$ mutation had no effect on the residues involved in contact interface; they were identical to those found in the wild-type but, additionally, included I193$^{5.54}$ in the interaction, despite the loss of the methionine at position 193. Mutation of Y197$^{5.58}$A/F had no effect on the way in which the two helices came together, emphasizing that the Y197$^{5.58}$ may not be a key residue in the specificity of TM5-TM5 dimer formation but rather, may only a part of the interacting motif in the wild-type homodimer.

These results clearly demonstrate that sufficient conformational sampling is required in coarse-grained MD to obtain reproducible and reliable results. In the eCG-MD simulations, it was noted that several of the replicas within the ensemble failed to show any interactions and that a number of others began to interact late in the simulation at a point when accurate estimates of distance could no longer be achieved. A single trajectory simulation, particularly if either of these circumstances were to occur, would give inaccurate and potentially misleading results. Error analysis was used to determine appropriate choices for ensemble size and run length. For ensemble size, the rate of change in the standard deviation of the mean distance between helices decreased with increasing replica size and found that approximately 30 replicas was sufficient per ensemble to obtain reproducible results. For run length, the rate of increase in the standard deviation of the mean distance between helices increased with increasing run length, but that the rate of increase slowed substantially after approximately 300 ns. Interestingly, the negative control (TM2-TM2) included in the eCG-MD simulations showed no variation in the standard deviation of the mean distance between helices as a function of run length and a low standard deviation with a very rapid decrease to a constant value at an ensemble size of ~15 replicas. This behavior was notably different from simulations in which interactions were identified and provides a means of confirming the absence of interaction.

Chapter 4
Computational comparison with experimental structural data

4.1 Introduction

In order to assess the validity of the computational method, it is necessary to compare the results with experimental values that exist. The computational results obtained for the A_{2A} receptor homodimers (see Chapter 3) closely match the experimental biophysical data and provide information about the nature of the contact interface between the two helices that cannot be determined experimentally. This chapter aims to determine whether the eCG-MD method developed in Chapter 3 gives findings in agreement with experimentally-obtained structural data. To do this, CG structures will be converted to atomistic structures and compared with existing structural data. Although no structural data exist yet for A_{2A} dimers, this approach may allow hypotheses to be drawn about the molecular nature and possible role of the interactions between the dimeric TM5-TM5 helices studied in Chapter 3.

Dimerisation in Class A GPCRs involves the transmembrane domains, as opposed to Class C GPCRs, where dimerisation is mediated by the large N terminal domain of the protein[178]. Three Class A GPCR dimeric crystal structures were identified that fulfilled the following criteria: 1) the

crystallographic asymmetric unit is a dimer; 2) the software-determined (PISA) quaternary structure is a dimer; and, 3) the dimeric quaternary structure has been confirmed functionally. Three class A GPCRs fit this criteria and have experimentally-obtained structural data. These were chosen evaluate the method developed, these include: rhodopsin, the CXCR4 chemokine receptor and the β_1 adrenergic receptor (β_1AR), their corresponding TM helices (listed in Table 2.3) were constructed as described in section-2.2.3 of materials and methods and used in ensemble-based simulations.

4.2 Computational identification of interacting interfaces: rhodopsin

Rhodopsin has been shown to exist in a native oligomeric form[29] and an atomic model of the rhodopsin dimer has been proposed as a working model for G protein-coupled receptors[28]. This model, 1N3M, is the only GPCR dimer for which structural data exist. Three contact points between the rhodopsin monomers have been reported. The first is considered to be the strongest, with the largest contact area (578Å^2) and is located between TM4 and TM5. The second exhibits a contact area of 333Å^2 and is located between TM1 and TM2. The third contact point is considered the weakest interaction and is found between rows of dimers at the extracellular ends of TM1 with a contact area of 146Å^2[28, 44].

Two heterologous simulations sets were run to identify whether contact interfaces could be identified for either: rhodopsin helices TM1 and TM2 and

rhodopsin helices TM4 and TM5. For each set, the same method used to analyse the $A_{2A}R$ simulation sets was applied to investigate and analyse the helix-helix dimerisation in the simulations of rhodopsin First, the final mean distance between the two helices in an ensemble of 50 replicas in each set was used to identify the specific interactions between the rhodopsin homodimers. Following the application of the 10Å cut-off, 14% of the ensemble formed stable dimers in the TM1-TM2 simulations. In the TM4-TM5 simulation set 32% of the ensemble formed stable dimers, confirming that our computational method is able to produce results in agreement with structural data. In each set, the mean distance between helices was ~7.6-8Å (see Table 4.1). The mean distance between specific interacting residues in the TM1-TM2 simulation (see Figure 4.1a) is further apart than the mean distance between specific interacting residues in the TM4-TM5 simulation (see Figure 4.1b).

Receptor	Helices	Interacting residues	Crystal structure distance (Å)	Mean Distance (Å) ± Standard deviation
Rhodopsin	TM1-TM2	F127$^{1.47}$ - L218$^{2.44}$	15.28	14.2 ± 4.07
		M119$^{1.39}$ - D224$^{2.50}$	9.01	10.32 ± 3.1
	TM4-TM5	F418$^{4.48}$ - L521$^{5.51}$	12.07	9.18 ± 3.24
		T419$^{4.49}$ - L521$^{5.51}$	14.77	8.13 ± 1.96
		G415$^{4.45}$ - F525$^{5.55}$	16.44	7.5 ± 2.3
		H411$^{4.41}$ - Q530$^{5.60}$	17.64	9.06 ± 3.34
CXCR4	TM5-TM5	F201$^{5.40}$ - V198$^{5.37}$	7.37	15.55 ± 3.34
		F201$^{5.40}$ - Q200$^{5.39}$	11.03	13.7 ± 1.46
		F201$^{5.40}$ - F201$^{5.40}$	7.91	13.6 ± 2.89
		F201$^{5.40}$ - Q202$^{5.41}$	8.72	14.93 ± 3.13
		F201$^{5.40}$ - I204$^{5.43}$	12.14	13.11 ± 3.37
		F201$^{5.40}$ - M205$^{5.44}$	10.6	12.6 ± 4.96
β$_1$AR	TM1-TM1	W40$^{1.31}$ - A42$^{1.33}$	12.21	15.28[*]
		W40$^{1.31}$ - S45$^{1.36}$	12.41	17.5[*]
		W40$^{1.31}$ - L46$^{1.37}$	13.84	18.3[*]
		M44$^{1.35}$ - L46$^{1.37}$	9.8	13.46[*]
		A49$^{1.39}$ - M48$^{1.38}$	8.86	5.39[*]
		L53$^{1.44}$ - M48$^{1.38}$	12.19	5.01[*]
		L53$^{1.44}$ - V51$^{1.40}$	10.75	4.9[*]
		L53$^{1.44}$ - V52$^{1.41}$	11.13	5.2[*]
		L54$^{1.45}$ - V51$^{1.40}$	13.9	5.09[*]
	TM4-TM5	L159$^{4.43}$ - Y231$^{5.58}$	ND[**]	9.01 ± 2.22
		W166$^{4.50}$ - Y227$^{5.62}$	ND[**]	7.9 ± 1.99

Table 4.1 The comparison of distance of the identified interacting residues[†] from contact matrix graphs.

The comparison of distance of the identified interacting residues[†] from contact matrix graphs of the rhodopsin, CXCR4 and β$_1$AR helices and the crystal rhodopsin dimer (1N3M), CXCR4 dimer (4GPO) and the β$_1$AR dimer (3ODU).

[†] Distances are measured from backbone to backbone.

[*] Interactions were detected in only one replica in the ensemble.

[**] Not Determined (ND): The distances between TM4 and TM5 could not be measured due to the orientation of the dimer in the 4GPO crystal structure, which is submitted showing the TM1-TM2 dimer interface.

4.2.1 Identification of the contact interface for the rhodopsin homodimer

Following screening, extending the cutoff to 15Å in the averaged interacting simulation revealed an increase in the number of interactions observed and presented the strongest interacting residues. The most prominent interactions within the 12Å cutoff with about ~8-9Å interhelical distances present were between the residues was found in the Y118$^{1.40}$MFxLIxxxF127$^{1.47}$ motif in TM1 with L218$^{2.44}$NLxxxDL225$^{2.51}$ motif in TM2 (see Figure 4.1a). These motifs are found in the middle of the TM1 and TM2 helices. The most prominent interactions within the 12Å cutoff with about ~7.6-8Å interhelical distances present were between the residues was found in the H411$^{4.41}$xxMGVxFT419$^{4.49}$ motif in TM4 with residues L521$^{5.51}$xxxFxxYGQ530$^{5.60}$ motif in TM5 (see Figure 4.2b). TM4 motif is found in the upper third of the TM at cytoplasmic side and the TM5 motif is found in the bottom third of the N-terminal end of TM5.

These findings are consistent with the first physical results of rhodopsin receptor dimerisation that emerged from atomic-force microscopy[29], that structure showed that the contact area between TM4-TM5 is the strongest and largest contact area, in our results it showed the shortest distance between TM helices.

Figure 4.1 Contact matrices (heat maps) between two rhosopsin helices

Contact matrices (heat maps) between two rhosopsin helices showing specific interactions between TM1 (Helix 1) and TM2 (Helix 2) (a) and between TM4 (Helix 1) and TM5 (Helix 2) (b). Results shown are the average for each ensemble. The color scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: 1) $F127^{1.47}$ with $L218^{2.44}$; 2) $L122^{1.42}$ with $L220^{2.46}$, $I123^{1.43}$ with $L218^{2.44}$ and $N219^{2.45}$ with $L220^{2.46}$; 3) $Y118^{1.38}$ with $D224^{2.50}$, $M119^{1.39}$ with $D224^{2.50}$, $F120^{1.40}$ with $D224^{2.50}$ and $F120^{1.40}$ with $L225^{2.51}$; 4) $F418^{4.48}$ with $L521^{5.51}$ and $T419^{4.49}$ with $L521^{5.51}$; 5) $F418^{4.48}$ with $F525^{5.55}$; 6) $M414^{4.44}$ with $F525^{5.55}$, $G415^{4.45}$ with $F525^{5.55}$ and $V416^{4.46}$ with $F525^{5.55}$; 7) $H411^{4.41}$ with $Y528^{5.58}$, $H411^{4.41}$ with $G529^{5.59}$ and $H411^{4.41}$ with $Q530^{5.60}$.

4.3 Computational identification of interacting interfaces:

CXCR4

The crystal structure of the CXCR4 chemokine receptor (3ODU) bound to an antagonist small molecule IT1t has been reported and reveals a homodimer with an interface involving TM helices 5 and 6. They report that this dimer is mainly held by hydrophobic interactions involving residues: L194$^{5.33}$, V197$^{5.36}$, V198$^{5.37}$, F201$^{5.40}$, M205$^{5.44}$ and L210$^{5.49}$[150]. The first step was to investigate the interactions between TM helices in the CXCR4 receptor and to identify the formation of stable dimers with specific interactions (see Figure 4.2). A heterologous simulation between TM5 and TM6 was performed and there were no interactions identified (see Figure 4.2a) this can be explained by the fact that CXCR4 is able to form homodimers in the absence of ligand[179] that are unable to be dissociated by a peptide derived from TM6[180], suggesting that in unliganded CXCR4, the dimer interface may reside between TM5 and TM5, in a manner analogous to the A$_{2A}$ receptor. Next, a CXCR4 TM5-TM5 simulation was run. In the TM5-TM5 simulation set 8% of the ensemble formed stable dimers. The mean distance between helices was ~11-13Å (see Table 4.1).

4.3.1 Identification of the contact interface for the CXCR4 homodimer

Following screening, the cutoff to 15Å was extended in the averaged interacting simulations, which revealed an increase in the number of interactions observed and presented the strongest interacting residues. Identification of the contact interface for the CXCR4 homodimers where the most prominent interactions was established from the averaged interhelix contact matrices, the formation of dimers with specific interactions between F201$^{5.40}$ on one TM and

the following six residues in the other TM: V198$^{5.37}$, Q200$^{5.39}$, F201$^{5.40}$, Q202$^{5.41}$, I204$^{5.43}$ and M205$^{5.44}$ (see Figure 4.2b). These residues are found in the upper third of the N-terminal end of TM5.
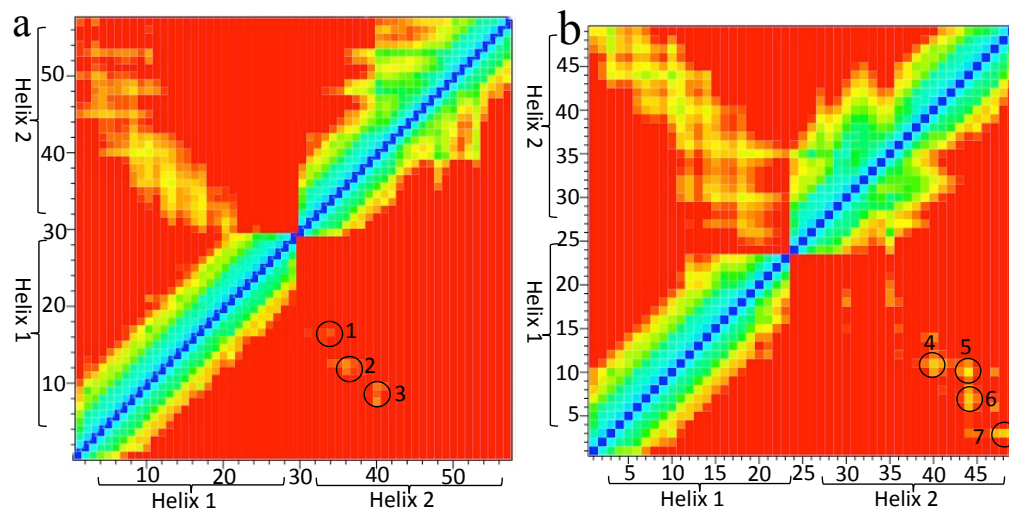

Figure 4.2 Contact matrices (heat maps) between two CXCR4 helices

Contact matrices (heat maps) between two CXCR4 helices showing specific interactions between TM5 (Helix 1) and TM6 (Helix 2) (a) and between TM5 (Helix 1) and TM5 (Helix 2) (b). The identified amino acid interactions are numbered as follows: 1) F201$^{5.40}$ with V198$^{5.37}$, Q200$^{5.39}$, F201$^{5.40}$, Q202$^{5.41}$, I204$^{5.43}$ and M205$^{5.44}$.

4.4 Computational identification of interacting interfaces: β_1AR

Two alternating dimer interfaces have been proposed from the crystal structure of the ligand-free basal state of the β_1 adrenergic receptor (β_1AR) (4GPO) that were packed in parallel arrangement within the lipid bilayer. The first involves TM1, TM2, extracellular loop 1 and the C-terminal H8; the second involves TM4, TM5 and the intracellular loop 2. In the first dimer, the TM1-TM2-H8 dimer, the interactions identified within the TM helices were mainly through TM1 and the interacting residues included: $Q38^{1.29}$, $Q39^{1.30}$, $A42^{1.33}$, $L46^{1.37}$, $L49^{1.40}$, $L50^{1.41}$, $V52^{1.43}$, $L53^{1.44}$ and $L54^{1.45}$ and from TM2: $P96^{2.59}$, $A99^{2.62}$, $T100^{2.63}$ and $V103^{2.66}$ these interactions are mainly hydrophobic and van dar Waals. The second dimer interface included residues from both TM4; $L171^{4.44}$; and TM5: $R205^{5.36}$, $A206^{5.37}$, $A210^{5.41}$, $I218^{5.49}$ and $R229^{5.60}$[151].

Two heterologous simulations between the β_1AR helices: TM1 and TM2 and between TM4 and TM5 were run to identify whether contact interfaces could be identified for either. No stable dimers were formed in the TM1-TM2 simulation (see Figure 4.3a) that led to the possibility that the contact interface was maybe between two TM1's from two receptors as the interactions were mainly from TM1. This possibility was investigated and a TM1-TM1 simulation was run and identified a stable dimer in only one replica in the ensemble, in a 2% of simulations (see Figure 4.3b). In the TM4-TM5 simulation, stable dimers were formed at 14% of simulations (see Figure 4.3c).
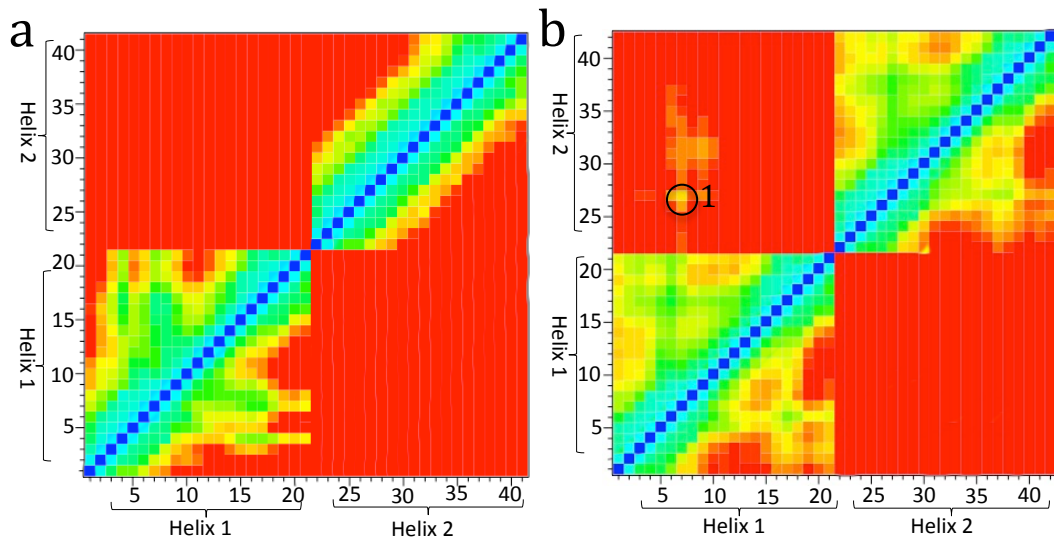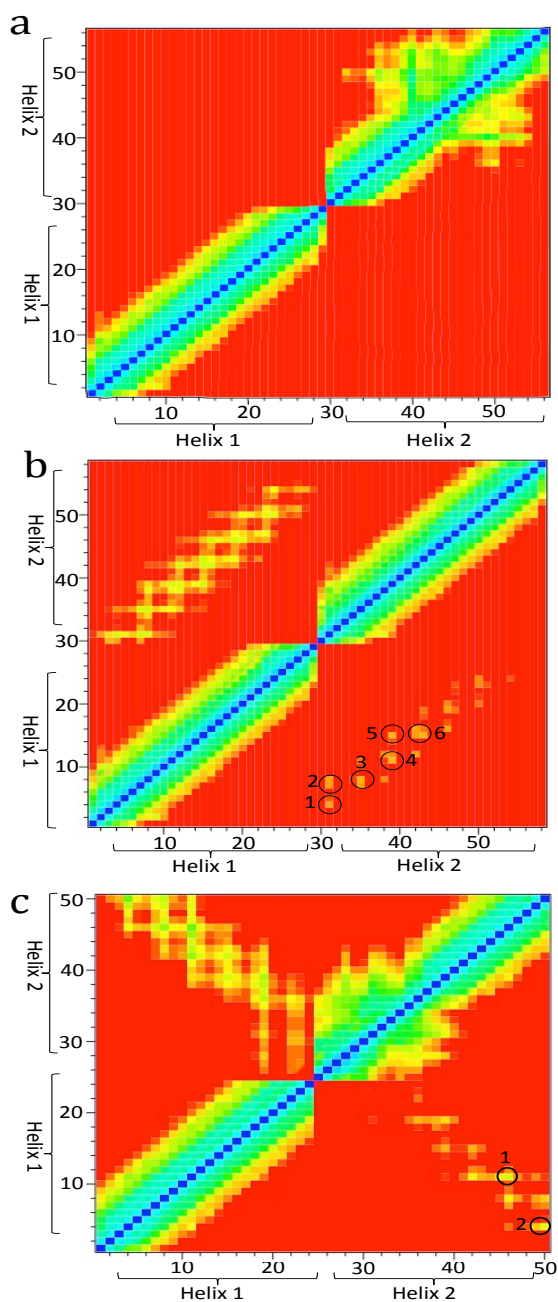
Figure 4.3 Contact matrices (heat maps) between two β₁AR helices

Contact matrices (heat maps) between two β_1AR helices showing specific interactions between TM1 (Helix 1) and TM2 (Helix 2) (a), between TM1 (Helix 1) and TM1 (Helix 2) (b) and between TM4 (Helix 1) and TM5 (Helix 2) (c). Results shown are the average for each ensemble. The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: in (b) 1) W40$^{1.31}$ with A42$^{1.33}$, S45$^{1.36}$ and L46$^{1.37}$; 2) M44$^{1.35}$ with L46$^{1.37}$; 3) A49$^{1.39}$ with M48$^{1.38}$; 4) L53$^{1.44}$ with M48$^{1.38}$; 5) L53$^{1.44}$ with M48$^{1.38}$, V51$^{1.40}$ and V52$^{1.41}$; 6) L54$^{1.45}$ with V51$^{1.40}$; (c) 1) K159$^{4.43}$ with Y227$^{5.58}$; 2) W166$^{4.50}$ with Y231$^{5.62}$.

4.4.1 Identification of the contact interface for the β_1AR homodimer

In the interacting replica of TM1-TM1 ensemble the contact interface was identified by measuring the average interhelical contact distance between the two TM1 helices the mean distance between helices was ~5-18Å. The interface included residues W40$^{1.31}$, M44$^{1.35}$, A49$^{1.39}$, L53$^{1.44}$ and L54$^{1.45}$ from one TM1 with A42$^{1.33}$, S45$^{1.36}$, L46$^{1.37}$, M48$^{1.38}$, V51$^{1.40}$ and V52$^{1.41}$ in the other TM1 (Figure 4.3,b) these residues are located in the center of TM1.

In the TM4-TM5 ensemble, the contact interface was identified (the contacting residues) by measuring the average interhelical contact distance between the TM4-TM5 helices which was ~7-8Å (Figure 4.3c) and the specific interactions identified were between K159$^{4.43}$ and Y227$^{5.58}$ and between W166$^{4.50}$ and Y231$^{5.62}$ these interactions are located in the middle of TM4 and at the end of TM5.

4.5 Atomistic representation and proposed nature of interactions

CG simulations lack the specific details needed to describe the nature and type of the interactions that might take place when the two TM helices are within 10Å of each other because in the CG simulation a small group of atoms is treated as a single particle in a 4:1 ratio. To try to understand what type of interactions were involved in the formation of stable dimers the combined representative CG dimer structures were converted into representative atomic structures to enable a measurement of distance between atoms[171] allowing hypotheses to be drawn about the molecular nature and possible role of the interactions between dimeric helices. The occurrence of C_α-H·····O contacts as a function of the interhelical

axial distance in the database are between 6 and 12Å with an average of 8.9Å[175].

4.5.1 The TM5-TM5 homodimer of A_{2A}

A representation of the converted atomistic wild-type A_{2A} TM5 dimer is seen in Figure 4.4. Using this atomistic representation, the presence of possible electrostatic interactions or hydrogen bonding was investigated by measuring the distance between the specific interacting residues. The interaction between the two methionine residues (M193$^{5.54}$-M193$^{5.54}$) and between valine and tyrosine (V196$^{5.57}$-Y197$^{5.58}$) are likely to correspond to van der Waals interactions. Y197$^{5.58}$ in helix 1 and Y197$^{5.58}$ in helix 2 each interact as hydrogen donor and acceptor in the dimer, forming bonds between the peptide backbone and the tyrosine side chain (see Figure 4.4). As the measurement of these distances is longer than the optimal hydrogen bond distance, 2.7Å, such hydrogen bonds are more likely to be formed backbone-to-side-chain because their interhelical distance of 8Å is above the 7.6Å limit of backbone-to-backbone interactions[175]. There are presently no structural data with which to compare these eCG-MD results.
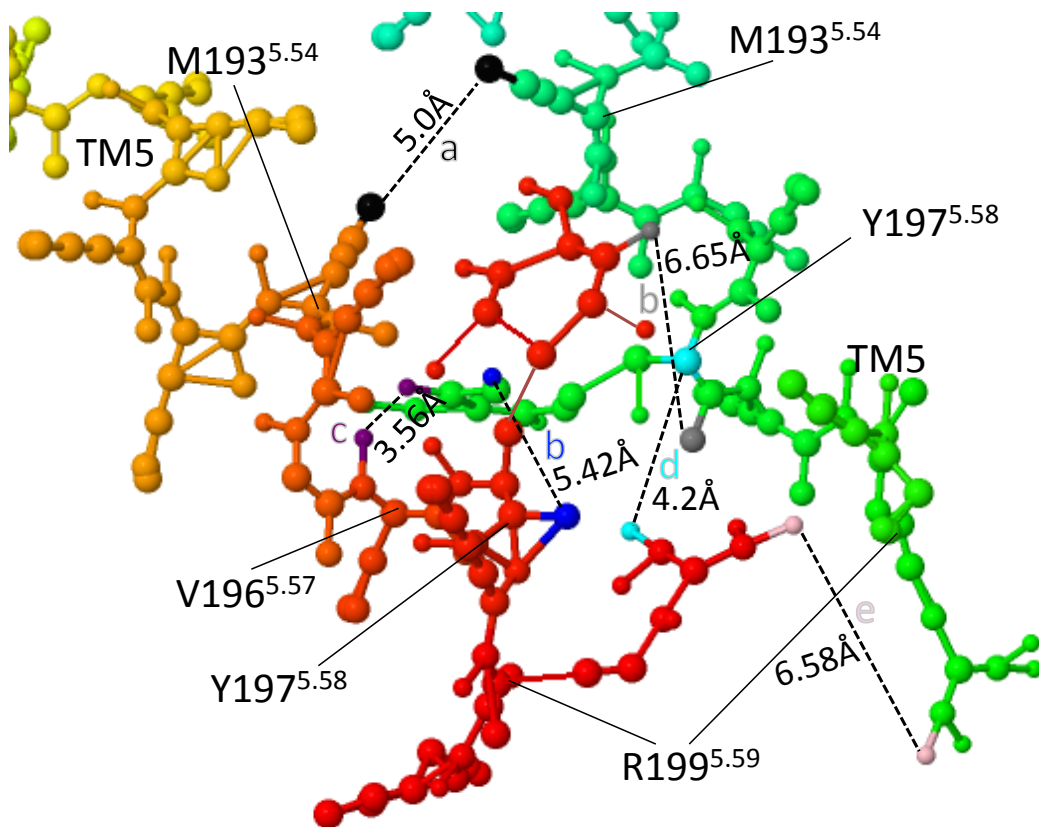
Figure 4.4 Atomistic representation of the pairwise interactions identified from the wild-type TM5-TM5 ensemble

The representative mean distance is shown in the figure and the mean distance ± SD for all hits detected per pair is shown in Table 4.1. All distances between interacting amino acids are calculated from side chain to side chain.

4.5.2 The rhodopsin homodimer

The rhodopsin dimer model (1N3M), shown in Figure 4.5a, reveals that there is a greater interface area between TM4 and TM5 than between TM1 and TM2. The specific interacting residues identified from the atomistic representation obtained using the eCG-MD computational method developed in Chapter 3 are distributed throughout the length of TM1 and TM2 but restricted to the bottom third of TM4 and TM5, with respect to the intracellular face of the receptor (see Figure 4.5b-e). A comparison of the eCG-MD results with the TM1 and TM2 contact interface of 1N3M is shown in Figure 4.5b and c, respectively. The measured distance between the hydrogen on the COOH group of M119$^{1.39}$ and the double-bonded oxygen of the COOH group on the side chain of D224$^{2.50}$ is 10.32 $\pm$ 3.21Å in the eCG-MD model (see Figure 4.5b), similar to 9.01Å in 1N3M (Figure 4.5c). Measurement of the distance between F127$^{1.47}$ and L218$^{2.44}$ is 14.20 $\pm$ 4.07Å in the model and 15.28Å in 1N3M. F127$^{1.47}$ and L218$^{2.44}$ are located towards the bottom of their respective helices, a position that is constrained by the first intracellular loop of rhodopsin in 1N3M, but not in the eCG-MD model. Similar conservation of distance was identified between interacting residues in TM4 and TM5 (see Table 4.1).

Figure 4.5 Atomistic structures of the rhodopsin dimer

Atomistic structures of the rhodopsin dimer (a) Atomistic structure of the rhodopsin dimer model (1N3M) viewed from above, with the TMs used for simulations identified by color as follows: TM1 (blue), TM2 (red), TM4 (purple) and TM5 (orange). Representative TM structures were obtained from the means of all replicas in which interactions were detected. The representative and model structures of TM1-TM2 are shown in (b) and (c), respectively. The representative and model structures of TM4-TM5 are shown in (d) and (e), respectively. Specific interactions were identified in TM1-TM2 simulations (M119$^{1.39}$ with D224$^{2.50}$ and F127$^{1.47}$ with L218$^{2.44}$) and in TM4-TM5 simulations (H411$^{4.41}$ with Q530$^{5.60}$, G415$^{4.45}$ with F525$^{5.55}$, F418$^{4.48}$ with L521$^{5.51}$ and T419$^{4.49}$ with L521$^{5.51}$). Table 4.1 shows a comparison of the distances between specific atoms in interacting residues of the representative structures and the distances between the same atoms in the model structure.

4.5.3 The CXCR4 homodimer

The eCG-MD studies of CXCR4 identified novel interactions in the homodimer between TM5 and TM5; these were visible but previously unidentified in the CXCR4 (3ODU) dimer (see Figure 4.6). This is similar to what was seen for A_{2A}, but the interacting residues in CXCR4 are closer to the extracellular side of the membrane than in A_{2A}. A comparison of the mean distance between interacting residues obtained from the simulations with the distance measured between the same residues in the crystal structure shows a similar conservation of distance, particularly between interacting residues further down the helix. This suggests a contribution of the loops in influencing interactions towards the ends of the helices, as was seen for rhodopsin.
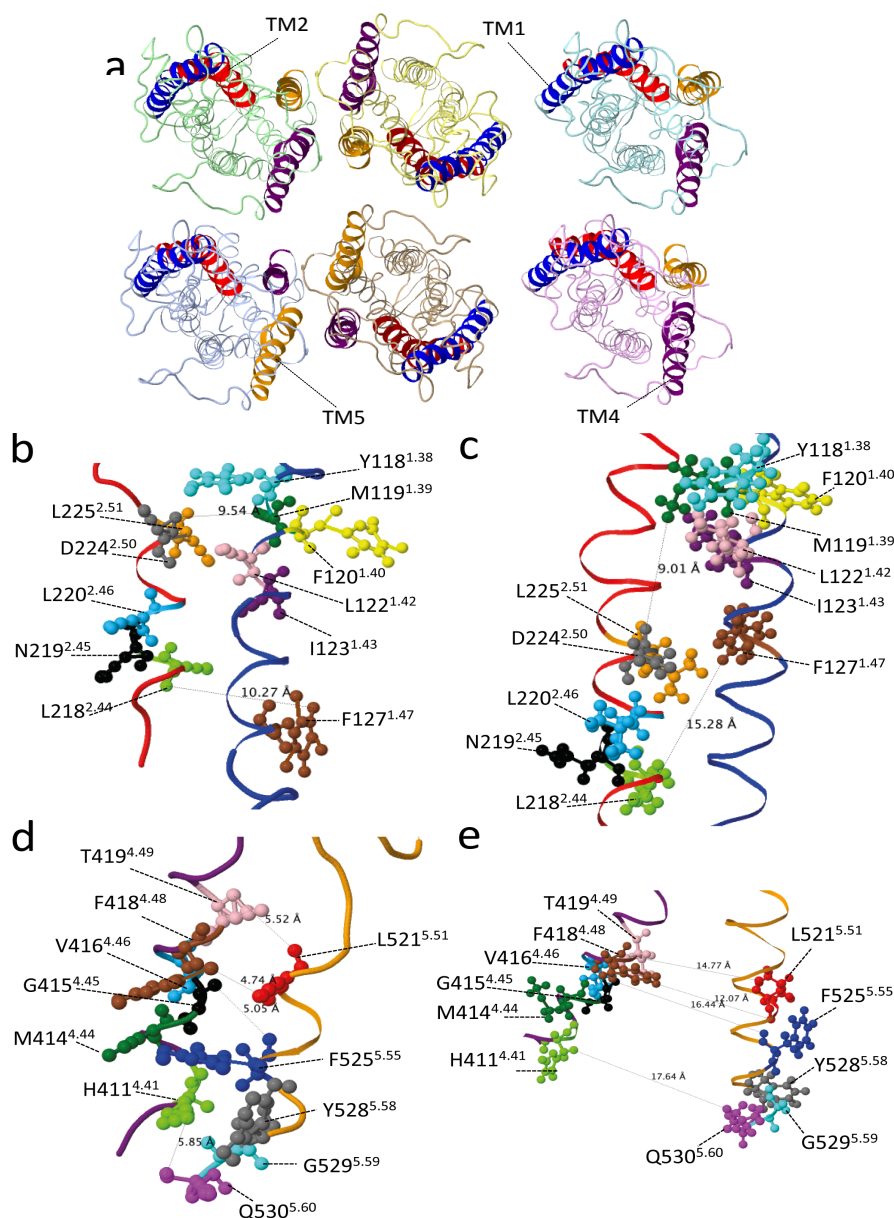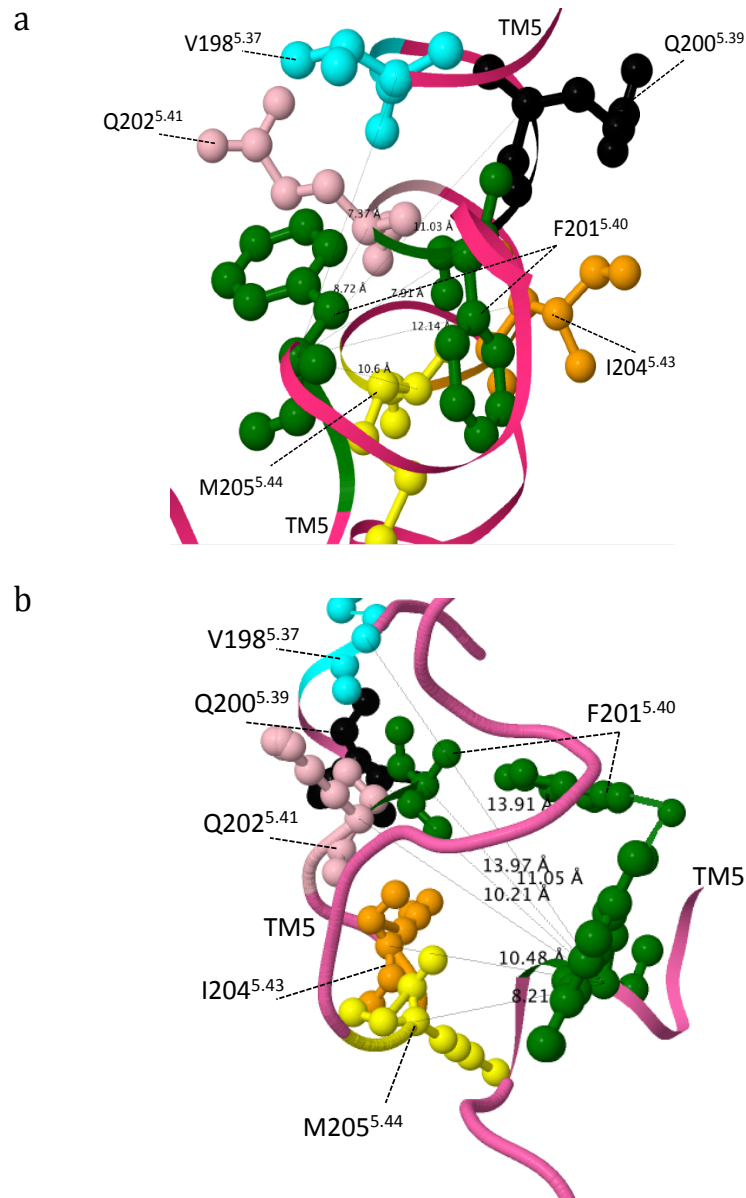
Figure 4.6 Atomistic structures of the CXCR4 dimer

Atomistic structures of the CXCR4 dimer (a) Atomistic structure of the CXCR4 dimer model (3ODU), with the TMs used for simulations identified by colour where is TM5 (pink). Representative TM structures were obtained from the means of all replicas in which interactions were detected. The representative and model structures of TM5-TM5 are shown in (a) and (b), respectively. Specific interactions were identified in TM5-TM5 simulations (F201[5.40] with V198[5.37], Q200[5.39], F201[5.40], Q202[5.41], I204[5.43] and M205[5.44]). Table 4.1 shows a comparison of the distances between specific atoms in interacting residues of the representative structures and the distances between the same atoms in the model structure.

4.5.3 The β₁AR homodimer

Like rhodopsin, contact interfaces between TM1 and TM2 and between

TM4 and TM5 had been proposed for the β_1 adrenergic receptor. However, using

our method it was possible to identify a contact interface between TM1 and

TM1, rather than between TM1 and TM2. The conversation of distance between

interacting residues and the distance between the same residues in the crystal

structure is difficult to compare given the interactions were only identified in one

replica, however, the measurements are in agreement with those of the crystal

structure within the limits of error calculated from other simulations. The eCG-

MD data suggest that the TM4-TM5 contact interface, and the four specific

amino acids identified within it, may constitute the principal dimer interface in

β_1AR homodimers. It was not possible to compare the distances obtained in the

TM4-TM5 simulation with those measured in the crystal structure 4GPO, which

had been submitted with the orientation of the dimer showing the proposed TM1-

TM2 interface.
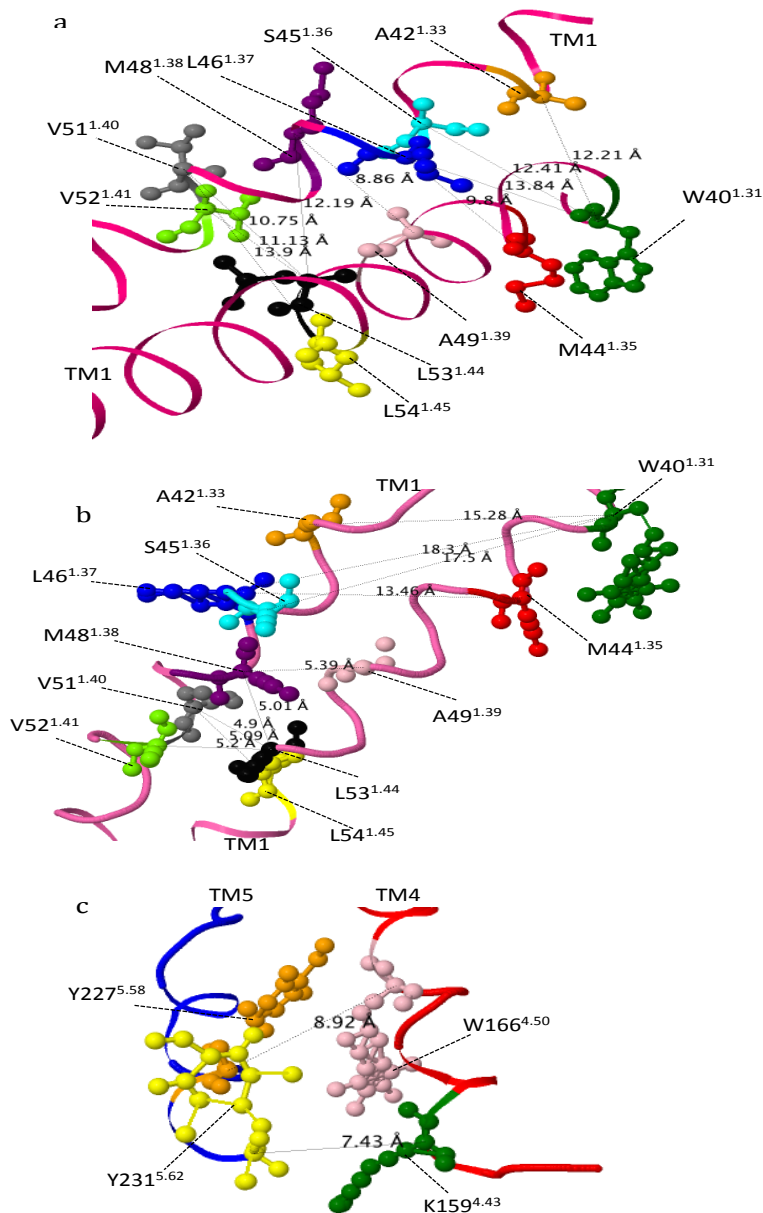
Figure 4.7 Atomistic structures of the β₁AR dimer

Atomistic structures of the β₁AR dimer (a) Atomistic structure of the β₁AR dimer model (4GPO), with the TMs used for simulations identified by colour as follows: TM1 (pink), TM4 (red) and TM5 (blue). Representative TM structures were obtained from the means of all replicas in which interactions were detected. The representative and model structures of TM1-TM1 are shown in (a) and (b), respectively. The representative and model structures of TM4-TM5 are shown in (c). Specific interactions were identified in TM1-TM1 simulations (W40$^{1.31}$ with A42$^{1.33}$, S45$^{1.36}$ and L46$^{1.37}$; M44$^{1.35}$ with L46$^{1.37}$; A49$^{1.39}$ with M48$^{1.38}$; L53$^{1.44}$ with M48$^{1.38}$; L53$^{1.44}$ with M48$^{1.38}$, V51$^{1.40}$ and V52$^{1.41}$; L54$^{1.45}$ with V51$^{1.40}$) and in TM4-TM5 simulations (K159$^{4.43}$ with Y227$^{5.58}$; W166$^{4.50}$ with Y231$^{5.62}$.).

4.6 Summary

To validate the results obtained using the ensemble-based CG-MD computational methodology, it was necessary to compare the computational results with the structural experimental data that existed for other GPCRs. In this chapter the eCG-MD method was tested on three GPCRs that fit these criteria. First, the rhodopsin dimer for which crystallographic data had identified contact interfaces between TM1 and TM2 and between TM4 and TM5. Ensemble CG-MD confirmed dimerisation and the identification of specific interactions within each of these heterologous TM pairs. There is a striking convergence between the distances predicted computationally and those calculated from 1N3M, particularly for specific interactions between TM1 and TM2, showing that the method is able to provide accurate and precise predictions in agreement with experimental findings. The eCG-MD method is also able to identify novel interfaces as seen in the second (CXCR4) and third (β_1AR) cases studied, where a novel interface in CXCR4 between TM5 and TM5 and a novel interface in β_1AR between TM1 and TM1 were identified, in addition to confirming the previously identified contact interface between TM4 and TM5 in the β_1AR. The β_1AR has been shown to form transient interactions whereas the β_2 adrenergic receptor can form stable oligomers[181]. The ability of the eCG-MD methodology to detect a stable dimer of TM1-TM1 in the β_1AR shows the value of ensemble-based simulations for the identification of transient interactions. This computational methodology was also able to provide insights into the molecular nature of the contact interface between the two TM5 helices of the A_{2A} receptor, something that was unable to be provided by the available biophysical data.

Chapter 5
Creation and computational analyses of TM helix-helix interactions in A$_{2A}$

5.1 Introduction

As mentioned previously, a computational method for analyzing GPCR interfaces, eCG-MD, has been developed. It was tested on specified TM helices of four types of GPCRs, all which had good experimental data in which to validate the method. The first case tested was that of homomeric A$_{2A}$ receptors. The fifth transmembrane domain (TM5) that was identified to be in involved in the self-association by far-UV CD spectroscopy and SDS-PAGE using synthetic peptides corresponding to the different transmembrane domains[64] was studied through the developed computational method, which gave results consistent with the experimental data. This chapter aims to apply the eCG-MD methodology described in Chapter 3 and validated in Chapter 4, to identify additional helix-helix interactions between the seven TM helices of A$_{2A}$ and to compare these results to the experimentally-obtained data for the self-association of the helices of the A$_{2A}$ adenosine receptor[147].

There are 21 possible heterologous pairs and seven possible homo-interactions; the TM sequences used for this study are found in Table 2.2. All combinations of possible helix-helix interactions were examined *in silico* in 26 different ensemble sets. The final mean distance between the two helices in an

ensemble of 50 replicas in each set was used to identify the specific interactions between the A_{2A} homo/heterodimers in all the simulation sets run.

5.2 Interacting interfaces of the TM5 pairwise ensembles

Six simulation sets were tested that included: 1) TM5-TM1; 2) TM5-TM2; 3) TM5-TM3; 4) TM5-TM4; 5) TM5-TM6 and 6) TM5-TM7. Following the application of the 10Å cut-off, 18% of TM5-TM1 ensemble formed stable dimers interacted and spontaneously formed homodimers, whereas in the TM5-TM2 ensemble 32% formed stable dimers. Interactions in the TM5-TM3 ensemble were seen in 8% of simulations. 16% of TM5-TM4 ensemble was found to interact and form stable dimers. The TM5-TM6 ensemble stable dimers were seen in 14% of the total simulations. In TM5-TM7 ensemble 12% were discovered to interact and form stable dimers. For all simulations, of those pairwise combinations in which dimerisation was identified after the cut-off 10Å had their trajectories combined and the results with heat maps of interactions observed at 12Å and 15Å compared (see Figure 5.1). The location of the contact interface in TM5 with the other helices was then mapped by comparison with the crystal structure of $A_{2A}R$ (3EML) on which the helices were modelled.
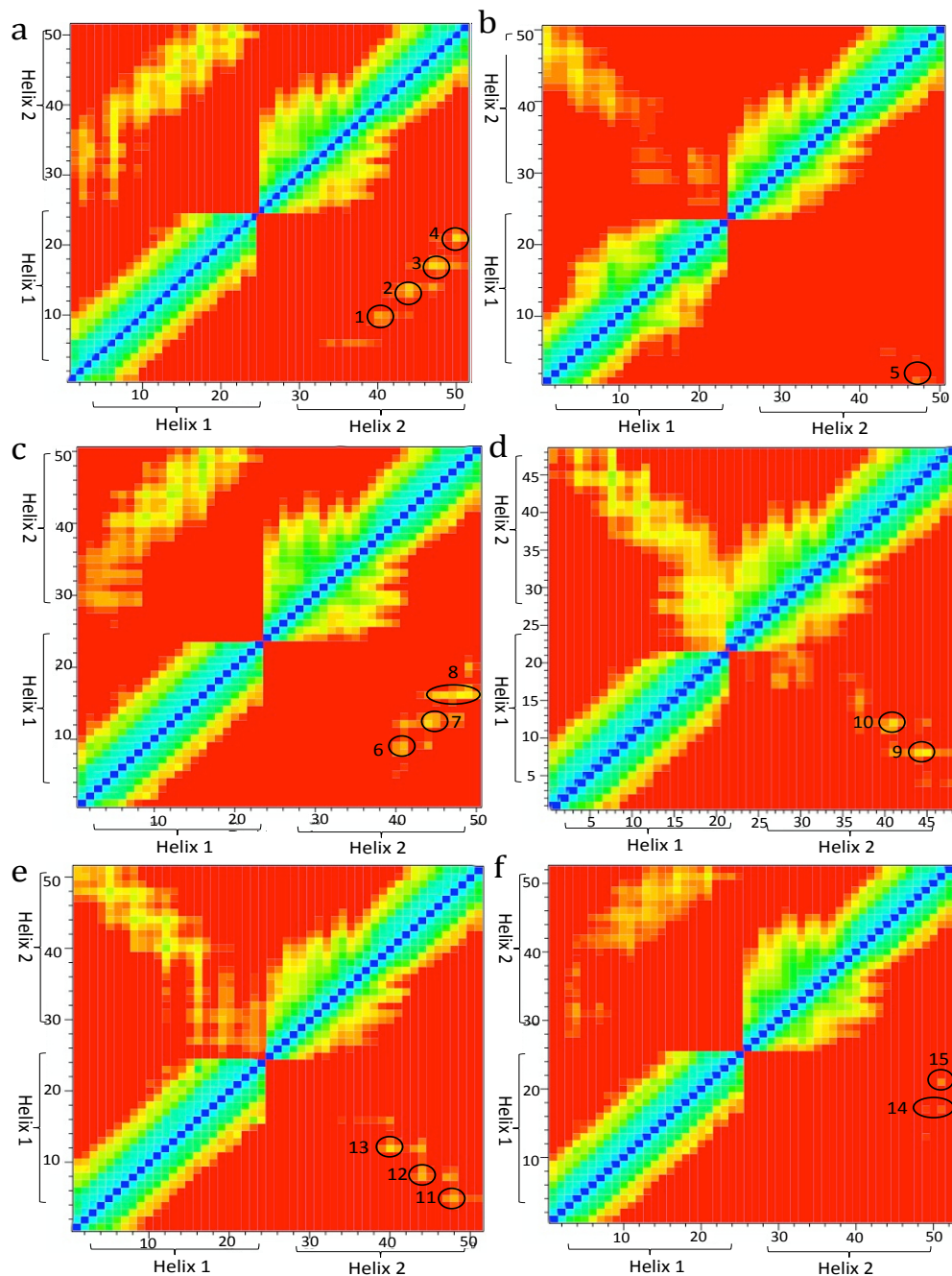
Figure 5.1 Contact matrices showing specific pairwise interactions between TM5 and an A$_{2A}$ helix

Contact matrices (heat maps) between two A$_{2A}$ helices showing specific interactions between TM1 (Helix 1) and TM5 (Helix 2) (a), between TM2 (Helix 1) and TM5 (Helix 2) (b). between TM3 (Helix 1) and TM5 (Helix 2) (c), between TM4 (Helix 1) and TM5 (Helix 2) (d), between TM6 (Helix 1) and TM5 (Helix 2) (e) and between TM7 (Helix 1) and TM5 (Helix 2) (f). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts.

5.2.1 Identification of the contact interfaces of the TM5 pairwise ensembles

After identifying the interacting replicas in all sets, their trajectories were combined and the contact interface residues were identified by measuring the average interhelical contact distance between the two helices as explained previously in section 3.7.2 and results plotted at the two different distances, 12Å and 15Å as heat maps of interactions.

The most prominent interactions within the averaged 12Å cutoff with an interhelical distances of about ~8-10Å in the TM5-TM1 ensemble involved the proline and methionine residues of the P189$^{5.50}$xxxM193$^{5.54}$ conserved motif were part of the interacting residues. The identified interacting amino acids between the TM1 and TM5 helices are: 1) A17$^{1.43}$ with P189$^{5.50}$; 2) A20$^{1.46}$, I21$^{1.47}$ with M193$^{5.54}$; 3) N24$^{1.50}$ with V196$^{5.57}$, Y197$^{5.58}$; 4) C31$^{1.54}$ with R199$^{5.60}$ (see Figure 5.1a). In TM5-TM2 ensemble no strong interactions were detected and only one interaction weak interaction between the two helices was detected between 5) F44$^{2.42}$ with Y197$^{5.58}$ (see Figure 5.1b). The identified interactions in the TM5-TM3 ensemble were between residues: 6) L87$^{3.36}$ with L191$^{5.52}$; 7) Q89$^{3.38}$ with L191$^{5.52}$ and L194$^{5.55}$; and 8) F93$^{3.42}$ with G195$^{5.56}$, V196$^{5.57}$, Y197$^{5.58}$, L198$^{5.59}$, R199$^{5.60}$ and I200$^{5.61}$ (see Figure 5.1c). Residues in TM5-TM4 ensemble are: 9) W129$^{4.50}$ with V196$^{5.57}$ and Y197$^{5.58}$; 10) F133$^{4.54}$ with M193$^{5.54}$ (see Figure 5.1d). Both the proline and methionine in the P189$^{5.50}$xxxM193$^{5.54}$ motif were involved in the TM5-TM6 ensemble interactions; the interactions present were between residues: 11) V239$^{6.51}$ with Y197$^{5.58}$; 12) A243$^{6.55}$ and L244$^{6.56}$ with M193$^{5.54}$; 13) L247$^{6.59}$ with P189$^{5.50}$ (see Figure 5.1e). In the TM5-TM7 ensemble the interacting residues were: 14) N284$^{7.49}$ with Y197$^{5.58}$ and R199$^{5.60}$; 15) Y288$^{7.53}$ with R199$^{5.60}$ (see Figure 5.1f).

In all six of the TM5 pairwise ensemble sets, the detected interactions took place at the same position within the TM5 helix indicating that a defined orientation is needed to establish a specific interaction in TM5 with other TM helices. The most prominently interacting residues in TM5 are: M193$^{5.54}$, V196$^{5.57}$, Y197$^{5.58}$, R199$^{5.60}$, and I200$^{5.61}$. The identification of the presence of M193$^{5.54}$ in the contact interface suggested that this residue may play a significant role in how the TM5 helix interacts with the other transmembrane helices.

The occurrence of C$_\alpha$-H⋯O contacts as a function of the interhelical axial distance in the database are between 6 and 12Å with an average of 8.9Å[175] was then investigated. The shortest distance between interactions were found between TM5-TM1 and TM5-TM3 ensembles, which was about ~8Å. TM5-TM4 and TM5-TM6 distances were in the range of 8.5 to 9Å. The TM5-TM2 ensemble showed that the interactions were at long distances, more than 10Å, suggesting that these two helices do not favour heterodimerisation. In the TM5-TM7 ensemble, the interactions were also very long and were about ≥ 10Å. Not all the result observed were consisted with the experimental FRET and CD results on the A$_{2A}$ receptor[147], where they have only identified one heterodimer that was between the TM5-TM6 helices but in our method we have identified the same pair along with more interacting pairs, namely the TM5-TM1, TM5-TM3 and TM5-TM4.

5.3 Interacting interfaces of the TM1 pairwise ensembles

Six simulation sets were tested. Following the application of the 10Å cut-off, which included: 1) TM1-TM1; 2) TM1-TM2; 3) TM1-TM3; 4) TM1-TM4;

5) TM1-TM6 and 6) TM1-TM7. No stable dimers were detected in the TM1-TM1 ensemble. 10% of TM1-TM2 and TM1-TM6 ensembles formed stable dimers in both simulations. In the TM1-TM3 and TM1-TM7 simulations, both sets were found to interact in 12% of ensembles. In the last set, the TM4-TM1 simulation 14% of ensemble formed stable dimers.

As explained previously, all the simulations in which dimerisation was identified had their trajectories combined and the results with heat maps of interactions observed at 12Å and 15Å compared (see Figure 5.2). The location of TM1 contact interface with the other helices was then mapped by comparison with the crystal structure of A$_{2A}$R (3EML) on which the helices were modelled on to.

5.3.1 Identification of the contact interfaces of the TM1 pairwise ensembles
After identifying the interacting replicas in all sets, their trajectories were combined and the contact interface residues were identified by measuring the average interhelical contact distance between the two helices as explained previously in section 3.7.2 and results plotted at the two different distances, 12Å and 15Å as heat maps of interactions.

In TM1-TM1 ensemble, it can be seen that TM1-TM1 does not come close enough to form potential specific interactions and none were detected in the combined trajectories, no results were obtained at the 12Å and also at the 15Å cutoff (see Figure 5.2a). The interactions occurring within the averaged 12Å cutoff with an interhelical distances of about ~8Å in TM1-TM2 simulations were

between the residues: 1) F44$^{2.42}$, V46$^{2.44}$ and S47$^{2.45}$ with G23$^{1.49}$; 2) S47$^{2.45}$ with L19$^{1.45}$ and A20$^{1.46}$; 3) A50$^{2.48}$ with L19$^{1.45}$; 4) A51$^{2.49}$ with A17$^{1.43}$; 5) A54$^{2.52}$ with A17$^{1.43}$; 6) A54$^{2.52}$, V55$^{2.53}$ and G56$^{2.54}$ with E13$^{1.39}$ (see Figure 5.2b). In TM1-TM3, interactions were found at a distance of $\geq$ 10Å between the following residues: 7) F79$^{3.28}$ and C82$^{3.31}$ with E13$^{1.39}$; 8) F83$^{3.32}$ with I16$^{1.51}$; 9) S90$^{3.39}$ with N24$^{1.50}$ and the last interactions were 10) F93$^{3.32}$ and S94$^{3.33}$ with N24$^{1.50}$ (see Figure 5.2c). From TM1-TM4 ensemble, the interaction interface at about 8.5 to 9Å was between the residues: 11) W129$^{4.50}$ with N24$^{1.50}$; 12) F133$^{4.54}$ with A17$^{1.43}$ and A20$^{1.46}$; 13) G136$^{4.57}$ and L137$^{4.58}$ with I16$^{1.51}$ (see Figure 5.2d). In the TM1-TM6 ensemble the interacting interface in the simulations is at about ~ 8Å, the interacting residues are: 14) F242$^{6.43}$ with L19$^{1.45}$, A20$^{1.46}$ and I21$^{1.47}$; 15) W246$^{6.47}$ with L14$^{1.49}$ and I16$^{1.51}$ (Figure 5.2,e). In TM1-TM7 simulations, the interactions interface present is between the residues: 16) M270$^{7.35}$ and Y271$^{7.36}$ with T11$^{1.36}$; 17) H278$^{7.43}$ with V18$^{1.44}$; 18) S281$^{7.46}$ with L22$^{1.48}$ (see Figure 5.2f).

The TM1-TM7 ensemble had the longest distance between the interactions $\geq$ 10Å. In the TM1-TM3 and TM1-TM4 ensembles, the distances between the two helices were in the range of 8.5 to 9Å. The lowest distance between the interacting residues was found between residues in the TM1-TM2 and the TM1-TM6 ensembles at about ~8Å. In most of the six TM1 pairwise ensemble sets, the detected interactions took place at the same position within the TM1 helix indicating that a defined orientation is needed to establish a specific interaction in TM1 with other TM helices. The most prominently interacting residues in TM1 are: E13$^{1.39}$, I16$^{1.51}$, A17$^{1.43}$, A20$^{1.46}$, and N24$^{1.50}$ they are found

at the upper third of TM1 helix at the N-terminal side. The identification of the presence of E13$^{1.39}$ in the contact interface suggested that this residue may play a significant role in how the TM1 helix interacts with the other transmembrane helices as well as a contact point between two A_{2A} receptors, as it is found at the exterior side of TM1 helix of the receptor. To investigate this possibility, we introduced one ensemble set that included a mutated TM1 helix (see Table 2.2) with a point mutation at the E13$^{1.39}$ where the substitution of glutamic acid to alanine was established.

5.3.2 Identification of the mutated contact interfaces of the TM1-TM2 ensemble

Following the application of the 10Å screening cut-off, 8% of the TM1-E13$^{1.39}$A-TM2 ensemble formed stable dimers. All the replicas, of those pairwise combinations in which dimerisation was identified after the 10Å cut-off had their trajectories combined and the results with heat maps of interactions observed at 12Å and 15Å compared (see Figure 5.3).

In the TM1-E13A$^{1.39}$-TM2 mutation the contact interface between the helices was completely changed and most of the key interacting residues were lost and a new contact interface at a longer distances > 10Å at a new location was identified. The key interacting residues of the new contact interface are: 1) D52$^{2.50}$ with T11$^{1.37}$ and 2) D52$^{2.50}$ with A15$^{1.41}$. Mutation of the E13$^{1.39}$ in TM1 causes major change in the way in which the two helices come together that may lead to a distribution of formation of A_{2A} receptors homodimers.

Figure 5.2 Contact matrices showing specific pairwise interactions between TM1 and an A$_{2A}$ helix

Contact matrices (heat maps) between two A$_{2A}$ helices showing specific interactions between TM1 (Helix 1) and TM1 (Helix 2) (a), between TM2 (Helix 1) and TM1 (Helix 2) (b), between TM3 (Helix 1) and TM1 (Helix 2) (c), between TM4 (Helix 1) and TM1 (Helix 2) (d), between TM6 (Helix 1) and TM1 (Helix 2) (e) and between TM7 (Helix 1) and TM1 (Helix 2) (f). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts.

Figure 5.3 Contact matrices showing specific pairwise interactions between mutated TM1-E13$^{1.39}$ and TM2

Contact matrices (heat maps) between two A$_{2A}$ helices showing specific interactions between TM2 (Helix 1) and TM1 (Helix 2). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: 1) D52$^{2.50}$ with T11$^{1.37}$ and 2) D52$^{2.50}$ with A15$^{1.41}$.

5.4 Interacting interfaces of the TM2 pairwise ensembles

Five simulation sets were tested, which included: 1) TM2-TM2; 2) TM2-TM3; 3) TM2-TM4; 4) TM2-TM6 and 5) TM2-TM7. At the 10Å screening cutoff, 24% of the TM2-TM2 ensemble resulted in the formation of stable dimers but there were no specific interactions identified between residues as explained previously in section 3.7.2 (see Figure 3.4d-f). In TM2-TM3 and TM2-TM7 simulations, no stable dimers were seen between the two helices in both ensembles and no specific interactions were identified (see Figure 5.4a and d). In the TM2-TM4 ensemble, stable dimers were found to form in 8% of simulations (see Figure 5.4b) and 10% of the TM2-TM6 ensemble was observed to interact and form stable dimers (see Figure 5.4c).

As explained previously, all replicas in which dimerisation was identified had their trajectories combined and the results with heat maps of interactions observed at 12Å and 15Å compared (see Figure 5.4). The location of TM2 contact interface with the other helices was then mapped by comparison with the crystal structure of A$_{2A}$R (3EML) on which the helices were modelled on to, it was located at the upper third of TM2 at the N-terminal side. The most prominent helix-helix interactions within the 12Å cutoff with an interhelical distance of more than $\geq$ 10Å in both ensembles where stable dimers were seen. In the TM2-TM4 simulations, the identified helical interface was between the residues: 1) P139$^{4.60}$ and M140$^{4.61}$ with F62$^{2.60}$ (see Figure 5.4b). In TM6-TM2 simulations, the interaction interface was found between 2) H250$^{6.52}$ with V55$^{2.53}$, G56$^{2.54}$, V57$^{2.55}$ and F62$^{2.60}$ (see Figure 5.4c). In these two simulations, the interactions between the two helices were far apart that is further than the

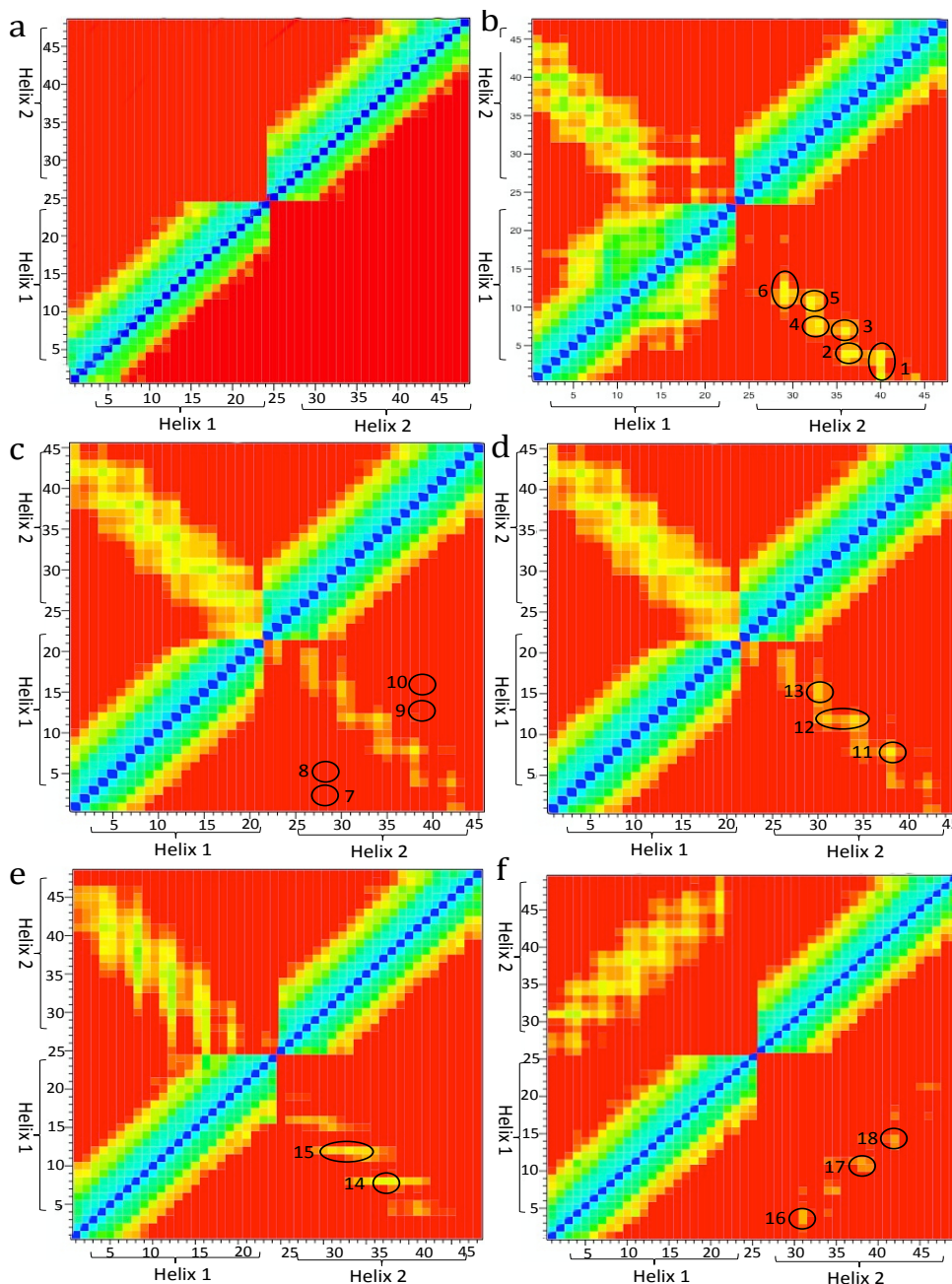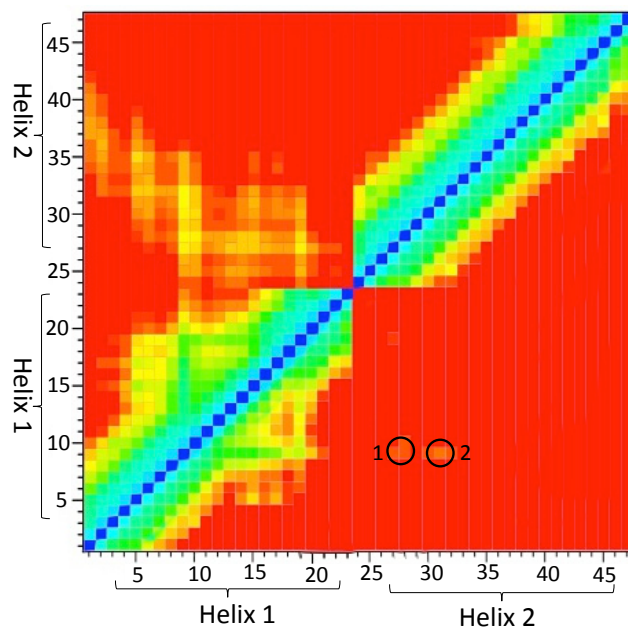suggested length of the C_α-H⋯O bonds.
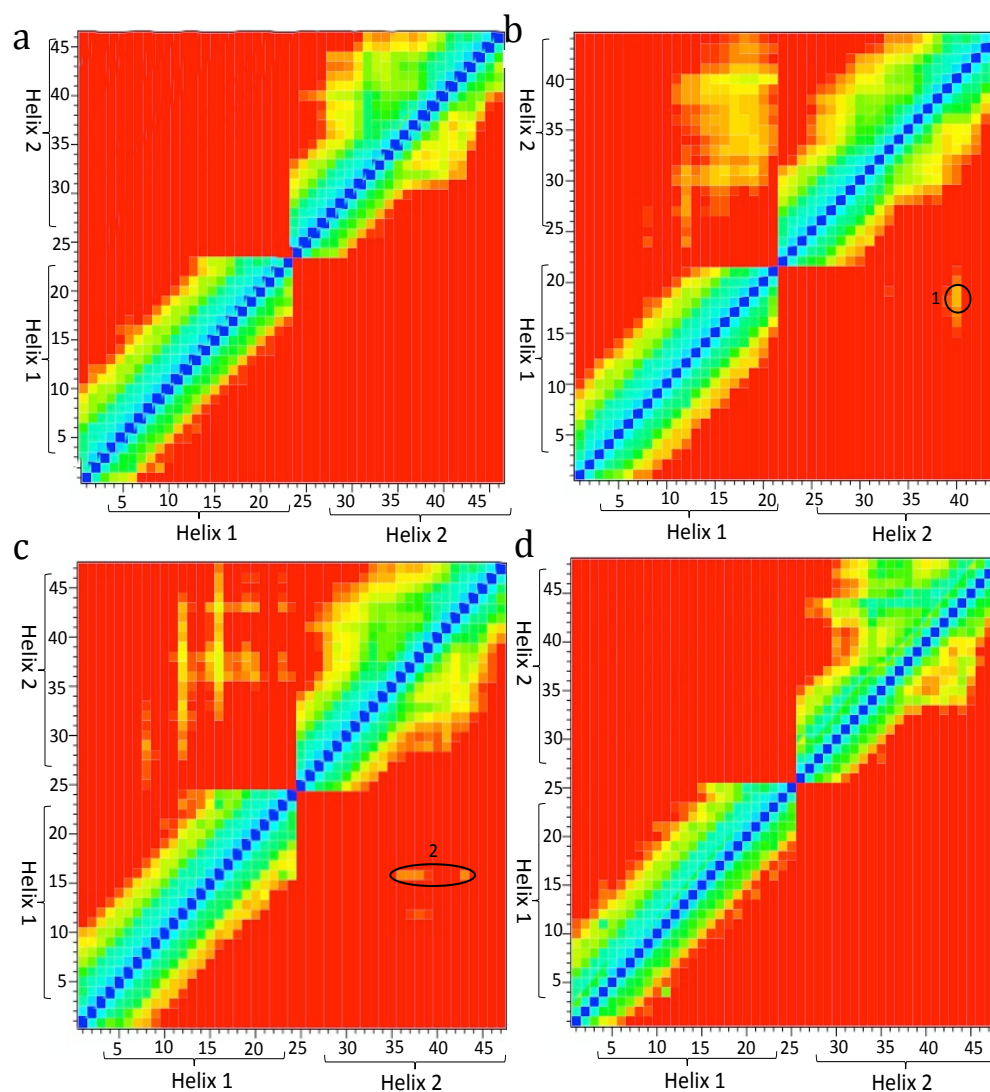
Figure 5.4 Contact matrices showing specific pairwise interactions between TM2 and an A_{2A} helix

Contact matrices (heat maps) between two A_{2A} helices showing specific interactions between TM3 (Helix 1) and TM2 (Helix 2) (a), between TM4 (Helix 1) and TM2 (Helix 2) (b), between TM6 (Helix 1) and TM2 (Helix 2) (c) and between TM7 (Helix 1) and TM2 (Helix 2) (d). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts.

5.5 Interacting interfaces of the TM3 pairwise ensembles

The remaining four possible pairwise simulation sets were tested that included: 1) TM3-TM3; 2) TM3-TM4; 3) TM3-TM6 and 4) TM3-TM7. Following the application of the 10Å cut-off, 14% of the TM3 homodimer ensemble was identified to form stable dimers. In the TM3-TM4 ensemble, 12% of simulations where found to interact and form stable dimers. Simulations of TM3-TM6, stable dimers were found in 2% of simulations only. In the TM7-TM3 simulations, no interaction and no stable dimers were seen between the helices in all of the replicas.

5.5.1 Identification of the contact interfaces of the TM3 pairwise ensembles

After identifying the interacting replicas in all sets, their trajectories were combined and the contact interface residues were identified by measuring the average interhelical contact distance between the two helices as explained previously in section 3.7.2 and results plotted at the two different distances, 12Å and 15Å as heat maps of interactions.

The interactions occurring within the averaged 12Å cutoff with an interhelical of about ~8-10Å in the TM3-TM3 ensemble were between the residues found in the V86$^{3.35}$xxxSxxFS94$^{3.43}$ motif in the two helices, (see Figure 5.5a). In the TM3-TM4 simulations was found between: 4) F133$^{4.54}$ with V86$^{3.35}$ and 5) W129$^{4.50}$ with S90$^{3.39}$ and F93$^{3.42}$ (see Figure 5.5b). In the TM3-TM6 set, the interactions were found between: 6) H250$^{6.52}$ with F83$^{3.32}$; 7) W246$^{6.48}$ and H250$^{6.52}$ with L87$^{3.36}$; 8) F242$^{6.44}$ with S90$^{3.39}$ and the last one is between 9) V239$^{6.41}$ with S94$^{3.43}$ (see Figure 5.5c). All the TM3 pairwise combination

simulations, except with the TM7 set where no stable dimers were identified, showed an interhelical distance between the two helices at the range of 8.5 to 9Å. In most of the TM3 pairwise ensemble sets, the interactions took place at the same position within the TM3 helix indicating that a defined orientation is needed to establish a specific interaction in TM3 with other TM helices. The most prominently interacting residues in TM3 are: V86$^{3.35}$, S90$^{3.39}$, F83$^{3.32}$, F93$^{3.42}$ and S94$^{3.43}$ they are found at the middle third of TM3 helix.
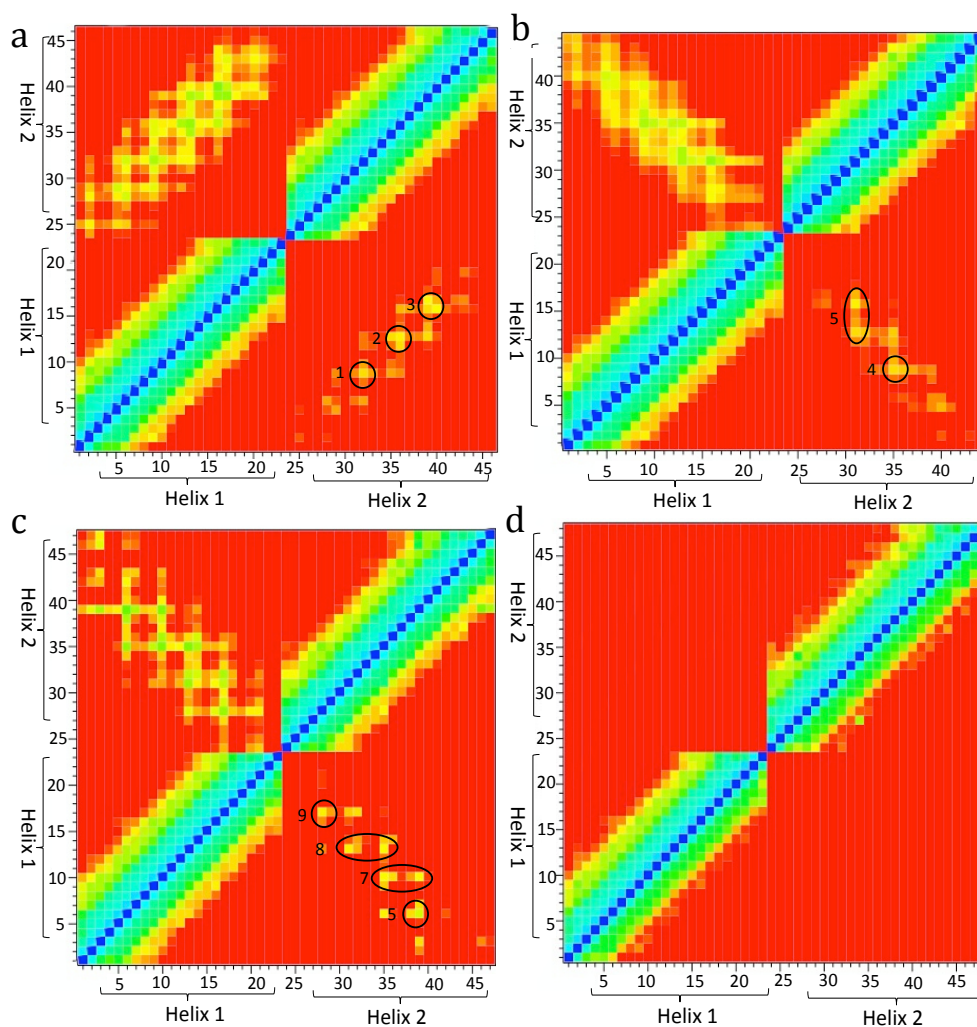
Figure 5.5 Contact matrices showing specific pairwise interactions between TM3 and an A_{2A} helix

Contact matrices (heat maps) between two A_{2A} helices showing specific interactions between TM3 (Helix 1) and TM3 (Helix 2) (a), between TM3 (Helix 1) and TM4 (Helix 2) (b), between TM3 (Helix 1) and TM6 (Helix 2) (c) and between TM3 (Helix 1) and TM7 (Helix 2) (d). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts.

145

5.6 Interacting interfaces of the TM4 pairwise ensembles

The remaining three possible pairwise simulation sets were tested that included: 1) TM4-TM4; 2) TM4-TM6 and 3) TM4-TM7. Following the application of the 10Å cut-off to the pairwise combinations of A$_{2A}$R TM4 pairwise ensembles revealed that in the TM4-TM4 and TM4-TM7 ensembles, 6% of simulations were identified to have interacting helices and forming stable dimers. In the TM4-TM6 simulations, 14% of the 50 simulations where found to interact and form the stable dimers.

5.6.1 Identification of the contact interfaces of the TM4 pairwise ensembles

After identifying the interacting replicas in all sets, they had their trajectories combined and the contact interface residues were identified by measuring the average interhelical contact distance between the two helices as explained previously in section 3.7.2 and results plotted at the two different distances, 12Å and 15Å as heat maps of interactions.

The interaction interface between the two helices in the averaged 12Å cutoff with interhelical distances of about ~8Å in the TM4-TM4 homodimer ensemble were between the residues found in the I125$^{4.36}$xxCWxxS132$^{4.53}$ motif on one helix with the residues found in the W129$^{4.50}$xxxxA134$^{4.55}$ motif on the other (see Figure 5.6a). The TM4-TM6 and TM4-TM7 pairwise combination simulations, both sets showed an interhelical distance of $\geq$ 10Å between the two helices. In the TM4-TM6, the interaction with the shortest distance was found between: 5) L247$^{6.49}$ with F133$^{4.54}$ (see Figure 5.6b). In the TM4-TM7 simulations, the interactions were found between: 6) residues found in the

N280$^{7.45}$SxVN284$^{7.49}$ motif in TM7 with W129$^{4.50}$ on TM4 and 7) the residues in the S277$^{7.42}$HTNS281$^{7.46}$ motif of TM7 with S132$^{4.53}$ on TM4 (see Figure 5.6c).

In the TM4 pairwise ensemble sets, the interactions took place at the same position within the TM4 helix indicating that a defined orientation is needed to establish a specific interaction in TM4 with other TM helices. The most prominently interacting residues in TM4 are: W129$^{4.50}$, S132$^{4.53}$ and F133$^{4.54}$ that are found at the lower third of TM4 helix from the N-terminal side.

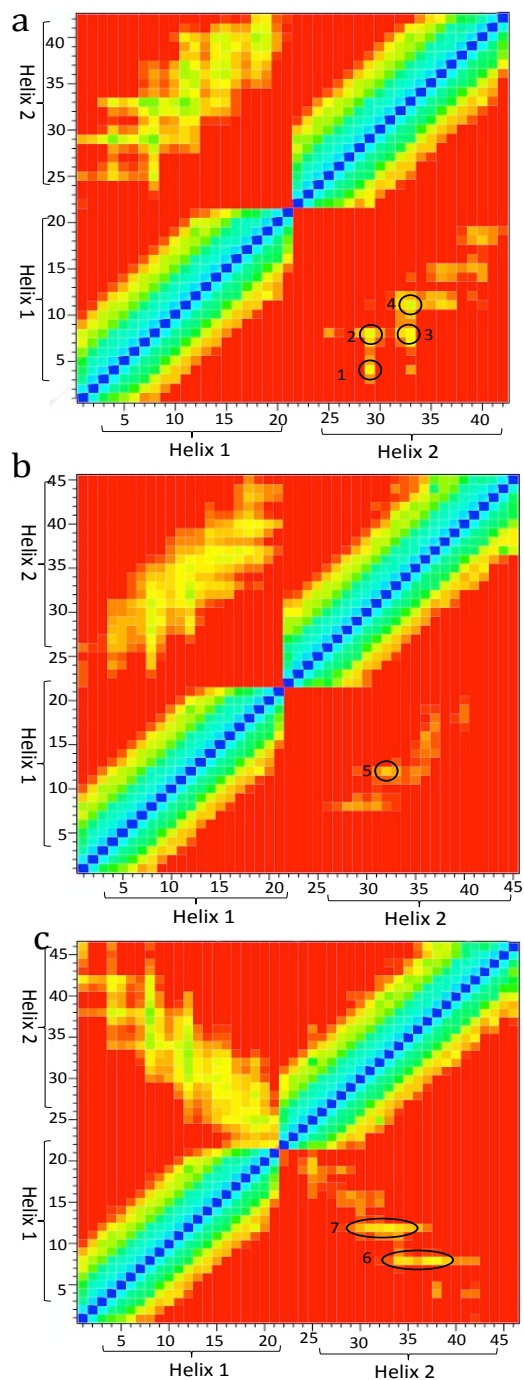**Figure 5.6 Contact matrices showing specific pairwise interactions between TM4 and an A$_{2A}$ helix**

Contact matrices (heat maps) between two A$_{2A}$ helices showing specific interactions between TM4 (Helix 1) and TM4 (Helix 2) (a), between TM4 (Helix 1) and TM6 (Helix 2) (b) and between TM4 (Helix 1) and TM7 (Helix 2) (c). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts.

5.7 Interacting interfaces of the TM6 pairwise ensembles

The remaining two possible pairwise simulation for sets of TM6 were tested they included: 1) TM6-TM6 homodimer and 3) TM6-TM7 heterodimer. Following the application of the 10Å cut-off, in the TM6-TM6 ensemble, stable dimers were observed in only 4% of simulations. In the TM6-TM7 simulations, 12% of simulations were identified as interacting helices and forming stable dimers.

After identifying the interacting replicas in all sets, they had their trajectories combined and the contact interface residues were identified by measuring the average interhelical contact distance between the two helices as explained previously in section 3.7.2 and results plotted at the two different distances, 12Å and 15Å as heat maps of interactions. The interaction interface between the two helices in both ensemble sets at the 12Å cutoff was about ~10Å for both the TM6-TM6 and TM6-TM7 helix-helix interacting simulations. The most dominant interactions present between TM6-TM6 helices were between residues: 1) F242$^{6.44}$ with F242$^{6.44}$; 2) A243$^{6.45}$ and W246$^{6.48}$ on the other and 3) W246$^{6.48}$ with C245$^{6.47}$, W246$^{6.48}$ and L249$^{6.51}$ (see Figure 5.7a). In the TM6-TM7, the prominent interaction was found between: 4) F242$^{6.44}$ with S277$^{7.42}$ and S281$^{7.46}$ and 5) W246$^{6.48}$ and S277$^{7.42}$ (see Figure 5.7b).

In the TM6 pairwise ensemble sets, the interactions took place at the same position within the TM6 helix indicating that a defined orientation is needed to establish a specific interaction in TM6 with other TM helices. The most prominently interacting residues in TM6 are: F242$^{6.44}$, W246$^{6.48}$ and

H250$^{6.52}$ that are found at the middle of TM6 helix.

5.8 Interacting interfaces of the TM7 pairwise ensembles

The remaining homodimer pairwise ensemble set of the TM7-TM7 was tested. Following the application of the 10Å cut-off to the ensemble, stable dimers were detected in only 10% of simulations.

After identifying the interacting replicas in the set, their trajectories were combined and the contact interface residues were identified by measuring the average interhelical contact distance between the two helices as explained previously in section 3.7.2 and the result plotted at the two different distances, 12Å and 15Å as heat maps of interactions. The interaction interface between the two helices in the ensemble set at the 12Å cutoff exhibited weak interactions with interhelical distances of about $\geq$ 10Å between the two helices. From that, the most dominant interaction between the two helices was present between the S277$^{7.42}$ residue on one helix with the H278$^{7.43}$ residue on the other helix (Figure 5.8).

In all the TM7 pairwise ensemble sets, the interactions took place at the same position within the TM6 helix indicating that a defined orientation is needed to establish a specific interaction in TM6 with other TM helices. The most prominently interacting residues in TM6 are: S277$^{7.42}$, H278$^{7.43}$, N280$^{7.45}$ and S281$^{7.46}$ that are found at the middle of TM7 helix.

Figure 5.7 Contact matrices showing specific pairwise interactions between TM6 and an A$_{2A}$ helix

Contact matrices (heat maps) between two A$_{2A}$ helices showing specific interactions between TM6 (Helix 1) and TM6 (Helix 2) (a) and between TM6 (Helix 1) and TM7 (Helix 2). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts.
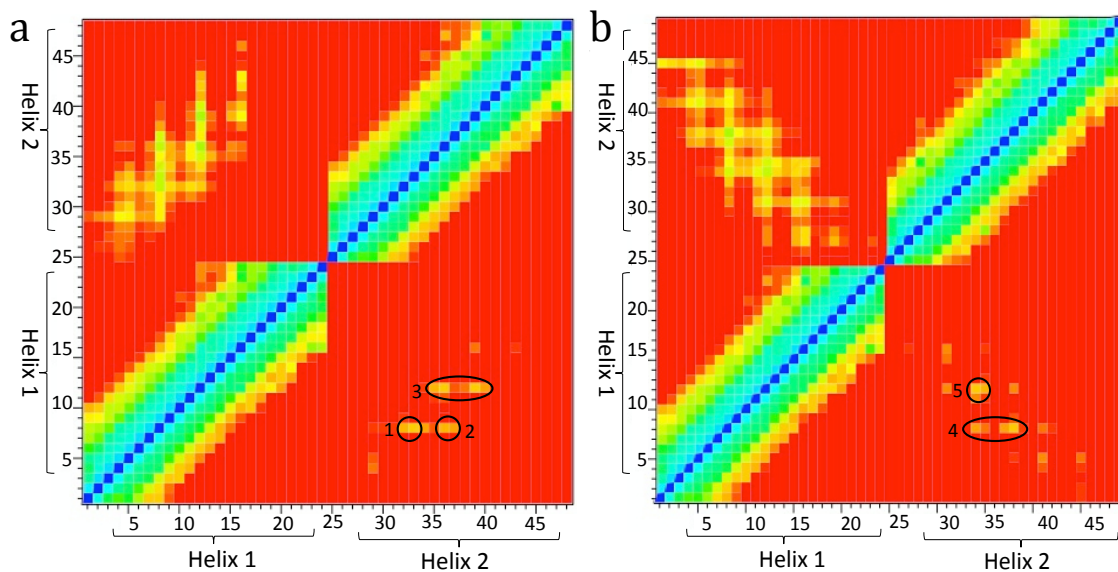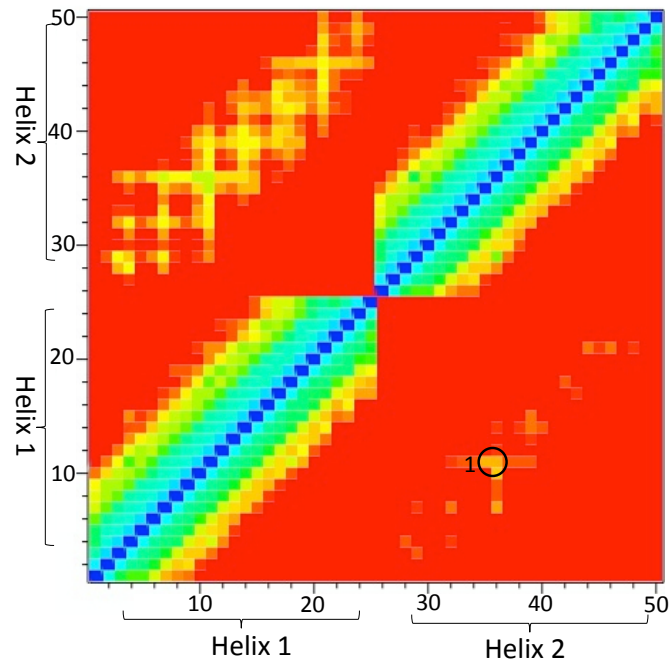
Figure 5.8 Contact matrices showing specific pairwise homodimer interactions between two TM7 helices

Contact matrices (heat maps) between two A$_{2A}$ helices showing specific interactions between TM7 (Helix 1) and TM7 (Helix 2). The colour scale is as indicated in Figure 3.4. Circles indicate areas with key interhelical contacts.

5.9 Summary

The aim of this chapter was to perform a comprehensive analysis *in silico* of every possible pairwise interaction between the seven different TM helices of the human A$_{2A}$ receptor. In all the A$_{2A}$ TM pairwise ensembles, the helices appeared to diffuse randomly within the lipid bilayer in the simulations sets and their positions along the bilayer were observed to be normal for both helices in the pair.

The analysis of the ensemble sets demonstrates that there are 23 out of 28 possible pairs of helices that formed dimers with varying interhelical distances. Not all the pairs of the ensemble sets showed a favor to form homo- or hetero-dimers. Stable contacts with an interhelical distance of ~8Å occurred in five pairs, namely: TM1-TM2, TM5-TM1, TM1-TM6 and TM4-TM4. Eight pairs establish a distance in the range of ~8.5 to 9Å, namely: TM1-TM3, TM1-TM4, TM3-TM3, TM3-TM4, TM3-TM6, TM5-TM3, TM5-TM4, and TM6-TM5. Seven pairs showed interhelical distance of ≥ 10Å, namely: TM1-TM7, TM2-TM4, TM2-TM6, TM4-TM6, TM4-TM7, TM5-TM7 and TM7-TM7. One pair, the TM5-TM2, displayed an interhelical distance of mare than 10Å. The rest of the pairs exhibited no interactions between them, Table 5.1. These interhelical distances occur within the range that was found in between the bacteriorhodopsin helices, the 7.8-10.6Å interhelical distances range[175] it seems that these interactions are formed C$_\alpha$-H⋯O bonds as they are above the approximate limit for backbone-to-backbone C$_\alpha$-H⋯O bonds formation which is 7.6Å. The occurrence of C$_\alpha$-H⋯O contacts between backbone-to-side-chain as a function

of the interhelical axial distance are between 6 and 12Å with an average of 8.9Å[175].

Two of the transmembrane domains (TM1 and TM5) displayed the highest affinity to form dimers while TM2 exhibited the lowest tendency to form dimers, which may be due to its high hydrophobicity. Notably, a greater number of interacting pairs have been identified than were identified by through FRET and CD spectroscopy[147]. This may be due to the slight difference in TM sequences used in their study and the sequences used here or it may be due to the fact that they have stated that in SDS micelles only the strongest helix-helix interactions can be detected through FRET.

Helices	TM1	TM2	TM3	TM4	TM5	TM6	TM7
TM1	None	~ 8 Å Interaction	8.5 to 9Å Interaction	8.5 to 9 Å Interaction	~ 8Å Interaction	~ 8Å Interaction	≥ 10Å Interaction
TM2		None	None	≥ 10Å Interaction	> 10Å Interaction	≥ 10Å Interaction	None
TM3			8.5 to 9Å Interaction	8.5 to 9Å Interaction	8.5 to 9Å Interaction	8.5 to 9Å Interaction	None
TM4				~8Å Interaction	8.5 to 9Å Interaction	≥ 10Å Interaction	≥ 10Å Interaction
TM5					~ 8 Å Interaction	8.5 to 9Å Interaction	≥ 10Å Interaction
TM6						~10Å Interaction	~10Å Interaction
TM7							≥ 10Å Interaction

Table 5.1 Summary of the TM helix-helix interactions for the 28 studied helix pairwise combinations *in silico*.

The interacting interfacial residues between the helices is usually found in pairs of small residues spaced at i, $i + 4$ which are present on all helices, it is these small residues that have an increased occurrence in transmembrane helices and are usually involved in the interhelical C_α-H·····O bonds formation[175]. The interactions interface we found between the transmembranes with the highest affinity to form dimers, at the ~8Å interhelical distance, is between the following motifs: 1) In TM1-TM2 dimer: E13$^{1.39}$xxIAxLAxxG23$^{1.49}$ motif that is mapped on the exterior on the TM1 helix with F44$^{2.42}$xVSxxAAxxAVG56$^{2.54}$ motif that is mapped to the exterior of the TM2 helix; 2) In TM5-TM1 dimer: A20$^{1.46}$xxAIxxNxxxC31$^{1.54}$ motif that is mapped on the interior of TM1 with P189$^{5.50}$xxxMxxVYx R199$^{5.60}$ motif that is mapped on the interior of TM5; 3) TM1-TM6 dimer: L14$^{1.49}$xIxxLAI21$^{1.47}$ motif that is mapped on the interior of TM1 with F242$^{6.43}$xxx W246$^{6.47}$ motif that is mapped on the interior of TM6 and 4) TM4-TM4 dimer the interactions were found within the I125$^{4.36}$xxCWxxS132$^{4.53}$ motif that is mapped on the interior of one helix with the residues W129$^{4.50}$xxxxA134$^{4.55}$ found on the exterior of the other helix.

The C_α-H·····O bonds formed between the dimers are more likely to be in the form of backbone-to-side-chain rather than backbone-to-backbone as the interhelical distance between the two helices is above the limit of the backbone-to-backbone interactions another reason is that the interacting interfacial residues in dimers do not consists of glycine, serine or threonine which is another feature observed in the backbone-to-backbone interaction C_α-H·····O=C[175]. The mapping of the interaction motifs reveals that some of these residues were on the inner side of the helix and would only come in contact with each other if the

receptor would open-up like in the suggest second model of dimerisation where the hinge loop of the dimer opens, causing the helical domains to exchange and produce a domain-swapped dimer.

Furthermore, from the simulations the involvement of serine and threonine residues in the interaction interface in a number of dimers formed, for example in the TM1-TM7 and in the TM4-TM7 dimers were observed at a longer interhelical distance $\geq$ 10Å, this is consistent with the other findings of that serine and threonine residues are the frequently found residues in the interhelical C_{α}-H⋯⋯O$_{\gamma}$ bonds, it has been establish that they each constitute 5% of the amino acid composition of transmembrane helices because of the donor potential of their polar side chains that can be satisfied by forming O-H⋯⋯O hydrogen bonds to the carbonyl at i - 4 or i - 3 on the same helix[175, 182].

Additionally, a number of interactions between the formed stable dimers in the simulations involved the β-branched amino acids (either isoleucine, valine and leucine) on one helix with a small residue (either glycine, alanine and serine) on the other, this is coherent with the established finding of Senes *et al*, they have validated that the β-branched amino acids (isoleucine and valine) at the *i, i* + 4 position, and to a less extent leucine, form pairs with the small residues (glycine, alanine and serine) at the position *i, i* + 1 and *i, i* + 2[182].

The eCG-MD results from the mutated TM1-E13[1.39]A-TM2 ensemble set may provide a molecular explanation for the cause that alters the monomer:dimer ratio at a level of detail that could not be determined biophysically and would require structural biology studies to confirm experimentally.

Chapter 6
Conclusion and Future Work

6.1 Conclusions

In this project, a method of ensemble-based coarse-grained classical molecular dynamics (eCG-MD) that would be used to predict protein-protein interactions between TM helices of dimeric GPCRs was developed and assessed. The method was applied and tested in two parts:

1) Four different cases of homodimeric GPCRs, namely the A_{2A} receptor, rhodopsin, CXCR4 and β_1AR receptors for which experimental data exist were used to enable a comparison of the eCG-MD results with published experimental data. In each of these cases, the ensemble-based CG-MD methodology provides a reproducible measurement of the distance between interacting helices that corresponds well with the experimental data and is within the range of distances at which protein-protein interactions occur was identified.

2) A comprehensive study of the seven TM of the A_{2A}R helices was conducted. There were 21 possible heterologous pairs, seven possible homo-interactions. All possible combinations helix-helix interactions were examined in the 26 different ensemble sets to identify novel interacting residues within the A_{2A} receptor TM helices and a comparison was conducted to identify whether these results match the experimental data of the self-association of the helices of

157

the A_{2A} adenosine receptor. The final mean distance between the two helices in every ensemble composed of 50 replicas in each set was used to identify the specific interactions between the helices in all of the ensemble sets.

In the first part of the work presented in this thesis, the first case included the A_{2A} adenosine receptor that had been experimentally shown to form homodimeric receptors through interactions between the TM5 helices of the two protomers. The results identified specific interactions involving the PxxxM motif of TM5 and, specifically, at the M193$^{5.54}$ residue within that motif. The method accurately identified residues shown experimentally to be involved in TM5 homodimerisation. In parallel with work done experimentally, the role of M193$^{5.54}$ was investigated by two types of mutations the M193I$^{5.54}$ and the M193$^{5.54}$A. Characterizing the M193I$^{5.54}$ mutation had no effect on the residues involved in contact interface, they were identical to those found in the wild-type but included I193$^{5.54}$ in the interaction, despite the loss of the methionine at position 193. Characterizing the M193$^{5.54}$A mutation, it was identified that the contact interface of the helices was completely changed and that the key interacting residues identified in the wild-type conformation had moved to a new position, preventing formation of TM5 homodimers. The results also provide a molecular explanation for the experimental finding that the M193$^{5.54}$A mutation alters the monomer:dimer ratio at a level of detail that could not be determined biophysically and would require structural biology studies to confirm experimentally. The most prominently occurring interactions identified in the wild-type TM5-TM5 ensemble involved the Y197$^{5.58}$ residue. The role of the Y197$^{5.58}$ was investigated by mutating the Y197$^{5.58}$ into alanine and

phenylalanine. Neither the $Y197^{5.58}A$ nor the $Y197^{5.58}F$ mutations had any effect on the way in which the two helices came together, suggesting that the $Y197^{5.58}$ may not be a key residue in the specificity of TM5-TM5 dimer formation but rather only a part of the interacting motif in the wild-type homodimer.

The second case examined was that of the rhodopsin dimer, for which crystallographic data had identified contact interfaces between TM1 and TM2 and between TM4 and TM5. Ensemble CG-MD confirmed dimerisation and the identification of specific interactions within each of these heterologous TM pairs. There is a striking convergence between the distances predicted computationally and those calculated from 1N3M, particularly for specific interactions between TMs 1 and 2, showing that the eCG-MD method is able to provide accurate and precise predictions in agreement with experimental findings. The method was also able to identify novel interfaces as seen in the third (CXCR4) and fourth (β_1AR) cases studied, where a novel interface in CXCR4 between TM5 and TM5 and a novel interface in β_1AR between TM1 and TM1 were identified, in addition to confirming the previously identified contact interface between TM4 and TM5 in the β_1AR. The β_1AR has been shown to form transient interactions whereas the β_2 adrenergic receptor can form stable oligomers[181]. The ability to detect a stable dimer of TM1-TM1 in the β_1AR shows the value of ensemble-based simulations for the identification of transient interactions.

In all of the cases studied, there appears to be a pattern emerging of the nature and location of the contact interfaces. It was observed as either a single interface, at TM5 in both the A_{2A} and the CXCR4, or of two contact interfaces,

as seen in rhodopsin and the β_1AR, one of which involves TM1 and the other between TM4 and TM5. Interestingly, interactions in TM5 are observed in all the of cases studied. As more dimeric GPCR crystal structures with corresponding biophysical and functional data become available, the conservation of this novel pattern will become clearer.

In the second part of the thesis work, the comprehensive characterisation of the seven TM of the A_{2A}R helices, eCG-MD of the pairwise ensemble sets revealed clearly that not all the pairs of the ensemble sets have a tendency to form homo- or heterodimers. There were 23 out of the 28 possible pairs of helices that formed dimers at varying interhelical distances and with different facing contact interfaces. The interactions with shortest distances appear to be between backbone-to-side-chain of the helices. Interestingly, the two helices displaying the highest affinity to form dimers were TM1 and TM5 of the A_{2A} receptor, both identified in the emerging pattern of the nature and location of the contact interfaces in GPCRs mentioned previously, whilst TM2 exhibited the lowest affinity to form dimers.

The method also showed that the different interfacial positioning of the interacting motifs can be recognized that can be either intradimeric contacts with exterior facing contact interfaces as seen between TM1-TM2 that can be found between two interacting receptors or interhelical interactions found internally within a receptor to stabilize the conformation of the structure such interactions interface were seen in the TM1-TM5 and TM1-TM6 dimers. Although the absence of the loops from the model system used in this method, it seems that it

is not detriment to the aim of study, as the ECL-1 in the A_{2A} receptor between TM2 and TM3 is very short no interaction between the two helices was detected. In the case of TM5 and TM6, the ICL-3 is fairly long, interaction between these two helices was detected by the method as well as it was identified in the literature[147]. Thus it can be considered that the loops are not absolutely required for bringing the helices together nor are they required for holding them together (as suggested by other studies) although the fact that they may contribute to thermodynamic stability and are commonly required for the correct folding and function of integral membrane proteins and assist in the intradimeric contact is important to note[28].

The results from the eCG-MD method unequivocally demonstrate that sufficient conformational sampling is required in coarse-grained MD to obtain reproducible and reliable results. In the simulations preformed, several of the replicas within the ensemble failed to show any interactions and that a number of others began to interact late in the simulation at a point when accurate estimates of distance could no longer be achieved. A single trajectory simulation, particularly if either of these circumstances were to occur, would give inaccurate and potentially misleading results. Indeed, ensembles are required to obtain accurate and precise results. Error analysis was used to determine appropriate choices for ensemble size and run length. For ensemble size, it was observed that the rate of change in the standard deviation of the mean distance between helices decreased with increasing replica size and found that approximately 30 replicas was sufficient per ensemble to obtain reproducible results. For run length, it was observed that the rate of increase in the standard deviation of the mean distance

between helices increased with increasing run length, but that the rate of increase slowed substantially after approximately 300 ns. Interestingly, the negative control, the TM2-TM2 ensemble, included in the eCG-MD simulations showed no variation in the standard deviation of the mean distance between helices as a function of run length and a low standard deviation with a very rapid decrease to a constant value at an ensemble size of ~15 replicas. This behavior was notably different from simulations in which interactions were identified and provides a means of confirming the absence of interaction

In conclusion, a systematic, reproducible and reliable computational protocol for determining the specific points of interaction between GPCR dimers, eCG-MD, was developed and validated using experimental data. The method discriminates between residues in TM helices that form specific interactions and residues that are in close proximity but do not interact. The work here extends the recent findings of ensemble-based fully atomistic MD studies, which have shown that an ensemble-based approach is required to generate predictions of protein properties that correlate well with experimental data.

Lastly, the use of eCG-MD simulations has provided a valuable tool to the time and length scales of systems used compared to what is possible with traditional all-atom models. This is easily explained by the required turnaround times of days, week, or up to months. Our tests showed that CG-simulations (one ensemble) run on 16 CPU cores on Legion for 500 ns completed within approximately 150 hours (an actual week). The same CG-simulations (one ensemble) run on 16 CPU cores on Grace for 500 ns completed within

approximately 72 hours (3 days). That is a much shorter run time than all-atom MD simulations, where it was reported that a 6 ns replica of an all-atom MD simulation was produced within 15 hours (using 64 CPU cores)[183] on a much smaller system.

6.2 Future work

The main objective behind the work of this thesis was to develop a method that can identify or predict the helix-helix interactions of GPCRs transmembrane domains. The method provided an ability to predict contact interfaces computationally and the error analysis enabled the identification of non-interacting regions in the A_{2A} receptor. This method can enhanced by providing and writing an automated pipeline software for performing high throughput eCG-MD simulations of α-helical peptides in lipid bilayer membranes to facilitate the running and analysis of such simulations ensembles.

Running the A_{2A} TM5, TM2 wild-types and mutated M193$^{5.54}$ ensembles sets for a longer time from 500 ns to μs would constitute a logical follow-up on these experiments to see if longer times might change the results observed in the types and location of the interactions involved between the helices.

Furthermore, running simulations with an intact A_{2A} receptor could be implemented to help in further understanding of the types of interactions produced in the pairwise comprehensive analysis of TM helices simulations

performed. Ensembles performed with two or more intact A_{2A} receptors preformed *in silico* would provide a means of confirming whether the proposed interfaces identified from interacting TM helices, as performed in this study, form part of the interface. The confirmation that computational simulations run with intact receptor proteins, rather than with isolated receptor TM domains, would enable the development of methodologies to rapidly identify interacting GPCR partners and contact interfaces at greatly reduced computational cost, compared with running much larger simulations with intact receptors and could potentially offer insight into the means by which GPCR protomers form homodimers.

As described in Chapter 3, the unsuccessful attempts to access the dimerisation interface experimentally and the effect of mutation on TM5 M193$^{5.54}$ by the BiFC fluorescent protein and the measurement of the cAMP levels due to problematic constructs and the difficulty in determining the cell number can be continued by first establishing a stably transfected cell line with the wild-type and mutated A_{2A} receptor followed by the functional analysis.

Additionally, it is of great interests to take forward the novel findings of the eCG-MD work and investigate experimentally the residue identified on the exterior side of TM1 part of the interacting interface between TM1 and TM2, This residue, E13$^{1.39}$, when mutated computationally, abolished the contact interface between TM1 and TM2, which is the proposed interface between two A_{2A} protomers. A number of different means of characterizing the effects of E13$^{1.39}$ could be employed, including peptide studies, co-immunoprecipitation,

FRET and other functional experiments.

In summary, the eCG-MD methodology developed here is of great utility in further understanding GPCR function and also has broad applicability to many different types of membrane proteins, including receptor tyrosine kinases, ion channels, transporters and oligomeric complexes of various combinations of these.

In addition, in chapter 5, the comprehensive simulations between the seven different TM helices of the human A_{2A} receptor produced several pairwise interactions. Identification if any of these amino acid residues are part of conserved motif is an important next step and also to distinguish whether they are part of sequence or structural motifs. This could be verified by one of the covariation analysis methods[184] that can detect conserved residues involved in dimer interfaces; including: mutual information or statistical coupling analysis. This can be followed by multiple sequence alignment to verify if the identified amino acid were conserved. Some of the bioinformatical tools used in multiple sequence alignment include: Clustal Omega; Kalign; MUSCLE; and T-Coffee among others.

Finally, there are a number of different approaches available that aim to identify or predict the protein-protein interaction interfaces. In this thesis, we have developed an ensemble-based coarse-grained molecular dynamics method that was used to predict the protein-protein interactions between TM helices of dimeric GPCRs. It was used to investigate the helix-helix dimerisation interface

of four types of interacting GPCRs. One of the investigated GPCRs, A_{2A}, has no dimeric structure available in the PDB until now. So the results obtained through this method for the A_{2A} receptor were assessed by comparing its results to the results obtained from the other three types of GPCRs that had dimeric structures in the PDB. From the method, the measurement of the distance between interacting helices corresponded well with the experimental data and is within the range of distances at which protein-protein interactions occur. As more dimeric GPCR crystal structures become available, their interacting interfaces can be assed by the eCG-MD method and then compare the results with the results obtained from the CAPRI method. CAPRI is a blind protein-protein prediction method, which uses the 3D structure of a protein or atomistic coordinates to predict the form or ability of two proteins to associate using docking algorithms. CAPRI prediction methods are a useful complement to experimental data. It uses X-ray or NMR protein structure of protein-protein complexes[185].

Chapter 7
Bibliography

1.   Abbracchio, M.P., et al., *International Union of Pharmacology LVIII: update on the P2Y G protein-coupled nucleotide receptors: from molecular mechanisms and pathophysiology to therapy.* Pharmacol Rev, 2006. **58**(3): p. 281-341.

2.   Shonberg, J., et al., *GPCR crystal structures: Medicinal chemistry in the pocket.* Bioorg. Med. Chem., 2015. **23**: p. 3880–3906.

3.   Foord, S.M., S. Jupe, and J. Holbrook, *Bioinformatics and type II G-protein-coupled receptors.* Biochem Soc Trans, 2002. **30**(4): p. 473-9.

4.   Palczewski, K., et al., *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science, 2000. **289**(5480): p. 739-45.

5.   Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-65.

6.   Rasmussen, S.G., et al., *Crystal structure of the human beta2 adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-7.

7.   Gomperts, B.D., Kramer, I.M., Tatham, P.E.R. , *Signal transduction.* Elsevier Inc., 2009(2nd Edition).

8.   Hill, S.J., *G-protein-coupled receptors: past, present and future.* Br J Pharmacol, 2006. **147 Suppl 1**: p. S27-37.

9.   Latek, D., et al., *G protein-coupled receptors--recent advances.* Acta Biochim Pol, 2012. **59**(4): p. 515-29.

10.   Katritch, V., V. Cherezov, and R.C. Stevens, *Diversity and modularity of G protein-coupled receptor structures.* Trends Pharmacol Sci, 2012. **33**(1): p. 17-27.

11.   Rasmussen, S.G., et al., *Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor.* Nature, 2011. **469**(7329): p. 175-80.

12.   Costanzi, S., et al., *Architecture of P2Y nucleotide receptors: structural comparison based on sequence analysis, mutagenesis, and homology modeling.* J Med Chem, 2004. **47**(22): p. 5393-404.

13. Sciences, T.R.S.A.o., *Scientific Background on the Nobel Prize in Chemistry 2012. Studies of G-protein coupled receptors*

14. Dupré, D.J.H.T.E., Jockers R., *GPCR signaling complexes- Synthesis, Assembly, Trafficking and Specificity.* Subcellular Biochemistry, 2012. **63**.

15. Marks, F., U. Klingmüller, and K. Müller-Decker, *Cellular signal processing: An introduction to the molecular mechanisms of signal transduction.* . Garland Science, Taylor and Francis Group, 2009.

16. Jones, K.A., et al., *GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2.* Nature, 1998. **396**(6712): p. 674-679.

17. Jordan, B.A. and L.A. Devi, *G-protein-coupled receptor heterodimerization modulates receptor function.* Nature, 1999. **399**(6737): p. 697-700.

18. Fotiadisa, D., et al., *The G protein coupled receptor rhodopsin in the native membrane.* FEBS Lett., 2004. **564**(3): p. 281–288.

19. Agnati, L.F., et al., *On the expanding terminology in the GPCR field: the meaning of receptor mosaics and receptor heteromers.* J Recept Signal Transduct Res, 2010. **30**(5): p. 287-303.

20. Fuxe, K., et al., *The changing world of G protein-coupled receptors: from monomers to dimers and receptor mosaics with allosteric receptor-receptor interactions.* J Recept Signal Transduct Res, 2010. **30**(5): p. 272-83.

21. Hillion, J., et al., *Coaggregation, cointernalization, and codesensitization of adenosine A2A receptors and dopamine D2 receptors.* J Biol Chem, 2002. **277**(20): p. 18091-7.

22. Canals, M., et al., *Adenosine A2A-dopamine D2 receptor-receptor heteromerization: qualitative and quantitative assessment by fluorescence and bioluminescence energy transfer.* J Biol Chem, 2003. **278**(47): p. 46741-9.

23. Kearn, C.S., et al., *Concurrent stimulation of cannabinoid CB1 and dopamine D2 receptors enhances heterodimer formation: a mechanism for receptor cross-talk?* Mol Pharmacol, 2005. **67**(5): p. 1697-704.

24. Marcellino, D., et al., *Antagonistic cannabinoid CB1/dopamine D2 receptor interactions in striatal CB1/D2 heteromers. A combined neurochemical and behavioral analysis.* Neuropharmacology, 2008. **54**(5): p. 815-23.

25. Baragli, A., et al., *Heterooligomerization of human dopamine receptor 2 and somatostatin receptor 2 Co-immunoprecipitation and fluorescence resonance energy transfer analysis.* Cell Signal, 2007. **19**(11): p. 2304-16.

26. Scarselli, M., et al., *D2/D3 dopamine receptor heterodimers exhibit unique functional properties.* J Biol Chem, 2001. **276**(32): p. 30308-14.

27. Rocheville, M., et al., *Receptors for dopamine and somatostatin: formation of hetero-oligomers with enhanced functional activity.* Science, 2000. **288**(5463): p. 154-7.

28. Liang, Y., et al., *Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes.* J Biol Chem, 2003. **278**(24): p. 21655-62.

29. Fotiadis, D., et al., *Atomic-force microscopy: Rhodopsin dimers in native disc membranes.* Nature, 2003. **421**(6919): p. 127-8.

30. Wertman, J. and D.J. Dupre, *G protein-coupled receptor dimers: look like their parents, but act like teenagers!* J Recept Signal Transduct Res, 2013. **33**(3): p. 135-8.

31. Rocheville, M., et al., *Subtypes of the somatostatin receptor assemble as functional homo- and heterodimers.* Journal of Biological Chemistry, 2000. **275**(11): p. 7862-7869.

32. Smith, N.J. and G. Milligan, *Allostery at G protein-coupled receptor homo- and heteromers: uncharted pharmacological landscapes.* Pharmacol Rev, 2010. **62**(4): p. 701-25.

33. Salahpour, A., et al., *Homodimerization of the beta2-adrenergic receptor as a prerequisite for cell surface targeting.* J Biol Chem, 2004. **279**(32): p. 33390-7.

34. Bulenger, S., S. Marullo, and M. Bouvier, *Emerging role of homo- and heterodimerization in G-proteincoupled receptor biosynthesis and maturation.* TRENDS in Pharmacological Sciences, 2005. **26**(3).

35. Robbins, M.J., et al., *Characterization of the dimerization of metabotrobic glutamate receptors using an N-terminal truncation of mGluR1a.* J Neurochem, 1999. **72**: p. 2539-47.

36. Rios, C.D., et al., *G-protein-coupled receptor dimerization: modulation of receptor function.* Pharmacology & Therapeutics, 2001. **92**(2-3): p. 71-87.

37. George, S.R., B.F. O'Dowd, and S.P. Lee, *G-protein-coupled receptor oligomerization and its potential for drug discovery.* Nat Rev Drug Discov, 2002. **1**(10): p. 808-20.

38. Breitwieser, G.E., *G protein-coupled receptor oligomerization: implications for G protein activation and cell signaling.* Circ Res, 2004. **94**(1): p. 17-27.

39. Bai, M., *Dimerization of G-protein-coupled receptors: roles in signal transduction.* Cellular Signalling 2004. **16**: p. 175-186.

40. Vila-Coro, A.J., et al., *HIV-1 infection through the CCR5 receptor is blocked by receptor dimerization.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(7): p. 3388-3393.

41. Issafras, H., et al., *Constitutive agonist-independent CCR5 oligomerization and antibody-mediated clustering occurring at physiological levels of receptors.* Journal of Biological Chemistry, 2002. **277**(38): p. 34666-34673.

42. Cvejic, S. and L.A. Devi, *Dimerization of the delta opioid receptor: implication for a role in receptor internalization.* J Biol Chem, 1997. **272**(43): p. 26959-64.

43. Fukushima, Y., et al., *Oligomer formation of histamine H2 receptors expressed in Sf9 and COS7 cells.* FEBS Lett, 1997. **409**(2): p. 283-6.

44. Fotiadis, D., et al., *Structure of the rhodopsin dimer: a working model for G-protein-coupled receptors.* Curr Opin Struct Biol, 2006. **16**(2): p. 252-9.

45. Maurel, D., et al., *Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization.* Nat Methods, 2008. **5**(6): p. 561-7.

46.   Rashid, A.J., et al., *D1-D2 dopamine receptor heterooligomers with unique pharmacology are coupled to rapid activation of Gq/11 in the striatum.* Proc Natl Acad Sci U S A, 2007. **104**(2): p. 654-9.

47.   Lee, S.P., et al., *Dopamine D1 and D2 receptor Co-activation generates a novel phospholipase C-mediated calcium signal.* J Biol Chem, 2004. **279**(34): p. 35671-8.

48.   Pei, L., et al., *Uncoupling the dopamine D1-D2 receptor complex exerts antidepressant-like effects.* Nat Med, 2010. **16**(12): p. 1393-5.

49.   O'Dowd, B.F., et al., *Separation and reformation of cell surface dopamine receptor oligomers visualized in cells.* Eur J Pharmacol, 2011. **658**(2-3): p. 74-83.

50.   Marcellino, D., et al., *Identification of dopamine D1-D3 receptor heteromers. Indications for a role of synergistic D1-D3 receptor interactions in the striatum.* J Biol Chem, 2008. **283**(38): p. 26016-25.

51.   Ciruela, F., et al., *Presynaptic control of striatal glutamatergic neurotransmission by adenosine A1-A2A receptor heteromers.* Journal of Neurochemistry, 2007. **102**: p. 97-97.

52.   Ecke, D., et al., *Hetero-oligomerization of the P2Y11 receptor with the P2Y1 receptor controls the internalization and ligand selectivity of the P2Y11 receptor.* Biochem J, 2008. **409**(1): p. 107-16.

53.   Diaz-Cabiale, Z., et al., *Adenosine A2A agonist CGS 21680 decreases the affinity of dopamine D2 receptors for dopamine in human striatum.* Neuroreport, 2001. **12**(9): p. 1831-4.

54.   Navarro, G., et al., *Detection of heteromers formed by cannabinoid CB1, dopamine D2, and adenosine A2A G-protein-coupled receptors by combining bimolecular fluorescence complementation and bioluminescence energy transfer.* ScientificWorldJournal, 2008. **8**: p. 1088-97.

55.   Carriba, P., et al., *Detection of heteromerization of more than two proteins by sequential BRET-FRET.* Nat Methods, 2008. **5**(8): p. 727-33.

56.   Navarro, G., et al., *Detection of receptor heteromers involving dopamine receptors by the sequential BRET-FRET technology.* Methods Mol Biol, 2013. **964**: p. 95-105.

57. Carriba, P., et al., *Striatal adenosine A2A and cannabinoid CB1 receptors form functional heteromeric complexes that mediate the motor effects of cannabinoids.* Neuropsychopharmacology, 2007. **32**(11): p. 2249-59.

58. Ferre, S., et al., *Synergistic interaction between adenosine A2A and glutamate mGlu5 receptors: Implications for striatal neuronal function.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(18): p. 11940-11945.

59. Cabello, N., et al., *Metabotropic glutamate type 5, dopamine D2 and adenosine A2a receptors form higher-order oligomers in living cells.* J Neurochem, 2009. **109**(5): p. 1497-507.

60. Gonzalez-Maeso, J., et al., *Identification of a serotonin/glutamate receptor complex implicated in psychosis.* Nature, 2008. **452**(7183): p. 93-7.

61. Yoshioka, K., et al., *Hetero-oligomerization of adenosine A1 receptors with P2Y1 receptors in rat brains.* FEBS Lett, 2002. **531**(2): p. 299-303.

62. Yoshioka, K., O. Saitoh, and H. Nakata, *Heteromeric association creates a P2Y-like adenosine receptor.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7617-22.

63. Nakata, H., et al., *Functions of heteromeric association between adenosine and P2Y receptors.* J Mol Neurosci, 2005. **26**(2-3): p. 233-8.

64. Thevenin, D., et al., *Oligomerization of the fifth transmembrane domain from the adenosine A2A receptor.* Protein Sci, 2005. **14**(8): p. 2177-86.

65. Hebert, T.E., et al., *A Peptide Derived from a beta2-Adrenergic Receptor Transmembrane Domain Inhibits Both Receptor Dimerization and Activation.* J Biol Chem 1996. **277**: p. 44925 – 31.

66. Mercier, J.F., et al., *Quantiative assessment of beta1- and beta2-adrenergic receptor homo- and heterodimerization by bioluminescence resonance energy transfer.* J Biol Chem, 2002. **277**: p. 44925 – 31.

67. Angers, S., et al., *Detection of beta2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET).* Proc Natl Acad Sci 2000. **97**: p. 3684 – 9.

68. Angers, S., A. Salahpour, and M. Bouvier, *Biochemical and biophysical demonstration of GPCR oligomerization in mammalian cells.* Life Sci, 2001. **68**(19-20): p. 2243-50.

69. Hebert, T.E., et al., *Functional rescue of a constitutively desensitized β2AR through receptor dimerization.* Biochem J, 1998. **330**: p. 287-93.

70. Blanpain, C., et al., *Multiple active states and oligomerization of CCR5 revealed by functional properties of monoclonal antibodies.* Mol Biol Cell, 2002. **13**(2): p. 723-37.

71. Benkirane, M., et al., *Mechanism of transdominant inhibition of CCR5-mediated HIV-1 infection by ccr532.* J Biol Chem, 1997. **272**: p. 30603-6.

72. Ramsay, D., et al., *Homo- and hetero-oligomeric interactions between G-protein-coupled receptors in living cells monitored by two variants of bioluminescence resonance energy transfer (BRET): hetero-oligomers between receptor subtypes form more efficiently than between less closely related sequences.* Biochemical Journal, 2002. **365**: p. 429-440.

73. Zeng, F.Y. and J. Wess, *Identification and molecular characterization of m3 muscarinic receptor dimers.* J Biol Chem, 1999. **274**(27): p. 19487-97.

74. Shyu, Y.J., C.D. Suarez, and C.D. Hu, *Visualization of ternary complexes in living cells by using a BiFC-based FRET assay.* Nat Protoc, 2008. **3**(11): p. 1693-702.

75. Magliery, T.J., et al., *Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: Scope and mechanism.* J Amer Chem Soci, 2005. **127**(1): p. 146-157.

76. O, S., J. FH, and S. Y., *Extraction, Purification and Properties of Aequorin,a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea'.* J Cell Comp Physiol, 1962. **59**: p. 223– 39.

77. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.

78. Ghosh, I., A.D. Hamilton, and L. Regan, *Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein.* J. Am. Chem. Soc., 2000. **122**: p. 5658-5659.

79. Hu, C.D. and T.K. Kerppola, *Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis.* Nature Biotechnology, 2003. **21**(5): p. 539-545.

80. Shyu, Y.J., et al., *Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions.* Biotechniques, 2006. **40**(1): p. 61-6.

81. Hu, C.D., Y. Chinenov, and T.K. Kerppola, *Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation.* Molecular Cell, 2002. **9**(4): p. 789-798.

82. Rizzo, M.A., et al., *An improved cyan fluorescent protein variant useful for FRET.* Nat Biotechnol, 2004. **22**(4): p. 445-9.

83. Nagai, T., et al., *A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.* Nat Biotechnol, 2002. **20**(1): p. 87-90.

84. Vidi, P.A., et al., *Visualization of G Protein−Coupled Receptor (GPCR) Interactions in Living Cells Using Bimolecular Fluorescence Complementation (BiFC).* Curr Protoc in Neurosci, 2010: p. 5.29.1-5.29.15.

85. Vidia, P.A., et al., *Fluorescent protein complementation assays: new tools to study G protein-coupled receptor oligomerization and GPCR-mediated signaling.* Molecular and Cellular Endocrinology, 2011. **331**: p. 185-193.

86. Canals, M., et al., *Homodimerization of adenosine A2A receptors: qualitative and quantitative assessment by fluorescence and bioluminescence energy transfer.* J Neurochem, 2004. **88**(3): p. 726-34.

87. Vidia, P.A., et al., *Adenosine A2A receptors assemble into higher-order oligomers at the plasma membrane.* FEBS Letters, 2008. **582**: p. 3985-3990.

88. Gandia, J., et al., *Detection of higher-order G protein-coupled receptor oligomers by a combined BRET–BiFC technique.* FEBS Lett., 2008. **582**: p. 2979-2984.

89. Kamiya, T., et al., *Oligomerization of adenosine A2A and dopamine D2 receptors in living cells.* Biochemical and Biophysical Research Communications, 2003. **306**: p. 544-49.

90. Vidi, P., et al., *Ligand-Dependent Oligomerization of Dopamine D2 and Adenosine A2A Receptors in Living Neuronal Cells.* Mol Pharmacol, 2008. **74**: p. 544–551.

91.   Guo, W., et al., *Dopamine D2 receptors form higher order oligomers at physiological expression levels.* EMBO j., 2008. **27**: p. 2293–2304.

92.   Filizola, M. and H. Weinstein, *The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics.* FEBS J, 2005. **272**: p. 2926–2938.

93.   Simpson, L.M., et al., *Bioinformatics and molecular modelling approaches to GPCR oligomerization.* Curr Opin Pharmacol, 2010. **10**(1): p. 30-7.

94.   Witt, M., M.J. S´lusarz, and J. Ciarkowski, *Molecular Modeling of Vasopressin V2 Receptor Tetramer in Hydrated Lipid Membrane.* QSAR Comb. Sci., 2008. **27**(6): p. 684-693.

95.   Bruno, A., A.E. Entrena Guadix, and G. Costantino, *Molecular Dynamics Simulation of the Heterodimeric mGluR2/5HT2A Complex. An Atomistic Resolution Study of a Potential New Target in Psychiatric Conditions.* J. Chem. Inf. Model., 2009. **49**: p. 1602-16.

96.   Vanni, S., et al., *Observation of "Ionic Lock" Formation in Molecular Dynamics Simulations of Wild-Type β1 and β2 Adrenergic Receptors.* Biochemistry 2009. **48**: p. 4789–4797.

97.   Rodríguez, D., A. Ranganathan, and J. Carlsson, *Strategies for Improved Modeling of GPCR-Drug Complexes: Blind Predictions of Serotonin Receptors Bound to Ergotamine.* J. Chem. Inf. Model., 2014. **54**: p. 2004-2021.

98.   Periole, X., et al., *G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers.* J. AM. CHEM. SOC., 2007. **129**: p. 10126-10132.

99.   Provasi, D., J.M. Johnston, and M. Filizola, *Lessons from Free Energy Simulations of δ-Opioid Receptor Homodimers Involving the Fourth Transmembrane Helix.* Biochemistry, 2010. **49**: p. 6771–6776.

100.   Periole, X. and S.J. Marrink, *The Martini coarse-grained force field.* Methods Mol Biol, 2013. **924**: p. 533-65.

101.   Mondal, S., et al., *Membrane Driven Spatial Organization of GPCRs.* Scientific Reports, 2013.

102. Prasanna, X., A. Chattopadhyay, and D. Sengupta, *Cholesterol Modulates the Dimer Interface of the b2-Adrenergic Receptor via Cholesterol Occupancy Sites.* Bioph. J. Vol., 2014. **106**: p. 1290–1300.

103. Ghosh, A., U. Sonavane, and R. Joshi, *Mutliscale modelling to understand the self-assembly mechanism of human beta2-adrenergic receptor in lipid bilayer.* Comput. Bio. & Chem., 2014. **48**: p. 29-39.

104. Marrink, S.J., et al., *The MARTINI force field: coarse grained model for biomolecular simulations.* J. Phys. Chem. B, 2007. **111**(27): p. 7812-24.

105. Monticelli, L., et al., *The MARTINI coarse-grained force field: Extension to proteins.* J. Chem. Theo. Comp., 2008. **4**(5): p. 819-834.

106. Psachoulia, E., et al., *Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin a helix dimerization.* Biochemistry, 2008. **47**(40): p. 10503-12.

107. Prakash, A., L. Janosi, and M. Doxastakis, *Self-association of models of transmembrane domains of ErbB receptors in a lipid bilayer.* Biophys J, 2010. **99**(11): p. 3657-65.

108. Filipek, S., et al., *A concept for G protein activation by G protein-coupled receptor dimers: the transducin/rhodopsin interface.* Photochem Photobiol Sci, 2004. **3**(6): p. 628-38.

109. Han, Y., et al., *Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation.* Nature chemical biology, 2009. **5**(9).

110. Kaczor, A.A., et al., *Modeling complexes of transmembrane proteins: systematic analysis of protein-protein docking tools.* Mol. Inf., 2013. **32**: p. 717-733.

111. Filizola, M., S.X. Wang, and H. Weinstein, *Dynamic models of G-protein coupled receptor dimers: indications of asymmetry in the rhodopsin dimer from molecular dynamics simulations in a POPC bilayer.* J Comput Aided Mol Des, 2006. **20**(7-8): p. 405-16.

112. Cordomi, A. and J.J. Perez, *Structural rearrangements of rhodopsin subunits in a dimer complex: a molecular dynamics simulation study.* J Biomol Struct Dyn, 2009. **27**(2): p. 127-47.

113. Neri, M., et al., *Role of aggregation in rhodopsin signal transduction.* Biochemistry, 2010. **49**(23): p. 4827-32.

114. Periole, X., et al., *Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers.* J Am Chem Soc, 2012. **134**(26): p. 10959-65.

115. Fanelli, F., et al., *Theoretical study on receptor-G protein recognition: New insights into the mechanism of the alpha 1b-adrenergic receptor activation.* International Journal of Quantum Chemistry, 1999. **73**(2): p. 71-83.

116. Soulier, J.L., et al., *Design and synthesis of specific probes for human 5-HT4 receptor dimerization studies.* J Med Chem, 2005. **48**(20): p. 6220-8.

117. Russo, O., et al., *Synthesis of specific bivalent probes that functionally interact with 5-HT4 receptor dimers.* Journal of Medicinal Chemistry, 2007. **50**(18): p. 4482-4492.

118. Berthouze, M., et al., *Two transmembrane Cys residues are involved in 5-HT4 receptor dimerization.* Biochem Biophys Res Commun, 2007. **356**(3): p. 642-7.

119. Gorinski, N., et al., *Computational and Experimental Analysis of the Transmembrane Domain 4/5 Dimerization Interface of the Serotonin 5-HT1A Receptor.* Molecular Pharmacology, 2012. **82**(3): p. 448-463.

120. Rodriguez, D. and H. Gutierrez-de-Teran, *Characterization of the homodimerization interface and functional hotspots of the CXCR4 chemokine receptor.* Proteins, 2012. **80**(8): p. 1919-28.

121. Casciari, D., D. Dell'Orco, and F. Fanelli, *Homodimerization of neurotensin 1 receptor involves helices 1, 2, and 4: insights from quaternary structure predictions and dimerization free energy estimations.* J Chem Inf Model, 2008. **48**(8): p. 1669-78.

122. Johnston, J.M., et al., *Making structural sense of dimerization interfaces of delta opioid receptor homodimers.* Biochemistry, 2011. **50**(10): p. 1682-90.

123. Fanelli, F. and A. Felline, *Dimerization and ligand binding affect the structure network of A2A adenosine receptor.* Biochimica et Biophysica Acta, 2011: p. 1256–1266.

124. Kim, S.K. and K.A. Jacobson, *Computational prediction of homodimerization of the A3 adenosine receptor.* J Mol Graph Model, 2006. **25**(4): p. 549-61.

125. Fanelli, F., *Dimerization of the lutropin receptor: insights from computational modeling.* Mol Cell Endocrinol, 2007. **260-262**: p. 59-64.

126. Woolf, P.J. and J.J. Linderman, *An algebra of dimerization and its implications for G-protein coupledreceptorsignaling.* Journal of Theoretical Biology, 2004. **229**: p. 157–168.

127. Liu, X., et al., *Computational study of the heterodimerization between l and d receptors.* J Comput Aided Mol Des (2009) 23:321–332, 2009. **23**: p. 321–32.

128. Barki-Harrington, L., L.M. Luttrell, and H.A. Rockman, *Dual inhibition of beta-adrenergic and angiotensin II receptors by a single antagonist: a functional role for receptor-receptor interaction in vivo.* Circulation, 2003. **108**(13): p. 1611-8.

129. Szidonya, L., M. Cserzo, and L. Hunyady, *Dimerization and oligomerization of G-protein-coupled receptors: debated structures with established and emerging functions.* J Endocrinol, 2008. **196**(3): p. 435-53.

130. Abbracchio, M.P. and G. Burnstock, *Purinoceptors: are there families of P2X and P2Y purinoceptors?* Pharmacol Ther, 1994. **64**(3): p. 445-75.

131. Ralevic, V. and G. Burnstock, *Receptors for purines and pyrimidines.* Pharmacol Rev, 1998. **50**(3): p. 413-92.

132. Burnstock, G., *Introduction: P2 receptors.* Curr Top Med Chem, 2004. **4**(8): p. 793-803.

133. Burnstock, G., *Review Purine and pyrimidine receptors.* Cell Mol Life Sci, 2007.

134. Broad, L.M., et al., *Role of the phospholipase C-inositol 1,4,5-trisphosphate pathway in calcium release-activated calcium current and capacitative calcium entry.* J Biol Chem, 2001. **276**(19): p. 15945-52.

135. North, R.A., *Molecular physiology of P2X receptors.* Physiol Rev, 2002. **82**(4): p. 1013-67.

136. Terrillon, S. and M. Bouvier, *Roles of G-protein-coupled receptor dimerization - From ontogeny to signalling regulation.* Embo Reports, 2004. **5**(1): p. 30-34.

137.   Jaakola, V.P., et al., *The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist.* Science, 2008. **322**(5905): p. 1211-7.

138.   Fredholm, B.B., R.A. Cunha, and P. Svenningsson, *Pharmacology of adenosine A2A receptors and therapeutic applications.* Curr Top Med Chem, 2003. **3**(4): p. 413-26.

139.   Olah, M.E., *Identification of A2a adenosine receptor domains involved in selective coupling to Gs. Analysis of chimeric A1/A2a adenosine receptors.* J Biol Chem, 1997. **272**(1): p. 337-44.

140.   Le, F., et al., *Characterization and chromosomal localization of the human A2a adenosine receptor gene: ADORA2A.* Biochem Biophys Res Commun, 1996. **223**(2): p. 461-7.

141.   Gubitz, A.K., et al., *Dual Signalling by the Adenosine A2a Receptor Involves Activation of Both N- and P-Type Calcium Channels by Different G Proteins and Protein Kinases in the Same Striatal Nerve Terminals.* J Neurochem, 1996. **67**(1): p. 374-81.

142.   Fredholm, B.B., et al., *International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors.* Pharmacol Rev, 2001. **53**(4): p. 527-52.

143.   Warne, T., et al., *Structure of a b1-adrenergic G-protein- coupled receptor.* Nature, 2008. **454**: p. 486-92.

144.   Rapoport, T.A., et al., *Membrane-protein integration and the role of the translocation channel.* Trends Cell Biol., 2004. **14**: p. 568-575.

145.   Cymer, F., A. Veerappan, and D. Schneider, *Transmembrane helix-helix interactions are modulated by the sequence context and by lipid bilayer properties.* Biochim Biophys Acta, 2012. **1818**(4): p. 963-73.

146.   Thevenin, D., et al., *Identifying interactions between transmembrane helices from the adenosine A2A receptor.* Biochem., 2005. **44**: p. 16239-45.

147.   Thevenin, D. and T. Lazarova, *Stable interactions between the transmembrane domains of the adenosine A2A receptor.* Protein Science, 2008. **17**: p. 1188–1199.

148.   Senes, A., D.E. Engel, and W.F. DeGrado, *Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs.* Curr Opin Struct Biol, 2004. **14**(4): p. 465-79.

149.   Gouldson, P.R., et al., *Dimerization and domain swapping in G-protein-coupled receptors: a computational study.* Neuropsychopharmacology, 2000. **23**(4 Suppl): p. S60-77.

150.   Wu, B., et al., *Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists.* Science, 2010. **330**(6007): p. 1066-71.

151.   Huang, J., et al., *Crystal structure of oligomeric beta1-adrenergic G protein-coupled receptors in ligand-free basal state.* Nat Struct Mol Biol, 2013. **20**(4): p. 419-25.

152.   Nurisso, A., Daina, A., Walker, R.C. , *A Practical Introduction to Molecular Dynamics Simulations: Applications to Homology Modeling. .* Methods Mol Biol, 2012. **857**: p. 137-73.

153.   Markt, P., S. Herdlinger, and D. Schuster, *Virtual Screening Against Obesity.* Curr Med Chem, 2011. **18**(14): p. 2158-2173.

154.   Young, D.C., *Computational Drug Design: A Guide for Computational and Medicinal Chemists.* Book, 2009. **308**.

155.   Berman, H.M., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**: p. 235-242.

156.   Kobilka, B.K., Deupi, X, *Conformational complexity of G-protein-coupled receptors.* Trends Pharmacol Sci, 2007.

157.   Buchwald, H., *Bariatric surgery: A systematic review and meta-analysis (vol 292, pg 1724, 2004).* Jama-Journal of the American Medical Association, 2005. **293**(14): p. 1728-1728.

158.   Schlick, T., *Molecular Modeling and simulation: An interdisciplinary guide.* Springer 2010., 2010(2nd Edition).

159.   Chiu, M. and M.C. Herbordt, *Molecular Dynamics Simulations on High-Performance Reconfigurable Computing Systems.* ACM Trans Reconfigurable Technol Syst, 2010. **3**(4).

160.   Perlmutter, J.D., et al., *All-atom and coarse-grained molecular dynamics simulations of a membrane protein stabilizing polymer.* Langmuir, 2011. **27**(17): p. 10523-37.

161. Marrink, S.J., A.H. de Vries, and A.E. Mark, *Coarse Grained Model for Semiquantitative Lipid Simulations.* J. Phys. Chem. B 2004. **108**: p. 750-760.

162. Nielsen, S.O., et al., *Coarse grain models and the computer simulation of soft materials.* J Phys-Conden Matter, 2004. **16**(15): p. R481-R512.

163. Caron, M., et al., *Measuring Performance of an Automated and Miniaturized LANCE Ultra cAMP Assay for the Gi-coupled 5-HT1A Receptor– a Comparative Study.* 2010.

164. Sadiq, S.K., et al., *Accurate ensemble molecular dynamics binding free energy ranking of multidrug-resistant HIV-1 proteases.* J Chem Inf Model, 2010. **50**(5): p. 890-905.

165. De Jong, D.H., et al., *Improved Parameters for the Martini Coarse-Grained Protein Force Field.* J. Chem. Theory Comput., 2012. **9**(1): p. 687-697.

166. Peetla, C., A. Stine, and V. Labhasetwar, *Biophysical interactions with model lipid membranes: applications in drug discovery and drug delivery.* Mol Pharm, 2009. **6**(5): p. 1264-76.

167. A., B.J. and W. H., *Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors.* Methods in Neurosciences, 1995. **25**.

168. Sali, A., et al., *Evaluation of comparative protein modeling by MODELLER.* Proteins, 1995. **23**(3): p. 318-26.

169. Fiser, A. and A. Sali, *Modeller: generation and refinement of homology-based protein structure models.* Methods Enzymol, 2003. **374**: p. 461-91.

170. Wassenaar, T.A., et al., *Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations.* J. Chem. Theory Comput. , 2015. **11**: p. 2144−2155.

171. Wassenaar, T.A., et al., *Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models.* J. Chem. Theory Comput. , 2014. **10**: p. 676−690.

172. Daura, X., et al., *Peptide folding: When simulation meets experiment.* Angewandte Chemie-International Edition, 1999. **38**(1-2): p. 236-240.

173. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

174. Sahoo, D., et al., *Lipid-triggered conformational switch of apolipophorin III helix bundle to an extended helix organization.* J Mol Biol, 2002. **321**(2): p. 201-14.

175. Senes, A., I. Ubarretxena-Belandia, and D.M. Engelman, *The C-H...O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions.* Proc. Natl. Acad. Sci., 2001. **98**: p. 9056-61.

176. Vaugh, A., *Graphing with Gnuplot and Xmg.* Linux Journal, 1996. **28**.

177. Gandia, J., et al., *Detection of higher-order G protein-coupled receptor oligomersby a combined BRET–BiFC technique.* FEBS Letters 2008. **582**: p. 2979–2984.

178. Levitz, J., et al., *Mechanism of Assembly and Cooperativity of Homomeric and Heteromeric Metabotropic Glutamate Receptors.* Neuron, 2016. **92**(1): p. 143-159.

179. Babcock, G.J., M. Farzan, and J. Sodroski, *Ligand-independent dimerization of CXCR4, a principal HIV-1 coreceptor.* J Biol Chem, 2003. **278**(5): p. 3378-85.

180. Percherancier, Y., et al., *Bioluminescence resonance energy transfer reveals ligand-induced conformational changes in CXCR4 homo- and heterodimers.* J Biol Chem, 2005. **280**(11): p. 9895-903.

181. Dorsch, S., et al., *Analysis of receptor oligomerization by FRAP microscopy.* Nat Methods, 2009. **6**(3): p. 225-30.

182. Senes, A., M. Gerstein, and D.M. Engelman, *Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions.* Journal of Molecular Biology, 2000. **296**(3): p. 921-936.

183. Wright, D.W., et al., *Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors.* J. Chem. Theory Comput., 2014. **10**: p. 1228−1241.

184. Pritchard, L., et al., *Evaluation of a novel method for the identification of coevolving protein residues.* Protein Eng, 2001. **14**(8): p. 549-55.

185. Janin, J., *Welcome to CAPRI: A Critical Assessment of PRedicted Interactions.* PROTEINS: Structure, Function, and Genetics, 2002. **47**: p. 257.