



**University of
Nottingham**

UK | CHINA | MALAYSIA

**Computational and experimental studies of novel β -
adrenoceptor ligands**

Keverne Anthony Louison

BSc MSc

Thesis submitted to the University of Nottingham for the degree of Doctor of
Philosophy

September 2021

Abstract

With over 30% of available drugs targeting them, G protein-coupled receptors (GPCRs) are of significant pharmaceutical interest. Efforts to understand protein-ligand interactions among this group of proteins have been aided by the increase in available x-ray crystallography structures. β_1 -Adrenergic receptor (β_1 -AR) antagonists are used as treatment in patients with cardiovascular and airway conditions. However, current widely used medications are considerably prone to off-target side effects due to the lack of selectivity between the β_1 - and β_2 -AR subtypes. Therefore, a deeper understanding into the structural differences in characteristics is necessary to utilise them as a means of increasing ligand selectivity and therefore reducing the prevalence of off-target side effects. Here, two characteristics of the β_1 -AR are targeted as a means of increasing receptor selectivity. The first being receptor plasticity - recent research has shown that β -ARs contain a fissure between transmembrane helices 4 and 5 (TM4, TM5) (dubbed the 'keyhole') that differ slightly between β -AR subtypes that may accommodate for extended moieties or ligand entry and exit via the intramembrane space. The second characteristic being receptor dimerization. Receptor dimerization among GPCRs remains an active area of research, that so far has many pharmacological implications. Targeting receptor homodimerization has been proposed to be a method of improving receptor specificity within GPCRs. Research into β -AR dimers and findings from X-ray crystallography have shown that β_1 -AR homodimers may indeed align with a TM4/TM5 interface, aligning the 'keyhole'. By combining and exploring both characteristics, we designed and computationally validated bivalent ligands capable of taking advantage of two unique β_1 -AR structural features as a means of improving ligand selectivity. Most current attempts at bivalent ligands in GPCRs explore using the extracellular space as a spacing route, leading to longer ligands, undesirably affecting molecular weight, lipophilicity, and viability. However, to validate our ligand design, we computationally demonstrate – by analyzing all-atom molecular dynamics (MD) simulations – that those ligands long enough to extend beyond the receptor via the keyhole can bind canonically and maintain key interactions that have

previously been pharmacologically verified, as well as investigate structure activity relationships (SARs) of differing steric and electronic configurations of ligand components exposed to the intramembrane space. Bivalent ligand linkers were designed and computationally investigated within a GPCR dimer system to determine whether flexibility of the linker impacts the pharmacophores' ability to maintain key canonical interactions. Long timescale coarse grained simulations of a membrane-bound β_1 -AR dimer showed the dimer interface to be stable, so shorter all-atom simulations could be used with confidence to aid bivalent ligand design. Bivalent ligands of the nature discussed in this work required at least one pharmacophore to enter/exit the receptor orthosteric binding site via the keyhole route. In house enhanced sampling computational methods were developed to study and validate the feasibility of this entry and exit route. Ligand exit pathways were generated by performing self-avoiding walk MD on protein-ligand complexes, then used to define starting and end points for weighted ensemble molecular dynamics (WEMD) to predict kinetic rate constants. These rate constants were then verified against pharmacologically derived β -AR kinetics data to validate the method, model, and ligand entry/exit pathway. The designed ligands would then lead to shorter and less hindering spacing between orthosteric sites.

Acknowledgements

I would like to thank my supervisors Professor Charlie Laughton and Dr Shailesh Mistry for their support throughout the duration of my postgraduate studies as well as their guidance and advice in many aspects within that time.

The computational work within this project required knowledge and expertise that I was fortunate enough to be able to work towards with the help of Professor Charlie Laughton – who helped enormously with my deeper technical understanding, programming knowledge and scripting - as well as the molecular recognition group and the CCPBIOSIM consortium.

The medicinal chemistry and computational work could not have been completed without the work of Guillermo Baltazares Segura and Roaa Shireen, who with the help of Dr Shailesh Mistry produced the series of ligands that inspired the directions taken within this project. Pharmacological assays were also a key aspect of this study, which without, would have limited the value of the results obtained. I'd like to thank Professor Jillian Baker for carrying out the *in-vitro* cell-based assays and related data processing.

I thank the medicinal chemistry group on the C floor for their support and for offering their knowledge, experience, and time, as well as technicians Lee Hibbett and Dr Ian Withers for their round-the-clock support in the medicinal chemistry labs and the molecular modelling lab respectively.

Lastly, but certainly not least, I'd like to thank Natasha Dawson, my family and my loved ones for their patience, support and encouragement over the duration of my studies.

This work was funded by the BBSRC.

Publications related to this thesis

GLIMPS: A Machine Learning Approach to Resolution Transformation for Multiscale Modelling

Authors: Keverne A. Louison, Ian L. Dryden, and Charles A. Laughton

Status: Published in *ACS Journal of Chemical Theory and Computation*

(Louison et al., 2021)¹

Abbreviations

Ar	aromatic ring
ASL	atom specific language
ASP	active site pressurization
ATP	adenosine triphosphate
β_1 -AR	β_1 -adrenoceptor
β_2 -AR	β_2 -adrenoceptor
β_3 -AR	β_3 -adrenoceptor
BBSRC	Biotechnology and Biological Sciences Research Council
BiFC	bimolecular fluorescence complementation
Boc	tert-butyloxycarbonyl
Boc ₂ O	di-tert-butyloxy dicarboxylate
Bn	benzyl
br	broad
calcd	calculated
cAMP	cyclic adenosine monophosphate
CDCl ₃	deuterated chloroform
CGMD	course-grained molecular dynamics
CNGs	cyclic nucleotide-gated ion channels
COM	centre of mass
conc	concentrated
CYP	cyanopindolol
d	doublet
DCM	dichloromethane
dd	doublet of doublets
DMF	<i>N, N</i> - dimethylformamide
DMSO	dimethyl sulphoxide
DMSO-d ₆	deuterated dimethyl sulfoxide
DPPC	Dipalmitoylphosphatidylcholine
dt	doublet of triplets
ECL	extracellular loop
eq	equivalents
ES	electrospray
Et ₃ N	triethylamine
Et ₂ O	diethylether
EtOAc	ethylacetate
FCC	flash column chromatography
G-protein	guanine nucleotide binding protein
GABA	γ -amino butyric acid
GPCR	G-protein coupled receptor
GDP	guanine diphosphate
GPK	G-protein kinase
GTP	guanine triphosphate

HPLC	high performance liquid chromatography
HRMS	high resolution mass spectroscopy
ICL	intracellular loop
ISA	intrinsic sympathetic activity
<i>J</i>	coupling constant
J_{CF}	carbon fluorine coupling constant
LCMS	liquid chromatography mass spectroscopy
m	multiplet
<i>m</i>	meta
MAPK	mitogen-activated protein kinase
MD	molecular dynamics
MeCN	acetonitrile
MeOH	methanol
mp	melting point
MS	mass spectroscopy
MW	microwave
<i>m/z</i>	observed ion
NMR	nuclear magnetic resonance
NPT	Isothermal-Isobaric (NPT): ensemble: moles (N), pressure (P) and temperature (T)
NVT	Canonical ensemble (NVT): moles (N), volume (V) and temperature (T)
<i>o</i>	ortho
<i>p</i>	para
PC	progress coordinate
PCA	principal component analysis
PDB	protein databank
PDE	phosphodiesterase
PE	petroleum ether 40 – 60
Piv	pivaloyl
PKA	protein kinase A
PLC	preparative layer chromatography
PMA	phosphomolybdic acid
PMB	<i>para</i> -methoxybenzyl
ppm	parts per million
^c Pe	cyclopentyl
^c Pr	cyclopropyl
QSAR	quantitative structure activity relationship
q	quadruplet
<i>rac</i>	racemic

t_R	retention time
s	singlet
SAR	structure activity relationship
SAWMD	self-avoiding walk molecular dynamics
S.E.M	standard error of the mean
SNS	sympathetic nervous system
SPAP	secreted placental alkaline phosphate
t	triplet
TEA	triethylamine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TM	transmembrane
TMS	tetramethylsaline
TOF	time of flight
WEMD	weighted ensemble molecular dynamics

Table of Contents

ABSTRACT	I
ACKNOWLEDGEMENTS	III
PUBLICATIONS RELATED TO THIS THESIS	IV
ABBREVIATIONS	V
1 INTRODUCTION	12
1.1 G PROTEIN-COUPLED RECEPTORS	12
1.2 GPCR FUNCTION	13
1.3 GPCR STRUCTURE	16
1.3.1 Extracellular surfaces and ligand binding sites	17
1.3.2 Cytoplasmic surfaces of the GPCR structures	19
1.4 GPCR PHARMACOLOGY	20
1.5 BINDING KINETICS AND PATHWAYS	23
1.5.1 Theory of kinetics	24
1.5.2 Experimental approaches to investigating kinetics	24
1.6 GPCR DIMERIZATION	25
1.6.1 GPCR dimerization and mechanisms	25
1.6.2 Biochemical approaches to detecting GPCR oligomerization	27
1.6.3 Resonance energy transfer approaches.....	28
1.6.4 GPCR dimer functions	30
1.6.5 GPCR dimers as therapeutic targets	32
1.7 X-RAY CRYSTALLOGRAPHY OF GPCRS	34
1.7.1 Methods in crystallising GPCRS	34
1.7.2 β -adrenergic receptor crystal structures.....	37
1.8 B-ADRENERGIC RECEPTOR LIGANDS	42
1.8.1 GPCR ligand classification.....	42
1.8.2 Orthosteric binding	42
1.8.3 Biased ligands	44
1.8.4 β -blocker pharmacology	47
1.8.5 Ligand interactions	49
1.8.6 The keyhole	53

1.9 COMPUTATIONAL TOOLS FOR BIOMOLECULAR MODELLING AND DYNAMICS.....	56
1.9.1 Homology modelling.....	56
1.9.2 Docking.....	56
1.9.3 Molecular dynamics.....	57
1.9.4 Classical Molecular modelling force fields.....	60
1.9.5 Analysis techniques for molecular dynamics simulations.....	62
1.9.6 Self-avoiding walk molecular dynamics.....	62
1.9.7 Weighted ensemble enhanced sampling.....	66
1.10 RESEARCH AIMS.....	70
2 EXPLORING MODES OF BINDING OF EXTENDED B-AR LIGANDS IN THE B₁-AR.....	73
2.1 AIMS AND OBJECTIVES.....	73
2.2 METHODS.....	74
2.2.1 Homology modelling using Prime.....	74
2.2.2 Ligand series to be docked.....	75
2.2.3 Glide docking.....	77
2.2.4 MD of protein systems.....	78
2.3 RESULTS.....	78
2.3.1 Docking results.....	78
2.3.2 MD results.....	82
2.4 DISCUSSION.....	89
2.5 CONCLUSIONS.....	92
3 EXPLORING STRUCTURE ACTIVITY RELATIONSHIPS OF STERIC AND ELECTROSTATIC LIGAND ANALOGUES AT THE INTRAMEMBRANE SPACE IN THE B₁-AR.....	93
3.1 AIMS AND OBJECTIVES.....	93
3.2 METHODS.....	93
3.2.1 Design and docking of compound 47 analogues.....	93
3.2.2 Synthesis of series of compound 47 analogues.....	95
3.2.3 Pharmacology.....	97
3.3 RESULTS AND DISCUSSION.....	97
3.3.1 Docking.....	97
3.3.2.1 Attempted synthesis of <i>o</i> - and <i>p</i> - methoxybenzene analogues.....	99

3.3.2.2 <i>Attempted synthesis of pyridine analogues</i>	100
3.3.3 <i>Pharmacology</i>	101
3.4 DISCUSSION.....	103
3.5 CONCLUSION	106
3.6 EXPERIMENTAL.....	107
4 INVESTIGATING INTRA-MEMBRANE ROUTES FOR LIGAND BIVALENCY USING COVALENTLY LINKED PHARMACOPHORES IN THE B₁-AR HOMODIMER	122
4.1 AIMS AND OBJECTIVES	122
4.2 METHODS	125
4.2.1 <i>Homology modelling</i>	125
4.2.2 <i>Preparing the protein-model systems</i>	125
4.2.3 <i>Dimer sustainability testing</i>	126
4.2.4 <i>Spacer length investigation and dimer-ligand interaction</i>	126
4.2.5 <i>Linker stability, interaction and dimer interface investigation</i>	127
4.3 RESULTS	128
4.3.1 <i>Dimer stability investigation</i>	128
4.3.2 <i>Spacer length investigation and dimer-ligand interaction</i>	128
4.3.3 <i>Linker Stability Investigation</i>	131
4.3.4 <i>Ligand-residue interactions</i>	133
4.3.5 <i>Investigating potential contact between ligands at the transmembrane space</i>	137
4.3.6 <i>Contact analysis of residues at the dimer interface</i>	139
4.4 DISCUSSION.....	142
4.5 CONCLUSION	146
5 ESTIMATING LIGAND BINDING AND UNBINDING KINETICS AT B₁-ARS.....	147
5.1 AIMS AND OBJECTIVES	147
5.2 METHODS	149
5.2.1 <i>System preparation</i>	149
5.2.2 <i>Obtaining ligand unbinding pathways using SAWMD</i>	150
5.2.3 <i>Obtaining rate kinetic estimates using WEMD</i>	150
5.2.4 <i>Initial iterations of the proposed methods</i>	151
5.3 RESULTS	153
5.3.1 <i>Obtaining ligand unbinding pathways using SAWMD</i>	153

5.3.2 <i>Obtaining rate kinetic estimates using WEMD</i>	155
5.4 DISCUSSION	158
5.5 CONCLUSION	160
6 CONCLUSIONS AND FUTURE WORK	161
6.1 GENERAL CONCLUSIONS.....	161
6.2 FUTURE WORK	162
APPENDICES	165
APPENDIX 1	165
APPENDIX 2 – LIGAND 47 AMBERFF99SB PARAMETERS IN GROMACS .ITP FORMAT	167
APPENDIX 3 - LIGAND 58 AMBERFF99SB PARAMETERS IN GROMACS .ITP FORMAT	193
APPENDIX 4 - LIGAND 59 AMBERFF99SB PARAMETERS IN GROMACS .ITP FORMAT	232
APPENDIX 5	269
APPENDIX 6 – SELF-AVOIDING WALK SCRIPT	270
APPENDIX 4 – WEIGHTED ENSEMBLE WITH PRAGMATIC ADAPTIVE BINNING SCHEME SCRIPT	274
APPENDIX 7 – EXAMPLE WEIGHTED ENSEMBLE POST-PROCESSING SCRIPT	282
APPENDIX 8 - EXAMPLE WEIGHTED ENSEMBLE POST-PROCESSING SCRIPT FOR RATE EQUILIBRATION GRAPH	285
REFERENCES	289

1 Introduction

1.1 G Protein-Coupled Receptors

G-protein coupled receptors (GPCRs) make up the largest family of proteins found in the human genome. These proteins are located on the membranes of cells, and their role is to transduce extracellular signals into physiological effects². Since the inception of their discovery, understanding the structure and functions of GPCRs has been fundamental in therapeutic research, and many diseases could be attributed to GPCRs including depression, cancer, Alzheimer's disease and type 2 diabetes mellitus³. Once activated by external signals, GPCRs elicit downstream intracellular responses through coupling to different arrestins and G-proteins⁴.

When the first tertiary structure of bovine rhodopsin was published in the 1980s^{5,6}, several characteristics were learned of GPCRs, such as their 7 transmembrane domain structure (7TMs). Further technical advances in analytical methodology allowed for further characterisation of GPCRs, thus leading to an expansion in the understanding of the GPCR superfamily^{6,7}.

Technical advances in protein engineering have enabled production of large amounts of protein at high purity, while structure determination has been improved with the growing use of nuclear magnetic resonance (NMR) spectroscopy and cryo-EM. After the revelation of the structural information of bovine rhodopsin⁸, the structure of β_2 -adrenergic receptor (β_2 -AR) spelled a breakthrough for GPCR X-ray structure determination⁹⁻¹¹. This breakthrough led to the discovery of methods to aid the crystallisation and structure determination process. The crystallisation of membrane proteins became easier when using conjugations, residue mutations and fusion proteins designed to reduce movement of highly mobile protein regions and introduce thermostability. Further discoveries allowed for discovery of the solved structure of the β_2 -AR/Gs

complex¹² (giving significant information about the specific interactions the GPCR makes with the G-protein) resulting in a Nobel prize¹³.

Crystallisation of membrane bound proteins had since improved due to the utilisation of companion proteins both covalently bound (e.g., BRIL, T4 Lysozyme (T4L)) at opposing sides of the 7TM (both incorporated into the amino acid sequence of the N-terminus or spliced into intracellular loop 3 (IL3)), non-covalently (such as antibodies), high affinity ligand binding, selective stability mutations and methods involving lipidic cubic phase (LCP)¹⁴.

Structure based drug discovery (SBDD) has improved remarkably since the first GPCR structures were reported and structures of GPCRs from several subclasses and activation states had since become used routinely for drug discovery and in-depth functional studies (both of which are explained in section 1.4).

1.2 GPCR function

GPCRs are important in most physiological responses to neurotransmitters, hormones, and the stimulants within our environment (such as light) and as discussed previously, have very high therapeutic target potential in several conditions. Through the increase in structural information, molecular understanding of agonist, inverse-agonist and antagonist recognition has grown significantly¹⁵.

The orthodox role of GPCRs is to bind to agonists and stimulate the activation of a heteromeric G protein, which then leads to modulation of a downstream cascade of effector proteins. However, it is important to note that recent studies have shown that some receptors do not bind to G proteins^{16–18}.

The β_2 -AR – found predominantly on the smooth muscle cells of the airways and modulates bronchoconstriction and dilation - would bind to noradrenaline and adrenaline in target tissues within the sympathetic nervous system, which would then lead to the activation of a heteromeric G protein ($G\alpha_s$) (this step involves the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) that then leads to the dissociation from the dimeric $\beta\gamma$ complex). Following this, the separated components continue leading to the stimulation of adenylyl cyclase, accumulation of cyclic AMP (cAMP), activation of protein kinase A (which is cAMP dependent) and the phosphorylation of the components involved in muscle contraction¹⁹. But the $G\alpha$ subunit catalyses the hydrolysis of GTP back to GDP. Once this happens, association of the α subunit and the $\beta\gamma$ subunits occur, restoring the inactive GPCR complex. A simplified illustration of this process can be found in Figure 1.

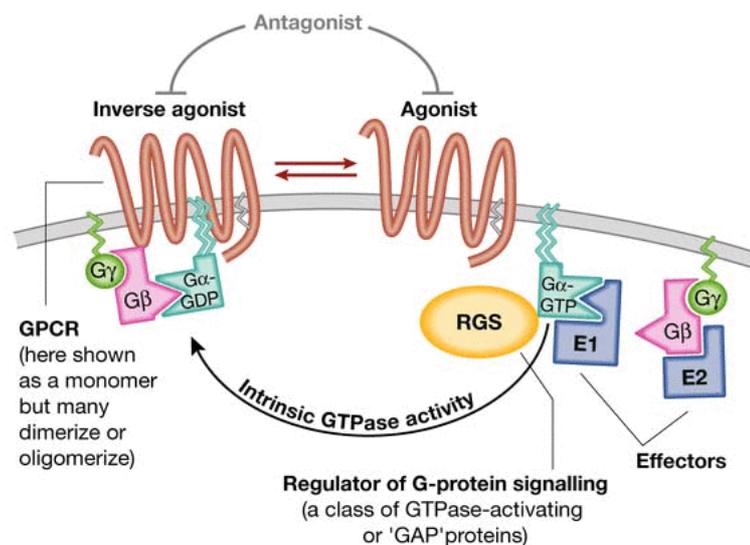


Figure 1 - Simplified schematic of cycle of heteromeric G protein activity. Source: Jones and Assmann, 2004²⁰.

The $G\alpha$ subunit can also be broken down into further classes; $G_s\alpha$ (G stimulatory), $G_i\alpha$ (G inhibitory), $G_o\alpha$ (G other), $G_{q/11}\alpha$ and $G_{12/13}\alpha$ ²¹. Receptors within the β -adrenergic receptor (β -AR) family can signal via the stimulatory pathway using the G_s heterotrimer. The β_2 -AR specifically, can activate the G_i inhibitory pathway. G_s and G_i can regulate adenylyl cyclase. It is the α_s subunit that is responsible for the initiation of the conversion of cAMP to adenosine triphosphate (ATP). This process

is required for the activation of PKA, as mentioned previously, and other targets. Using the muscle contraction example stated previously, the role of PKA involves the regulation of several cellular proteins including L-type Ca^{2+} ion channels¹⁹, which are responsible for the phenomenon, and the AR itself, regulated by means of phosphorylation. It is worth noting that the $\beta\gamma$ subunit can also act on Ca^{2+} ion channels. Signalling response amplification is achieved by the activation of several components of the cascade activating their succeeding components. Signalling is therefore a highly regulated process, as is its cessation, which is also regulated. Specific phosphodiesterase proteins (PDEs) are responsible for the downregulation of Second messenger cAMP. C-terminal phosphorylation by GPCR kinase (GRKs) can also occur because of activation. This, in turn, results in receptor coupling to arrestin, a regulatory protein.

Although β -ARs are capable of signalling via G proteins, it is not exclusively done so. They are also capable of being activated via the mitogen-activated protein kinase (MAPK) pathway. The resulting activation of extracellular signal-regulated kinases (ERK) is due to the involvement of arrestin recruitment and receptor phosphorylation. Clathrin-mediated endocytosis, done as a result of the presence of arrestin binding, can result in the degradation or recycling of the protein^{21,22}.

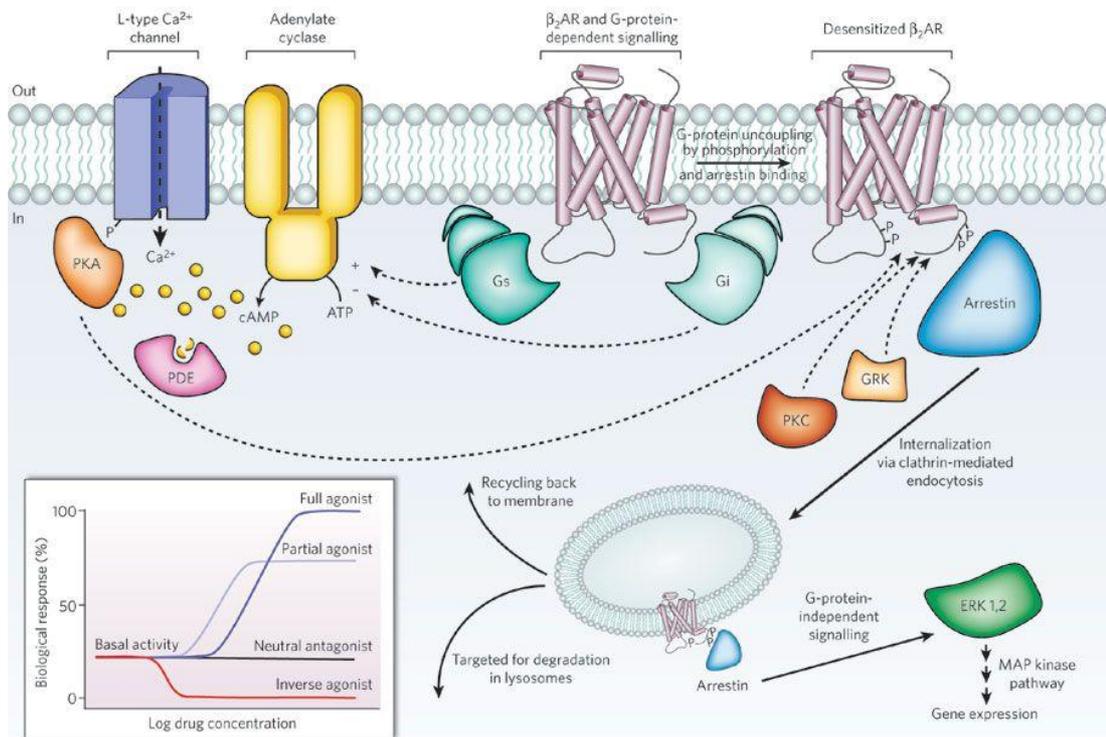


Figure 2 - Schematic diagram depicting the components of the GPCR cell signalling cascade. From left to right: PKA, protein kinase A; PDE, phosphodiesterase; Gs, G stimulatory signalling; Gi, G inhibitory signalling; PKC, protein kinase C; -P, phosphorylation; ERK 1,2, extracellular signal-regulated kinases; MAP, mitogen activated protein kinase. Source: Rosenbaum et al., 2014¹⁵.

1.3 GPCR structure

Over time, our knowledge of GPCR structure has increased, and as a result so has our understanding of how they function. Although GPCRs share similar characteristics, there have been methods to classify them further. As a result, GPCRs are now grouped into at least five categories, rhodopsin (701 members), adhesion (24 members), frizzled/taste (24 members), glutamate (15 members) and secretin (15 members).²¹ In the International Union of Basic and Clinical Pharmacology Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR), they group the receptors differently. Instead, they exclude light, odour and taste receptors and categorise the non-sensory GPCRs with respect to their pharmacological properties. This system uses class A rhodopsin-like receptors,

class B secretin-like receptors, class C metabotropic glutamate/pheromone receptors and frizzled receptors.²³

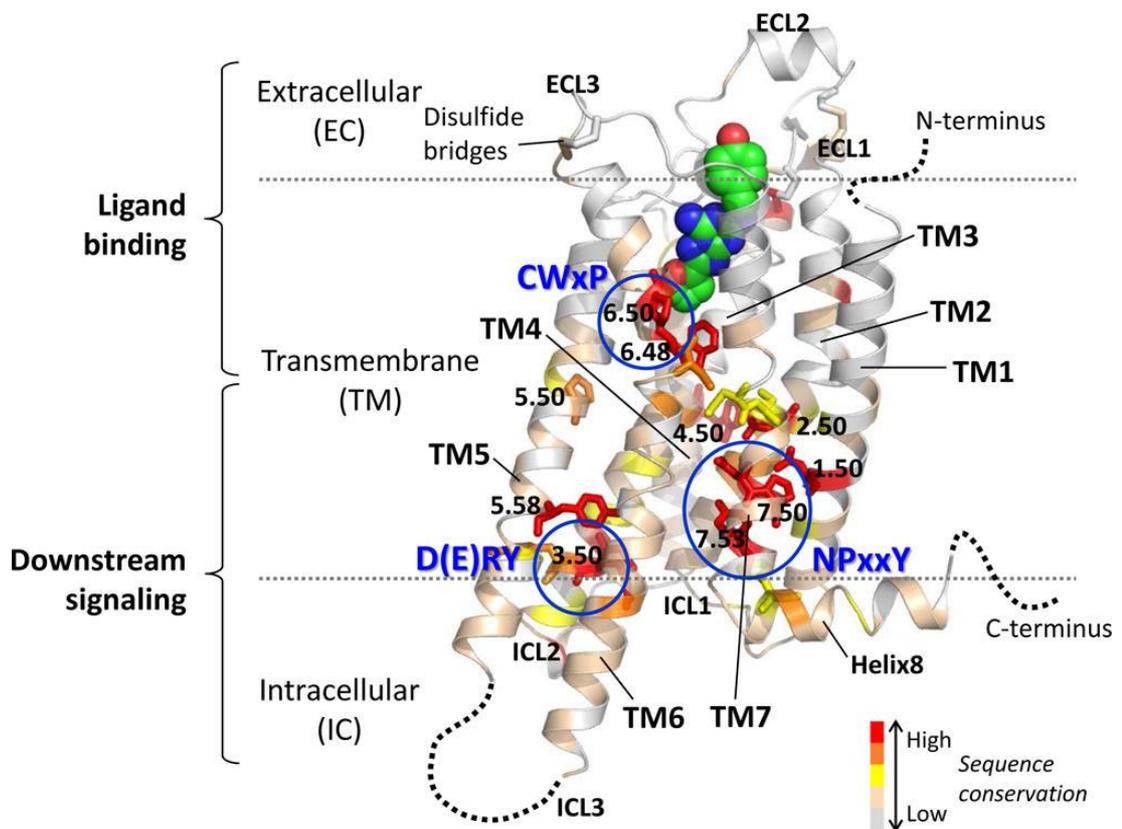


Figure 3 - General structural features of GPCRs. The diagram shows the tertiary structure comprised of seven transmembrane helices (7TM), three extracellular loops (ECL1-3) and three intracellular loops (ICL1-3). This diagram uses the $A_{2A}AR$ (PDB code: 4EIY) bound to ZM241385 (depicted in green carbons) as an example. The colour scheme represents the level of sequence conservation between class A GPCRs. The extracellular module (EC) can vary from short unstructured domains (class A GPCRs) to long globular structures (class B, C and F). The intracellular module (IC) is responsible for interactions with the downstream effectors. This component is also accompanied by a large, disordered C-terminal, that isn't readily accessible for crystallographic studies. This region also encompasses an amphipathic eighth helix containing a palmitoylation site that's role is to anchor this region to the membrane. Also shown in the diagram are the highly conserved CWxP, D(E)RY and NPxxY region. Disulphide bridges connecting the ECLs are denoted as sticks. Source: Lee, Y. et al., 2018²⁴.

1.3.1 Extracellular surfaces and ligand binding sites

Most of the structural divergences between the receptors exist within the extracellular loops and ligand-binding region. Extracellular loop 2 (ECL2) in

rhodopsin forms a short β -sheet that covers the covalently bound 11-*cis* retinal, preventing bulk solvent from entering the chromophore, thus preventing Schiff base hydrolysis. The glycosylated amino terminus further shields a covalently bound ligand due to its structured conformation at the extra-cellular apex.^{8,25} However the ECL2 in the β_1 - and β_2 -ARs contain a short α -helix, stabilised by disulphide bonds in the inter- and intra-loops, and the cytoplasmic N-terminal regions are disordered^{9,10,15,26}. Unlike the β -ARs, other receptors such as the A_{2A} adenosine receptor lack an objective secondary structure at the ECL2, but instead contains several disulphide bonds with ECL1 that constrain its conformation and exposes the ligands-binding cavity to bulk solvent outside of the cell.²⁷

The binding sites between different GPCRs in the same class can also differ. Carazolol, 11-*cis* retinal and cyanopindolol have overlapping positions in β_1 -, β_2 - and rhodopsin when compared (Figure 3). However, bound ligands tend to the extracellular side more in β_1 - and β_2 -ARs than in rhodopsin. Trp 286^{6.28} (human β_2 -AR - Ballesteros/Weinstein numbering²⁸ in superscript) and Trp 303^{6.48} (avian β_1 -AR) undergo conformational rotamer transitions in the activation of GPCRs. This is referred to as the 'rotamer toggle switch'²⁹.

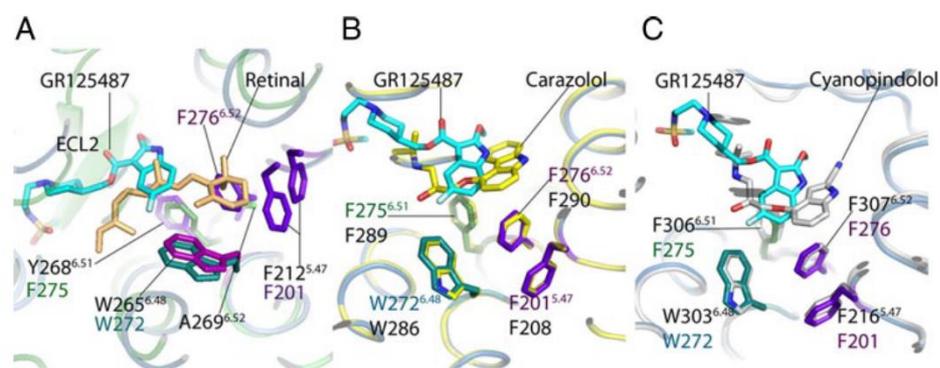


Figure 4 - Ligand binding site comparison in GPCRs. Docked antagonist GR125487 with the 5-HT₄R model superimposed with: (A) 11-*cis*-retinal bound to rhodopsin (1F88, (B) carazolol bound to β_2 -AR (2RH1) and (C) cyanopindolol bound to β_1 -AR. The β_2 -ARs toggle switch mechanism comprises of the movement of F2906.52, which is placed between F2896.51 and F2086.52 (yellow trace). During the activation process, W2866.48 is seen to rotate into the position occupied by F2906.52 in the receptors inactive form. It is from this where it can be observed that the steric constraints imposed by F2906.52 mimic retinal's β -ionone ring with rhodopsin's W2656.48 and F2125.47. The toggle switch of the β_1 -AR are also conserved and line the sides of

the binding pocket (Silver trace). The similarity of the bound complexes suggest that antagonist configurations are similarly inhibitory to toggle switch signalling in all three receptors. Source: Padayatti, P.S., et al., 2013³⁰.

In the β_2 -AR, partial agonism can be achieved by a ligand without interacting with the toggle switch, however, it is suggested that full agonism requires this conformational change.^{15,31} In rhodopsin, the ionone ring in retinal comes into contact with the analogous Trp residue, however cyanopindolol in β_1 -AR and carazolol in β_2 -AR pack against binding site shielding aromatic residues. This relationship between inverse-agonist binding interactions and the rotamer toggle switch's inactive conformation may explain the elevated basal activity of the β_2 -AR relative to rhodopsin.¹⁵ Ligand binding in rhodopsin and the adrenergic receptors are mediated by interactions with hydrophobic and polar residues in transmembrane helices 3, 5, 6 and 7 (TM3, TM5, TM6 and TM7).

The binding pockets identified in the avian β_1 -AR bound to cyanopindolol²⁶ and the human β_2 -AR bound to carazolol^{9,10} are almost identical. This could be explained by the proteins' related functions their tendency to completely conserve binding-site contact residues.³² However this conservation is not exclusive to GPCRs in this subfamily, it is also observed in others (such as histamine, serotonin and dopamine), a further potential reasoning for the difficulty in identifying subtype specific ligands.³³ Ligand binding selectivity is observed between β_1 -AR and β_2 -AR^{34,35}, however the differences cannot be attributed solely to the amino acids within the binding pocket, but rather to the subtype-specific conformational preferences in relevant surrounding residues. It has been shown that binding pathways can influence binding selectivity between β_1 -AR and β_2 -AR.³⁶

1.3.2 Cytoplasmic surfaces of the GPCR structures

A major structural difference between ligand activated rhodopsin and other GPCRs is the 'ionic lock' at the cytoplasmic surface between a glutamate residue on TM6 and the highly conserved E/DRY motif on TM3. It is suggested that disruption of

the ionic lock is not directly related to the rotamer toggle switch³⁷, due to catechol, which is capable of activating it, was not able to.³⁸ In contrast, salbutamol, which is known not to activate this switch³¹, was able to fully activate the ionic lock³⁸.

Catechol and dopamine have a similar binding affinity for β_2 -AR³⁸. The affinity for catechol determined by a conformational assay is 160 μm , whereas the affinity for dopamine is 350 μm .³⁸ The interaction of the primary amine of dopamine and Asp113 make the strongest binding contributions, therefore making this finding surprising. Only dopamine disrupts the ionic lock, but still has a similar binding affinity to catechol, suggesting that the binding energy associated between the interaction between dopamine and Asp113 may be utilised to offset the energetic cost of breaking the ionic lock.³⁷

1.4 GPCR pharmacology

Pharmacology and biological activity of ligands that bind to GPCRs are defined by several parameters including affinity and efficacy. These parameters are useful when comparing the varying effects between ligands. The techniques used to measure each of these parameters may differ depending on the information being investigated, for example, radio ligand binding can tell you information about affinity whereas cAMP and other mediator assays such as Ca^{2+} can give information about intracellular signalling. It is common to see pharmacological assays used in conjunction with mutagenesis studies to further identify the structural characteristics (e.g. residues) involved in the proteins further processes (receptor signalling, expression and binding)^{39,40}.

Expanding on what was briefly mentioned in Section 1.2, GPCR ligands can be classified into different categories depending on the effect they have, such as the ability to stabilise or bring upon the active or inactive conformations, on said GPCR.

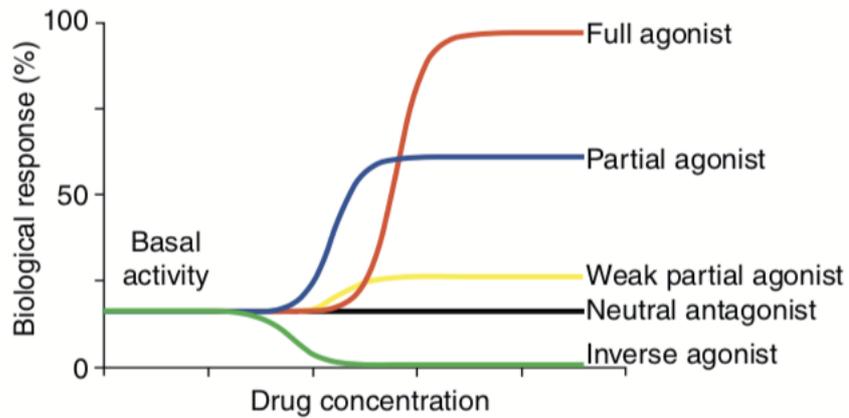


Figure 5 - Example of graph trends demonstrated by ligands of different pharmacological classes. Source: Tate, 2012⁴¹.

It is worth noting, that while neutral agonists do exist (e.g SB 242084 in the serotonin 2C receptor⁴²) and their biological response is represented in Figure 5, they are rare in β -ARs as most ligands that demonstrate antagonistic activity, tend to also show either weak agonistic or inverse agonistic behaviour (capable of inducing an effect opposing those of agonists)^{15,41,43}. The graph is also only a demonstration of the activity when a ligand is bound to the 'orthosteric' site of a GPCR. The orthosteric site is the site at which endogenous ligands, regardless of activity, bind. Allosteric binding is also observed in GPCRs and is thought to play a role in modulating binding at the orthosteric site^{44,45}.

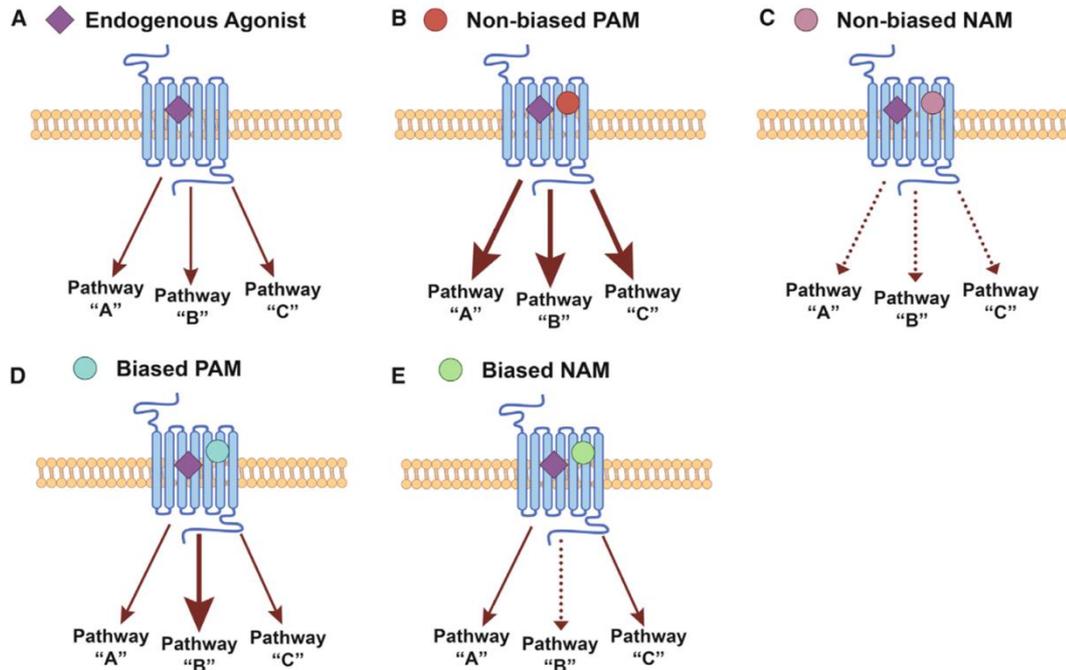


Figure 6 - Examples of pathways activated based on conformations induced by different types of ligand binding (A). Non-biased ligands can equally potentiate (B) or have inhibitory effects (C) on all signalling pathways traditionally induced by an agonist. Some positive allosteric modulators (PAMs) and negative allosteric modulators (NAMs) can have selectively modulate signalling pathways without affecting others by conferring bias (D and E)⁴⁶. Source: Foster et al., 2017⁴⁷.

Ligands that bind to GPCRs may be agonists and antagonists for more than one signalling pathway. An example of this is when antagonists activate arrestin mediated pathways while inhibiting the cAMP pathway^{43,48}. Several ligands have stabilising effects on different conformations that are each known to activate different signalling pathways. This is known as 'ligand-directed signalling bias'⁴⁸⁻⁵¹.

Current research has shown that some compounds including zinc could be allosteric modulators in β -ARs⁵²⁻⁵⁵. In addition, the presence of other potential low affinity allosteric sites in β -ARs have complicated matters more^{56,57}. It is also suggested that the low affinity allosteric site may be in close proximity to orthosteric site of rhodopsin family GPCRs⁵⁸. Dimerization could be a heavily influential factor in conformational change and consequently, pharmacological activity^{47,59-61}. This is discussed in greater detail in Section 1.6.

1.5 Binding kinetics and pathways

Kinetics, also known in chemistry as the study of the rate of chemical reactions, is the quantitative description of the speed of a chemical reaction. Kinetics aims to identify what factors affect key aspects of reaction pathways. Some pathways can involve just simple one-step mechanisms, and others feature complex, multi-step mechanisms. Through this study of kinetics, it is possible to identify the most important (generally the slowest) steps in a complex process.

Kinetics has recently captured the interest of those in the drug discovery discipline, mainly because of its potential significance in the prediction of drug efficacy in vivo and in humans by measuring unbinding kinetics^{62–64}. Mechanistically, drug binding and unbinding can be thought of as a multistep reaction mechanism, both on and off rates known as the macroscopic k_{on} and k_{off} respectively, could be influenced by microscopic rate constants associated with the binding/unbinding processes, isomerization, protein conformational changes etc. The overall lifespan of a receptor-ligand complex is the result of both the association rate (k_{on}) and the dissociation rate (k_{off}). The residence time of a drug (the time a drug resides within its target = $1/k_{off}$) is increasingly considered a key parameter for optimization due to evidence that it can be used to predict the in vivo drug efficacy^{65,66}.

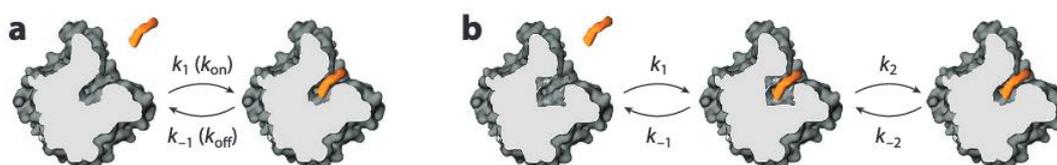


Figure 7 - Schematic of protein-ligand binding and associated rate constants. (a) A single-step drug-target binding interaction characterized by the association and dissociation constants $k_1(k_{on})$ and $k_{-1}(k_{off})$. (b) A more general, complex representation of the association process. Receptor conformational changes and changing of other properties are considered. This multi-step mechanism of the on and off rates are the product of the microscopic constants k_1 , k_{-1} , k_2 , and k_{-2} . Source: Bernetti M. et al., 2019⁶⁷.

1.5.1 Theory of kinetics

Historically, rate process theories regarding predicting kinetic rates have primarily focussed on the height of separating energy barriers. The linear correlation between the logarithm of reaction rates and the inverse temperature was identified by prominent works in the area^{68,69}. The overcoming of these high energy barriers led to theories describing fluctuations, which had itself derived from the idea that leaving a metastable state could have been due to noise-assisted hopping. Farkas⁷⁰ was the first to describe the expulsion rate from a metastable state as a flux of particles that escape through the bottleneck separating products and reactants.

1.5.2 Experimental approaches to investigating kinetics

There are several strategies for investigating kinetics experimentally (not all are discussed here), labelled methods being one of the more widely used approaches⁷¹.

There are two predominant types of labelled approach, radiolabelling and spectroscopic labelling. In radiolabelled techniques, the molecule of interest is tagged with a radioactive isotope. The isotopes emit radiation during decay and can therefore be detected. This approach to kinetics investigations is not limited to this cause, the approach can also perform competition and saturation assays⁷², capable of determining binding affinity as K_d , and the equilibrium inhibitor constant K_i . Radioligand binding can be further divided into two categories, direct and indirect, where the direct approach (the most straightforward approach) monitors the concentration of the radioligand bound to the receptor of interest over time, allowing for measuring of the association and dissociation constants (k_{on} and k_{off})⁷². The indirect approach uses an unlabelled ligand of interest, and the kinetic rates are determined by the competitive binding of a radioligand.

Spectroscopy approaches utilise fluorescently tagged ligands. Fluorophores emit at characteristic and generally longer wavelengths after electromagnetic radiation is absorbed and diminishes once the exposure is stopped. The fluorescence is typically observed as the ligand reaches the bound state. Fluorescent resonance energy transfer (FRET) uses a fluorescent group present on the receptor that is capable of transferring energy via a nonradiative mechanism to the acceptor label on the ligand^{73,74}. The fluorescent tag on the ligand then emits a fluorescent signal.

1.6 GPCR dimerization

The full extent of the role of GPCR dimerization is yet to be understood, however studies have shown that dimerization can affect pharmacology, signal transduction and cellular trafficking. As a result of this, some recent efforts at designing receptor selective ligands have involved bivalent or monovalent dimer-specific ligands, that can ultimately result in more selective or potent ligands, reducing side-effects⁷⁵.

1.6.1 GPCR dimerization and mechanisms

The class A GPCRs are typically described as monomers, however recently there is growing evidence from a broad range of approaches that it is possible that they can form dimers and/or higher-order oligomers. These structures can contain two of the same protomer (homodimers/oligomers) or with other GPCRs of the same sub-family or that respond to different ligands (hetero dimers/oligomers). When dimerization is discussed within this thesis, it is in the context of homodimers.

As mentioned earlier, there is still much to be understood about the functional role, molecular basis of, the extent and even the existence of class A GPCR

oligomerization⁷⁶. The lack of knowledge is due to several reasons including that many investigations have been conducted with inadequate controls, many investigations were limited to being done in simple transfected cell systems, and although structurally similar in that they possess seven linked transmembrane domains, there is no single pattern of structural organization or with equivalent self-avidity. In addition, many studies are qualitative and do not report on the proportion of receptors that might be present as dimers or oligomers at steady-state or how this could be regulated. Expression levels of GPCRs vary in different physiological environments, from a few hundred per cell in some, and several times this number in the central nervous system⁷⁷. Therefore it is suggested that expression levels alongside intrinsic affinity could determine the extent of interactions via mass-action^{78,79}. Nenasheva et al., in 2013 using total internal reflection fluorescence microscopy, observed the M2 receptor (M2R) of Chinese Hamster Ovary cells exist as a monomer, but was able to form reversible dimers at the plasma membrane⁸⁰.

There are currently two proposed mechanisms of dimerization: contact dimerization and domain-swapping dimerization. The former is also known as 'lateral packing' where the individual protomers quaternary structures are preserved, and the latter comprises of protomers that individually separate and recombine between each other⁸¹.

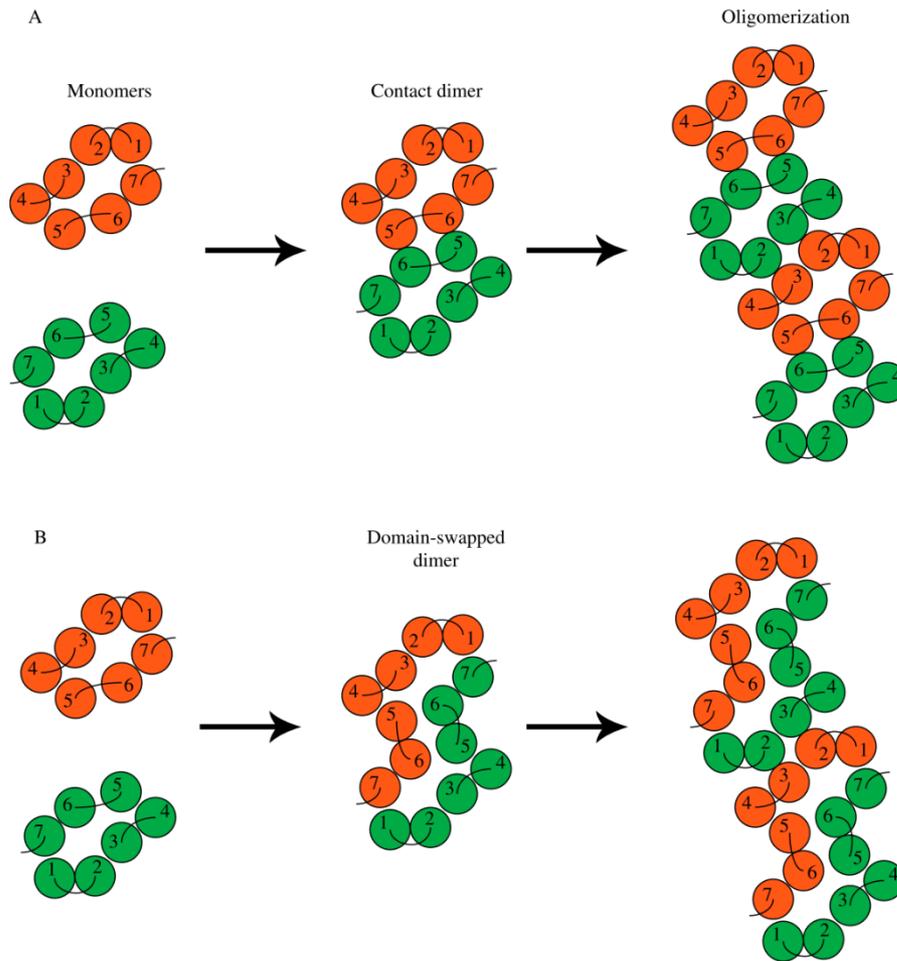


Figure 8 - Proposed methods of GPCR oligomerization. (A) Contact dimer – protomers make contact via exterior residues of the transmembrane helices. (B) Domain-swapping – Independently folded transmembrane helices are swapped between monomers. Source: Szidonya et al., 2008⁸².

1.6.2 Biochemical approaches to detecting GPCR oligomerization

The most frequently used method of detecting oligomerization in GPCRs is co-immunoprecipitation, where differentially epitope-tagged molecules are expressed in recombinant systems. The first receptor studied with this technique was the β_2 -AR tagged with influenza hemagglutinin (HA)- and myc-tags were also expressed in sf9 cells. Once the immunoprecipitation with an anti-myc antibody was complete, the dimer was detected using an anti HA antibody⁸³. Selectivity is controlled by co-expressing myc-tagged M2 Muscarinic Ach receptors with the HA- β_2 -AR, the two could not be co-immunoprecipitated.

The technique works by solubilizing cells expressing the two receptors, the lysate is incubated with an antibody against one of the receptors or epitope tag fused to one of the receptors. The complexes formed are then bound to an appropriate medium, electrophoresed, and then blotted. They are then visualised using an antibody against the desired receptor or its tag, which then shows GPCRs which form oligomers.

1.6.3 Resonance energy transfer approaches

Resonance energy transfer (RET) methods sensitively detect protein-protein interactions in real-time in live cells, making it possible to observe the complexes kinetics and dynamics⁸⁴.

RET is the nonradiative transfer of energy from a donor molecule to an acceptor molecule via electro-magnetic dipole-dipole coupling⁸⁵.

The two major methods are fluorescent RET (FRET) and bioluminescence RET (BRET). In the former, the energy donor is a fluorescent molecule that becomes excited by being exposed to light of a characteristic wavelength. This energy is then transferred to an acceptor molecule. In the latter, an enzyme is used as a donor molecule, Renilla luciferase (Rluc), which, when its substrate becomes oxidized (coelenterazine to coelenteramide), transfers energy to a fluorescent acceptor molecule. Both approaches are demonstrated in Figure 9.

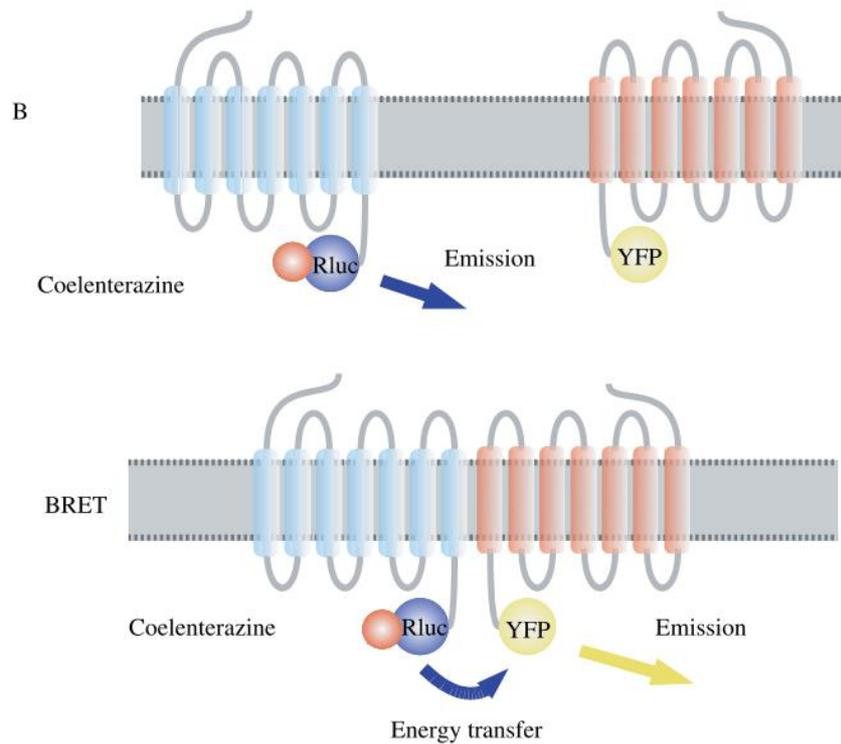
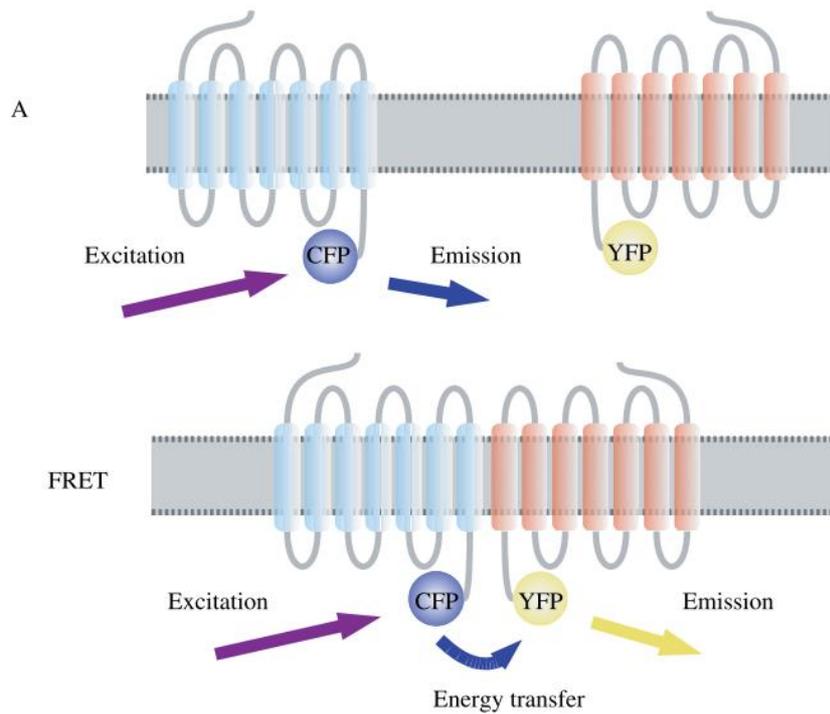


Figure 9 - Depiction of resonance energy transfer methods. (A) FRET – upper panel – A donor fluorophore becomes excited by external light source; blue light is then emitted in the absence of an acceptor molecule. Lower panel – In the presence of an appropriate acceptor molecule, a yellow shifted light is emitted. (B) BRET – upper panel – Donor molecule (Rluc enzyme) emits blue light upon the oxidation of its substrate. Lower panel – In the presence of an acceptor molecule, the resonance energy is transferred, which results in the emission of light from the acceptor molecule. Source: Szidonya et al., 2008⁸².

1.6.4 GPCR dimer functions

The idea of receptor dimerization/oligomerization has been supported not only by investigations into their formation, but also observing their functions. There are five predominant functions that receptor oligomerization have a proposed impact in⁸⁶: ontogeny, ligand-promoted regulation, pharmacological diversity, signal transduction and internalization.

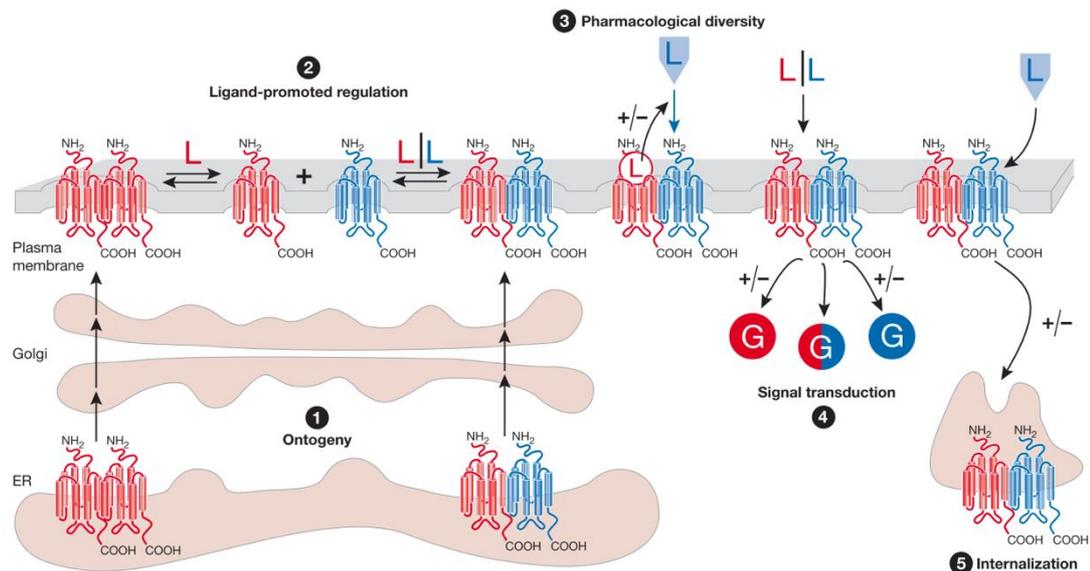


Figure 10 - Proposed roles of GPCR dimerization. (1) There is evidence that dimerization plays a role in the maturation of receptors and facilitates the transport from the endoplasmic reticulum (ER) to the cell surface. (2) Dimers could also be dynamically regulated by ligands at the plasma membrane. (3) Heterodimerization could also positively (+) or negatively (-) impact ligand binding cooperativity. (4) Signal transduction or prevention thereof during signalling or changing G-protein selectivity. (5) Co-internalization could also be promoted or inhibited (when accompanied by an agonist-promoted endocytosis resistant protomer) by heterodimerization after stimulation of one protomer. Source: Terrillon et al., 2004⁸⁶.

Ontogeny

GPCRs exit from the endoplasmic reticulum (ER) is a crucial step in controlling their surface expression⁸⁷. Petaja-Repo et al., showed that receptors that had failed to

fold correctly were retained and then eventually degraded⁸⁸. It's therefore important that this step is regulated. An example of this type of regulation is with the γ -aminobutyric acid b receptor (GbR), which is made up of two subunits, a GbR₁ and a GbR₂⁸⁹. GbR₁ when expressed alone is retained in the ER due to its carboxy-terminal ER retention motif, GbR₂ at the cell surface has no function. However, when co-expressed, the GbR₂ masks the retention motif in the GbR₁ which allows transport of a fully functioning GbR heterodimer.

Ligand-promoted regulation

It is also suggested that a dimer could be dynamically regulated by ligands, by promoting or inhibiting the dimer formation. However, this area of dimer behaviour remains unclear. There have been studies to show ligands promoting⁹⁰, inhibiting⁹¹, and showing that the dimer formation process is independent of ligand presence⁹².

Pharmacological Diversity

Jordan and Devi in 1999 found that the dimerization of the δ - and κ -opioid receptors could have a role in pharmacological diversity. The co-expression of the two receptors demonstrated a low affinity for the δ - and κ -selective ligands separately, but an increased affinity of the two when they are both present⁹³. The full extent of pharmacological diversity within GPCR dimers are still unclear, but studies have been shown to demonstrate both positive and negative cooperativity⁹⁴⁻⁹⁶.

Signal Transduction

GPCR dimerization evidence in the form of signal transduction within the GbR receptor was first convincing case for their existence. It was shown by Margeta-Mitrovic et al., 2000 and Galvez et al., 2001 that although GbR₁ expressed the binding site for γ -aminobutyric acid (GABA), GbR₂ was necessary for the functionality of the G-protein signalling cascade^{94,97}. It was shown that their co-expression wasn't only necessary for the GbR₁ to reach the cell surface from the ER, but it was necessary for signal transduction.

Internalization

Agonist-promoted endocytosis of GPCRs has been suggested to be affected by heterodimerization. It has been documented that stimulation of only one of the protomer was required to promote the co-internalization of both receptors^{95,98-100}.

1.6.5 GPCR dimers as therapeutic targets

Bivalent ligands have become a means of interrogating dimer function or promote the dimerization process¹⁰¹. They are comprised of two pharmacophores covalently linked with a spacer, allowing simultaneous interaction with two protomers^{102,103}. Bivalent ligands containing two identical pharmacophore units are referred to as homobivalent, whereas those that contain different pharmacophore units are referred to a heterobivalent. This class of ligand has been a growing area of interest in the field of GPCRs^{75,104}.

Shonberg et al., 2011 also suggested that pharmacophores to be used in bivalent ligands should possess a low to medium molecular weight because the addition of the linker will increase the molecular bulk significantly¹⁰⁴. It was also mentioned that the chosen pharmacophores should already display high affinity and subtype

selectivity in the low nanomolar range. Ensuring favourable binding and selectivity profiles for the bivalent counterparts¹⁰⁴.

Work by Decker et al., 2009 and Zhang et al., 2011 suggested that the ideal moieties to link the pharmacophores are hydroxyl, amine and carboxylic groups^{105,106}. It is important that linkers are added without hindering the affinity or potency of the pharmacophore's functional groups.

Spacer length is also a vital aspect in the design of bivalent ligands. A limiting factor of some bivalent ligands that have a linker length too short is the ability for it to span to the second protomer. However, this is an ideal feature in bitopic ligands capable of interacting with a receptors orthosteric and allosteric sites simultaneously. In opioid receptors, it is found that an optimal spacer length between pharmacophores is in the 18 - 25 atoms range¹⁰⁷⁻¹¹¹. This range is a general indication but is not necessarily true for other dimer interfaces and species. The length must also be taken into consideration with flexibility, as the pharmacophores ideally should have conformational mobility to bind within the binding pockets of the protomers. However, rigidification of the spacers could improve knowledge of the required shape to occupy the binding pockets while reducing the entropic penalty for the conformation restriction upon binding⁷⁵. Lastly the spacer polarity will have a significant impact on the solubility of the ligand. Long hydrocarbon spacers run the risk of becoming insoluble especially if the pharmacophore is also hydrophobic. To counter this issue, many spacers contain polyamide or polyethylene glycol (PEG)^{112,113}. Spacers also have the potential to affect the ligands overall binding affinity and function, and therefore it is important to understand the possible effects of the spacer on the receptors.

Therefore, by finding an alternative route for bivalent ligands that could reduce the spacer length significantly could also reduce the ability for the spacer to have major undesirable effects and physical attributes regarding ligand properties, leading to a potentially new approach to ligand bivalency in GPCRs.

1.7 X-ray crystallography of GPCRs

1.7.1 Methods in crystallising GPCRs

Since the crystal structure of rhodopsin was solved, it was seven years before the first human GPCR (β_2 -AR) with a diffusible ligand was solved.¹⁰ The crystallisation of membrane proteins is challenging, some of these challenges include the availability of polar surfaces of receptors for crystal contacts, the availability of functional protein, the inherent conformational flexibility and low expression levels in hosts.¹¹⁴ However, as mentioned in section 1.1 G Protein-Coupled Receptors, vast improvements have been made in the visualisation of GPCRs and their signalling complexes at a structural level.

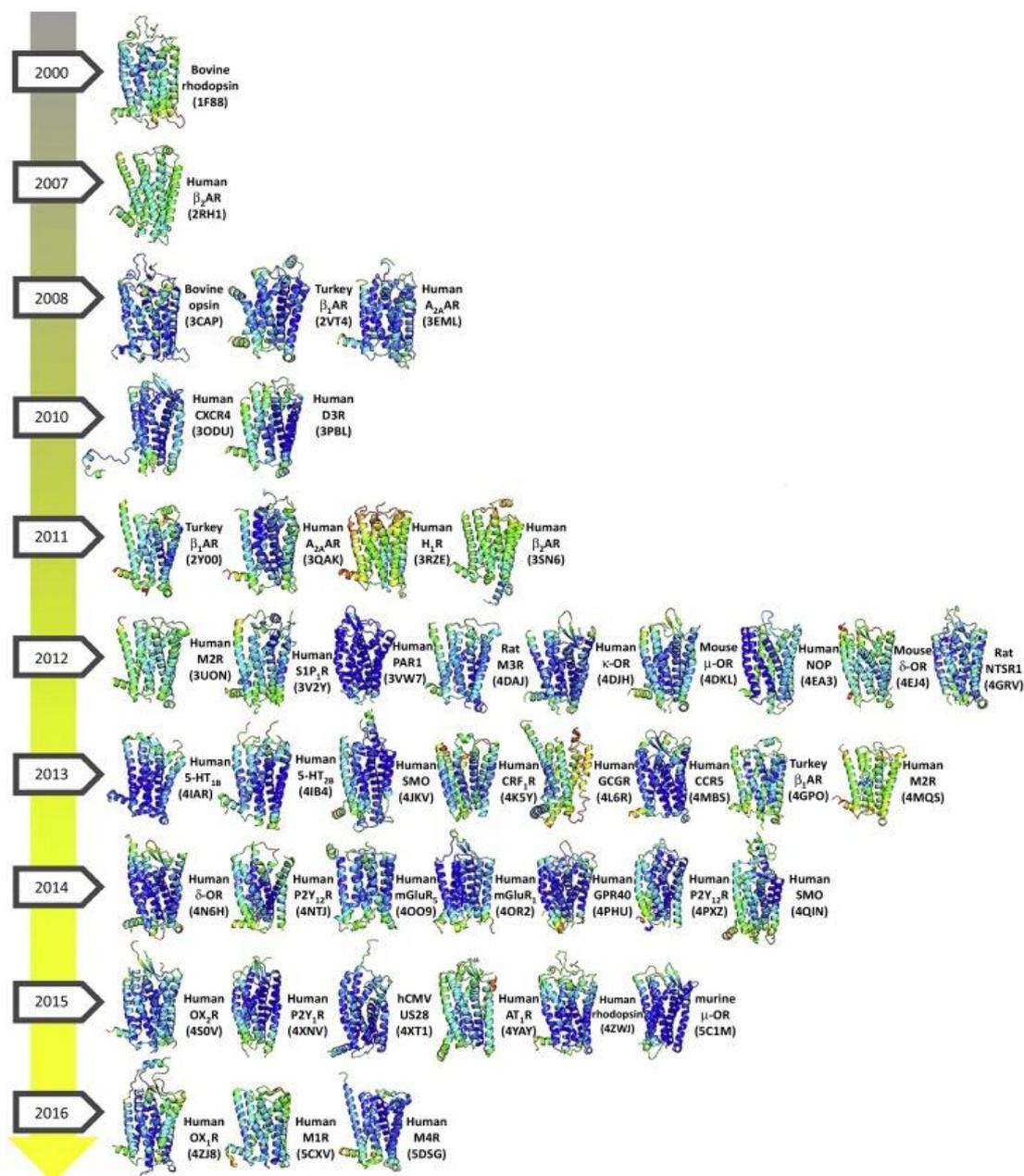


Figure 11 - Timeline of the first crystal structures of each GPCR type discoveries between 2000 – 2016. PDB codes are denoted in parentheses below the GPCR type name. Source: Xiang, J. et al., 2016¹¹⁵.

GPCRs are known to have a high degree of flexibility in their structure and conformation, essential for receptor coupling and signalling methods. Being membrane bound means that the hydrophilic components of the protein that are required to make lattice contacts are relatively small. This can also make the crystallisation of GPCRs more difficult.¹¹⁵ Truncation of the long, flexible loop regions can decrease the polar surface contacts. Replacing this truncated

component with smaller, more stable crystallisable protein domains (e.g. T4 lysozyme (T4L) and b562RIL (BRIL)) can improve crystal contacts.¹¹⁵ In addition to the fusion approach, the addition of nanobodies and antibodies have also been done to reduce polar surface area and improve stability.¹¹⁵ Another approach that is used is thermostabilisation, also known as the StaR (stabilised receptor) approach.¹¹⁶ The method was not used exclusively for thermostabilisation, but also for improved protein expression.¹¹⁵ Thermostabilisation can also be used in conjunction with the other methods, as is being increasingly for solving new structures.

Extracting membrane proteins from their lipid membranes is required before purification, therefore the use of a suitable detergent is used to achieve this. N-dodecyl- β -D-maltoside (DDM) and cholesteryl hemisuccinate (CHS) are a combined detergent frequently used to extract GPCRs from their membranes.¹¹⁵ The use of a ligand within the GPCR is typically used for constraining the receptor at a specific conformation and therefore conserving functional stability. In the absence of a ligand, GPCRs can take multiple conformations, increasing the degree of heterogeneity and therefore hindering successful crystallisation. Nucleation is the initial step in crystallisation and crystal packing, which can be achieved more easily when GPCRs adopt one major population.¹¹⁴ Most crystallised GPCR structures to date are in complexes containing a high affinity ligand. Covalent ligands are often used as they do not diffuse away from the sample, and therefore assisting in the crystallisation of the agonist-bound conformation, which are known to be less stable.²⁴ StaRs can be used without the presence of such potent ligands, as the conformations are trapped by mutagenesis.²⁴

Vapour diffusion can be used to crystallise membrane proteins. As mentioned previously in 1.1 G Protein-Coupled Receptors, LCP can also be used for the crystallisation of membrane proteins. The LCP method could be used to provide crystals of GPCR-ligand complexes in a more amenable and well crystallisable conformation compared to vapour diffusion, where the large size of detergent

micelle could occlude small hydrophilic surface areas of GPCRs and therefore restricting the formation of crystal lattice contacts.¹¹⁵

LCP (also known as *meso*-crystallisation) methods allow membrane proteins to be embedded in membrane-mimetic lipid environments that interact with each other, thus improving crystal contacts with both hydrophobic and hydrophilic areas of the receptor.¹¹⁵ LCP is capable of mimicking lipid bilayers, therefore higher protein stability can be achieved compared to using detergent micelles. Although LCP methods are widely used in the crystallisation of GPCRs, it is worth noting that thermostabilised GPCRs can be crystallised in detergent micelles using vapour diffusion.¹¹⁴

The latest method used in the crystallisation of GPCRs is X-ray free electron laser (XFEL), known as femtosecond crystallography.¹¹⁴ Using microcrystals can be used to overcome the difficulties of having to obtain high-quality single crystals, which are often hard to obtain. Advanced synchrotron sources provide tailored microfocus beamlines that can overcome the radiation sensitive nature of microcrystals. XFEL exposes the microcrystals to ultrashort and highly intense X-ray pulses allowing for collection of diffraction data from single points on the microcrystal.¹¹⁴ Primary advantages of such method is the ability for the microcrystal to evade radiation damage and the increased performance of data collection at room temperatures, allowing recreation of the native state conditions.

1.7.2 β -adrenergic receptor crystal structures

The adrenergic receptors can be grouped into two main classes α and β . Each of these groups have their own subgroups, α_1 , α_2 , β_1 , β_2 and β_3 . Therapeutically, antagonists of β -adrenergic receptors are referred to as ' β -blockers' and are used in medicine to treat conditions including cardiovascular, migraine and anxiety. The agonists are typically used to treat respiratory-related diseases. There have been

more than 35 crystal structures of β -adrenergic receptors reported thus far, providing high-quality insight to their ligand-receptor interactions.

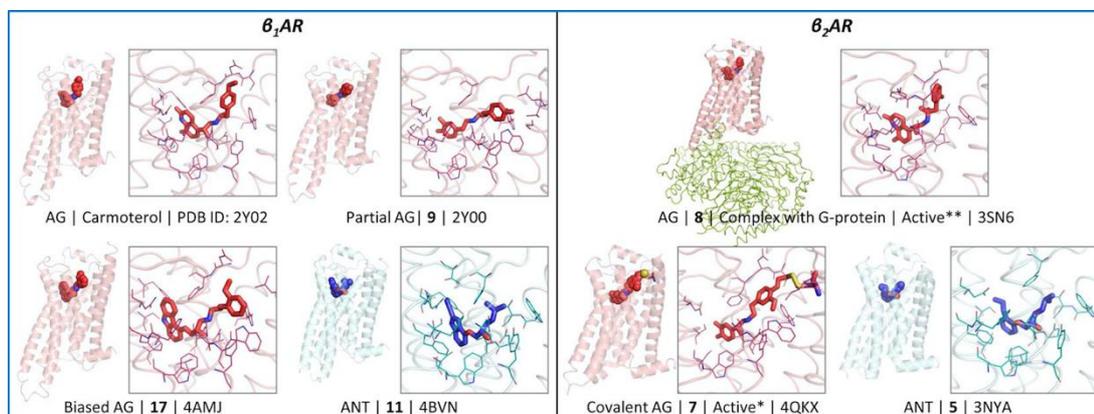


Figure 12 - Ligand binding pockets in β_1 - and β_2 -adrenergic receptors. Agonist- and antagonist-bound structures are displayed in pink and cyan ribbons, and bound ligands are represented in sticks or spheres. Residues within 4 Å from the bound ligand are represented with sticks. * - Active state conformations stabilised by G protein mimicking antibody, ** - fully active structure coupled with G protein. Source: Lee, Y et al., 2018²⁴.

Of the adrenergic receptors (and the aminergic receptors) β_2 -AR is the most well studied from a structural perspective, due to the number of structures available in its inactive, fully active, and partially active state with a G protein. In a high resolution β_2 -AR-T4L, bound to a high-affinity inverse agonist was solved at 2.4 Å.¹⁰ It was discovered that the inverse agonist made hydrogen bonds with S203^{2,42}, D113^{3,32}, N312^{7,39} and Y316^{5,32}. Hydrophobic interactions were also seen with V114^{3,33}, F290^{6,52} and F193^{5,32}. The hydrogen bond interactions involving D113^{3,32}, N312^{7,39}, and Y316^{7,43} are preserved in the interactions between most known ligands and β_2 -ARs, but additional hydrogen interactions and hydrophobic interactions vary depending on the chemical structure of the ligand, which could influence the strength of inverse agonism.²⁴

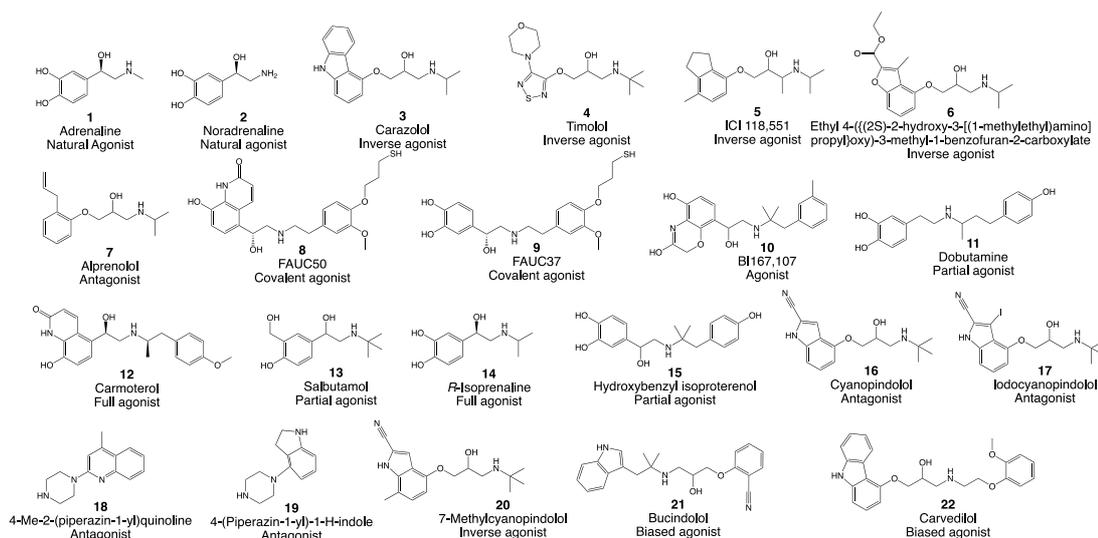


Figure 13 - Chemical structures of drugs and endogenous ligands of β -adrenergic receptors, some of which have been used in obtaining crystal structures. Source: Lee, Y et al., 2018²⁴.

The inactive structure of avian β_1 -AR in a complex with the agonist cyanopindolol (**16**)²⁶ was determined in 2008. Inactive crystal structures of avian β_1 -AR bound to iodocyanopindolol (**17**), arylpiperazine derivatives **18** and **19**^{26,117}, and inverse agonists **3** and **20** (7-methylcyanopindolol)^{26,118} were also obtained. The agonist bound β_1 -AR structures in a complex containing several full and partial agonists was also discovered, but displayed inactive state-like receptor conformations.¹¹⁹ The active state of β_2 -AR demonstrated a significant outward movement of TM5 and TM6, which was absent in β_1 -AR. This difference is said to potentially be caused by agonist binding in β_2 -AR that imparts 1 Å contraction in the orthosteric binding pocket which is associated with the conformational changes in the S212^{5,43} and S215^{5,46} residues. These interactions weaken TM4-TM5 interactions while strengthening TM5-TM6 interactions, which could lead to these differences in helices movement between the two subtypes.²⁴ The three ECLs are almost identical between the two subtypes, along with most of the extracellular surface, allowing access for the diffusible ligands to the ligand binding pocket (ECL2, TM3, TM5, TM6 and TM7).

The thermostabilised avian β_1 -AR was found using the biased agonists **21** and **22**.⁵⁰ The biased agonist's head groups (defined in Figure 14) adopted a similar

conformation to the β -AR that were antagonist-bound. The extended ligand binding site comprising of residues forming TM2, TM3, TM7 and ECL2 also elucidate the structural requirements of biased ligands. In 2013¹²⁰, Huang et al., reported a basal, ligand-free state of avian β_1 -AR, which would've been the first, class A non-rhodopsin structure to have been crystallised in its ligand-free state. However, in 2015 Leslie et al.,¹²¹ repeated the structure deduction and discovered significant unmodeled electron densities for a ligand occupying the orthosteric site.

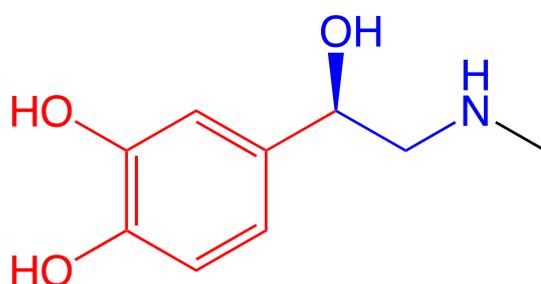


Figure 14 – Ligand structural component definitions using adrenaline (**1**) as the base. The ethanolamine core (backbone) is coloured in blue, while the variable head group is coloured in red, and the variable tail group is coloured in black.

There is a 67% similarity between β_1 -AR and β_2 -AR in the transmembrane regions. As a result, the structures are similar in their inactive states. A major similarity between the two receptors is their orthosteric binding pocket. The ammonium salt bridge formation between the ligand and D^{3.32} and two hydrogen bonds (between the aromatic head group of the ligand and S^{5.42} and the aliphatic OH-group of the ethanolamine backbone of the ligand and N^{7.39}) (Figure 15).

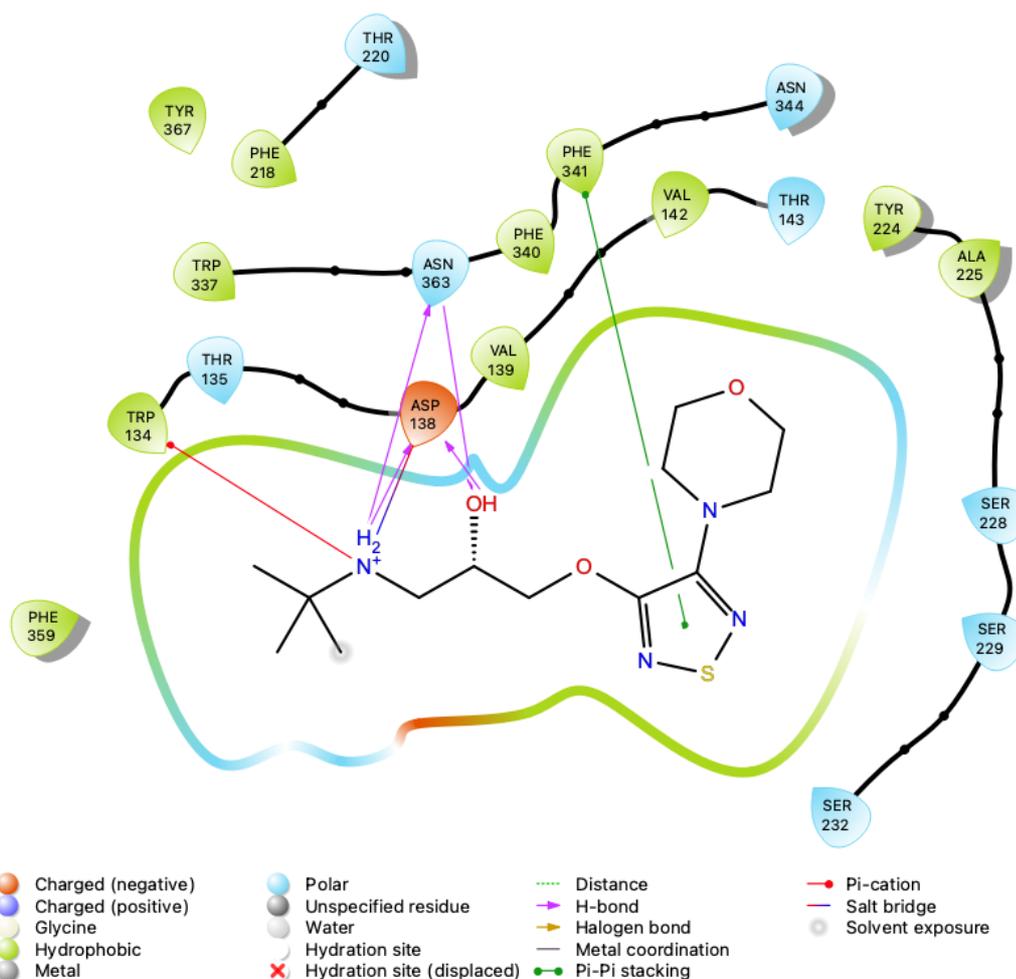


Figure 15 - Diagram showing interactions between timolol (**4**) and the β_1 -AR crystal structure. Key is displayed below schematic, explaining each interaction. The diagram clearly shows the hydrogen interactions between the ethanolamine backbone and the Asp^{3.32} and Asn^{7.39} residues, along with the salt bridge between the charged nitrogen and Asp^{3.32}. Image generated using Schrodinger Suite Maestro.

ELC1-3 regions show very high variability between the two subgroups, and it is believed that the difference in sequence at ECL2 could have a major impact on ligand specificity. Although investigations revealed similarities in conformation with ICL1 between the subtypes, there were distinct differences between ICL2 and ICL3. Differences in ICL3 between the subtypes is due to the presence of insertions and stabilising mutations. In the ICL2 region of β_1 -AR, an α -helix was observed, whereas in β_2 -AR, a loop was observed. An open conformation of the ionic lock was observed in both receptors. The ionic interactions of the D(E)RY motifs mentioned previously, are believed to stabilise the conformation of GPCRs,

however, antagonist-bound β_1 -AR and inverse-agonist bound β_2 -AR challenge this belief due to the ionic lock being absent in these complexes, but present in its ligand-free basal state.²⁴

1.8 β -Adrenergic receptor ligands

1.8.1 GPCR ligand classification

Most ligands resulting from traditional medicinal chemistry and drug discovery methods have the chemical foundations of endogenous substrates present within the body that activate GPCR signalling pathways (agonists) or bind to the orthosteric site without inducing an activating effect (antagonist). Inverse agonists illicit a reverse response to the classical agonist response and therefore demonstrates negative efficacy.

The advances in our knowledge of receptor structure and function have unveiled other potential targets of GPCRs.¹²² Allosteric ligands that can make endogenous ligands more potent are known as positive allosteric modulators (PAMs), those that reduce potency are negative allosteric modulators (NAMs), those that make no change to potency are silent allosteric modulators (SAMs) and those that have potency of its own are known as ago-allosteric. Bitopic ligands act by targeting both the allosteric and orthosteric binding sites, and bivalent ligands are those that bind simultaneously to adjacent receptors. Biased ligands are increasingly being used to stabilise receptor conformation due to their ability to selectively activate one signalling pathway over another.

1.8.2 Orthosteric binding

Ligands that bind to the orthosteric site of the hepta-helical GPCRs, bind towards the extracellular side. The receptors display a vast flexibility in their ability to bind

to ligands of various sizes, structure, and that elicit different responses. However, this flexibility varies significantly between receptor subtypes. Most ligand-contacting residues are on the inner side of the transmembrane helices with the exception of peptidic ligands.²³ TM3, TM6 and TM7 make the most interactions with the ligands in most of the receptors. TM1 however, does not interact directly with the ligand. The residues that are involved in consensus ligand contact in class A GPCRs include 3.32, 3.33, 3.36, 6.48, 6.51 and 7.39. The residue pairs 3.36-6.48 and 6.51-7.39 are seen to form inter transmembrane contact (Figure 16).²³ The area between TM3, TM6 and TM7 where the ligands interact is sometimes referred to as the ‘ligand binding cradle’. In addition to these residues, water molecules are present in several structures mediating indirect interactions between the ligand and the receptor.²⁴ Mutagenesis studies in the β -ARs have shown that replacing these residues with other amino acids decrease ligand binding.^{39,123}

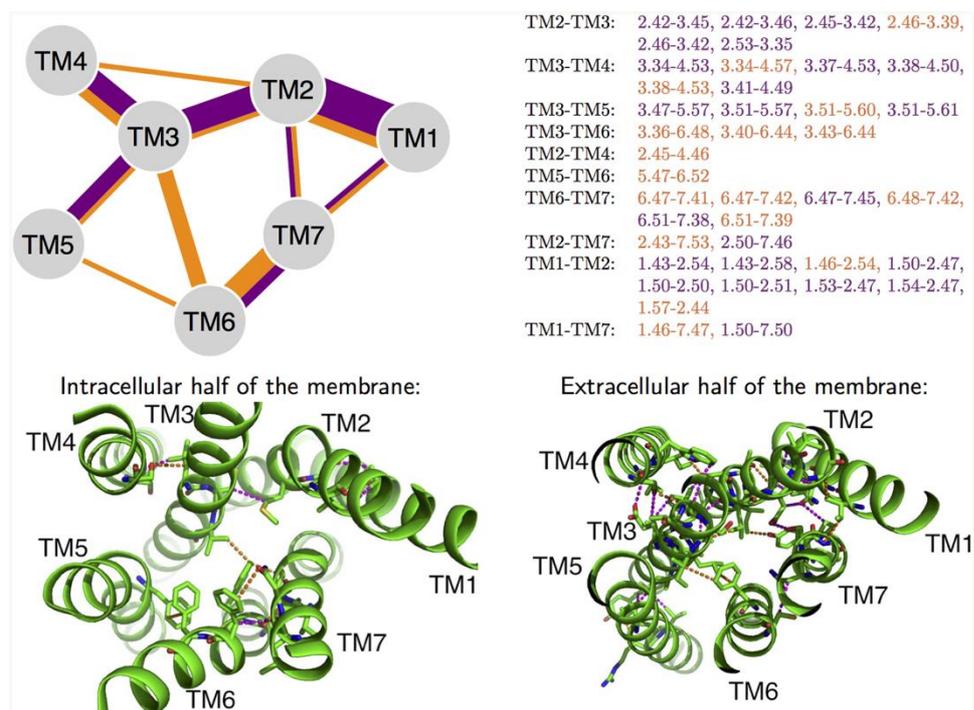


Figure 16 - Top left: Schematic of 40 conserved inter-helical contacts (CHICOs) present in the majority (23 out of 24) studied class A structures. Purple denotes contacts present in all classes, while orange denotes contacts present in class A only. Top right: Interactions shown in Ballesteros-Weinstein numbering. Bottom: Extracellular perspectives of the contacts present within the β_2 -AR crystal structure, showing both the

intracellular half and extracellular half of the receptor on the left and right respectively. Source: Cvicsek, V. et al., 2016¹²⁴.

1.8.3 Biased ligands

By the mid-1990s, it was suggested that GPCRs are more complex than just being binary switches. Biased ligands are distinct ligands that activate one downstream cascade preferentially over another. This phenomenon became known as 'biased agonism'. Biased agonism can produce distinct physiological outcomes.^{125,126} This phenomenon can also be arise from a receptor favouring coupling to β -arrestin over a G protein. Most investigation into biased signalling focus on biased, orthosteric and allosteric ligands that bind at the extracellular region. However, recently, efforts to utilise biased GPCR signalling have been focused on intracellular targets.^{127,128}

1.8.4 Clinical indications of β -AR ligands

β -AR agonists and antagonists are of clinical significance; agonists are sympathomimetic and are often used in treating shock and post-cardiac surgery. Adrenaline (epinephrine) (**1**) and dobutamine (**11**) and other inotropes increase heart contractility and can transiently interact with α -ARs, increasing blood pressure by vasoconstrictor action. Adrenaline (epinephrine) (**1**) is also used in case of emergency in the event of allergic reactions and anaphylaxis. β -AR agonists including salbutamol (**13**) and terbutaline (**23**) (Figure 17) are used in the event of bronchospasm in asthma and chronic obstructive pulmonary disease (COPD). These agents are β_2 -AR selective and have a therapeutic significance due to their ability to stimulate smooth muscle dilation in the airways. Formoterol (**24**) and salmeterol (**25**) are commonly used daily for prevention of bronchospasm.

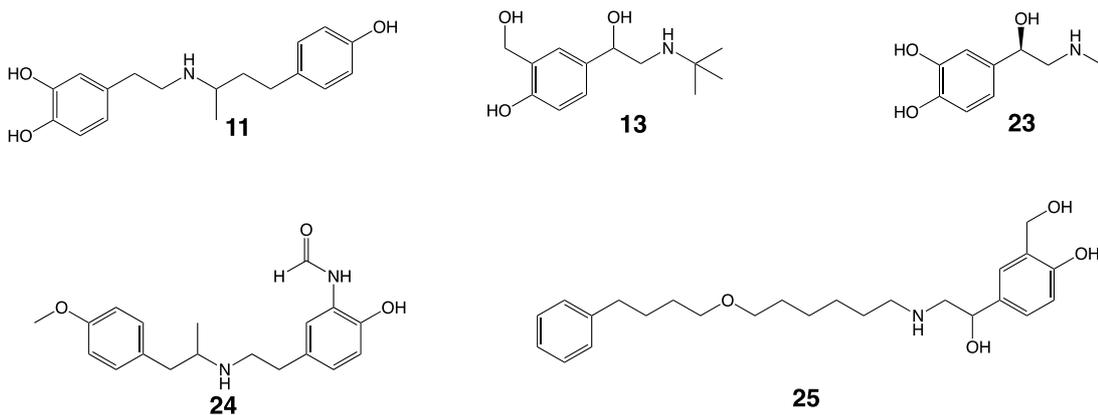


Figure 17 - Clinically used β_2 - agonists- Dobutamine (**11**), salbutamol (**13**), terbutaline (**23**), formoterol (**24**), salmeterol (**25**).

Between 1958 and 1964, Sir James Black discovered compounds that would be called ' β -blockers'. This class of drug would exhibit antiarrhythmic and antihypersensitive properties and thus reduce heart rate and cardiac output. Propranolol is an example that is still largely used today.

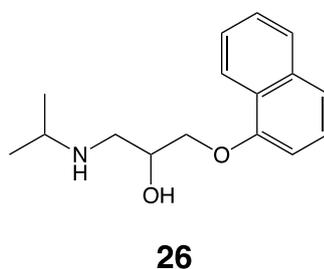


Figure 18 - Chemical structure of propranolol (**26**).

β -blockers are a diverse family of drugs that can range from being used for management of glaucoma, anxiety, migraine, tremor, thyrotoxicosis and a vast array of cardiovascular conditions (myocardial infarction (MI), angina pectoris, hypertension, heart failure and cardiac arrhythmia).^{56,129} It is also worth noting that β -blockers may have applications in malaria, osteoporosis and cancer.⁴³

Table 1 – Current indications for β -blockers. Source: Adapted, from British National Formulary 2018¹²⁹.

β -blocker	Clinical indication
Acebutolol §	Hypertension, angina, post-myocardial infarction (MI), arrhythmias
Atenolol ±	Hypertension, angina, post-MI, arrhythmias, migraine
Betaxolol	Glaucoma
Bisoprolol	Hypertension, angina, heart failure
Carteolol	Glaucoma
Carvedilol	Hypertension, angina, heart failure
Celiprolol § ±	Hypertension
Esmolol	Arrhythmias (short-term)
Labetolol	Hypertension
Levobunolol	Glaucoma
Metoprolol	Hypertension, angina, post-MI, arrhythmias, migraine, hyperthyroidism
Nadolol ±	Hypertension, angina, arrhythmias, migraine, thyrotoxicosis
Nebivolol	Hypertension, heart failure
Oxprenolol §	Hypertension, angina, arrhythmias, anxiety
Pindolol §	Hypertension, angina
Propranolol	Hypertension, angina, post-MI, arrhythmias, portal hypertension, anxiety, essential tremor, migraine, thyrotoxicosis
Sotalol ±	Arrhythmias
Timolol	Glaucoma, hypertension, angina, post-MI, migraine
§ Intrinsic sympathomimetic activity (ISA, partial agonist activity) ± Water-soluble β -blockers	

The primary therapeutic objective of β -blocking in cardiovascular disease is antagonism of the endogenous catecholamines at β_1 -ARs in cardiac tissue. This results in a reduction in myocardial workload, which in turn, alleviates symptoms of several clinical indications falling under cardiovascular disease. Atherosclerotic plaque is usually the cause of MI or heart attacks due to the fatal lack of blood supply caused by blockage. Mortality in conditions that fall under ischaemic heart disease is also improved by β -blocker therapy.⁴³ MI is closely linked to hypertension, or high blood pressure, which often precedes heart attack or stroke or other ischemic events. Increased blood pressure places strain on arteries or can lead to kidney conditions. Angina (*angina pectoris*) is characterised by chest pain caused by restricted blood flow to the heart. Irregular heartbeat (arrhythmia) is caused by the misfiring of electrical impulses. This can increase the risk of stroke due to lack of coordinated blood pumping and oxygen supply. Inefficient pumping of the heart (heart failure) can arise because of the previously mentioned conditions, or heart valve related problems. β -blocker therapy is often lifelong, and is usually a method of managing symptoms and secondary prevention of events.⁴³

Generally, β -blockers have poor selectivity, and are contraindicated in patients with respiratory diseases, most notably COPD and asthma.¹²⁹ Bronchospasm can be caused by the blocking of β_2 -AR, worsening symptoms. Non-selective antagonists can also work against agonist treatment.¹³⁰ This can lead to denial of treatment. The World Health Organization reports that cardiovascular disease is the leading cause of death worldwide, and COPD is to be the third leading cause of death by 2030.¹³¹

β_1 -ARs are almost exclusively found in cardiac tissue, and a highly selective β_1 -AR antagonist would potentially reduce the onset of adverse effects of current β -blocker therapy. As mentioned previously, some antagonists display weak partial agonism and therefore have intrinsic sympathomimetic activity (ISA) (Table 1). The agonism of these treatments can be detrimental in the case of migraine or heart failure.^{43,129} Water soluble β -blockers that have a low chance of crossing the blood-brain-barrier (BBB), could lead to patients having less sleep disturbance and nightmares, however, doses must be controlled in those with renal conditions.¹²⁹

1.8.4 β -blocker pharmacology

The clinical importance of β_1 -selective β -blockers has been a driving force for selectivity studies. These studies have shown that even the most selective β -blockers aren't significantly selective. Two examples of high selectivity ligands used in the clinical setting are bisoprolol (**27**) and nebivolol (**28**), which are only 13-fold and 14-fold more selective for β_1 - than β_2 -ARs in humans respectively.^{34,132,133}

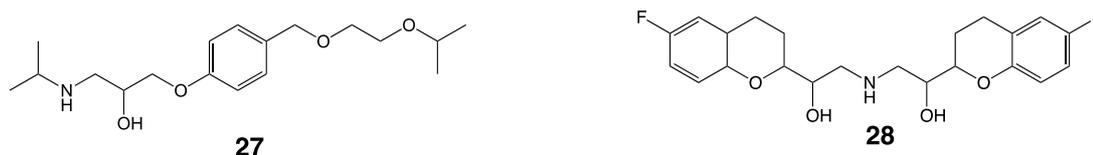


Figure 19 - Chemical structure of Bisoprolol (**27**) and Nebivolol (**28**).

Propranolol (**26**) is an example of a β -blocker that is slightly β_2 -selective. There are very highly selective ligands for both β_2 -AR (ICI118551 (**5**)) and β_1 -AR (CGP20712A (**13**)) that have become cornerstones of β -AR ligand research. The affinity of ICI118551 (**12**) for β_2 -AR is 550-fold larger than for β_1 -AR. CGP20712A (**13**) is 501-fold more selective for β_1 -AR than β_2 -AR.

Table 2 – Binding affinities of a selection of β -AR ligands in the β -ARs. Log K_D values were calculated from ^3H -CGP 12177 binding studies in three human β -ARs, the turkey β_1 -AR and the thermostabilised turkey mutant β_3 -6-m23 used for crystallographic structure determination. K_D is a measure of the dissociation constant to gauge binding affinity. This is the concentration of a ligand that when added displaces 50% of specifically bound radioligand. Receptors will favour high affinity ligands, therefore displacing the radioligands at lower concentrations, therefore giving more negative log K_D values. Source: Baker et al., 2005, 2011^{34,134}.

	Human β -ARs			Turkey β -ARs	
	β_1 -AR	β_2 -AR	β_3 -AR	β_1 -AR	β_3 -6-m23
Acebutolol	-6.46	-6.08	-4.41	-5.85	-4.72
Alprenolol	-7.83	-9.04	-6.93	-7.96	-7.35
Atenolol	-6.66	-5.99	-4.11	-5.40	-4.29
Betaxolol	-8.21	-7.38	-5.97	-6.90	-5.85
Bisoprolol	-7.83	-6.70	-5.67	-6.70	-5.38
Bupranolol	-8.51	-9.85	-7.04	-8.43	-7.61
Butoxamine	-4.85	-6.23	>-4.00	-5.24	-4.78
Carazolol	-9.69	-10.49	-8.35	-10.20	-9.23
Carvedilol	-8.75	-9.40	-8.30	-9.43	-8.72
CGP20712A	-8.81	-6.11	-5.11	-7.53	-6.38
S-Cyanopindolol	-10.39	-11.09	-8.36	-10.89	-10.01
ICI118551	-6.52	-9.26	-6.44	-7.20	-6.35
Labetolol	-7.63	-8.03	-6.18	-7.51	-6.48
Metoprolol	-7.26	-6.89	-5.16	-6.42	-5.27
Nadolol	-7.23	-8.60	-6.18	-7.99	-7.08
Nebivolol	-9.06	-7.92	-7.04	-8.25	-7.73
Pindolol	-8.57	-9.23	-7.08	-8.55	-8.01
Propranolol	-8.16	-9.08	-6.93	-8.51	-7.78
Sotalol	-5.77	-6.85	-5.85	-6.14	-5.21

Timolol	-8.27	-9.68	-6.80	-8.76	-7.82
---------	-------	-------	-------	-------	-------

Pharmacological investigations of β -ARs can be rather ambiguous, as the type of assay used can produce different definitions for a single ligand.^{135,136} Using different radioligands can give different binding affinities. LK204545 (**14**) (a highly selective β_1 -AR selective ligand) has been used as a lead compound in discovering other high affinity ligands.¹³⁷

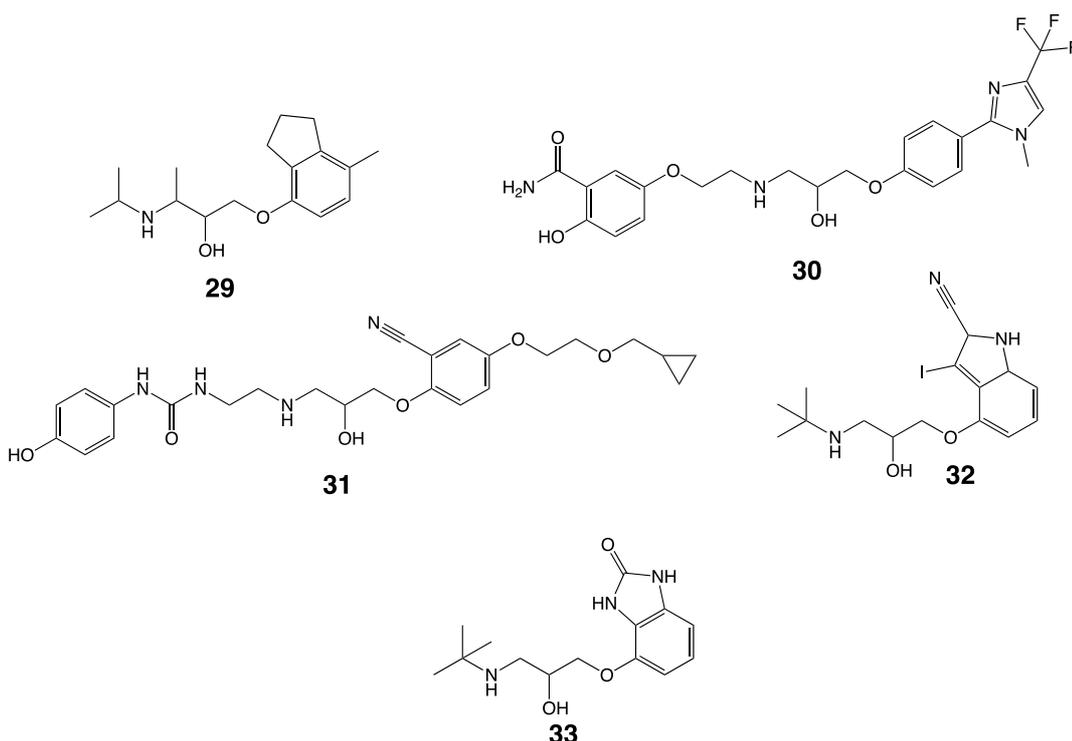


Figure 20 - Important β -AR ligands – ICI118551 (**29**), CGP20712A (**30**), LK204505 (**31**), Iodocyanopindolol (**32**) and CGP12177 (**33**).

Several studies have suggested that β_1 - and β_3 -ARs have a low affinity binding site in addition to their high affinity orthosteric site. CGP12177 (**33**) and carvedilol (**22**) are two ligands with high affinity binding to the orthosteric site, known to produce a biphasic concentration-response curve as a result of low affinity site binding.^{56,57}

1.8.5 Ligand interactions

Orthosteric binding in the β -ARs (including the turkey structure) occurs in a highly conserved binding pocket and are identical. Several crystal structures for the human β_2 - and turkey β_1 -ARs containing ligands of different pharmacological activity (agonists, antagonists and fragments) demonstrate a consistent method of binding with the receptors.^{12,119,138} β -AR ligands typically contain an ethanolamine core, with the nitrogen component being protonated at physiological pH.

In site directed mutagenesis studies, it was shown that Asp^{3.32} and Asn^{7.39} are fundamental in ligand binding throughout all human β -AR subtypes.^{139,140} The protonated nitrogen of the ethanolamine backbone typically forms a salt bridge with the Asp^{3.32}, and molecular dynamics studies have shown that it is this salt bridge that is seen to be the last interaction that is broken when the ligand leaves the receptor.^{48,141} Residues within the binding pocket are also seen to have stabilising effects on other residues, for example, Tyr^{7.43} is shown to have a stabilising effect on Asp^{3.32} via a hydrogen bond interaction that could also interact with the ethanolamine of a ligand. A chirally specific methyl group in the α to the amine could activate the arrestin pathway via this residue.¹⁴²

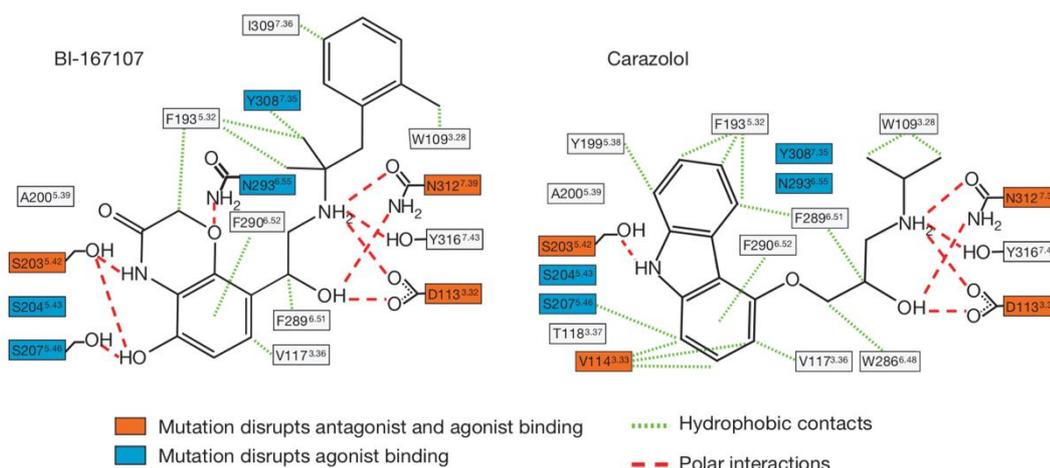


Figure 21 - Binding interactions between an agonist and inverse agonist in β_2 -AR – BI-167107 (left) and carazolol (**3**) (right). Ligands displayed have at least one atom within 4 Å of the crystal structure ligand. Ballesteros-Weinstein numbering is denoted in superscript. Residues that demonstrate disruptions in binding with agonists and antagonists when mutated are denoted in orange, and residues that only show disruptions

in binding with agonists are in blue. Potential polar interactions are denoted in red dashed lines, and hydrophobic interactions in green dashed lines. Source: Rasmussen et al., 2011¹².

The difference between a ligand acting as an agonist and an antagonist could be attributed largely to the head group on the β -hydroxyl side of the ethanolamine core. Ser^{5.42} and Ser^{5.46} on TM5 make hydrogen bond interactions with meta- and para- hydroxyl groups in catechol head groups present in ligands such as adrenaline.^{119,138,143} In most cases, a full agonist will interact with these residues, and ligands containing a phenoethanolamine without a hydrogen bond donor at the meta position will act as a partial agonist due to the absence of a Ser^{5.46} interaction. It is worth noting that although dobutamine contains the 3,4-dihydroxyphenyl group, it still acts as a partial agonist. It is suggested that this may be because of a lack of a chiral hydroxyl group within its core region.¹¹⁹ Crystallography and molecular modelling has shown that Ser^{5.42} is important in agonist binding, whereas Ser^{5.46} is important in both agonist and antagonist binding.^{39,144} It is when an agonist interacts with one or both of these serines that TM5 is brought inward, reducing binding pocket size and stabilising the active conformation.^{138,143,145} A third serine, Ser^{5.43} is said to be important in agonist binding despite not being shown to have interactions with the ligands in crystal structures. Asn^{6.55}, a residue shown to be important in agonist stereoselectivity of catecholamine-based ligands¹⁴⁶, is thought to be linked to Ser^{5.43}. Asn^{6.55} can form an interhelical interaction with Ser^{5.43} in the agonist bound state^{119,138,147}, and in some cases, an inactive state. Rendering it still unclear whether this trait is exclusive to the active conformation¹¹⁹. Canonical binding interactions between the ligand and Asp^{3.32} and Asn^{7.39} cause ligand head groups to tend towards helices 5 and 6, causing direct and indirect interactions with Asn^{6.55}¹³⁸.

Antagonist structures tend to include a methylether or a two-atom spacer between the ethanolamine backbone and the aromatic group, therefore making them generally longer at the head group. This reduces the receptors' ability to adopt an active conformation by reducing binding pocket space and by stabilising the inactive rotameric states of Ser^{5.46} and Ser^{5.42}^{119,138}.

Ligand head groups, regardless of their pharmacological characteristics tend to π - π stack with Phe^{6.52} and potentially Phe^{5.32} on the ELC2 along with hydrophobic interactions if the aromatic system is larger. This Phe^{6.52} is suggested to work as a 'gatekeeper' in ligand entry and exit^{141,144,148}. Val^{3.33} is also shown to contribute to this hydrophobic containment network. When mutated to alanine, it is shown to reduce agonist and antagonist binding affinities¹⁴⁹, and works alongside Phe^{6.51} and Val^{3.36} in hydrophobic head group interactions^{138,144}.

Tail groups of β -AR ligands vary in length as the head group does and tend to have van de Waals interactions with Trp^{3.28}, Thr^{3.29}, Phe^{5.32} and Tyr^{7.43}^{9,26,119}. Along with engaging in interactions within the binding pocket, longer tails are often associated with biased signalling as well as subtype selectivity. This is due to the extracellular surface having less conserved residues^{50,119,138}.

Cys199^{EL2} in the turkey β_1 -AR shows hydrogen bonding with the involvement of water to agonists that have been able to demonstrate biased signalling activation. Long ligands such as dobutamine, despite having a long tail group, do not demonstrate this biased signalling. This may be due to the lack of this hydrogen bond interaction¹¹⁹. β_2 -ARs contain non-conserved residues, namely Lys305^{7.32} and Asp192^{EL2}, at the extracellular surface that are thought to interact with the long allyl chains of β_2 -AR selective agonists such as salmeterol. These interactions are absent in β_1 -ARs due to the substitution of Lys^{7.32} with an aspartate at this position.

Pharmacological investigations and evidence have suggested that β_1 - and β_3 -AR harness a second, 'low affinity' binding site. Adrenaline and CGP12177 have shown to be able to occupy this space and bring about site specific responses. Longer ligands have been thought to be able to occupy both at a single given time.

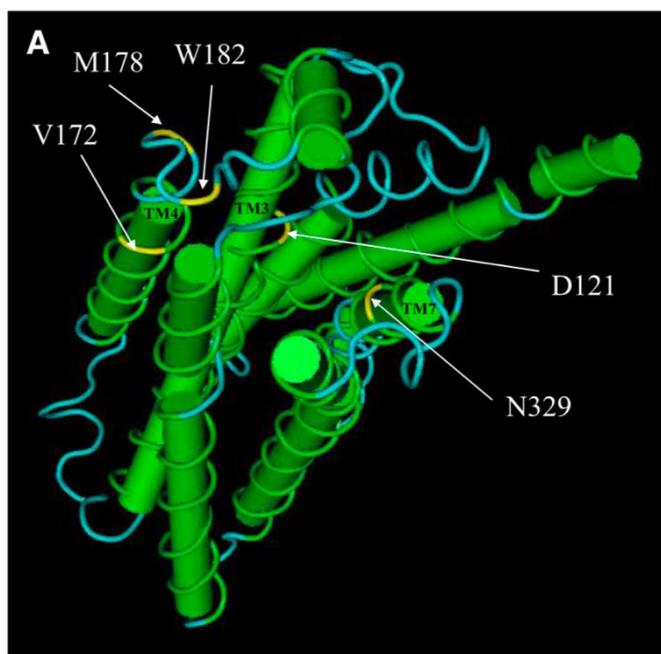


Figure 22 - Location of key residues (V172, M178, W182) in a turkey β_1 -AR crystal structure described by Huang et al., (2013). D121 and N329 (D138 and N363 in humans) along with the other main 'low affinity' site residues are highlighted in yellow. Baker, J. et al., 2014¹²³.

The residues involved in 'low affinity' site binding are near the orthosteric binding site. It is suggested that the β_1 -AR has two agonist conformations where the main residues responsible for secondary, 'low affinity' binding are located on TM4. When a V189T mutation was performed, the affinity of ICI118551 was impacted. L195Q and W199Y mutations abolished secondary site activity, where the W199Y mutation could significantly reduce activity alone¹²³.

1.8.6 The keyhole

While investigating ligand selectivity between β_1 - and β_2 -AR, Emtage et al., 2017 proposed the idea that a small fissure located between the TM4 and TM5, dubbed 'the keyhole', could be an important feature in receptor plasticity. Ligands with long enough head groups could still maintain canonical interactions between their ethanolamine backbone and the Asp^{3.32} and Asn^{7.39} residues within the binding pocket while being present within the keyhole. Polar groups of the head group

(typically ethers) forge hydrogen bond interactions with Ser^{5.42}, Ser^{5.46} and Tyr^{5.38}, which all line the keyhole. It is also believed that terminal hydrophobic groups that occupy the intra-membrane space can maintain hydrophobic interactions in the space¹⁵⁰.

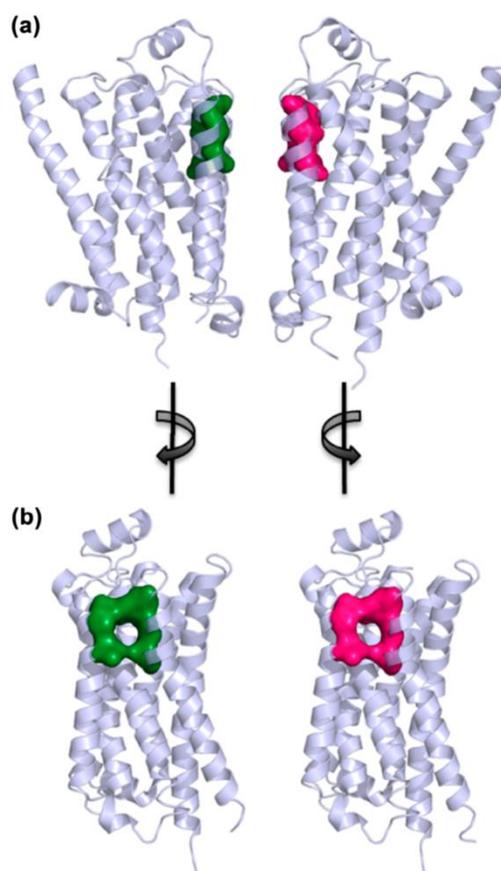


Figure 23 - Aligned keyholes in a dimer crystal structure of the turkey β_1 -AR. (a) depicts the dimer interface between the two receptors (PDB code: 4GPO; Huang et al., 2013). (b) depicts both interfaces rotated 90° to show keyholes.

Molecular dynamics studies in other investigations have suggested that the fissure between TM4 and TM5 is the second most common egress route when using carazolol (20 and 27% in the two data sets)¹⁵¹. Further supporting the importance of this location. The same authors had found in a separate study that this feature facilitated the exit of the ligand retinal in bovine rhodopsin in most cases¹⁵². Although it has been shown that this intramembrane space isn't the preferred

route for entry, it was suggested that the extracellular mode of entry was greatly preferred^{148,153}.

It was also believed that the different rotamer conformations of Ser^{5.46}, and its dependence upon whether it can make hydrogen bonding interactions with what's in the binding pocket further supports the plausibility of the keyhole¹⁵⁰.

As shown in Figure 23, the keyhole positions juxtapose each other in the dimer crystal structure, giving some insight into a potential role the keyhole may play in the bivalence of ligands^{102,148}.

Where the keyhole isn't a definitive explanation for β -AR ligand selectivity, since this feature is present in both β_1 - and β_2 -ARs, it is also worth noting that there is a water molecule occupying the keyhole space in several β_1 -AR crystal structures to date; a feature that could impact keyhole occupancy between subtypes^{50,154,155}. Interactions between residues in the keyhole could influence the selectivity of ligands. In addition to this, it is suggested that the entry and exit mechanisms of ligands vary between subtypes¹⁴⁸, and the keyhole could help to explain how and why that is. However, it remains unclear. Therefore, since the work conducted by Emtage et al., 2017^{150,156}, the keyhole hypothesis is that β -AR ligands with extended moieties at the aryloxy region demonstrate higher selectivity for the β_1 -AR versus the other β -AR subtypes.

Although the keyhole is present in some, but not all, of the crystal structures of the β -ARs, Active Site Pressurisation was used to pronounce the potential effect a wider fissure could have¹⁵⁰. The keyhole hypothesis is one to be considered and could potentially aid in explaining receptor plasticity, and therefore ligand specificity in β -ARs.

1.9 Computational tools for biomolecular modelling and dynamics

For over 30 years, one of the core areas of focus for use of computers within the field was the prediction of structures and properties of biomolecules. Initially, the field mainly focused on force-field based studies of structures, dynamics of biomolecules and development of solvated mechanics. But since then, has become capable of producing increased accuracy calculations and generating far more data due to the increase in accurate quantum chemistry approaches and more powerful hardware. The method used for computational analysis of biomolecular systems depends on the questions being asked and the availability of computational resources.

1.9.1 Homology modelling

When protein structures have not yet been solved experimentally, the process of homology modelling can be a useful tool. Structural templates can be aligned with the amino acid sequence of the protein of interest. Typically, at least 30% of the two sequences should be identical, with motifs and conserved residues being key anchor points for alignment. There are now several computational tools capable of performing homology modelling, in this study, Prime^{157,158} was used.

1.9.2 Docking

The conformational changes that occur because of a ligand being bound to its target can help to explain and further understand structure-activity relationships (SARs). Often, targets are in the form of structures solved crystallographically and are static; ligands can be rigid or allowed a degree of structural flexibility. Molecular docking can be used to rationalise these conformations, however

different programs have different methods for ranking conformations. What most share, is their output, that attempts to score the conformations by affinity and activity of small molecules. The software used for this study was Glide^{159,160}, which has had its protocol validated by several groups, and has been concluded to be one of the best protocols for re-docking small molecules from solved crystal structures^{159–161}. The software has also been used for β -AR ligand docking^{154,162–165}.

The user can modify constraints within Glide to best fit their investigation. Modifications include ligand rigidity (which instructs the protocol to either dock ligands rigidly, or flexibly), binding pocket flexibility (similar to ligand flexibility but performs small molecular dynamics (MD) simulations to produce alternative binding pocket conformations), SMARTS patterns (atom patterns ligands are required to contain), area grid and essential interactions (including filtering docking results and removing poses that do not match the defined mandatory interactions). The software will assess results based on these constraints and filter them to produce suitable outcomes. The docking procedure is based upon the empirical ChemScore¹⁶⁶ function, where the flexible docking option also includes an energy minimisation procedure to allow for ligand movement and torsional activity to improve scoring.

1.9.3 Molecular dynamics

Molecular dynamics (MD) can be used to provide an understanding of macro systems on a molecular level. The behaviour of each individual atom with respects to their external conditions can be examined. Examples of these conditions are temperature, pressure, and molecular environment. MD can give an insight into systems too small to be observed with traditional biophysical techniques applying theory from thermodynamics, statistical mechanics, and structural properties of the system of interest.

The method computes the movement of atoms along time using the integration of Newton's equations of motions (classical mechanics), which can be represented by the following equation¹⁶⁷.

$$\frac{d^2r_i(t)}{dt^2} = \frac{F_i(r(t))}{m_i} \quad \# (1)$$

where $F_i(r(t))$ is the force exerted on atom i at time t , $r_i(t)$ is the vector position of atom i at time t , and m_i is the mass of the atom.

These calculations are conducted for each step of a systems dynamics, and the exact functional form varies between the available software. The approaches are commonly split the calculations into two types of term: bonded interactions and non-bonded interactions. Bonded interactions are then subdivided into bonds, angles, and dihedrals; modelled as harmonic potentials with parameters from experimental values and/or QM calculations, and dihedral rotations modelled as Fourier series. Non-bonded interactions, which describes all atom pairs that are separated by at least three bonds, are subdivided into van der Waals interactions which are generally modelled with Lennard-Jones potentials, and electrostatic interactions modelled with Coulombic potentials. The particle mesh Ewald (PME)¹⁶⁸ smoothing function is frequently used to compute the impact of long-range forces, and cut-off radii allow for faster calculations¹⁶⁹.

Atoms in a system that have been parameterised by molecular mechanics move between timesteps. The forces acting upon each of these atoms are calculated using Newton's equation of motions (Equation # (1)). These steps are generally within the femtosecond (fs) range and are governed by the fastest degree of motion in a system (bonded hydrogen vibrations), and the coordinates are used to compute the forces acting upon each atom. The new coordinates and velocities are then calculated using deterministic calculations, and the atoms are moved to their new positions for the next cycle¹⁶⁹. Users can determine the frequency of

coordinate outputs, these outputs are compiled into a trajectory, which can be visualized with compatible software.

MD simulations are comprised of several steps: beginning with energy minimization, followed by an MD simulation run under NVT conditions (where N = moles, V = volume and T = pressure). This is then followed by an MD simulation run under NPT conditions (where N and T represent the same as in NVT, and P = pressure) and finally the production run. The energy minimisation step attempts to produce the lowest energy conformation within the system, relieving it of any unfavourable, high-energy interactions. Typically, steepest descent and conjugant gradient energy minimisation methods are used to relax the structure to the nearest energy minima.

In the NVT ensemble, a thermostat is used to regulate the kinetic energy in the system. The target temperature is used to produce atom velocities representative of the behaviour of the atoms at the said temperature. In the NPT ensemble, a barostat is applied to the system, that works in tandem with temperature to obtain an isothermal-isobaric ensemble, to represent laboratory conditions as closely as possible. Semi-isotropic pressure coupling is used for membrane simulations to maintain membrane area pressure, that isotropic pressure control is unable to achieve.

AMBER^{170,171} and GROMACS¹⁷²⁻¹⁷⁴ are two increasingly popular MD packages and are the two packages utilised in the investigations within this thesis. The package used for a particular investigation depended upon the problem attempting to be solved. For most investigations, regardless of the MD engine, the AMBER all-atom forcefield ff14SB¹⁷⁵ was used. Not only was it important to consider hydrogens explicitly for our investigation, but this forcefield works very well with the lipid and solvent model we employed in the systems used. The general AMBER force field (gaff)¹⁷⁶ was used for ligand parameterisation. Coarse-grained systems were run using the Martini¹⁷⁷ forcefield within GROMACS. These systems consider groups of atoms as an 'atom-group', where they are represented by a single 'bead'. This

approach can significantly speed up simulation running time at the expense of resolution. AMBER was used when incorporating advanced sampling methods to MD simulations, discussed further below.

1.9.4 Classical Molecular modelling force fields

Forcefields are mathematical expressions detailing the relationship between the energy of a system and its atomic coordinates. Interatomic potential energies are calculated as a function of a set of parameters¹⁷⁸.

$$U = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{\text{improper}} V_{\text{imp}} + \sum_{\text{LJ}} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}},$$

(2)

where the first four terms refer to the intramolecular (bonded) or local contributions to the system's total energy (stretching of bonds, bending of angles and both dihedral and improper torsions), and the last two describe the repulsive and Van der Waals (in this example, by means of a 12-6 Lennard-Jones potential) and Coulombic interactions (non-bonded).

1.9.4.1 Intramolecular terms

Atoms are bonded via spring-like bonds, acting as a harmonic oscillator. The bending within the bonds and angles are represented by a harmonic potential that controls the length of covalent bonds. X-ray diffraction experiments can yield reasonable values for r_0 , whereas Raman and IR spectra can give good information on the spring constant. Harmonic potentials are also usually used to represent angle bending; however, a trigonometric potential is preferred¹⁷⁸:

$$U_{\text{bending}} = \frac{1}{2}k_a(\cos \theta - \cos \theta_0)^2. \quad \# (3)$$

The Urey-Bradley potential is usually added to optimise the fitting to vibrational spectra^{178,179}:

$$U_{\text{UB}} = \sum_{\text{angles}} \frac{1}{2}k_{\text{UB}}(s - s_0)^2, \quad \# (4)$$

where s represents the distance between two external atoms forming the angle.

Dihedral or torsional terms are required to describe molecules containing four or more atoms in a row. These characteristics are usually several times stiffer than bond stretching motions and play a crucial role in the local structure of the macromolecule. They represent the degree of rigidity and rotations about bonds. A cosine function can be used to represent torsional energy, as seen in equation # (2), where φ represents the torsional angle, δ represents the phase, n is the number of minima and maxima between 0 and 2π , and V_n is the height of the potential barrier.

Improper torsions describe the positive contribution to the energy of the out-of-plane motions not described in the torsional parameters discussed above. These improper torsions are required to ensure the planarity of groups such as sp^2 hybridised or aromatic carbons. This can be described as¹⁷⁸:

$$U_{\text{imp}} = \sum_{\text{impropers}} \frac{k_{\text{imp}}}{2}(\omega - \omega_0)^2, \quad \# (5)$$

where ω is the improper angle corresponding to the deviation from planarity.

1.9.4.2 Intermolecular terms

The balance between repulsive and attractive forces gives rise to Van der Waals forces. The overlap of the two atoms' electron clouds leads to repulsion, whereas the attraction is caused by the interaction between induced dipoles that varies as r^{-6} . The Lennard-Jones potential is usually used to represent these intermolecular forces, described in equation # (2). Van der Waals forces are present between any pair of atoms part of different molecules and can intervene in atoms belonging to the same molecule.

Electrostatic interactions are represented by the final term in equation # (2). Coulomb's law is typically used to compute the contribution of atomic partial charges to the total energy. Fitting to experimental thermodynamic data can be used to derive these partial charges, but this approach is only practical for small molecules¹⁶⁹.

1.9.5 Analysis techniques for molecular dynamics simulations

Trajectories produced by MD can be analysed in a multitude of ways, the method chosen generally depends on the questions being asked. Root mean square deviation (RMSD) is a common starting place for trajectory analysis. This technique involves measuring the change in coordinates of defined atoms between a structure in a frame of a trajectory with a defined reference conformation. When assessing movement of individual residues or components, root mean square fluctuations (RMSF) can be insightful. This technique measures the fluctuation of a structure about its mean position, averaged overtime.

1.9.6 Self-avoiding walk molecular dynamics

Although computational power has increase dramatically in recent years, the ability to perform simulations long enough to observe biological processes such as ligand unbinding, and protein conformational changes are still very demanding on resources and are therefore expensive to investigate. Because of this, several enhanced sampling methods have been developed for the sole purpose of analysing rare events in biological systems.

The discovery of ligand binding and unbinding pathways is a potential application of targeted MD (TMD), and there have been several methods developed with this objective in mind. Different methods require different information from the user ranging from a particular progress coordinate (useful if the ligand expulsion route is known), levels of bias (including acting forces) and iteration cycles. Although TMD has been used for decades, an early example of a ligand expulsion route discovery method was the random expulsion molecular dynamics method (REMD)¹⁸⁰ by Lüdemann et al., in 2000. The method employed an artificial force in tandem with the standard MD forcefield to investigate the ligand expulsion route of substrates within cytochrome P450s. The direction of the force is random and is applied for several iterations of a specified length of time, and the force is maintained if the average velocity in a new direction is over a user specified number.

1.9.6.1 Proposed ligand expulsion method

The self-avoiding walk (SAW) MD method can produce ligand unbinding pathways by utilising a history dependent bias to prevent a ligand from returning to a region of conformational space it has previously occupied. The method requires a few specified parameters to produce these pathways (excluding atom coordinates and topologies), and the first is a selection of the atoms of interest within a particular system, the second is a force value to enact upon these atoms of interest. The process also requires a total number of cycles to run the procedure, and the cycles themselves must be defined as a length of time and a number of steps.

Traditionally, TMD takes a starting configuration and biases the system to produce a specified target configuration^{181,182}. In the case of SAW MD, a starting structure is provided, and the system is biased away from that structure, without an explicit target. In the case of REMD, the random force can be described as a harmonic potential¹⁸⁰. However, SAW MD implements a gaussian potential instead. This gaussian function prevents acting forces from tending to infinity, an outcome that would effectively cause a simulation to explode if achieved.

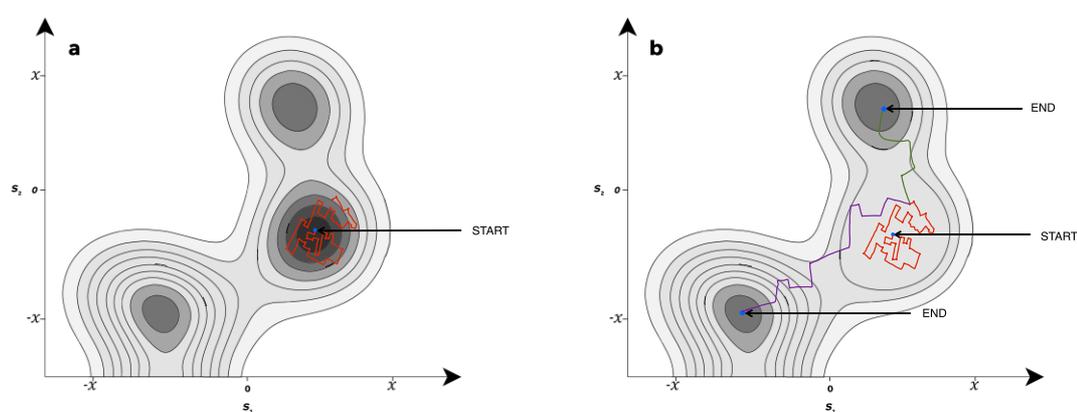


Figure 24 – Example schematic of the free energy landscape of a simple ligand unbinding SAW simulation. The trajectory begins from a defined start point (blue circle), but an end point is not defined. The trajectory is made to explore currently unexplored areas while being forced away from previously explored areas. In other words, the trajectory’s journey in particular direction is based solely on historic conformation occupation. Dark contours represent areas of low free energy and lighter contours represent areas of high free energy across the energy landscapes s_1 and s_2 . **a)** shows the trajectory beginning from a start point defined within an area with a very low relative free energy (deep energy well) (red line), meaning high energy barriers that would be difficult to overcome. While the trajectory explores conformations, the deep energy well becomes shallower as it is filled with hills and therefore becomes easier for the trajectory to exit what was once energetically difficult to exit. **b)** demonstrates that because of not having a specified end point, the trajectory can explore different routes and can therefore produce different end points. In some protein-ligand systems, the trajectory’s historic conformation occupation could lead to end points via different exit routes (as seen in **a** and **b** green and purple lines).

In principle, the SAW method works by building up ‘hills’ along regions of RMSD space, exploring them and eventually being forced out to explore a previously unexplored region. The area with a very low free energy (deep energy well that is

difficult to escape) becomes shallower. This concept of filling hills to explore regions of conformational space can be seen in Metadynamics (METAD)¹⁸³. The difference being that METAD builds up hills along a reaction coordinate or collective variable (CV) based on free energy minima, which could represent a free energy change within the distance between two points, a start, and an end. However, instead of having two defined points, in SAW MD, hills are built up along the pathway forged while avoiding previously occupied conformations. Although the free energies of the conformations aren't directly used in the SAW MD method, the acting force upon a structure that remains within an explored conformation for prolonged periods of time will be actively preventing the simulation from remaining within it longer. Conformations that are occupied for extended periods of time despite the presence of the acting force could represent regions of relatively low free energy.

The SAWMD method used within this study was engineered to produce 'waymarks' at user defined RMSD thresholds from a trajectory's previous waymark. Waymarks are .pdb files containing the atom coordinates of the system at those specific RMSD definitions between the user defined starting conformation (e.g., ligand bound state), and an end conformation (e.g., ligand unbound state).

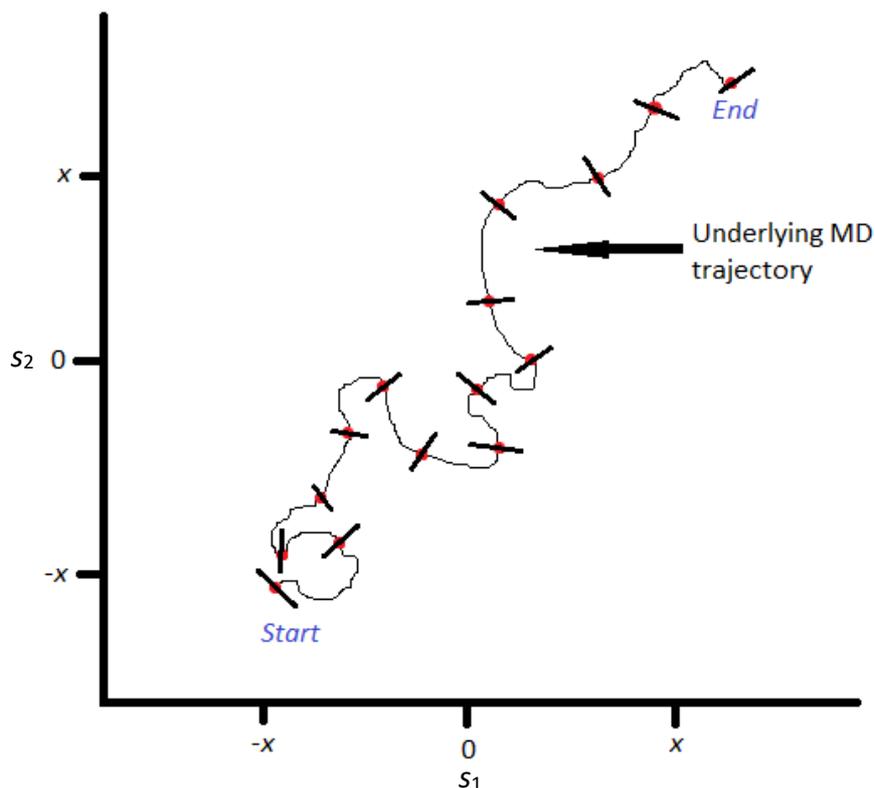


Figure 25 – Schematic of an example SAWMD trajectory beginning and ending at two different regions of conformational space (using the fundamentals described in this section and Figure 24), where the blue line represents the movement of the trajectory through different regions of conformational space, the red dot with a black line through it near the 'Start' and 'End' labels represent the defined start point and given end point of the trajectory respectively, and the red dots and black lines crossing them represent evenly spaced waymarks along the trajectory based on a user defined RMSD criteria.

1.9.7 Weighted ensemble enhanced sampling

As mentioned at the start of Section 1.9.4, traditional MD simulations are insufficient for observing biological processes such as ligand unbinding. We used an enhanced sampling method for discovering ligand unbinding pathways, in this chapter we discuss an enhanced sampling method capable of utilising this information to produce kinetics data.

Weighted ensemble (WE) could be a rediscovery of the splitting strategy capable of producing estimates of nonequilibrium observables. Kahn in 1951 described the replication of trajectories as an idea of von Neumann's. He explained that when a

particle moves from a less important region to a more important region, it is split and the resulting particles hold half the weight of the original¹⁸⁴.

In 1996, Huber and Kim¹⁸⁵ proposed a WE algorithm with roots in nonequilibrium trajectory physics. For a given system simulated over time t , the distribution will relax to a nonequilibrium steady state if there is a steady flux of particles from A to B, or relax to an equilibrium distribution if there is no equilibrium flux from A to B.

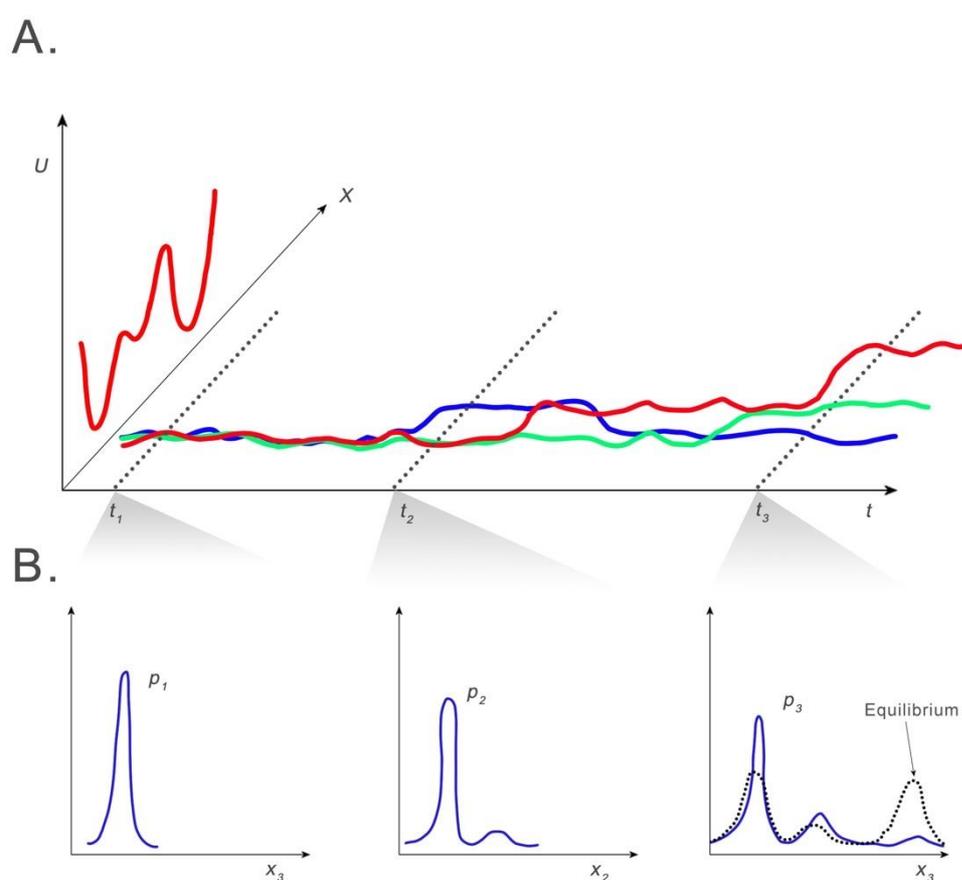


Figure 26 - A) example trajectories from schematic trajectory ensemble generated in potential U . B) Configuration distributions can be obtained by histogramming configurations at fixed time points t_1 , t_2 , t_3 . Configurational distributions will relax to the equilibrium distribution over long-time scales. Therefore, to effectively sample a free energy profile of a system, either a long-time scale trajectory or efficient sampling across the conformational space and using the probabilities of state occupation to derive a free energy profile. Adapted from: Zuckerman et al., 2017¹⁸⁶.

Ultimately, the approach is comprised of two steps, a simulation step, and a resampling step. Before the method is employed, the user defines a value for the target number of walkers (trajectories) per bin M . The total weight of every walker within the ensemble is equal to 1. Bins are regional divisions across a conformational pathway from a known start point (A) and a known end point (B). During the simulation step, walkers are simulated for a user defined simulation interval. At the end of the simulation step, the resampling begins, and walkers are assigned a bin dependent on a user defined calculation method. If walkers occupy a bin with less than the user defined M walkers within a bin, the trajectory is split to achieve M , and the resulting daughter trajectories carry the coordinates and total weight equal to the parent ($1/M$). In bins where the number of walkers exceeds the user defined M , trajectories are pruned to achieve M , and the resulting trajectory carries the total weight of the chosen trajectories and the coordinates of one of them.

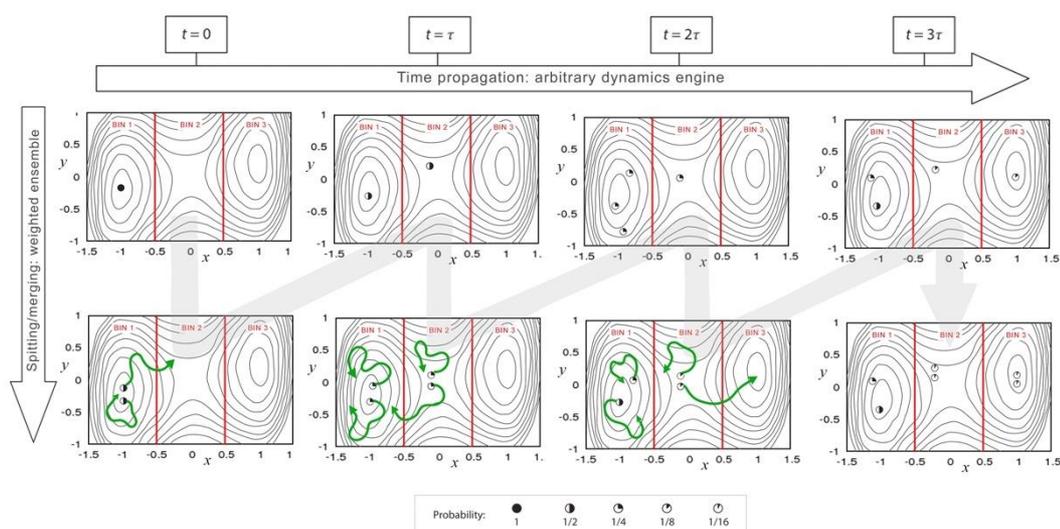


Figure 27 – Schematic of the weighted ensemble method using predefined bins. Trajectories are run for simulation intervals τ and resampled once complete. Circles and shading represent the walkers and their designated weights that always total 1. The prevention of bias is ensured by rigorous splitting and merging procedures. Adapted from Zuckerman et al., 2017¹⁸⁶.

A steady state flux can be obtained from the WE technique once the system has relaxed to a steady state. For complex systems, the time required to achieve this

may be relatively significant compared to a much simpler system, but still less time than a traditional MD simulation¹⁸⁶. The steady state flux can be calculated as the running mean of the weight of trajectories that reach a particular target bin. The rate constant can then be obtained using the Hill relation between flux and mean first passage time (MFPT)¹⁸⁷.

With this information, it is then possible to calculate the rate constants. In a protein-ligand binding/unbinding system, since the event is nonequilibrium, the unbinding and binding rate constants (k_{off} and k_{on} respectively) can be calculated by^{188–190}:

$$k_{\text{off}} = \frac{\langle f_{\text{bound to unbound}} \rangle}{\langle p_{\text{bound}} \rangle} \quad \# (6)$$

$$k_{\text{on}} = \frac{\langle f_{\text{unbound to bound}} \rangle}{\langle p_{\text{unbound}} \rangle v N_{\text{av}}} \quad \# (7)$$

Where for k_{off} , $\langle f_{\text{bound to unbound}} \rangle$ is the running average flux from the bound state to the unbound state and $\langle p_{\text{bound}} \rangle$ is the steady state probability of the initial state (bound), and for k_{on} , $\langle f_{\text{unbound to bound}} \rangle$ is the running average flux from the unbound state to the bound state, $\langle p_{\text{unbound}} \rangle$ is the steady state probability of the initial state (unbound), v is the volume of the simulation box at equilibrium, and N_{av} represents Avogadro's number. With this information, it is then possible to calculate the dissociation constant k_{D} ¹⁸⁷:

$$k_{\text{D}} = \frac{k_{\text{off}}}{k_{\text{on}}} \quad \# (8)$$

where k_{D} represents a unitless ratio relative to the concentration 1 M¹⁹¹.

1.10 Research Aims

The need for selective β_1 -/ β_2 -AR ligands has been discussed at length within this chapter, including the benefits it would bring to patients with cardiovascular and pulmonary conditions. It would be beneficial to understand the structural differences between the two receptors and the behaviour of current high affinity ligands to steer synthetic efforts.

Receptor plasticity within the β -AR subtypes has been proposed as a potential characteristic for determining ligand selectivity¹⁵⁰, particularly the compensation for ligands with extended moieties. What has been observed (discussed later in this thesis) is the role of this behaviour in the increase of binding affinity. Increased selectivity could be obtained by using ligands that can increase binding affinity based on known structural differences between the subtypes. Using what we know about the keyhole region in the β_1 -AR, it may be possible to design high-affinity ligands capable of manipulating this feature. Receptor dimerization may also be manipulated using this method of binding to improve receptor selectivity. The primary aims of these are therefore:

- To use computational approaches to understand modes of ligand binding at the β_1 -AR.
- To design, synthesise, and pharmacologically evaluate extended head-end ligands with differing properties to investigate their effects on binding affinity.
- To computationally investigate the conformational dynamics of the β_1 -AR homodimer.
- To computationally investigate the feasibility of utilising the keyhole as a route for ligand bivalency in the β_1 -AR.

Previous attempts at understanding the effect of extended ligand aryloxy head-ends on their affinity towards specific β -AR subtypes have been taken within the research group. The work previously undertaken will aid in the computational evaluation of the structural impact of ligands on their binding affinity for the β -AR subtypes.

Homology modelling will be completed to obtain a β_1 -AR model using a high-resolution avian crystal structure due to the unavailability of a human β_1 -AR crystal structure at the time of investigation.

Docking of a subset of high affinity ligands will be used to understand the preferred method of binding, and to be used for further MD and enhanced sampling investigations. The results from these investigations will be compared with pharmacological data to inform on whether the predicted mode of binding is likely.

A series of ligands previously synthesised within the group will be used to influence the design of high-affinity ligands. The highest affinity ligand of the previously synthesised series will be used as the basis of the new series. These ligands will change the component of the basis ligand that we believe is extended beyond the receptor and into the transmembrane space and will test whether binding affinity can be altered by modifying its electron density.

Once modelled computationally, the ligands will be synthesised and tested pharmacologically by a colleague for evaluation.

Although the exact physiological role of the β_1 -AR homodimer is poorly understood, its potential as a method of improving ligand selectivity is of significant interest, especially in the context of bivalent ligands. Information about the exact conformation is limited, however, crystal structures do exist of the homodimer with a contact interface on the TM4 and 5, aligning the keyholes of both protomers. This interface could allow for well-designed bivalent ligands,

capable of taking advantage of the shorter mode of entry than via the extracellular space. Molecular dynamics studies will be done to evaluate the potential effect introducing ligands of the kind will have.

The information we gain about the method of binding of extended head-end ligands and the proposed conformation of the β_1 -AR homodimer, will be used to investigate the potential for a bivalent ligand manipulating keyhole. As mentioned previously, targeting a homodimer may be a means of improving receptor selectivity. The feasibility of this will depend upon the ligand entry and exit pathways, which will be investigated computationally using enhanced sampling MD methods.

2 Exploring modes of binding of extended β -AR ligands in the β_1 -AR

2.1 Aims and objectives

Previous work within the group used molecular docking studies to predict how relatively large antagonist molecules with extended moieties might bind within the apparently limited confines of β -AR ligand binding sites. As mentioned in an earlier section, all crystal structures to date of the β_1 -AR show the ligand's ethanolamine component making interactions with Asp138^{7,39}/Asn363^{3,32,26,48,50,120,121,155}, and pharmacological mutagenesis studies reinforce the significance of this interaction in β_1 -AR activity¹⁹². The docking studies suggested that bulky ligands could display one of two behaviours¹⁵⁰. The first being that the receptor does not structurally adapt enough to allow the extended ligands to enter the binding pocket entirely, so they adopt a "v" shape pose in the receptor, where the head and tail groups of the molecule protrudes into the extracellular space and limited interactions are formed with normally important residues within the binding pocket (Asp138/Asn363). The second is that the receptor compensates for extended ligands with a degree of flexibility, create a fissure between TM4 and TM5 (the "keyhole") to accommodate parts of their structure, and therefore allowing for strong canonical interactions with the important residues to be maintained. In this pose the head group of the ligand is exposed to the intramembrane space.

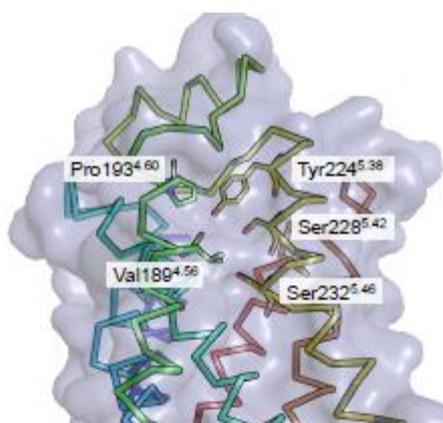


Figure 28 – Residues determined to be lining the keyhole as defined by Emtage et al., 2013. Source: Emtage et al., 2013¹⁵⁶.

In this chapter I extend this previous study to new ligands with established antagonistic pharmacology, and new, updated, molecular models of the β_1 -AR. To investigate the binding pose that an extended ligand undertakes within the β_1 -AR, docking will be performed to attempt to obtain binding positions that represent both previously suggested modes of binding. Once docking poses are obtained, one ligand will be taken forward, and molecular dynamics will be performed to analyse the behaviour of the system with respect to stability and interactions within a lipid membrane. Analysis of the dynamic stability of the system containing a ligand with this method of binding will allow assessment of the confidence we can have that the pose does not feature unfeasible interactions within the binding pocket and maintains favourable interactions with key residues, suggesting to what extent the model can be used as a reliable representation of the protein-ligand complex.

2.2 Methods

2.2.1 Homology modelling using Prime

Prime^{157,158} was used to produce a homology model and Maestro with Prime loop refinement was used for structure refinement. The β_1 -AR wildtype sequence was

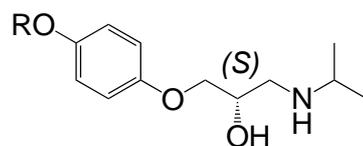
obtained by NCBI and aligned with the turkey crystal template 4BVN¹⁵⁵ using Prime (Alignment information can be found in Appendix 1). This crystal structure was chosen because it was the highest resolution β_1 -AR structure available. ClustalW¹⁹³ was used to assess the alignment. The N terminus begins at Leu53 and the C terminus ends at Cys393. The IL3 was truncated by 45 residues and a bond created between Asp259 and Arg305 to form a pseudo-loop.

The model was refined (hydrogen atoms only, OPLS 2005 force field) and prepared using protein preparation wizard within Maestro.

2.2.2 Ligand series to be docked

Since the original publication on the “keyhole” theory by our research group, the synthesis and pharmacological evaluation of a further series of ligands designed to come into contact with the keyhole region of the β_1 - and β_2 -AR has been carried out by other members of the research group.

Table 3 - Binding affinity results from previous keyhole ligand series including both β_1 - and β_2 - binding affinity data.



Compound	Compound ID	R	H β_1 - log k_D	N	H β_2 - log k_D	N
34	SMRS01	H	-8.03 \pm 0.03	6	-7.65 \pm 0.04	5
35	SMRS16	OH	-6.95 \pm 0.08	4	-6.42 \pm 0.06	3
36	SMRS02	OCH ₃	-6.53 \pm 0.04	6	-5.99 \pm 0.07	5
37	SMRS03 GBS007A*		-7.04 \pm 0.07	6	-5.95 \pm 0.14	5
38	SMRS09		-7.61 \pm 0.03	6	-6.05 \pm 0.03	5
39	SMRS13		-8.06 \pm 0.04	6	-6.53 \pm 0.05	5
40	GBS001*		-6.74 \pm 0.09	4	-5.57 \pm 0.05	4
41	GBS007*		-7.10 \pm 0.06	4	-5.79 \pm 0.08	4
42	GBS003*		-6.29 \pm 0.04	4	-5.33 \pm 0.07	4
43	SMRS12		-7.58 \pm 0.04	6	-6.16 \pm 0.07	5
44	SMRS08		-6.85 \pm 0.07	5	-5.47 \pm 0.06	5
45	SMRS14		-8.00 \pm 0.04	4	-6.29 \pm 0.07	4
46	GBS005*		-8.62 \pm 0.04	4	-7.36 \pm 0.05	4
47	GBS004*		-8.92 \pm 0.04	4	-7.78 \pm 0.02	4
48	GBS002*		-8.45 \pm 0.01	4	-7.05 \pm 0.03	3

*HCl salt

Structures were drawn using Maestro's 2D sketcher and LigPrep was used for ligand preparation. The (*S*)- enantiomer and protonated nitrogen were produced within the ethanolamine core.

2.2.3 Glide docking

Protein preparation wizard within Maestro was used to prepare the protein and docking grid boxes were produced using Glide. The centroid of the box was defined as the centre of mass of residues Asp138 and Asn363 and length of ligands to be docked (setting within Glide) was 38 Å. The inner box (centre of mass setting within Glide) was set to 40Å to allow for longer ligands with distributed mass. Hydrogen bond constraints were set as Asn363 OD1, Asn363 1HD2 and Asp138 OD1. When running the docking protocol using Glide SP, the hydrogen bond constraints were applied, and result must satisfy all specified interactions. Each ligand returned 50 poses that were subsequently minimized post procedure to produce a maximum of 20 poses. Default parameters were used unless otherwise stated.

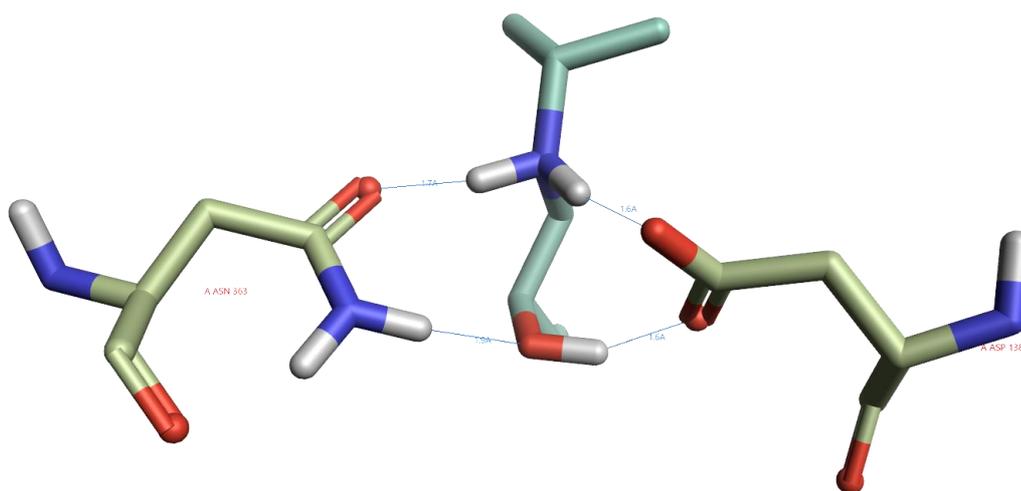


Figure 29 – Molecular visualisation of docking pose constraints. The ethanolamine backbone component of the ligand (OCCN+) had to satisfy hydrogen bonding interactions with the receptor's Asn363 OD1, Asn363 1HD2 and Asp138 OD1. Blue lines represent the interactions between the ligand and the receptor, and the visualisation demonstrates the conformation required (as seen in crystal structures) to satisfy the strict bonding constraints.

2.2.4 MD of protein systems

The selected protein-ligand complex and the apo state protein were exported and input into CHARMM-GUI¹⁹⁴. Here, the protein was subject to the membrane builder^{195–198} procedure to produce a protein-membrane system. For the ligand bound state, the ligand was parameterised using the Antechamber tool. The protein was embedded in a bilayer formed of 200 DPPC molecules in each leaflet, oriented in the xy plane. The length of the x and y box axes was set to 110 Å, and the z-axis to 148 Å, and outside the bilayer an ion concentration of 0.15M NaCl was used along with TIP3P as a water model. The AMBERFF99SB-LIDN¹⁹⁹ force-field was chosen, and the system was output to include GROMACS input files^{200,201}. GROMACS 2016.4 was used to perform all molecular dynamics simulations. The system first underwent energy minimization for 500 steps, NVT MD simulation was then performed for 100 ps, NPT MD equilibration was performed for 50 ns, and a production run was performed for 200 ns total for the ligand bound state to investigate whether the ligand can sustain key interactions over an extended period, and 200 ns for the apo state to compare receptor activity between the two states. Both systems were run in triplicate. For the ligand bound system, ligand-residue interactions were analysed using Flare Version 5²⁰².

2.3 Results

2.3.1 Docking results

The distance cut-off for hydrogen bonding identification was set at 2.5 Å between the donor hydrogen and acceptor heavy atom. Within the set of ligands, not all were able to produce docking poses. However, of the ligands that did, all the high-ranking poses demonstrated four hydrogen bonds between the ethanolamine

core of the ligand and the Asp 138 and Asp 363 residues. The Glide scores for these docking poses can be seen in Table 4.

Table 4 - Docking scores of keyhole ligand series in the β_1 -AR 4BVN crystal structure. Docking poses were not obtained for some ligands and is notated within the table as such.

COMPOUND	DOCKING SCORE
34	-4.34
35	-4.51
36	-3.48
37	No docking pose
38	No docking pose
39	No docking pose
40	No docking pose
41	-3.85
42	-6.42
43	No docking pose
44	No docking pose
45	No docking pose
46	-6.90
47	-8.12
48	-7.88

It is worth noting that poses with the lowest docking scores (most energetically favourable) all extended through the keyhole and into the transmembrane space. The best pose by docking score (-8.115) also maintained canonical binding interactions with known residues of importance within the orthosteric binding pocket.

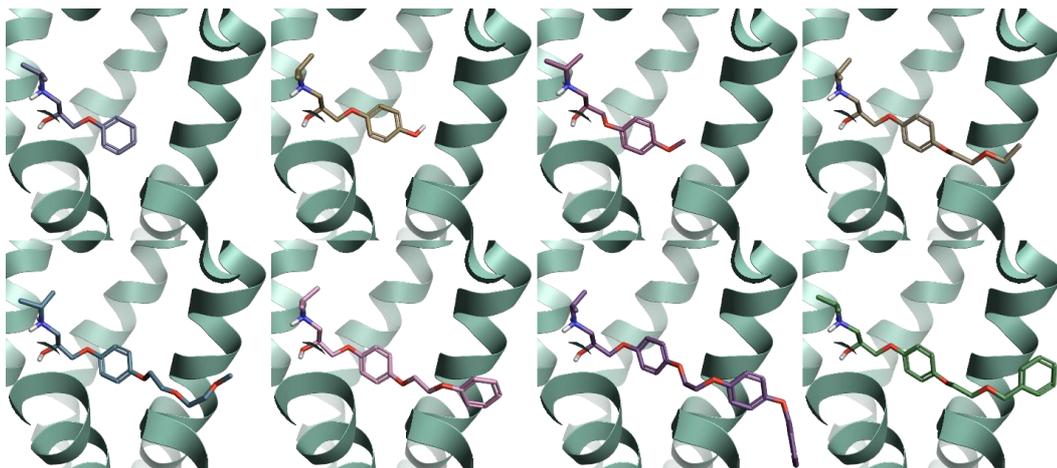
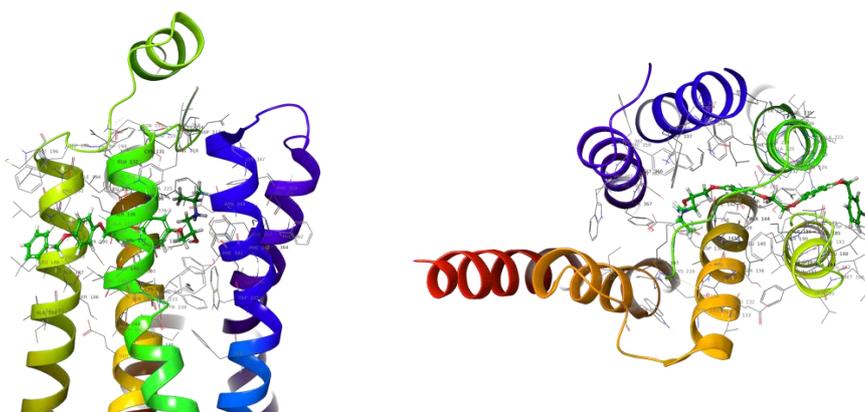


Figure 30 – Top binding poses of ligands detailed in Table 4 within the β_1 -AR (viewed from TM3/4). Top row from left to right: **34**, **35**, **36**, **41**. Bottom row from left to right: **42**, **46**, **47**, **48**.

It is also worth noting that the pharmacological assays show that ligands that contained the diethylene glycol ether feature and at least one aryl ring generally show higher binding affinities but not necessarily better selectivity between the subtypes. It is also apparent that ligands containing aryl rings in their head group generally show better binding affinity than their linear counterparts. Docking studies show that aryl rings in the head group could engage in pi-stacking interactions with Phe341, Tyr224 and Trp198 on the extracellular loop 2 (ECL2), therefore improving binding.



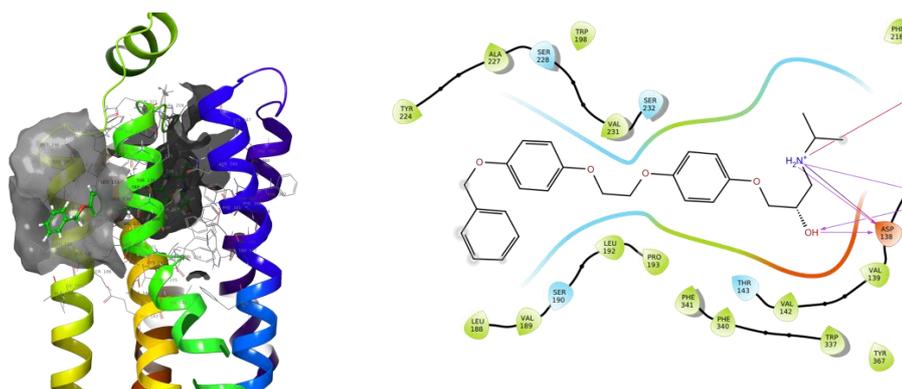


Figure 31 – Molecular representation of binding pose of best scoring binding conformation of **47** in homologised PDBID: 4BVN β_1 -AR from different perspectives and contact map of binding pocket interactions. (Top-left) TM5 direction. (Top-right) ECL2 direction. (Bottom-left) TM5 direction. (Bottom-right) Ligand interactions between ligand **47** and the residues within the binding pocket.

Ligand **47** docking studies show that the end aromatic ring can extend far enough out of the keyhole that it interacts with the intramembrane space. Due to the lipophilic nature of the head group of the ligand, it is likely that the ligand could also be interacting favourably with the lipid component of the transmembrane space. It was not possible to deduce for certain whether this is the case using Glide, however the ligand properties at the head group extending beyond the keyhole could still be of interest due to interactions with residues on the outer surface of the receptor.

The ligands that did not return docking poses are known to bind in nature (as shown in pharmacology), however, binding may require receptor conformational changes that cannot be replicated in static receptor docking. Flexible receptor docking was assessed within the research group previously and was shown that the increased computational costs did not yield the results to justify the method's use within this system¹⁵⁶. The docking constraints required to return poses were strict (Figure 29) to represent the ethanolamine backbone conformation seen in crystal structures as best as possible. Considering this, ligands that failed to return a docking pose may have done so because in order to satisfy the interactions within the constraint, several unfavourable interactions may have arisen between

the ligand and the receptor throughout the binding pocket in this protein conformation. Therefore, preventing poses from being output.

2.3.2 MD results

Ligand **47** was taken forwards for MD analysis. This was due to it being the longest ligand in length and the ligand with the highest pharmacological binding affinity and best docking score. The MD analysis was conducted to understand whether the protein backbone behaves differently in the ligand bound and apo states. In addition to this, analysis of the residues that line the keyhole as defined by Emtage et al., 2013¹⁵⁶ will be conducted to identify whether the presence of a ligand influences or prevents change in movement relative to the apo state. Ligand-protein interactions will also be analysed to gain further understanding of which residues are involved in ligand binding as well as investigating the ligand's capability of maintaining stable canonical interactions within the binding pocket.

Analysis of structural stability of ligand-bound and apo states

Comparing the structural stability of the receptor in the presence and absence of ligand **47** gave significant insight into the effect that an extended ligand has on the receptor's tertiary structure. Ideally, the structure of the receptor would behave similarly in both cases, to signify little disruption when the ligand is introduced. Both an RMSD and RMSF calculation of the alpha carbons of the protein backbone (N, CA, C, O) was calculated to demonstrate this.

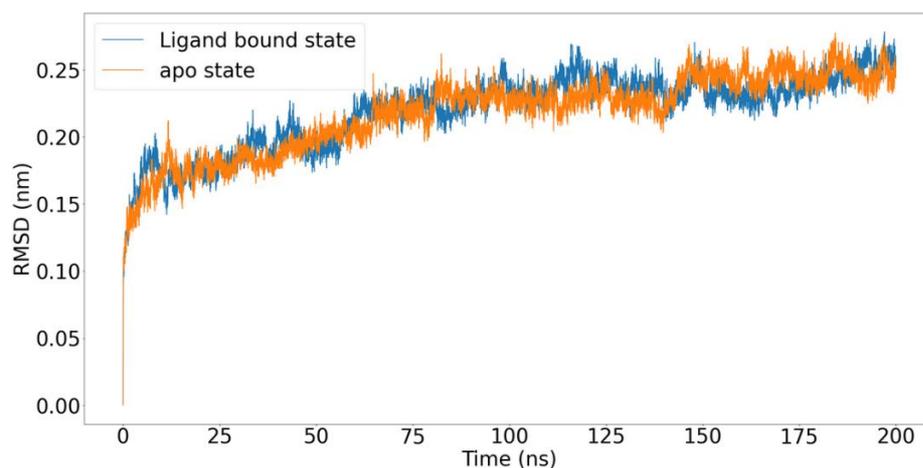


Figure 32 – RMSD of the protein backbone of the β_1 -AR over 200 ns for both the ligand bound and apo states of the receptor.

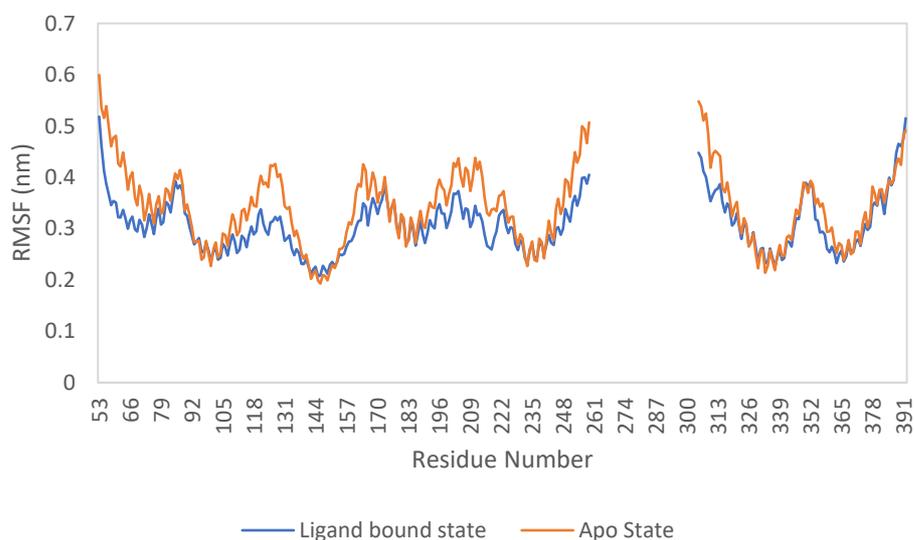


Figure 33 – RMSF of the protein residues of the β_1 -AR over 200 ns for both the ligand bound and apo states of the receptor.

Table 5 – Mean and standard deviation of the RMSD and RMSF of protein backbone of receptor in nm for both the ligand bound and apo states over 200 ns.

System backbone	Average RMSD (nm)	Average RMSF (nm)
Ligand bound	0.22 ± 0.03	0.31 ± 0.05
Apo	0.22 ± 0.03	0.34 ± 0.08

For both the ligand bound and apo states, the RMSD remained identical in both their mean values and standard deviation (0.22 ± 0.03). The RMSD values suggest that the backbone remains stable and identical regardless of the receptors ligand bound status, and both systems demonstrate very little deviation around their RMSDs, further suggesting that the receptor backbone's tertiary structure remains resilient.

The RMSF values follow a similar trend. The RMSF values for each residue remain identical regardless of whether it is ligand bound or apo state (0.31 ± 0.05 and 0.34 ± 0.08 respectively). Each RMSF remains within the standard deviation of their counterparts, and the overall fluctuation between residues for each state are identical. Both systems demonstrate relatively high RMSF values at their N- and C-terminus as well as ICL3. Largely the pattern in behaviour seen in the RMSF investigation supports the findings in the RMSD investigation, further supporting that the presence of ligand **47** has no adverse effects and induces imperceptible change overall on the protein's tertiary structure.

This data gives an idea of the overall structural stability of the protein backbone; however, additional information would be required to learn about the ligand's impact on the keyhole residues (defined in Figure 28).

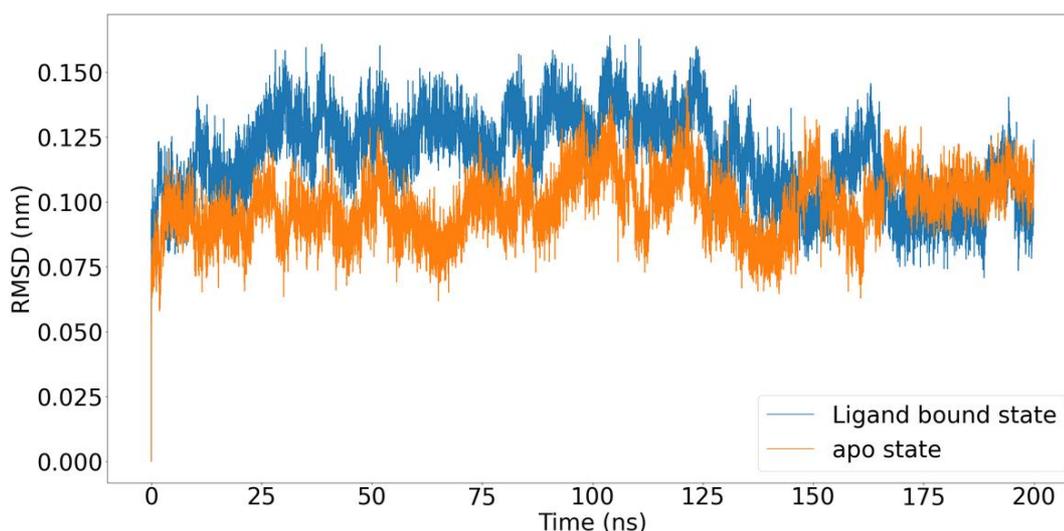


Figure 34 - RMSD of the protein keyhole residues as defined by Emtage et al., 2013 of the B1-AR over 200 ns for both the ligand bound and apo states of the receptor.

System backbone	Average RMSD (nm)
Ligand bound	0.12 ± 0.01
Apo	0.10 ± 0.01

An RMSD analysis of the keyhole residues over the duration of the simulation showed that its activity was largely identical between the ligand bound and apo states (0.12 nm and 0.10 nm respectively). For the first ~125 ns of the simulations, the ligand bound state shows a slightly higher and sustained RMSD for the keyhole residues, before becoming more in-line with the RMSD demonstrated in the apo state. The presence of ligand **47** within the keyhole region could have elicited minor repulsive interactions in this area before stabilising. However, the overall difference between RMSD for both systems is negligible, therefore it isn't unreasonable to consider that this may just be an artifact and that the protein's behaviour (including keyhole residues) remains identical in both the ligand bound and apo states.

Analysis of ligand-residue interactions

The interactions between the ligands and the residues within the receptor's orthosteric binding site were measured as distances and classified into the different interaction types. The atom identities are present within the table along with the specific atom numbers; however, the numbering is present not to show the specific atom's interactions, but the type of atoms involved and where they are found. The study was conducted to show the presence of the interactions between residues that the ligand as a percentage of their occurrence throughout the 200 ns simulation. To consider the interactions that occurred the most frequently, a cut-off of 30% was applied.

Table 6 - Residue interactions between ligand **47** and the residues within the receptor as a percentage of their occurrence throughout the 100 ns simulation (with a 30% threshold). The first column details the type of interaction taking place between an atom of the ligand and an atom of a residue within the receptor. The second details the ligand atoms (yellow fill: atoms of the ethanolamine backbone, brown: atoms of the ligand within the benzyloxy head end). The third column details the receptor atom involved in the interaction (blue: atoms of the ASP138 and ASN363 key residues, purple: atoms of residues surrounding the keyhole). The final column details the percentage occurrence over the duration of the trajectory (the darker the green, the more frequently the interaction occurs).

Bond Type	Ligand Atom	Protein Atom	% Frames Present
<i>Salt Bridge</i>	Ligand 47 N1	ASP 138 OD2	78.10%
<i>Hydrogen bond</i>	Ligand 47 H67	ASP 138 OD1	30.70%
<i>Hydrogen bond</i>	Ligand 47 H15	ASP 138 OD1	72.60%
<i>Hydrogen bond</i>	Ligand 47 H67	ASP 138 OD2	86.70%
<i>Hydrogen bond</i>	Ligand 47 H4	ASN 363 OD1	99.90%
<i>Hydrogen bond</i>	Ligand 47 O1	ASN 363 HD22	98.30%
<i>Aromatic-Aromatic</i>	Ligand 47 C23	TYR 224 HD2	34.30%
<i>Aromatic-Aromatic</i>	Ligand 47 C11	PHE 341 HE1	38.10%
<i>Aromatic-Aromatic</i>	Ligand 47 C11	PHE 341 HZ	42.10%

As expected, the ethanolamine nitrogen atom maintains the salt bridge interaction with the ASP138 residue of the orthosteric binding pocket (~80% of the frames). This is also supported by the hydrogen bonding interactions between the ethanolamine backbone of the ligand and the key Asp138 and Asn363 residues of the binding pocket (all over 70% of the frames). The distances of these hydrogen bonding interactions are investigated in the subsequent section.

An interesting interaction that occurs relatively frequently is the aromatic-aromatic interaction between the benzyloxy ring of the ligand and Tyr224 that lines the keyhole (~35% of the frames).

Analysis of atomic protein-ligand interactions

Gathering information on the effect of the ligand on the protein's tertiary structure is important, however understanding the atomic interactions within the binding pocket will reveal whether known key interactions are occurring, and whether those interactions are maintained. Distances between receptor atoms and their ligand atom pair were measured to determine whether hydrogen bonding interactions were taking place. The definition of a hydrogen bonding interaction between two atom pairs was defined as 0.25 nm for this study.

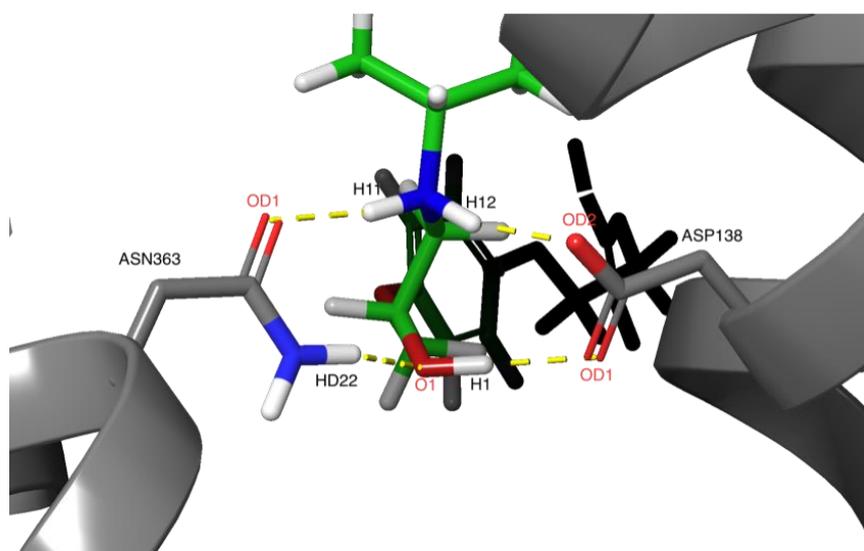


Figure 35 – Diagram of atoms of interest. Individual atoms of ligand **47** are labelled as well as the residues of which they are contained. The Yellow dotted line represents the interaction between the atoms involved in the key hydrogen bonding interactions between the ligand and the receptor.

The four specific hydrogen bonding interactions of interest are seen in Table 5.

The distances between the atom pairs were investigated to understand the cause of these sudden changes.

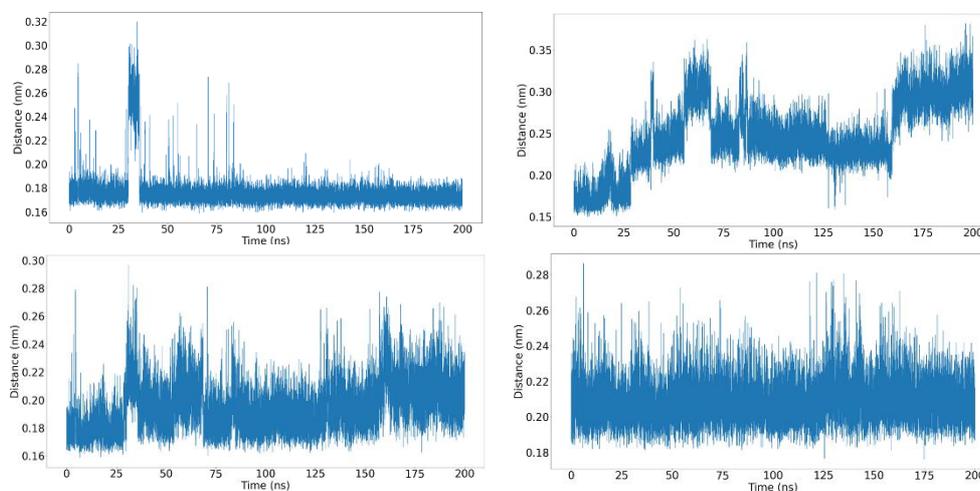


Figure 36 – Atom distances in nm between atom pairs involved in the key protein-ligand interactions over 200 ns. (Top left) H11 and Asn363-OD1 atom pair. (Top right) O1 and Asn363-HD22 atom pair. (Bottom left) H12 and Asp138-OD2 atom pair. (Bottom right) H1 and Asp138-OD1 atom pair.

Table 7 – Atom pairs and the distance between each atom in angstroms.

LIGAND - ATOM	RECEPTOR - ATOM	AVERAGE DISTANCE (nm)
H11	Asn363 – OD1	0.18
O1	Asn363 – HD22	0.21
H12	Asp138 – OD2	0.19
H1	Asp138 – OD1	0.25

All atom pair distances are generally consistent with the H1 and Asp138-OD1 pair having the most deviation. All replicates showed a sustained change in distance from ~0.2 nm to ~0.4 nm but at different time for a different duration. It also appears that these sudden but short events have no impact on the interactions of the other atom pairs. Despite these minor fluctuations in distance, all atom pairs-maintained average distances at 0.25 nm and below, showing sustained hydrogen bonding interaction.

2.4 Discussion

Homology modelling can prove to be a difficult process, and it is important to accurately represent the target receptor as best as possible. Loop prediction has proven to be difficult by previous researchers in the research group^{150,156}. It was found that truncated loop regions have no negative impact on the modelling of the β_1 -AR in the region of the orthosteric ligand binding site and produce reliable models. Therefore, the loop regions were treated the same way in this study, and the general homology modelling method was identical. Upon completion of the homology modelling, the model retained the same secondary structure as the template and as a result, the binding pocket was identical.

β_1 -AR selective antagonists generally contain extended head groups with a para-substituent at the phenoxy propanolamine core^{137,203}. Ligands co-crystallised with the β_1 -AR either show poor subtype selectivity or tend towards β_2 -AR selectivity^{34,204}, therefore it isn't possible at time of writing to extrapolate the expected mode of binding of our class of ligands from any current crystal structure data. It was therefore necessary to use docking methods to predict how the receptor might compensate for their extra bulk.

Of the 15 ligands designed to extend towards the keyhole, only 8 produced poses demonstrating key interactions known to be vital in binding. It was previously proposed that extended ligands could adopt a "v-shape" pose within the binding pocket¹⁵⁰, but they were shown to demonstrate poor or no contacts with the Asn^{7.39} and Asp^{3.32} residues in the β_2 -AR. It was also proposed that this type of binding was as a result of the size difference between the two subtypes binding pockets. The longer ligands capable of binding within the β_1 -AR were unable to bind canonically to the β_2 -AR due to its smaller binding pocket. Interestingly, no ligand in the current set demonstrated this mode of binding (including those that produced no result), suggesting that in the β_1 -AR, ligands in poses of this nature

are unable to make enough energetically favourable contacts, which is supported again by the importance of the Asn^{7.39} and Asp^{3.32} interactions.

Of all of the docked ligands, **47** and **48** were the only ones long enough to extend beyond the receptor and into the TM space. **42** and **46** were long enough to be present at the TM helix/ TM space interface. The final CH₃ of **42** and the final aryl ring of **46** are placed at the keyhole surrounded by the Ser residues lining it. All of the top binding poses for the ligands exhibited the key interactions with the Asn363 and Asp138 residues. The hydrophobic nature of this feature and the lipid bilayer were hypothesised to have had a positive impact on the ligands' binding affinity to the receptor. Unfortunately, it wasn't possible to assess the impact of the lipid bilayer on binding affinity using docking methods, but it is thought that the hydrophobic nature of the extended ligands could generate favourable interactions with the lipids, but it's effect on receptor binding is still unclear.

Docking scores do not explain the full story with respects to protein ligand interactions, however, it is worth noting that there was a correlation with the length of a ligand and the Glide score. As the ligands increased in length, the Glide score had also increased. This could be due to the increased number of interactions the ligand is able to make within the binding pocket.

The overall purpose of collecting ligand binding poses was to use one of the collected ligands and one pose to take forward for more extensive and potentially more accurate modelling investigations. Ligand **47** was the ligand of choice for two main reasons. The first is that the ligand is the longest of the set, and that binding canonically within the receptor would require the ligand to protrude into the TM space. Also, although the Glide scores weren't representative of the binding affinities, it had the lowest of the set, as well as having the highest experimental k_D .

The best docking pose for ligand **47** was used to analyse the behaviour of the protein and ligand with a binding pose of this nature. It was important that the

ligand could elicit and maintain the key hydrogen interactions with the Asn363 and Asp138 residues, and not affect the protein adversely (e.g., cause unfeasible structural changes) over a long period. Therefore 200 ns of production run simulation was performed on the protein-ligand system present in a lipid bilayer.

Upon analysing the structural activity of the receptor by means of RMSD and RMSF, it appears as though the receptor bound to ligand **47** does not show any evidence of unfavourable conformational changes, which was the ideal result. However, it was shown that the presence of ligand **47** has a very minimal effect, if any, on the receptor's tertiary structure. Both systems yielded very similar results for the RMSD of the overall protein backbone and the keyhole residues as well as for the RMSF of the protein backbone. The overall pattern of behaviour of the residues were similar between states, demonstrating that there were no residue movements that varied highly relative to all residues between states.

Ligand-residue interactions had shown that the ligand was not only capable of maintaining known key interactions within the orthosteric binding pocket, but the ligand was also capable of eliciting interactions with a residue lining the keyhole (Tyr224). This additional ligand-residue interaction in conjunction with the known ethanolamine, Asp138 and Asn363 interactions could help explaining the higher binding affinities seen in some extended ligands.

Using the hydrogen bond definition as any interaction between a hydrogen bond acceptor and hydrogen bond donor less than 0.25 nm in length, the key interactions with the Asn363 and Asp138 residues were maintained for the entirety of the simulation. It was always a possibility that favourable interactions between the ligand head group and the TM space could draw the ligand out of the binding pocket via the TM helices, but it appears that even if this force existed, the interaction between the ligand and the Asn363 and Asp138 residues were more potent, giving confidence in the ligand binding pose.

Lastly, it was important to assess the effect that binding in this method has on the tertiary structure of the receptor. An overall stability of the complex would demonstrate that there aren't any unfavourable interactions capable of causing the protein to dismantle its tertiary structure. The RMSD of the backbone atoms indeed demonstrated an overall stability of the protein within the bilayer. There can therefore be confidence that the binding pose of the extended ligand can be used as a representation of keyhole ligands in subsequent investigations.

2.5 Conclusions

The work discussed within this chapter build upon and consolidates our understanding of the mode of binding of extended ligands within the β_1 -AR. The keyhole remains a plausible explanation as to how extended ligands can occupy the orthosteric binding pocket while maintaining the key interactions known to take place when binding.

The resulting ligand docking pose for the ligand **47** was therefore sufficient and capable enough to represent keyhole ligands in subsequent investigations.

3 Exploring structure activity relationships of steric and electrostatic ligand analogues at the intramembrane space in the β_1 -AR

3.1 Aims and objectives

To maintain energetically favourable interactions with Asn^{7.39} and Asp^{3.32} in the β_1 -AR (two residues known to be critical in regulating activity³⁹), the aryloxy “head” end of ligands with extended moieties protrudes beyond the TM helices 4 and 5 and into the intramembrane space. This introduces the potential for ligand-lipid interactions to impact binding affinity/behaviour. After analysing the results of the docking studies of ligands synthesised by previous members of the research group, the longest and highest affinity ligand was used as the basis for the design of a series of analogues.

A small library of ligands was first designed and analysed via docking studies to observe the binding pose undertaken by the ligands and see whether these ligands utilise the keyhole while maintaining canonical interactions. The library was then synthesised and then tested pharmacologically to obtain binding affinity data. This data was used to assess whether the varying steric and electronic factors of the head group of the series of ligands impacts the binding affinity to the β_1 - and β_2 -AR.

All pharmacological investigations were conducted by Prof. Jillian G. Baker.

3.2 Methods

3.2.1 Design and docking of compound **47** analogues

In previous docking studies (chapter 1), compound **47** (Table 3), used as the basis for the design of subsequent analogues) was found to dock to the β_1 -AR partaking in canonical key interactions between the ethanolamine backbone of the ligand and Asp138^{7.39} and Asn363^{3.32}. Before synthesising the series of analogues, each was docked into the β_1 -AR, to see if the same method of binding was predicted.

In order to explore the steric and electronic factors, a library of compounds containing *ortho*-, *meta*- and *para*- substituted aryl rings containing electron withdrawing (fluoro), electron rich (methoxy) and bicyclic aromatic rings such as quinoline at the aryloxy (**R**) position of the ligand shown in Figure 32.

Fluorobenzyl rings were conjugated to ligands to represent electron deficient molecular properties. Fluoride atoms have electron withdrawing properties, which draws atoms within the aryl ring towards it, leaving the ring electron deficient. Methoxybenzyl rings were used to provide an electron rich aryl ring due to the methoxy oxygen atom's lone pair electrons are delocalised into the ring. A quinoline addition was included to explore the effect of a bicyclic ring at this final aryloxy position on the binding affinity at the β_1 - and β_2 -AR.

Aryloxy rings were chosen as the substitutions as opposed to a more hydrophilic component to maintain hydrophobicity within the intramembrane space. Introducing a hydrophilic component into this hydrophobic space could have a negative impact on the overall binding affinity.

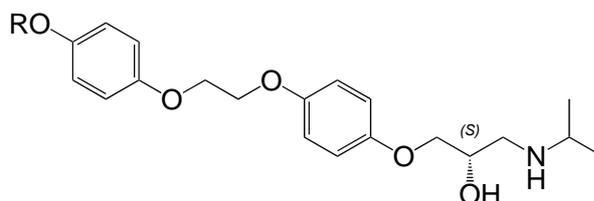


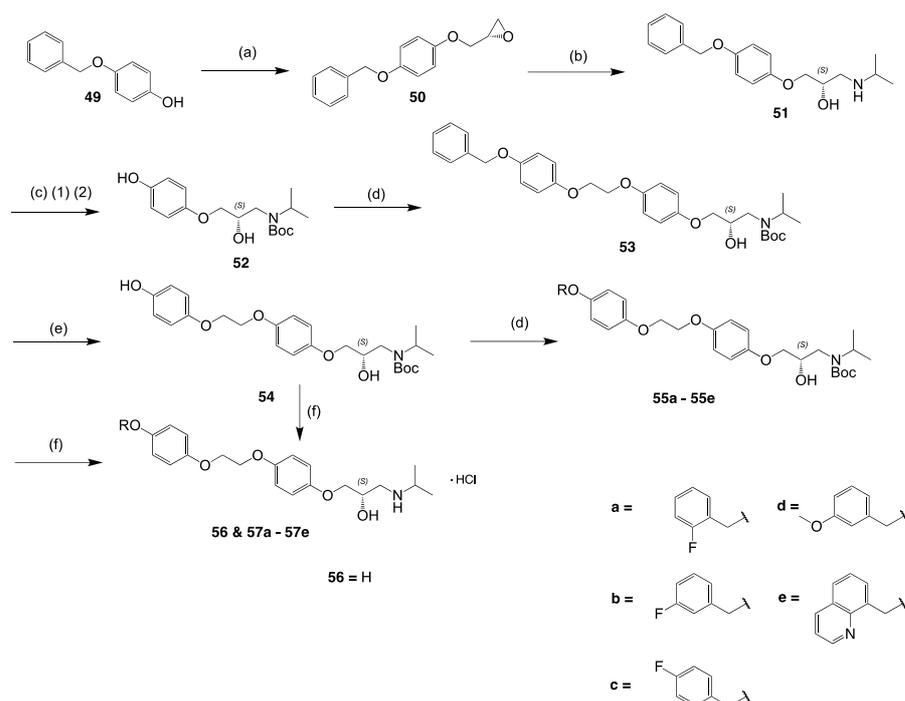
Figure 37 – General structure of the basis of the analogues varying at the point (R).

Receptor protein and ligand were prepared for docking using the methodology described in section 2.2.1. The centroid of the box was defined as the residues Asp138 and Asn363 and length of ligands to be docked was 38 Å. The inner box (centre of mass) was set to 40Å to allow for longer ligands with distributed mass. Hydrogen bond constraints were enabled at Asn363^{3.32} OD1, Asn363^{3.32} 1HD2 and Asp138^{7.39} OD1. When running the docking protocol using Glide SP, the hydrogen bond constraints were applied, and result were required to satisfy all specified interactions. Each ligand returned 50 poses that were subsequently minimized post procedure to produce a maximum of 20 poses. Default parameters were used unless otherwise stated.

3.2.2 Synthesis of series of compound **47** analogues

Each ligand began with the same general procedure until the conjugation of the aryl ring substitutes, where a variation the Williamson ester synthesis was used (Scheme 1: step d). The commercially available 4-(benzyloxy)phenol (**49**) was deprotonated and alkylated using (*S*)-oxiran-2-ylmethyl 3-nitrobenzenesulfonate (also commercially available). (*S*)-Oxiran-2-ylmethyl 3-nitrobenzenesulfonate was used as the initial starting material to ensure that the resulting oxirane was as enantiomerically pure to the *S*- enantiomer as possible. The endogenous β_1 - and β_2 -AR ligands belong to the aryethanolamine class and tend to be of the *R*-enantiomer. However, many administered β_1 - and β_2 -AR antagonists are of the aryloxypropanolamine class and are generally the *S*- enantiomer.

Scheme 1: Synthesis of compounds 49-65^a



^aReagents and conditions: (a) (i) NaH (60% in mineral oil), (S)-glycidyl nosylate, DMF, 60 °C; (ii); (b) isopropylamine, reflux 45 °C; (c)(i) H₂, 10% Pd/C, MeOH/H₂O/AcOH, room temperature (rt); (ii) Boc₂O/NaHCO₃, rt; (d) bromoalkane, CsCO₃, NaI, DMF, 80 °C; (e) H₂, 10% Pd/C, MeOH/H₂O/AcOH, rt; (f) 4M HCl in 1,4-dioxane/H₂O, r.t.

The epoxide was opened using isopropylamine (Scheme 1: step (b)) at the secondary carbon atom, leaving the desired aryloxypropanolamine. Opening at the tertiary carbon is technically possible, but the bulk of the isopropyl component of the isopropyl amine causes steric hinderance, therefore becoming a less favourable target.

Hydrogenation of the benzyl protecting group at the head group (Scheme 1: step (c)(1) & (e)) was done to expose the phenyl oxygen atom for alkylation. To prevent conjugation to the secondary amine of the ethanolamine component, it was boc-protected (Scheme 1: step (c)(2)). Selective alkylation was achieved by using a base (CsCO₃) to deprotonate the phenolic oxygen while not deprotonating the hydroxy oxygen of the ethanolamine component. This is due to the difference in pKa of these acidic protons (10 and 15 respectively (5-log units difference)). The alkylation reactions were conducted in the presence of NaI, to promote halogen

displacement of the bromoalkane.²⁰⁵ Increased reactivity of the electrophile is achieved by displacing the bromide with iodide, where the latter is a better leaving group than the former.

When the desired substituted aryloxy ring is conjugated, removal of the boc-protecting group on the amine was performed using 4M HCl in 1,4-dioxane/H₂O (Scheme 1: step (f)). The presence of the acid is capable of removing the acid labile carbamate. This results in the free amine product as a hydrochloride salt.

3.2.3 Pharmacology

Pharmacology was performed and analysed by Prof. Jillian G. Baker.

Chinese hamster ovary (CHO) cells stably expressing the human recombinant β_1 -AR (CHO- β_1 WT) and a reporter gene, Secreted Placental Alkaline Phosphatase (SPAP) under the transcriptional control of a six-cAMP response element (CRE) were used. In addition, Chinese hamster ovary (CHO) cells stably expressing the human recombinant β_2 -AR (CHO- β_2 WT) and a reporter gene, Secreted Placental Alkaline Phosphatase (SPAP) were also used. ³H-CGP12177 whole cell binding was carried out using human β_1 - and β_2 -ARs. The binding affinities of **57a – 57e** were measured at these two β -ARs. Binding affinities are reported as a logarithm of K_D which is a representation of the concentration of ligand required to occupy 50% of the receptors.

3.3 Results and discussion

3.3.1 Docking

The docking scores for the ligand series correlated relatively strongly with the pharmacological data, and all highest-ranking docking poses produced poses with

the substituted aryl ring position protruding into the TM space via the keyhole. It is important to note that these docking scores do not necessarily correlate with the true free energy of binding, and therefore the binding affinities cannot be accurately predicted using these scores.

Table 8 – Docking scores of compounds to homologised β_1 -AR (4BVN).

Compound	β_1 -AR docking score	human β_1 -AR log k_D
47	-8.12	-8.92±0.04
56	-4.997	-7.04±0.09
57a	-8.066	-9.22±0.07
57b	-7.806	-8.76±0.07
57c	-7.493	-8.77±0.01
57d	-6.903	-8.87±0.04
57e	-8.227	-8.65±0.07

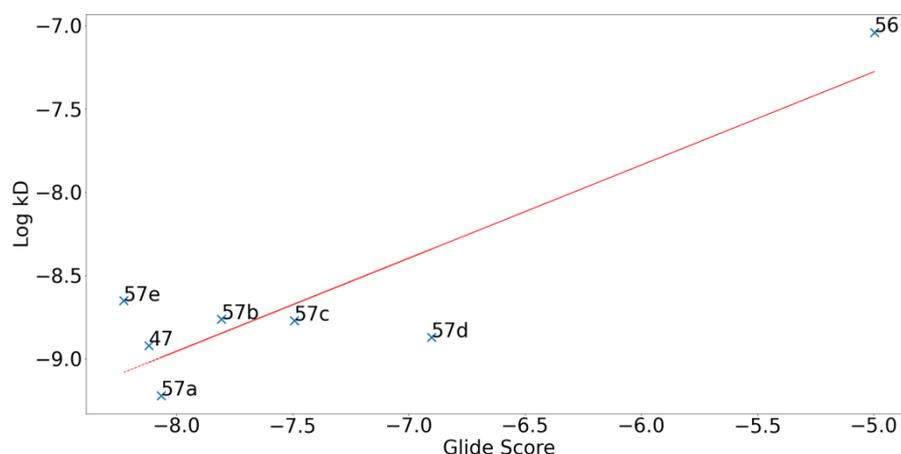


Figure 38 – Docking score data (x axis) plotted against pharmacological k_D (y axis) for compounds **57a** – **57e** in the β_1 -AR. Despite the trendline (red) demonstrating a strong positive correlation between docking scores and the pharmacological k_D , this may not necessarily be the case.

Despite the slight correlation between the docking scores and the pharmacological k_D , the docking poses themselves don't display any noticeable additional interactions within or outside of the binding pocket. Unfortunately, docking software limitations meant that it was not possible to investigate ligand interactions with membrane components at the time of writing, therefore it is

important to note that lipid-membrane interactions were not considered in the software docking scores.

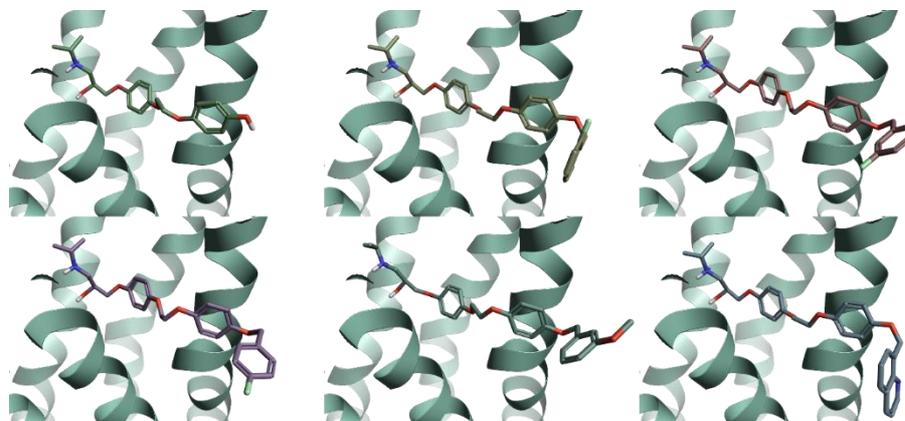
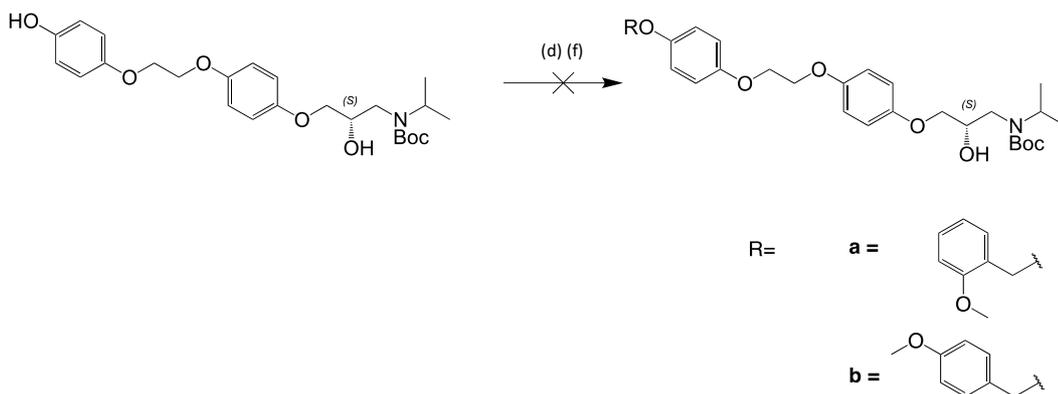


Figure 39 - Top binding poses of ligands detailed in Table 6 within the β_1 -AR (viewed from TM3/4). Top row from left to right: **56**, **57a**, **57b**. Bottom row from left to right: **57c**, **57d**, **57e**.

At first glance, Figure 33 may appear to demonstrate a strong positive correlation between docking scores and the pharmacological k_D . However, the ligand with the largest difference in both docking score and binding affinity relative to the ligand series is ligand **56**, which is also the ligand that has no substituent at the aryloxy region. The graph could therefore suggest that ligands with substituents at the aryloxy region generally show higher binding affinities and better glide scores than those without. The reasoning for this cannot be determined for certain, due to limitations discussed earlier regarding the absence of the capability to assess ligand-membrane interactions, but these interactions may have an impact.

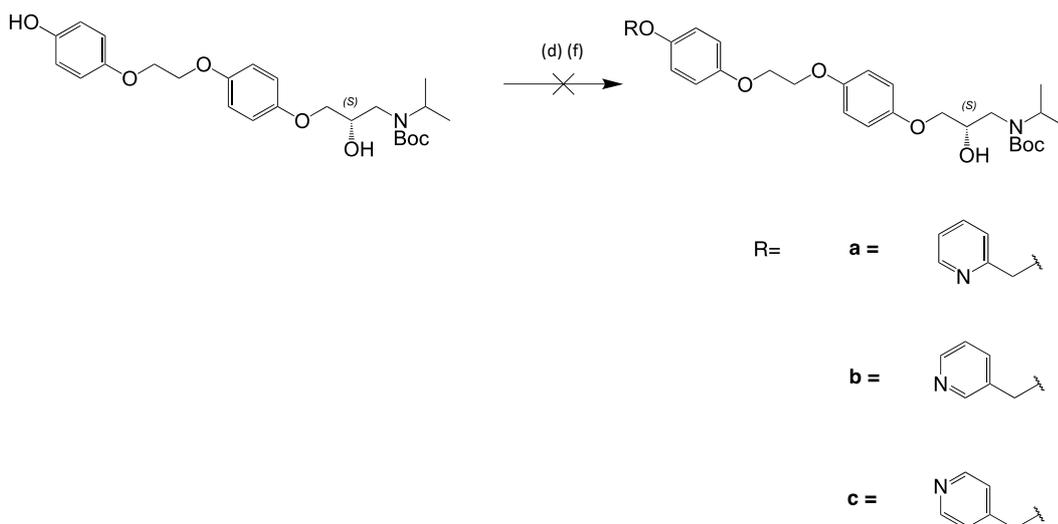
3.3.2.1 Attempted synthesis of *o*- and *p*- methoxybenzene analogues



Reagents and conditions: steps (d) and (f) from Scheme 1.

The synthesis of the *o*- and *p*- steric variations were attempted but were unfortunately unsuccessful. *p*-Methoxybenzene is an established protecting group^{206,207} that can be removed by oxidation or in the presence of an acid. In this case it is cleaved in the presence of HCl to produce *p*-methoxybenzylchloride^{208,209}. *o*-Methoxybenzene has a similar electron distribution to its *p*- counterpart, that also renders it acid labile. The final step of the synthesis of the library of ligands requires removal of the boc-protecting group, and to do this, 4M HCl in dioxane/H₂O is used. This step removes the *o*- and *p*-methoxybenzene and was therefore unable to be synthesised using the synthesis scheme used.

3.3.2.2 Attempted synthesis of pyridine analogues



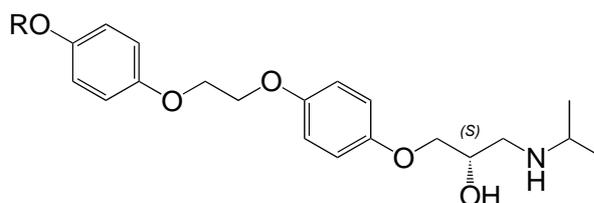
Reagents and conditions: steps (d) and (f) from Scheme 1.

Synthesis of steric analogues of pyridine (*o*-, *m*- and *p*-) were also unsuccessful. Alkylation steps within the synthesis scheme use haloalkanes, which have been shown to be able to alkylate pyridyl nitrogen atoms²¹⁰. After alkylation steps, it is possible that the alkylpyridines were alkylating at the nitrogen of other pyridines, forming repeating pyridinium rings.

3.3.3 Pharmacology

Binding affinities of **57a** – **57e** were determined at human β_1 - and β_2 -ARs. The lead compound **47** that the analogues were based on has a binding affinity to β_1 - and β_2 -AR of -8.92 ± 0.04 and -7.78 ± 0.02 respectively and showed a 14-fold selectivity for β_1 -AR over the β_2 -AR.

Table 9 - Binding affinity results for ligand series in both β_1 - and β_2 -AR. The binding affinities are recorded as a $\log K_D$.



Compound	R	HB ₁ - log <i>K_D</i>	n	HB ₂ - log <i>K_D</i>	n	B ₁ -/B ₂ - SELECTIVITY
47		-8.92±0.04	4	-7.78±0.02	4	13.8
56	H	-7.04±0.09	5	-6.03±0.05	5	10.2
57a		-9.22±0.07	5	-7.86±0.04	5	23.0
57b		-8.76±0.07	5	-7.28±0.05	5	30.8
57c		-8.77±0.01	5	-7.35±0.05	5	26.3
57d		-8.87±0.04	5	-7.50±0.03	5	24.3
57e		-8.65±0.07	5	-7.44±0.06	5	16.5

Of all the analogues tested, the *ortho*-fluorobenzyl analogue **57a** displayed the highest binding affinity to both receptors, while maintaining a 23-fold increase in

selectivity for the β_1 -AR over the β_2 -AR (0.6 nM and 13.8 nM respectively). The lack of extension to the head group **56** resulted in the lowest binding affinity of the series, and a lower binding affinity compared to the lead compound **47** at both receptors. It also had the lowest selectivity between the receptors (10.2-fold)

The *m*-F and *p*-analogues of the fluorobenzyl group (**57b** and **57c** respectively) showed similar binding affinities for both the β_1 -AR or the β_2 -AR and in both cases showed a lower binding affinity to both receptors than their *o*-F counterpart.

The *meta*-methoxybenzyl analogue **57d** displayed the second highest binding affinity of the series, placing itself below the *o*-F analogue and higher than the *p*-F and *m*-F analogues. **57d** also displayed a 24-fold selectivity for the β_1 -AR than the β_2 -AR.

The quinoline analogue **57e** displayed the lowest binding affinity to the β_1 -AR of the set when **56** (no substituent) is excluded (2.2 nM). It also showed the least selectivity (91.2 nM) when the same criteria are used (17-fold).

As seen in Table 7, although compound **47** demonstrates a high binding affinity to both receptors compared to all ligands in the series except **57a**, it also showed the second poorest selectivity between the subtypes. The least selective being **56**. Therefore, although some of the analogues displayed lower binding affinities to β_1 -AR, they show higher selectivity to β_1 -AR than β_2 -AR compared to **47**.

Ligands bearing a phenyl ring substituent showed higher selectivity ratios than those with no substituent **56** or a bicyclic aromatic ring **57e** as well as **47**.

3.4 Discussion

It was already clear that ligands with extended aryloxy head groups exhibit better binding affinities at the β_1 -AR compared to the β_2 -AR. However, it was unclear whether the length of the ligand alone was enough to improve selectivity at the β_1 -AR or whether the electron density of the aryloxy head group had a considerable impact.

Naturally, where there is an increase in favourable interactions between the ligand and the protein environment, binding affinity is likely to increase, therefore understanding which molecular properties at the extended head group lead to more favourable interactions and therefore higher binding affinities is a beneficial investigation.

One would expect that an extended ligand with a more hydrophobic head group would see less unfavourable interactions within the binding pocket compared to the hydrophilic ethanolamine component due to the lack of charged atoms and proton donors and acceptors, hence the reason for keeping the aryl ring as opposed to substituting in a more hydrophilic component. A more hydrophilic component could potentially cause the ligand to be drawn out of the binding pocket via the membrane due to the favourable interactions with the lipid beyond the membrane or cause the ligand to force itself into the protein and away from the membrane in an unfavourable way.

It appears the overall binding affinity is not significantly impacted with the substitutions used within this study, however, when a substitution is absent, the binding affinity to both receptors is decreased. This suggests that the presence of this aryl ring is important in not only increasing binding affinity to both receptors, but also increasing selectivity. It is not clear from this study however, the impact of the aryl ring as opposed to a linear chain, which would both demonstrate hydrophobic properties.

Another structural aspect of the aryl ring substitutions that was investigated was the positions of the additions to the aryl ring. Unfortunately, as explained earlier

in section 3.3.2.1, the effect of the positioning of the electron rich substitutions was unable to be analysed, however, the positioning of the electron deficient fluorobenzyl analogues appeared to make a minor difference to the binding affinity. The *o*-F analogue (**3n**) displayed the highest binding affinity at both receptors but not the highest selectivity. The *m*-F **57b** and *p*-F **57c** substituted analogues displayed a lower binding affinity at both receptors relative to **3n** but showed a higher selectivity towards the β_1 -AR. The *o*-F substituent could itself be making interactions with residues on the proteins surface, though, this wasn't directly observed in the docking study. These interactions could be more difficult for the *m*-F and *p*-F substitutes to achieve due to the flexibility of the ligands head group and rotational restraints. This could be the case for both receptors. Although all substituents could structurally engage in π - π stacking interactions with residues lining the receptor if available, the *o*-F substitute fluoride may be able to forge a stronger interaction elsewhere.

The difference between electron-rich and electron-poor substituents however becomes less clear. Not all substitutes were obtained for electron rich, therefore it is difficult to have confidence in the explanation for the difference in binding affinity between the electron rich and electron poor substitutes. Accordingly, it is still possible that the electron poorness or richness of the head group aryl ring does not in itself play a significant role in improving binding affinity or selectivity.

Introducing a bicyclic substitute improved binding affinity only slightly to the β_1 -AR compared to **47**, it also displayed the poorest selectivity of the analogues excluding ligand **56** which has no substituent. Only one bicyclic aryl ring was studied, therefore it is still unclear whether adding a bicyclic ring in general is a sensible way of improving binding affinity or selectivity compared to a single aromatic ring. Although in theory, a bicyclic ring has more opportunities to engage in π - π stacking with protein surface residues, the rigidity of the two rings could be a disadvantage in terms of flexibility and rotational freedom.

As explained previously (section 3.3.1), it is still unclear whether interactions between the head groups of the ligands and the lipids in the bilayer contribute to the increase in binding affinity. It is possible that in the series synthesised here all demonstrate a similar interaction with the bilayer due to their hydrophobicity, but it cannot be concluded with confidence due to the absence of hydrophilic head groups investigated in this study.

3.5 Conclusion

Synthesis undergone in this chapter afforded a series of ligands designed to protrude into the transmembrane space via the keyhole while preserving canonical interactions between the ethanolamine backbone of the ligand and the conserved Asn^{7.39} and Asp^{3.32} of the β_1 - and β_2 -ARs. The ligand series was based on the previously synthesised and pharmacologically tested ligand **47**. All ligands in the series demonstrated selectivity for the β_1 -AR over the β_2 -AR, suggesting that the β_1 -AR may have the better ability to structurally adapt for ligands of similar or longer length to those synthesised in this study.

The difference in electron density of the head group aryl ring does not appear to influence the binding affinity or selectivity of the ligand, however positioning of the aryl ring addition may have an impact but to what extent is still unclear. Also, the substitution with a bicyclic ring appears to have a negative effect on binding affinity and selectivity. The absence of that head group aryl ring saw a significant decrease in binding affinity and selectivity compared to all other ligands compared within this study. Suggesting that the length of the ligand and the head group aryl ring are important in improving both selectivity and binding affinity.

It appears single aryl rings of the head group positioned outside of the receptor of any electron richness produce the highest binding affinity and selectivity between the two subtypes.

3.6 Experimental

All reagents and solvents were purchased from commercial suppliers and used without further purification. All reactions were carried out at ambient temperature unless otherwise stated. Reactions were monitored by thin-layer chromatography on commercially available silica pre-coated aluminium-backed plates (Merck Kieselgel 60 F²⁵⁴). Visualisation was under UV light (254 nm and 366 nm), followed by staining with ninhydrin or KMnO₄ dips. Flash column chromatography was performed using silica gel 60, 230-400 mesh particle size (Sigma Aldrich). Automated flash column chromatography was performed on a Biotage Isolera system equipped with a dual wavelength UV detector (200-600 nm) using either silica high performance (HP) 50 µm, or C18-HP (30 µm) cartridges. NMR spectra were recorded on a Bruker-AV 400.

¹H NMR spectra were recorded at 400.13 MHz and ¹³C NMR spectra at 101.62 MHz. All ¹³C NMR are ¹H broadband decoupled. Solvents used for NMR analysis (reference peaks listed) were CDCl₃ (δH = 7.26 ppm, δC = 77.16) supplied by Cambridge Isotope Laboratories Inc., MeOD₄ (δH = 3.34 ppm, δC = 49.86) and DMSO-*d*₆ (δH = 2.50 ppm, δC = 40.45) supplied by Sigma-Aldrich (UK). Chemical shifts (δ) are recorded in parts per million (ppm) and coupling constants are recorded in Hz. The following abbreviations are used to describe signal shapes and multiplicities; singlet (s), doublet (d), triplet (t), quadruplet (q), broad (br), dd (doublet of doublets), ddd (double doublet of doublets), dtd (double triplet of doublets) and multiplet (m). Processing of the NMR data was carried out using the NMR software Mnova 12.0.4.

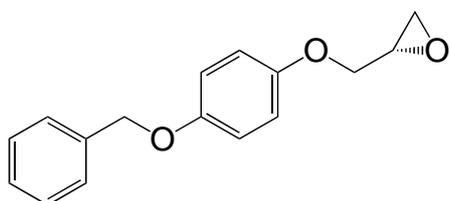
LC-MS spectra were recorded on a Shimadzu UFLCXR system coupled to an Applied Biosystems API2000 and visualised at 254 nm (channel 1) and 220 nm (channel 2). LC-MS was carried out using a Phenomenex Gemini-NX C18 110A, column (50 mm × 2 mm × 3 µm) at a flow rate 0.5 mL/min over a 5 min period. The retention time of the final product is reported using a gradient method of 5-98% solvent B in

solvent A over 5 min. (Solvent A = 0.01% formic acid in H₂O, solvent B = 0.01% formic acid in CH₃CN **(Method A)**).

Melting points (Mp) were recorded on a Reichert 7905 apparatus and were uncorrected. Optical rotation was measured on a Bellingham-Stanley ADP220 polarimeter. All high-resolution mass spectra (HRMS) were recorded on a Bruker microTOF mass spectrometer using MS electrospray ionization operating in positive ion mode.

Analytical RP-HPLC was performed using a YMC-Pack C8 column (150 mm × 4.6 mm × 5 μm) and a Phenomenex Gemini NX-C18 column (250 mm × 4.6 mm × 5 μm) at a flow rate of 1.0 mL/min. The retention time of the final product is reported using a gradient method of 10-98% solvent B in solvent A over 13 min. (Solvent A = 0.01% formic acid in H₂O, solvent B = 0.01% formic acid in CH₃CN **(Method B)**). Final products were one single peak and >95% pure.

(S)-2-((4-(Benzyloxy) phenoxy) methyl) oxirane (49)^{211,212}:



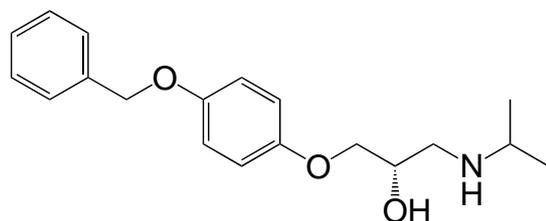
NaH (60% in mineral oil) (779.0 mg, 1.3 eq, 32.46 mmol) was placed in a flame-dried 250 mL round-bottomed flask (RBF) under an atmosphere of nitrogen gas, before adding anhydrous *N,N*-dimethylformamide (DMF) (10 mL) with stirring at room temperature. To this was added a solution of 4-(benzyloxy)phenol (**3a**) (5 g, 24.97 mmol) in anhydrous DMF (10 mL) resulting in a colour change from clear to brown. To this was added a solution of (*S*)-Oxiran-2-ylmethyl 3-nitrobenzenesulfonate (6.823 g, 1 eq, 24.97 mmol) in anhydrous DMF (10 mL), resulting in a colour change from brown to yellow. The reaction was left to stir at 60°C for 20 hours. The reaction was monitored by TLC (eluent EtOAc/PE 6:4) until starting material had been consumed, at which point, the reaction mixture was quenched with saturated NH₄Cl (50 mL) and extracted with EtOAc (3x200 ml). The combined organic extracts were then washed with saturated NaHCO₃ (200 ml), water (200

ml) then brine (200 ml), before drying over Na₂SO₄ and concentration under reduced pressure, to give 3.69 g of white solid (71% yield).

$[\alpha]_D^{20} = +20.30$ (c=0.011, dichloromethane (DCM)).

¹H NMR (CDCl₃) δ 7.49 – 7.29 (m, 5H, benzyl ring *H*), 7.00 – 6.77 (m, 4H, 1,4-disubstituted phenyl ring *H*), 5.02 (s, 2H, benzyl ring CH₂), 4.17 (dd, *J* = 11.1/3.2 Hz, 1H, CH(O)CH₂), 3.92 (dd, *J* = 11.0/5.6 Hz, 1H, CH(O)CH₂), 3.34 (ddt, *J* = 5.8/4.1/2.9 Hz, 1H, CH₂CH), 2.90 (dd, *J* = 5.0/4.1 Hz, 1H, epoxide CH₂), 2.74 (dd, *J* = 5.0/2.6 Hz, 1H, epoxide CH₂).

(S)-1-(4-(benzyloxy)phenoxy)-3-(isopropylamino)propan-2-ol (50)²¹³:



(S)-2-((4-(Benzyloxy) phenoxy) methyl) oxirane (**49**) (3.69 g, 14.39 mmol) was stirred with propan-2-amine (45 mL, ~30 eq) while heating

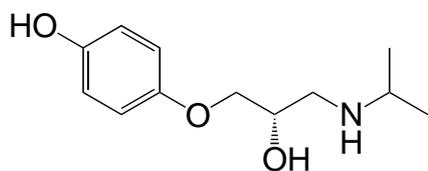
to 45°C under reflux conditions for 45h. The reaction was monitored by TLC (eluent 1N NH₃ in MeOH/DCM 7:93). Once the starting material was fully consumed, propan-2-amine was removed under reduced pressure leaving behind a white solid, which was purified using flash column chromatography (eluent 1N NH₃ in MeOH/DCM 7:93) to afford 4.2 g of a white crystalline solid (95% yield).

$[\alpha]_D^{21} = -4.67$ (c=0.009, MeOH (MeOH)).

¹H NMR (CDCl₃) δ 7.48 – 7.28 (m, 5H, benzyl ring *H*), 6.95 – 6.79 (m, 4H, 1,4-disubstituted phenyl ring ring *H*), 5.01 (s, 2H, benzyl ring CH₂), 4.09 – 3.97 (m, 1H, CH(OH)), 3.97 – 3.85 (m, 2H, NHCH₂CHCH₂), 2.96 – 2.78 (m, 2H, NHCH₂), 2.72 (dd, *J* = 12.1, 7.9 Hz, 1H, CHCH₃), 1.10 (d, *J* = 6.3 Hz, 6H, CHCH₃).

LCMS: C₁₉H₂₆NO₃ [MH]⁺ calcd 315.4; found 315.8.

(S)-4-(2-Hydroxy-3-(isopropylamino) propoxy) phenol (51)²¹³:

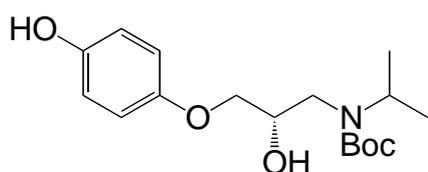


(S)-1-(4-(benzyloxy)phenoxy)-3-(isopropylamino)propan-2-ol (**50**) (4.2 g, 13 mmol) was dissolved in ~105 ml

MeOH. palladium on carbon (0.43 g, 0.30 Eq, 4.0 mmol) was added to ~10 ml H₂O

and was added to the reaction mixture. The vial containing palladium on carbon was washed successively with the remaining ~50 ml H₂O and added to the reaction mixture. Acetic Acid (30.0 mL) was then added. The vial was then sonicated after covering the flask with a rubber cap. The air in the flask was then pumped out and replaced with hydrogen x3. Upon completion of the final hydrogen replacement, the reaction was left to stir at room temperature overnight. The reaction was monitored by TLC (eluent 1N NH₃ in MeOH/DCM 7:93) and upon disappearance of the starting material, the reaction mixture was filtered using celite powder and MeOH was removed by reduced pressure. The compound was used without further purification.

***tert*-Butyl (S)-(2-hydroxy-3-(4-hydroxyphenoxy) propyl) (isopropyl) carbamate (52):**



Water (75 mL) and THF (75 mL) were added to the flask and left to stir. Sodium bicarbonate (4.2 g, 3.6 Eq, 50 mmol) was added gradually. This was left to stir for 30 mins. Boc-anhydride

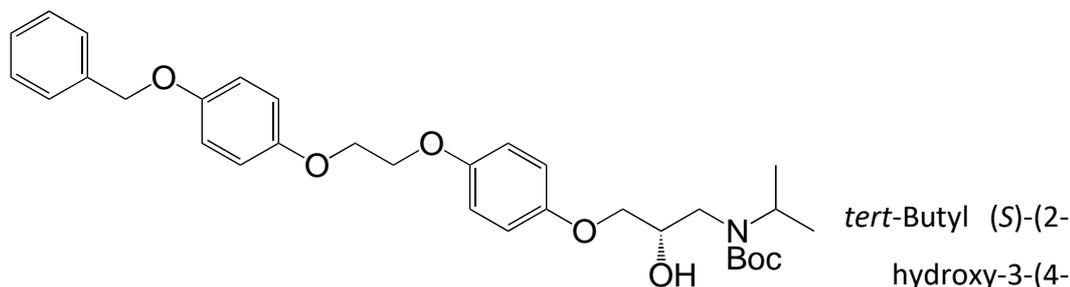
(3.3 g, 3.5 mL, 1.1 Eq, 15 mmol) was then added and the reaction was left to stir at room temperature for 20h. When the starting material was consumed, THF was removed under reduced pressure and reaction mixture was extracted using 3x 60 ml EtOAc and the combined organic layer was then dried with MgSO₄. Orange oil was then purified using silica gel at EtOAc:PE (6:4). The colourless oil was then redissolved in EtOAc and washed with NH₄Cl (50 ml) and then with brine (50 ml). The organic solution was then redried under MgSO₄, and EtOAc was removed under reduced pressure to afford 1.46 g of a yellow semi-solid (63% yield).

$[\alpha]_D^{21} = +4.67$ (c=0.009, MeOH);

¹H NMR (CDCl₃) δ 6.84 – 6.74 (m, 4H, Aryl ring *H*, phenyl *OH*), 4.22 – 4.11 (m, 1H, CHCH₃), 4.10 – 3.91 (m, 2H, NHCH₂CHCH₂), 3.84 (s, 1H, CHOH), 3.40 (d, *J* = 5.3 Hz, 2H, NCH₂), 1.51 (s, 9H, *tert*-butyl CH₃), 1.18 (dd, *J* = 12.0, 6.8 Hz, 6H, CHCH₃).

LCMS: C₁₇H₂₇NO₅Na [MNa]⁺ calcd 348.4; found 348.2.

tert-Butyl (S)-(3-(4-(2-(4-(benzyloxy)phenoxy)ethoxy)phenoxy)-2-hydroxypropyl)(isopropyl)carbamate (53):



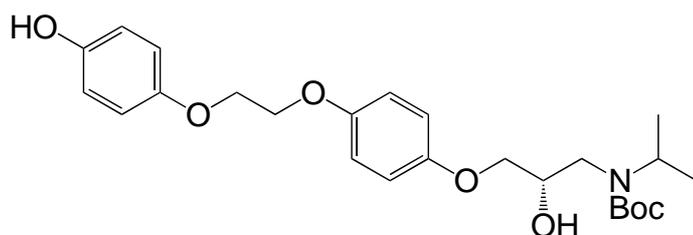
(1.46 g, 1 Eq, 4.49 mmol) was dissolved in Acetone (15 mL) while stirring. Potassium carbonate (930 mg, 1.5 Eq, 6.73 mmol) was then added periodically to the reaction mixture, followed by sodium iodide (67.3 mg, 0.1 Eq, 449 μ mol). 1-(benzyloxy)-4-(2-bromoethoxy)benzene (2.76 g, 2 Eq, 8.97 mmol) then added to the reaction mixture and it was left to stir at 80°C under reflux conditions for 144h. The reaction was monitored by TLC (eluent EtOAc/PE 3:7) until starting material had been consumed, at which point, acetone was removed by reduced pressure and EtOAc (50 ml) was added, and reaction mixture was dissolved. Organic phase was washed with H₂O (1x 50 ml), 1M NaOH (2x 50 ml), NH₄Cl (2x 50 ml) and brine (2x 50 ml). The organic phase was then dried under MgSO₄, and solvent was removed by reduced pressure leaving a brown/white solid mix. Product was extracted using column chromatography EtOAc:PE (3:7) to afford 1.0 g of a white solid (40% yield).

$[\alpha]_D^{21} = +22.40$ (c=0.007, DCM);

¹H NMR (CDCl₃) δ 7.46 – 7.27 (m, 4H, benzyl ring *H*), 6.95 – 6.81 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.02 (s, 2H, Aryl ring CH₂), 4.25 (s, 4H, CH₂CH₂), 4.14 (s, 1H, CHOH), 4.09 – 3.90 (m, 2H, NHCH₂CHCH₂), 3.83 (s, 1H, CHOH), 3.39 (d, *J* = 4.9 Hz, 2H, NCH₂), 1.49 (s, 9H, *tert*-butyl CH₃), 1.16 (dd, *J* = 12.8, 6.8 Hz, 6H, CHCH₃).

LCMS: C₃₂H₄₂NO₇ [MH]⁺ calcd 551.7; found 552.0.

tert-Butyl (S)-(2-hydroxy-3-(4-(2-(4-(benzyloxy)phenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate (54):



tert-Butyl (S)-3-(4-(2-(4-(benzyloxy)phenoxy)ethoxy)phenoxy)-2-hydroxypropyl)(isopropyl)carbamate (**53**) (1 g, 1 Eq, 2 mmol) was dissolved in MeOH (~28 ml). Palladium on carbon (20 mg, 0.1 Eq, 0.2 mmol) was added to water (7 ml) and added to the reaction mixture. The remaining water (7 ml) was used to rinse the flask and was added to the reaction mixture. The remaining MeOH (~28 ml) was added to the reaction mixture where it was then sonicated to remove as much air as possible. Air was removed from the flask via vacuum and flushed with hydrogen (x3). Reaction was then left to stir under hydrogen for 24h. The reaction was monitored by TLC (eluent EtOAc/PE 3:7) until starting material had been consumed, at which point, the reaction was filtered through celite and dried to afford 1.0 g of a white solid (81% yield).

$[\alpha]_D^{21} = -11.20$ (c=0.007, MeOH);

$^1\text{H NMR}$ (CDCl₃) δ 6.94 – 6.66 (m, 8H, Aryl ring *H*), 5.02 (s, 1H, Phenyl *OH*), 4.23 (s, 4H, CH₂CH₂), 3.99 (ddd, *J* = 36.1, 11.0, 6.6 Hz, 2H, NHCH₂CHCH₂), 3.82 (s, 1H, CHOH), 3.39 (d, *J* = 5.5 Hz, 2H, NCH₂), 1.49 (s, 9H, *tert*-butyl CH₃), 1.15 (dd, *J* = 12.1, 6.8 Hz, 6H, CHCH₃).

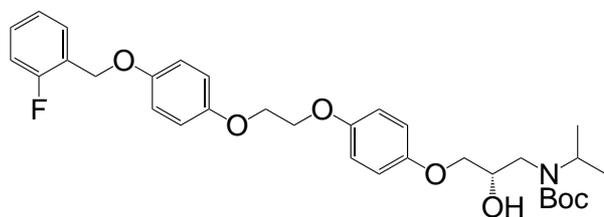
$^{13}\text{C NMR}$ (CDCl₃) δ 153.01 (COCH₂), 152.72 (COH Ph), 150.08 (1-C Ar), 116.08 (3-C and 5-C Ph), 115.99 (2-C and 6-C Ph), 115.83 (3-C and 5-C Ar), 115.31 (2-C and 6-C Ar), 80.76 (*tert*-butyl C), 71.99 (CH(OH)CH₂), 70.20 (CH(OH)CH₂), 67.43 (PhOCH₂), 67.39 (ArOCH₂), 61.61 (CH₂CH₂), 48.61 (NCH), 47.05 (NCH₂), 28.48 (*tert*-butyl CH₃), 20.53 (CH(CH₃)).

LCMS: C₂₅H₃₆NO₇ [MH]⁺ calcd 461.6; found 461.8.

General procedure A: Alkylation of *tert*-butyl (S)-(2-hydroxy-3-(4-(2-(4-hydroxyphenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate: (1.2 Eq) of alkylating agent and (3 Eq) of CsCO₃ and a catalytic amount of NaI was added to a

microwave vial containing (50mg, 1 Eq) of *tert*-butyl (*S*)-(2-hydroxy-3-(4-(2-(4-hydroxyphenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate (**54**) dissolved in 2 ml of DMF. Once the mixture was left to stir at room temperature for 10 mins, the vial containing the reaction mixture was placed in irradiation conditions for 30 mins at 80°C. Reaction was monitored by TLC (EtOAc:PE - 3:7). Once reaction is complete solvent was removed under reduced pressure, and the product was purified by column chromatography using the same eluent ratio as the TLC.

***tert*-Butyl (S)-(3-(4-(2-(4-((2-fluorobenzyl)oxy)phenoxy)ethoxy)phenoxy)-2-hydroxypropyl)(isopropyl)carbamate (55a):**



The title compound was prepared according to general procedure A, using 2-fluorobenzylbromide (16 μ l, 0.13 mmol) to give 36 mg

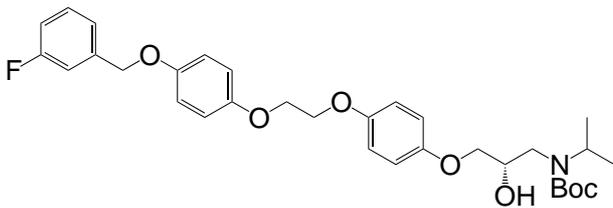
(58%) of white solid.

¹H NMR (CDCl₃) δ 7.53 (td, J = 7.5, 1.8 Hz, 1H, FCCCH), 7.33 (tdd, J = 7.5, 5.2, 1.8 Hz, 1H, FCCHCH), 7.18 (td, J = 7.6, 1.2 Hz, 1H, FCCCHCH), 7.10 (ddd, J = 9.7, 8.2, 1.2 Hz, 1H, FCCH), 6.99 – 6.83 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.12 (s, 2H, FBz CH₂), 4.91 (d, J = 15.1 Hz, 1H, CHOH), 4.27 (s, 4H, CH₂CH₂), 4.16 (s, 1H, CHOH), 4.03 (d, J = 36.7 Hz, 2H, NHCH₂CHCH₂), 3.85 (s, 1H, NCH), 3.41 (d, J = 5.2 Hz, 2H, NCH₂), 1.51 (s, 9H, *tert*-butyl CH₃), 1.18 (dd, J = 13.0, 6.8 Hz, 6H, CHCH₃).

¹³C NMR (CDCl₃) δ 153.2 (d, J_{CF} = 246.1 Hz, CF), 153.1 (Ar 1C and 4C), 129.7 (d, J_{CF} = 7.7 Hz, 2-fluorobenzyl ring 4-C), 129.7 (d, J_{CF} = 7.6 Hz, 2-fluorobenzyl ring 6-C), 124.3 (d, J_{CF} = 21.0 Hz, 2-fluorobenzyl ring 1-C), 124.2 (d, J_{CF} = 3.1 Hz, 2-fluorobenzyl ring 5-C), 115.9 (NCH₂CH(OH)CH₂OCCHCH), 115.8 (NCH₂CH(OH)CH₂OCCH), 115.80 (FBz CH₂OCCHCH), 115.4 (d, J_{CF} = 20.9 Hz, 2-fluorobenzyl ring 3-C), 115.3 (FBz CH₂OCCHCH), 80.6 (*tert*-butyl CH), 67.4 (NCH₂CH(OH)CH₂), 67.4 (NCH₂CH(OH)), 64.4 (FBz CH₂), 64.4 (CH₂CH₂), 28.5 (*tert*-butyl CH₃), 20.5 (CH(CH₃)).

LCMS: C₃₂H₄₁FNO₇ [MH]⁺ calcd 569.7; found 569.9.

***tert*-Butyl (S)-(3-(4-(2-(4-((3-fluorobenzyl)oxy)phenoxy)ethoxy)phenoxy)-2-hydroxypropyl)(isopropyl)carbamate (55b):**



The title compound was prepared according to general procedure A, using 4-fluorobenzylbromide (16 μ l, 0.13

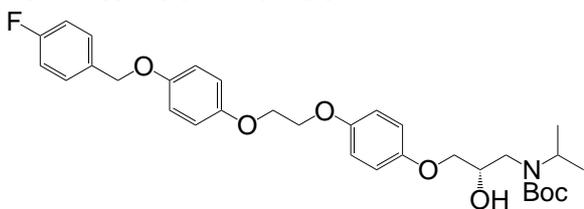
mmol) to give 15 mg (24%) of white solid.

$^1\text{H NMR}$ (CDCl_3) δ 7.34 (td, $J = 7.9, 5.8$ Hz, 1H, FCCHCH), 7.22 – 7.08 (m, 2H, FCCHCCH), 7.08 – 6.95 (m, 1H, FCCH), 6.95 – 6.73 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.01 (s, 2H, FBz CH_2), 4.85 (d, $J = 17.9$ Hz, 1H, CHOH), 4.25 (s, 4H, CH_2CH_2), 4.13 (s, 1H, CHOH), 4.00 (d, $J = 37.6$ Hz, 2H, $\text{NHCH}_2\text{CHCH}_2$), 3.83 (s, 1H, NCH), 3.39 (d, $J = 5.1$ Hz, 2H, NCH_2), 1.49 (s, 9H, *tert*-butyl CH_3), 1.16 (dd, $J = 12.9, 6.8$ Hz, 6H, CHCH_3).

$^{13}\text{C NMR}$ (CDCl_3) δ 163.0 (d, $J_{\text{CF}} = 246.2$ Hz, CF), 153.2 ($\text{CH}_2\text{CH}_2\text{OC}$), 153.1 ($\text{CH}(\text{OH})\text{CH}_2\text{OC}$), 153.0 (FBz CH_2OC), 139.9 (d, $J_{\text{CF}} = 7.8$ Hz, 3-fluorobenzyl ring 1-C), 130.1 (d, $J_{\text{CF}} = 7.7$ Hz, 3-fluorobenzyl ring 5-C), 122.7 (d, $J_{\text{CF}} = 3.1$ Hz, 3-fluorobenzyl ring 6-C), 115.9 (Ar 3-C and 5-C), 115.3 (Ar 2-C and 6-C), 114.6 (FBz CHOCCH), 114.1 (d, $J_{\text{CF}} = 21.0$ Hz, 3-fluorobenzyl ring 4-C), 80.3 (*tert*-butyl C), 70.3 (FBz CH_2), 69.9 ($\text{NCH}_2\text{CH}(\text{OH})\text{CH}_2$), 67.4 ($\text{CH}(\text{OH})$), 48.7 (NCH_2), 47.0 (NCH), 28.5 (*tert*-butyl CH_3), 20.5 ($\text{CH}(\text{CH}_3)$).

LCMS: $\text{C}_{32}\text{H}_{41}\text{FNO}_7$ $[\text{MH}]^+$ calcd 569.7; found 569.9.

***tert*-Butyl (S)-(3-(4-(2-(4-((4-fluorobenzyl)oxy)phenoxy)ethoxy)phenoxy)-2-hydroxypropyl)(isopropyl)carbamate (55c):**



The title compound was prepared according to general procedure A, using 3-fluorobenzylbromide (16 μ l, 0.13 mmol) to give 34 mg (54%)

of white solid.

$^1\text{H NMR}$ (CDCl_3) δ 7.43 – 7.34 (m, 2H, FCCHCH), 7.11 – 7.01 (m, 2H, FCCH), 6.95 – 6.78 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 4.97 (s, 2H, FBz CH_2), 4.25 (s, 4H,

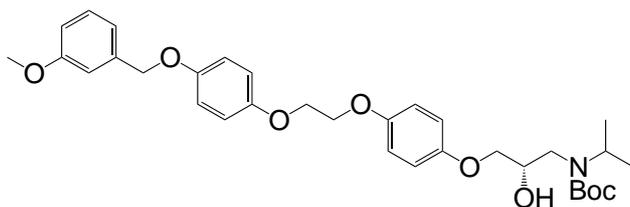
CH₂CH₂), 4.13 (s, 1H, CHOH), 4.08 – 3.92 (m, 2H, NHCH₂CHCH₂), 3.83 (s, 1H, NCH), 3.39 (s, 2H, NCH₂), 1.49 (s, 9H, *tert*-butyl CH₃), 1.16 (dd, *J* = 13.1, 6.8 Hz, 6H, CHCH₃).
¹³C NMR (CDCl₃) δ 162.5 (d, *J*_{CF} = 246.1 Hz, CF), 153.1 (FBz CH₂OCCHCHC), 153.1 (CH(OH)CH₂OCCHCHC), 153.1 (CH(OH)CH₂OC), 153.0 (FBz CH₂OC), 133.0 (4-fluorobenzyl ring 1-C), 129.3 (4-fluorobenzyl ring 2-C and 6-C), 115.7 (CH₂OAr 3-C and 5-C), 115.2 (d, *J*_{CF} = 20.9 Hz, 4-fluorobenzyl ring 3-C and 5-C), 80.6 (*tert*-butyl C), 70.3 (CH(OH)CH₂), 70.0 (FBz CH₂), 67.4 (CH₂CH₂), 67.4 (CH(OH)), 48.5 (NCH₂), 47.0 (NCH), 28.5 (*tert*-butyl CH₃), 20.5 (CH(CH₃)).

LCMS: C₃₂H₄₁FNO₇ [MH]⁺ calcd 569.7; found 570.0.

***tert*-Butyl**

(*S*)-(2-hydroxy-3-(4-(2-(4-(3-methoxybenzyl)oxy)phenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate

(55d):



The title compound was prepared according to general procedure A, using 3-methoxybenzylbromide (18 μl,

0.13 mmol) to give 29 mg (46%) of white solid.

¹H NMR (CDCl₃) δ 7.29 (dd, *J* = 8.3, 7.4 Hz, 1H, CH₃OCCHCH), 7.01 (dd, *J* = 1.6, 0.9 Hz, 1H, CH₃OCCHCHCH), 6.98 (dd, *J* = 3.3, 1.3 Hz, 2H, MeOBz CH₂OCCH), 6.93 – 6.81 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.00 (s, 2H, MeOBz CH₂), 4.84 (d, *J* = 19.8 Hz, 1H, CHOH), 4.25 (s, 4H, CH₂CH₂), 4.13 (s, 1H, CHOH), 4.05 (s, 1H, CHCH₃), 3.97 (d, *J* = 9.0 Hz, 1H, NCH), 3.82 (s, 3H, Aryl ring OCH₃), 3.38 (s, 2H, NCH₂), 1.49 (s, 9H, *tert*-butyl CH₃), 1.16 (dd, *J* = 13.1, 6.8 Hz, 6H, CHCH₃).

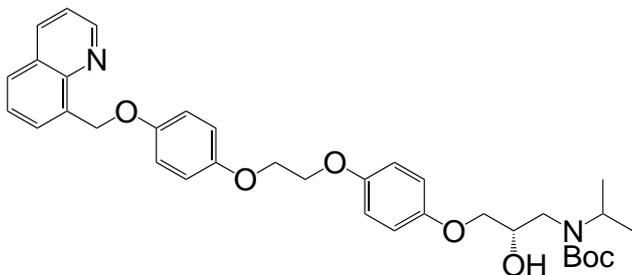
¹³C NMR (101 MHz, CDCl₃) δ 159.8 (CH₃OC), 153.1 (1,4-disubstituted phenyl ring ring 1-C and 4-C), 138.9 (3-methoxybenzyl ring 1-C), 129.6 (3-methoxybenzyl ring 5-C), 119.6 (3-methoxybenzyl ring 4-C), 115.9 (3-methoxybenzyl ring 6-C), 115.3 (1,4-disubstituted phenyl ring ring *C* excl. 1-C and 4-C), 113.5 (3-methoxybenzyl ring 2-C) 70.6 (MeOBz CH₂), 67.4 (CHOH), 67.4 (CH₂CH₂), 55.3 (COCH₃), 28.5 (*tert*-butyl CH₃), 20.5 (CH(CH₃)).

LCMS: C₃₃H₄₄NO₈ [MH]⁺ calcd 581.3; found 581.9.

tert-Butyl

(S)-(2-hydroxy-3-(4-(2-(4-(quinolin-8-

ylmethoxy)phenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate (55e):



The title compound was prepared according to general procedure A, using 8-(bromomethyl)quinoline (28.9 mg, 0.13 mmol) to give 37 mg

(57%) of red solid.

¹H NMR (CDCl₃) δ 8.94 (dd, *J* = 4.2, 1.8 Hz, 1H, CCNCH), 8.88 (dd, *J* = 4.3, 1.8 Hz, 1H, NCHCHCH), 8.20 (ddd, *J* = 8.3, 6.5, 1.8 Hz, 1H, NCHCHCHCCH), 7.93 (dt, *J* = 7.1, 1.3 Hz, 1H, NCHCH), 7.77 (ddd, *J* = 8.3, 4.6, 1.5 Hz, 1H, NCCCHCH), 7.62 – 7.40 (m, 2H, quinoline CH₂OCCH), 7.06 – 6.68 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.80 (s, 1H, CHOH), 5.21 (s, 2H, quinoline CH₂), 5.11 (s, 1H, CHOH), 4.24 (d, *J* = 5.4 Hz, 4H, CH₂CH₂), 4.00 (d, *J* = 38.8 Hz, 1H, CHCH₃), 3.82 (s, 1H, NCH), 3.39 (s, 2H, NCH₂), 1.48 (s, 9H, *tert*-butyl CH₃), 1.15 (dd, *J* = 12.9, 6.8 Hz, 6H, CHCH₃).

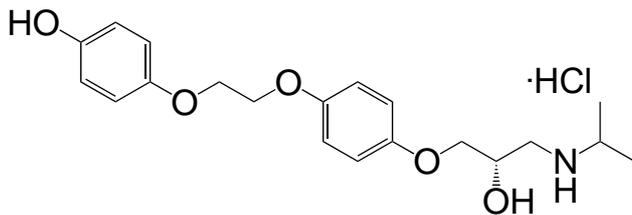
LCMS: C₃₅H₄₃N₂O₇ [MH]⁺ calcd 602.3; found 602.9.

General procedure B: *tert*-butyloxycarbonyl protecting group removal:

Compounds were dissolved in minimal MeOH and 2 ml 4N HCl/1,4-Dioxane was added. Solutions were left to stir for 2.5 hours. Reactions were monitored by TLC (EtOAc:PE - 4:6). Upon disappearance of starting material, solvent was removed by reduced pressure.

4-(2-(4-((2-Hydroxy-3-

(isopropylamino)propoxy)methyl)phenoxy)ethoxy)phenol hydrochloride (56):



The title compound was prepared according to general procedure B, using *tert*-Butyl (S)-2-hydroxy-3-(4-(2-(4-

hydroxyphenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate (**54**) (15 mg, 0.03 mmol) to give 13 mg (100%) of solid.

MP: 105 – 110°C;

[α]_D²³ = +16.00 (c=0.010, MeOH);

¹H NMR (DMSO-*d*₆) δ 8.93 (d, *J* = 8.8 Hz, 1H, NH), 8.59 (s, 1H, Phenyl OH), 6.96 – 6.87 (m, 4H, CH₂CH₂OCCH), 6.83 – 6.76 (m, 2H, NHCH₂CH(OH)CH₂OCCH), 6.76 – 6.63 (m, 2H, C(OH)CH), 5.86 (d, *J* = 4.9 Hz, 1H, ethanolamine OH), 4.24 – 4.16 (m, 4H, CH₂CH₂), 4.15 (s, 1H, CHOH), 3.99 – 3.83 (m, 2H, NHCH₂CH(OH)CH₂), 3.17 – 3.07 (m, 2H, NHCH₂), 2.96 (ddt, *J* = 13.3, 9.2, 4.5 Hz, 1H, NHCH), 1.26 (t, *J* = 6.4 Hz, 6H, CH₃).

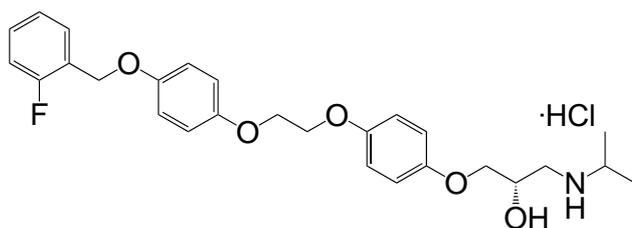
¹³C NMR (DMSO-*d*₆) δ 153.1 (CH(OH)CH₂OCCHCHC), 152.9 (CH(OH)CH₂OC), 151.9 ((OH)CCHCHC), 151.6 ((OH)C), 116.2 (CH(OH)CH₂OCCHCH), 116.0 ((OH)CCHCH), 115.9 (CH(OH)CH₂OCCH), 115.9 ((OH)CCH), 70.8 (CH(OH)CH₂), 67.3 (CH₂CH₂), 65.7 (CH(OH)), 50.3 (NHCH), 47.2 (NHCH₂), 18.6 (CH(CH₃)).

***m/z* HRMS** (TOF ES⁺) C₂₀H₂₈N₁O₅ [MH]⁺ calcd 362.1962; found 362.1982.

LCMS t_R: 2.8 min (Channel 1)

1-((4-(2-(4-((2-Fluorobenzyl)oxy)phenoxy)ethoxy)benzyl)oxy)-3-

(isopropylamino)propan-2-ol hydrochloride (57a):



The title compound was prepared according to general procedure B, using *tert*-butyl (S)-3-(4-(2-(4-((2-

fluorobenzyl)oxy)phenoxy)ethoxy)benzyl)oxy)-2-

hydroxypropyl)(isopropyl)carbamate (**55**) (36 mg, 0.06 mmol) to give 29 mg (97%) of solid.

MP: 115– 120°C;

$[\alpha]_D^{23} = -20.00$ (c=0.010, MeOH);

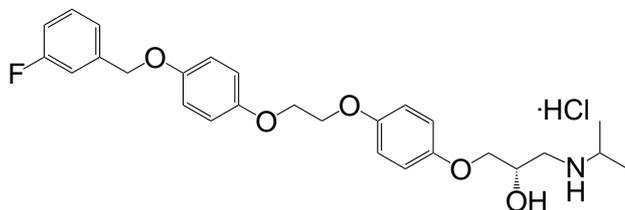
¹H NMR (DMSO-*d*₆) δ 8.96 (s, 1H), 8.60 (s, 1H, NH), 7.54 (td, *J* = 7.6, 1.7 Hz, 1H, FCCCH), 7.42 (tdd, *J* = 7.4, 5.4, 1.8 Hz, 1H, FCCH), 7.29 – 7.19 (m, 2H, FCCCHCH), 7.00 – 6.87 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.87 (d, *J* = 4.9 Hz, 1H, ethanolamine OH), 5.08 (s, 2H, FBz CH₂), 4.22 (s, 4H, CH₂CH₂), 3.93 (dt, *J* = 7.1, 3.4 Hz, 2H, CHOH), 3.12 (s, 2H, NHC₂), 2.96 (d, *J* = 9.8 Hz, 1H, NHCH), 1.27 (t, *J* = 6.4 Hz, 6H, CH₃).

¹³C NMR (DMSO-*d*₆) δ 160.8 (d, *J*_{CF} = 246.0 Hz, CF), 153.1 (CH₂CH₂OC), 153.0 (CH(OH)CH₂OC), 152.8 (FBz CH₂OC), 131.1 (d, *J*_{CF} = 7.7 Hz, 2-fluorobenzyl ring 6-C), 130.8 (d, *J*_{CF} = 7.8 Hz, 2-fluorobenzyl ring 4-C), 124.9 (FBz CH₂OCCHCHC), 124.9 (d, *J*_{CF} = 20.9 Hz, 2-fluorobenzyl ring 1-C), 116.2 (CH(OH)CH₂OCCH₂), 116.0 (FBz CH₂OCCH), 115.9 (CH₂CH₂OCCH), 115.7 (d, *J*_{CF} = 21.0 Hz, 2-fluorobenzyl ring 3-C), 70.8 (CH(OH)CH₂), 67.3 (CH₂CH₂), 65.7 (CH(OH)), 64.5 (FBz CH₂), 50.3 (NHCH), 47.2 (NHCH₂), 18.6 (CH(CH₃)).

***m/z* HRMS** (TOF ES⁺) C₂₇H₃₃F₁N₁O₅ [MH]⁺ calcd 470.2337; found 470.2357.

LCMS *t*_R: 4.4 min (Channel 1)

1-((4-(2-(4-((3-Fluorobenzyl)oxy)phenoxy)ethoxy)benzyl)oxy)-3-(isopropylamino)propan-2-ol hydrochloride (57b**):**



The title compound was prepared according to general procedure B, using *tert*-Butyl

(*S*)-3-(4-(2-(4-((3-fluorobenzyl)oxy)phenoxy)ethoxy)phenoxy)-2-hydroxypropyl)(isopropyl)carbamate (**56**) (15 mg, 0.03 mmol) to give 12 mg (100%) of solid.

MP: 120 – 130°C;

$[\alpha]_D^{21} = -8.00$ (c=0.010, MeOH);

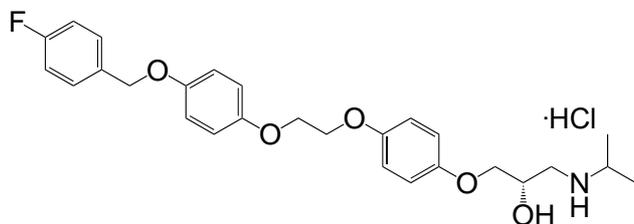
¹H NMR (DMSO-*d*₆) δ 9.05 (s, 1H), 8.65 (s, 1H, NH), 7.43 (td, *J* = 8.0, 6.0 Hz, 1H, FCCHCCHCH), 7.27 (dd, *J* = 10.7, 4.5 Hz, 2H, FBz CH₂OCCHCH), 7.15 (td, *J* = 8.7, 2.6 Hz, 1H, FCCHCHCHCCH), 7.02 – 6.82 (m, 6H, 1,4-disubstituted phenyl ring ring *H* excl. FBz CH₂OCCHCH), 5.87 (d, *J* = 4.9 Hz, 1H, OH), 5.08 (s, 2H, FBz CH₂), 3.93 (dt, *J* = 7.1, 3.4 Hz, 2H, CH₂CH₂), 3.17 – 3.07 (m, 2H, NHCH₂), 2.96 (ddt, *J* = 13.7, 9.4, 4.7 Hz, 1H, NHCH), 1.27 (t, *J* = 6.5 Hz, 6H, CH₃).

¹³C NMR (DMSO-*d*₆) δ 162.7 (d, *J*_{CF} = 243.6 Hz, CF), 153.1 (CH₂CH₂OC), 153.0 (CH(OH)CH₂OC), 152.7 (FBz CH₂OC), 140.8 (d, *J*_{CF} = 20.6 Hz, 3-fluorobenzyl ring 1-C), 130.9 (d, *J*_{CF} = 7.6 Hz, 3-fluorobenzyl ring 5-C), 123.89 (3-fluorobenzyl ring 6-C), 116.2 (CH(OH)CH₂OCCH), 116.0 (FBz CH₂OCCH₂), 115.9 (CH₂CH₂OCCH), 114.8 (CH₂CCH), 114.5 (CH₂CCHCHCH), 70.8 (CH(OH)CH₂), 69.3 (FBz CH₂), 67.3 (CH₂CH₂), 65.7 (CH(OH)), 50.3 (NHCH), 47.3 (NHCH₂), 18.6 (CH(CH₃));

***m/z* HRMS** (TOF ES⁺) C₂₇H₃₃F₁N₁O₅ [MH]⁺ calcd 470.2337; found 470.2371.

LCMS *t*_R: 4.4 min (Channel 1)

1-((4-(2-(4-((4-Fluorobenzyl)oxy)phenoxy)ethoxy)benzyl)oxy)-3-(isopropylamino)propan-2-ol hydrochloride (57c):



The title compound was prepared according to general procedure B, using *tert*-Butyl (S)-(3-(4-(2-(4-((4-fluorobenzyl)oxy)phenoxy)ethoxy)phenoxy)-2-hydroxypropyl)(isopropyl)carbamate (**57**) (34 mg, 0.06 mmol) to give 25 mg (89%) of solid.

MP: 130 – 135°C;

[α]_D²² = -14.00 (c=0.009, MeOH);

MP: 130 – 135°C;

[α]_D²² = -14.00 (c=0.009, MeOH);

¹H NMR (DMSO-*d*₆) δ 8.96 (d, *J* = 16.6 Hz, 1H), 8.59 (s, 1H, NH), 7.55 – 7.41 (m, 2H, FCCHCH), 7.29 – 7.13 (m, 2H, FCCH), 7.02 – 6.83 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.86 (d, *J* = 5.0 Hz, 1H, ethanolamine OH), 5.02 (s, 2H, FBz CH₂), 4.21 (s, 4H, CH₂CH₂), 4.18 (d, *J* = 8.7 Hz, 1H, NHCH₂CH(OH)), 3.92 (dt, *J* = 7.1, 3.4 Hz, 2H,

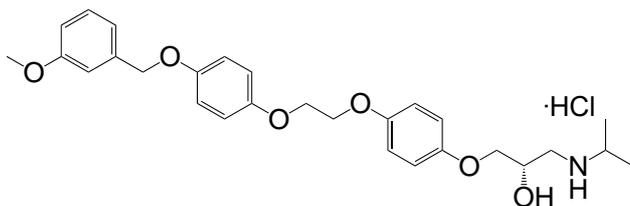
NHCH₂CH(OH)CH₂), 3.18 – 3.07 (m, 2H, NHCH₂), 3.02 – 2.89 (m, 1H, NHCH), 1.26 (t, *J* = 6.4 Hz, 6H, CH₃).

¹³C NMR (DMSO-*d*₆) δ 162.2 (d, *J*_{CF} = 243.5 Hz, CF), 153.5 (CH₂CH₂OC), 152.9 (d, CH₂CH₂OCCHCHC), 134.0 (CH₂C), 130.3 (d, *J*_{CF} = 20.5 Hz, 4-fluorobenzyl ring 3-C and 5-C), 116.2 (CH₂CH₂OCCH), 115.9 (CH(OH)CH₂OCCH), 115.9 (FBz CH₂OCCH), 115.6 (d, *J*_{CF} = 7.5 Hz, 4-fluorobenzyl ring 2-C and 6-C), 70.8 (CH(OH)CH₂), 69.4 (FBz CH₂), 67.3 (CH(OH)), 65.7 (CH₂CH₂), 50.3 (NHCH), 47.2 (NHCH₂), 18.6 (CH(CH₃)).

m/z HRMS (TOF ES⁺) C₂₇H₃₃F₁N₁O₅ [MH]⁺ calcd 470.2337; found 470.2365.

LCMS *t*_R: 4.4 min (Channel 1)

1-(Isopropylamino)-3-((4-(2-(4-((3-methoxybenzyl)oxy)phenoxy)ethoxy)benzyl)oxy)propan-2-ol hydrochloride (57d):



The title compound was prepared according to general procedure B, using *tert*-Butyl (S)-(2-hydroxy-3-(4-(2-(4-((3-

methoxybenzyl)oxy)phenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate (**58**) (29 mg, 0.05 mmol) to give 25 mg (100%) of solid.

MP: 110 – 120°C;

[α]_D²³ = -9.33 (c=0.009, MeOH);

¹H NMR (DMSO-*d*₆) δ 8.99 (s, 1H), 8.62 (s, 1H, NH), 7.30 (t, *J* = 8.1 Hz, 1H, CH₃OCCHCH), 7.03 – 6.98 (m, 2H, CH₃OCCHCH), 6.97 – 6.83 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.87 (d, *J* = 4.9 Hz, 1H, ethanolamine OH), 5.02 (s, 2H, MeOBz CH₂), 4.22 (s, 4H, CH₂CH₂), 4.19 (s, 1H, CHOH), 3.93 (dt, *J* = 7.2, 3.4 Hz, 2H, NHCH₂CH(OH)CH₂), 3.34 (s, 3H, COCH₃), 3.11 (s, 2H, NHCH₂), 2.96 (s, 1H, NHCH), 1.27 (t, *J* = 6.5 Hz, 6H, CH₃).

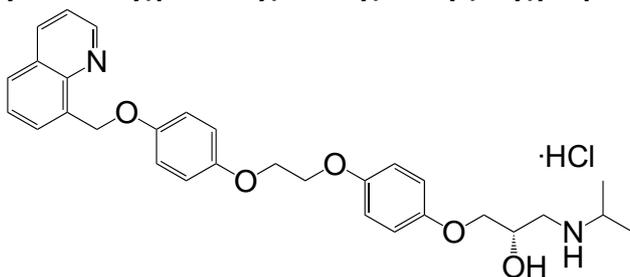
¹³C NMR (DMSO-*d*₆) δ 159.8 (CH₃(O)C), 153.1 (CH₂CH₂OC), 153.0 (CH₂CH₂OCCHCHC), 139.4 (3-methoxybenzyl ring 1-C), 130.0 (3-methoxybenzyl ring 5-C), 120.1 (3-methoxybenzyl ring 6-C), 116.0 (CH₂CH₂OCCH), 115.9 (CH₂CH₂OCCHCH), 113.5 (3-methoxybenzyl ring 2-C and 4-C), 70.8 (CH(OH)CH₂),

70.0 (MeOBz CH₂), 67.3 (CH₂CH₂), 65.7 (CH(OH)), 55.5 (CH₃(O)), 50.3 (NHCH), 47.3 (NHCH₂), 19.1 (CH(CH₃)).

m/z HRMS (TOF ES⁺) C₂₈H₃₆N₁O₆ [MH]⁺ calcd 482.2537; found 482.2566.

LCMS t_R: 4.4 min (Channel 1)

1-(Isopropylamino)-3-((4-(2-(4-(quinolin-8-ylmethoxy)phenoxy)ethoxy)benzyl)oxy)propan-2-ol hydrochloride (57e):



The title compound was prepared according to general procedure B, using *tert*-Butyl (S)-(2-hydroxy-3-(4-(2-(4-(quinolin-8-

ylmethoxy)phenoxy)ethoxy)phenoxy)propyl)(isopropyl)carbamate (**59**) (37 mg, 0.06 mmol) to give 31 mg (100%) of solid.

MP: 110 – 120°C;

[α]_D²³ = -5.60 (c=0.007, MeOH);

¹H NMR (DMSO-*d*₆) δ 8.96 (dd, *J* = 4.2, 1.8 Hz, 1H, NCH), 8.42 (dd, *J* = 8.3, 1.8 Hz, 1H, NCHCHCH), 7.96 (dd, *J* = 8.2, 1.5 Hz, 1H, NH), 7.87 (dd, *J* = 7.1, 1.4 Hz, 1H, NCHCHCHCCH), 7.62 (ddd, *J* = 9.9, 8.2, 5.7 Hz, 3H, NCHCHCHCCHCHCH), 6.95 – 6.83 (m, 8H, 1,4-disubstituted phenyl ring ring *H*), 5.70 (s, 2H, quinoline CH₂), 4.20 (s, 4H, CH₂CH₂), 3.91 – 3.77 (m, 3H, CH(OH)CH₂), 3.16 (d, *J* = 4.7 Hz, 2H, NHCH₂), 2.81 – 2.67 (m, 1H, NHCH), 1.00 (d, *J* = 6.2 Hz, 6H, CH₃).

¹³C NMR (DMSO-*d*₆) δ 153.1 (CH₂CH₂OC), 151.9 (CH(OH)CH₂OC), 151.6 (Quinoline CH₂OC), 147.6 (NC), 143.4 (NCH), 136.9 (NCHCHCH), 130.4 (CH₂CC), 128.9 (CH₂CCC), 128.6 (CH₂CCH), 128.0 (CH₂CCHCHCH), 127.6 (CH₂CCHCH), 122.3 (NCHCH), 116.2 (CH(OH)CH₂OCCH), 116.1 (Quinoline CH₂OCCH₂), 115.9 (m, CH₂CH₂OCCHCH), 70.8 (CH(OH)CH₂), 67.7 (CH(OH)), 67.3 (Quinoline CH₂), 65.7 (d, CH₂CH₂), 50.3 (NHCH), 47.3 (NHCH₂), 18.6 (CH(CH₃)).

m/z HRMS (TOF ES⁺) C₃₀H₃₄N₂O₅Na [MNa]⁺ calcd 525.2360; found 525.2360.

LCMS t_R: 4.1 min (Channel 1)

4 Investigating intra-membrane routes for ligand bivalency using covalently linked pharmacophores in the β_1 -AR homodimer

4.1 Aims and objectives

The purpose of the studies within this thesis is to investigate not only how current ligands are able to demonstrate high binding affinities in the β_1 -AR, but how to potentially improve selectivity between the β_1 - and β_2 -AR. As mentioned earlier in this thesis, targeting dimers has recently been explored as a potential method of improving receptor selectivity^{102,214,215} by means of bivalent ligands.

Previous attempts at targeting dimers in general have involved designing ligands capable of occupying the binding pockets of both protomers simultaneously, bridged by a linker using the extracellular space as the route^{107,109–111}. Discussed earlier were reasons why this approach may not be the most optimal in terms of drug delivery and bioavailability¹⁰⁴. The length and properties of the linkers can drastically impact the ligands solubility¹⁰⁴. By reducing the linkers' length, it may be possible to offset these immediate problems, while improving the ligands' selectivity.

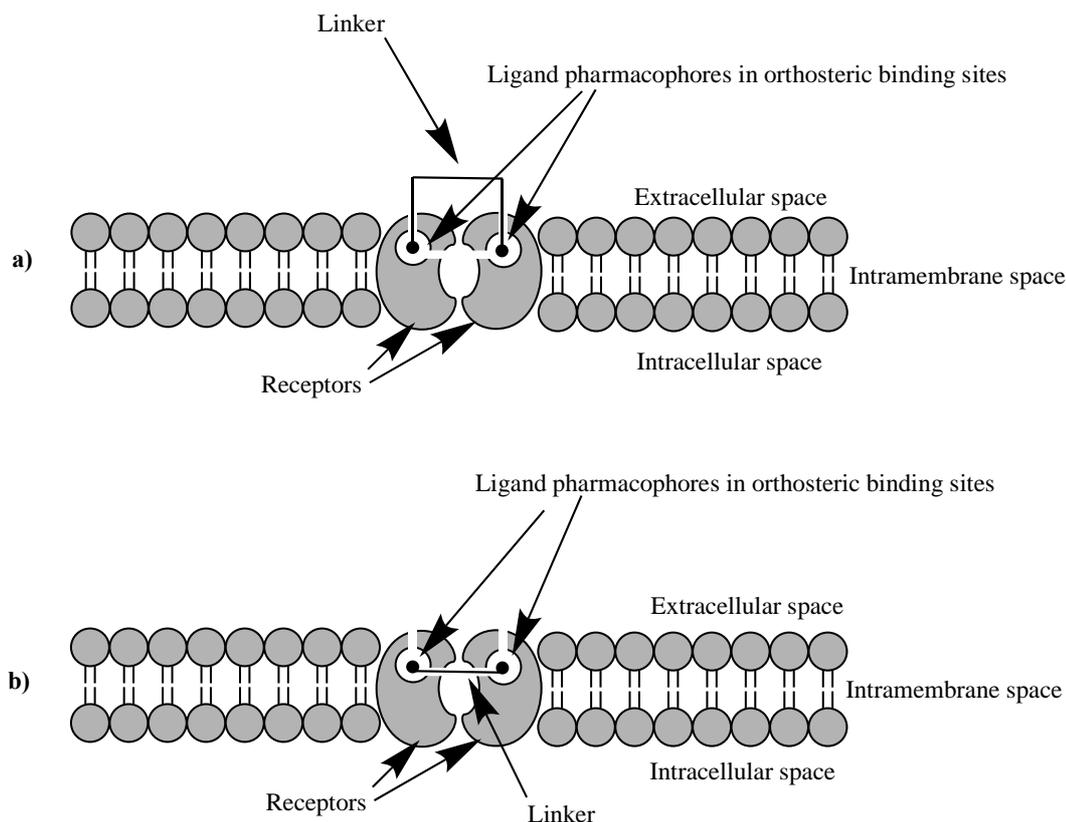


Figure 40 – Diagram showing both the current bivalent ligand linking strategy and the strategy proposed in this study. **a)** demonstrates the current linking strategy that involves using the extracellular space as a linker route. This method would only require both pharmacophores to enter the receptor conventionally at the expense of implementing potentially disadvantageous characteristics with regards to the ligand’s features and its bioavailability/activity. **b)** demonstrates this work’s proposed linker route. By using this route, linkers can be shorter and implement less features that can negatively impact the ligand’s bioavailability/activity but would instead require a pharmacophore’s entry via the intramembrane space. (Discussed further in this chapter and Chapter 5).

Another reason the selectivity may be significantly improved is because the proposed route of ligand bivalency (via the keyhole) is capitalising on a structural feature preferred by the β_1 -AR, potentially leading to receptor specific targeting.

The function of the β_1 -AR homodimer is unclear, and the role of the ligand in the dimerization process is also unknown and not the purpose of this investigation. The investigation undertaken in this study are aimed to understand the possibility of occupying both binding pockets of the β_1 -AR homodimer using the keyhole as the linkers’ route, based on the packing configuration seen in the crystal structure

PDBID: 5F8U¹²¹. This packing configuration was also observed by Mondal et al., 2013²¹⁶ using CGMD on the β_2 -AR. Mondal et al., 2013 also found that the β_1 -AR has a higher preference for forming dimers whereas the β_2 -AR has a higher preference for forming higher order oligomers. Due to the highly homologous nature of the β_1 -AR with respect to the β_2 -AR, it is suspected that it too readily forms dimers at the TM4/5 interface.

The main objectives for the investigations reported in this chapter are:

- To understand whether the distance between each protomer is maintained for a sustained period.
- To understand the length a linker would have to be in order to allow both ethanolamine cores to sustain interactions with the key residues within the binding pocket.
- To test computationally to bivalent ligands containing linkers with different levels of flexibility and evaluate their ability to maintain key interactions within the binding pocket and the effect this may have on residue interactions at the dimer interface.

Previous works in the field has shown that the β_1 -AR homodimer is transient and relatively short lived²¹⁷. It isn't properly understood what promoted the conformation or what dismantles it, however, some studies that assess the activity of the β_1 -AR homodimer have used bimolecular fluorescence complementation (BiFC) to 'lock' the dimer conformation in place²¹⁸ without affecting the rate of dimerization.

Work within this chapter was then used in works which have since been published. The publication details can be seen on page iv of this thesis.

4.2 Methods

4.2.1 Homology modelling

Prime^{157,158} was used to produce a homology model and Maestro with Prime loop refinement was used for structure refinement. The β_1 -AR wildtype sequence was obtained by NCBI and aligned with the turkey crystal template 5F8U¹²¹ using Prime (Alignment information can be found in Appendix 1). This crystal structure was chosen because it was the most recent, highest resolution β_1 -AR dimer structure available. ClustalW¹⁹³ was used to assess the alignment. The N terminus begins at Gln 56 and the C terminus ends at Cys 393. The IL3 was truncated by 62 residues and a bond created between Lys 256 and Glu 319 to form a pseudo-loop.

The model was refined (hydrogen atoms only, OPLS 2005 force field) and prepared using protein preparation wizard within Maestro.

4.2.2 Preparing the protein-model systems

4.2.2.1 Coarse-grained system

The dimer structure was converted into a coarse-grained model using MARTINIZE^{219,220}, and a DPPC lipid bilayer with an ion concentration of 0.15M NaCl and MARTINI water as a water model was constructed using INSANE²²¹.

4.2.2.1 All atom systems

The selected protein-ligand complexes were exported and input into CHARMM-GUI¹⁹⁴. Here, the protein was subject to the membrane builder^{195–198} procedure to produce a protein-membrane system. The ligand was parameterised using the Antechamber^{176,222} tool. The protein was embedded in a bilayer formed of 174 DPPC molecules in each leaflet for the monomeric system and 273 per leaflet for the dimeric system, oriented in the xy plane. The length of the x and y box axes

was set to 140 Å, and the z-axis to 140 Å, and outside the bilayer an ion concentration of 0.15M NaCl was used along with TIP3P as a water model.

4.2.3 Dimer sustainability testing

To assess the longevity of the dimer conformation, trajectories of long-time scale molecular dynamics simulations were performed. It is computationally expensive to run extended simulations in general, but more so when all atom systems are used. Because the variable of interest (the distance between the binding pocket vicinities of the two receptors) does not require a high resolution to be measured, it's possible to use coarse-grained models. Using this would significantly increase the accessible simulation time at the expense of resolution but was deemed a useful compromise for this investigation. GROMACS 2019.1 was used to perform molecular dynamics and the MARTINI v2.2 forcefield was used for coarse-grained modelling. Energy minimization was run for 50000 steps, NVT MD was run for 100 ps, NPT equilibration MD was run for 50 ns and the production MD run was done for 5 μs. The metric used for the dimer's stability was the distance between the two orthosteric binding pockets, as any significant change in the overall dimer's conformation would significantly impact this metric. This was also chosen as the metric of interest because it gives information as to what the optimal length for a bivalent ligand might be.

4.2.4 Spacer length investigation and dimer-ligand interaction

All atom dimer models containing a ligand within each protomer were run to study the average distance between the two ligands head group phenoxy oxygen atoms. This atom was chosen because it is the final atom of the ligand within the receptor at the keyhole interface. Linkers would then be designed based on this distance to bridge the ligands via the intramembrane space. This measurement will also be supported by the monitoring of key interactions within the binding pocket.

Simulations were run using GROMACS 2019.2 with the AMBER ff99sb-ildn all atom forcefield was used. Energy minimization was run for 50000 steps, NVT MD was run for 100 ps, NPT equilibration MD was run for 50 ns and the production MD run was performed for 200 ns. Snapshots were taken at 10 ps intervals.

4.2.5 Linker stability, interaction and dimer interface investigation

Once a suitable linker length was established, two linkers were designed that satisfied this, and bivalent ligands modelled. One linker option was designed to provide a flexible presence within the intramembrane space, while the other was designed to be much more rigid. This was to investigate whether rigidifying the linker could hinder or improve core interactions of the ethanolamine backbone with the important binding pocket residues. Linkers were created by using the structure of the dimer in the presence of the unlinked pharmacophores and manually modifying and connecting the ligands using the linkers proposed. The linkers were then subjected to the energy minimisation and equilibration protocol described below. These simulations were run for a relatively extended period of time to see whether key interactions were maintained and used GROMACS 2019.2 with the AMBER ff99sb-ildn all atom forcefield. Energy minimization was run for 50000 steps, NVT MD was run for 100 ps, NPT equilibration was run for 50 ns and production MD was 100 ns. Protein-ligand interactions were analysed using Flare Version 5²⁰², and the residues involved in the 5F8U dimer interface was identified using PISA²²³.

4.3 Results

4.3.1 Dimer stability investigation

The 5 μs trajectory was analysed upon completion and the centre of mass (COM) of the 2 coarse-grained beads that represented the Asp138 and Asn363 for each protomer was selected as the objects to measure the distance between.

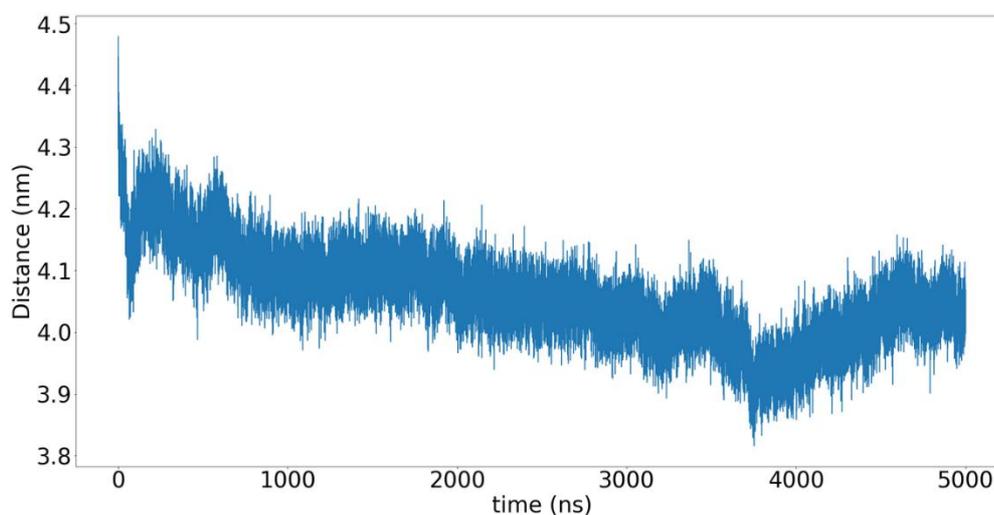


Figure 41 – Distance in nm between orthosteric binding pockets in the PDBID:5F8U dimer crystal conformation over 5 μs .

It was found that the distance between each orthosteric binding site remained relatively stable ranging between 3.8 to 4.3 nm and a mean distance of 4.09 nm. This gave confidence that 200 ns would be a reasonable timescale to investigate the distance between two ligands bound to each protomer to obtain a linker spacer length suitable for covalently bonding the two without any significant changes to the dimer conformation.

4.3.2 Spacer length investigation and dimer-ligand interaction

Spacer investigations were able to be conducted at atomistic resolution, and so for shorter timescales, but with confidence that the dimer's conformation wouldn't significantly change over the course of the trajectory, allowing for a fair measurement between the two ligands.

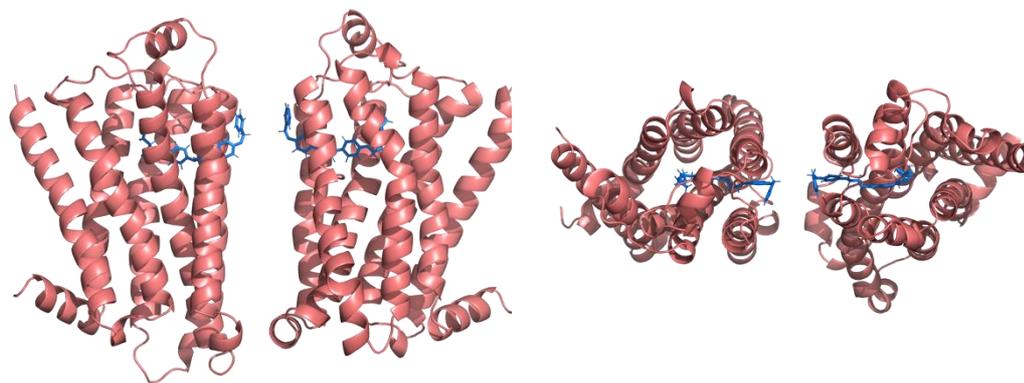


Figure 42 – Dimer conformation with both orthosteric sites occupied by the lead ligand **47**. Left: side view, right: arial view.

The atoms used to measure the distance between the area of the ligands occupying the keyhole were the O4 atoms (phenoxy oxygen). This atom was chosen due to its proximity to the keyhole.

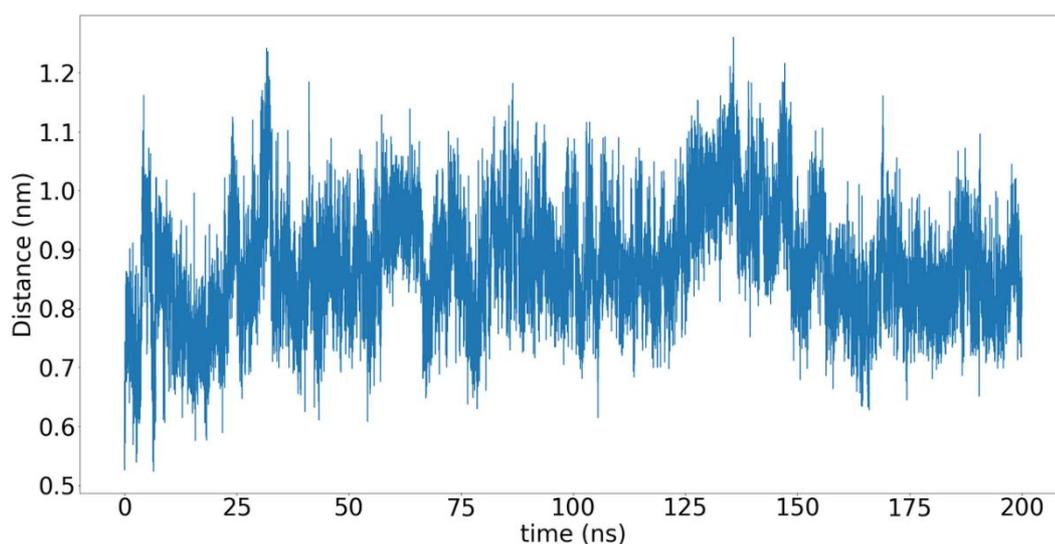


Figure 43 – Distance between phenoxy oxygen atoms of the **47** ligands occupying each protomer of the β_1 -AR homodimer conformation over the course of the 200 ns all-atom dimer simulation.

Over the 200 ns simulation, the average distance between the O4 atoms was 0.86 nm with very little observed deviation. The investigation showed that a bivalent ligand based on ligand **47** using the keyhole as a route would only require a spacer with the length of ~ 0.86 nm, significantly lower than the spacer lengths required in other homodimeric systems using extracellular space.

Another important thing to consider is whether the dimer conformation affects the ability of the ligand's ethanolamine core to maintain interactions with the key residues in the binding pocket. The ligand was considered to be within the orthosteric binding pocket if the ligand's ethanolamine backbone resided within the distance between the COM of the central carbons of the Asp 138 and Asn 363 binding 'clamp' and the COM of the residues lining the keyhole interface as defined by Emtage et al., 2017^{150,156}. This distance was calculated as 1.12 nm.

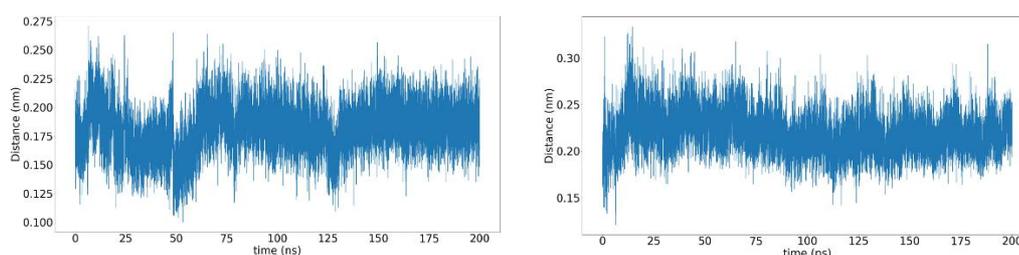


Figure 44 – Distance between COM of ethanolamine backbone atoms of ligand **47** and COM of central carbons of the key orthosteric binding pocket residues (Asn 363 and Asp 138). Left: Chain A. Right: Chain B.

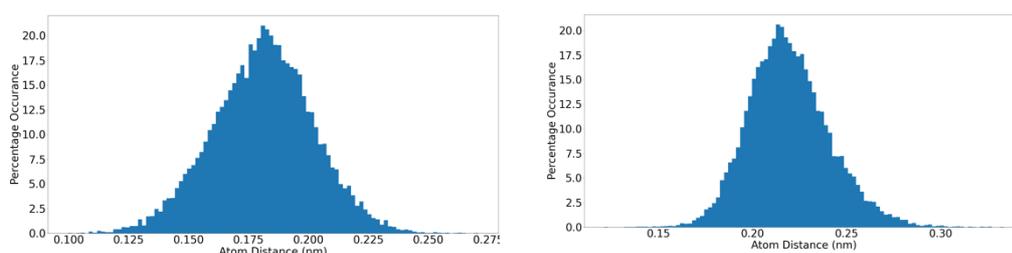


Figure 45 - Distance between COM of ethanolamine backbone atoms of ligand **47** and COM of central carbons of the key orthosteric binding pocket residues (Asn 363 and Asp 138) represented as a histogram. The percentage occurrence is measured along the y-axis. Left: Chain A. Right: Chain B.

Table 10 – Mean distance and standard deviation between COM of ethanolamine backbone atoms of ligand **47** and COM of central carbons of the key orthosteric binding pocket residues (Asn 363 and Asp 138).

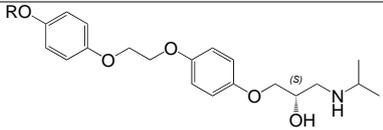
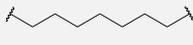
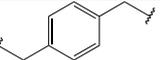
Chain	Mean distance (nm)
A	0.18 ± 0.02
B	0.22 ± 0.02

The ligand maintained strong and stable interactions within the binding pocket, suggesting that the dimer conformation of the protein does not have a negative impact on ligand **47**'s ability to maintain key contacts within the binding pocket.

4.3.3 Linker Stability Investigation

Using the distances calculated from the spacer investigation simulations, two linkers were proposed as a means of connecting the two ligand pharmacophores between two protomers. The two ligands represent two different levels of flexibility.

Table 11 – Proposed linkers with suitable length to bridge two ligands bound to each protomer.

Base ligand	Linker	Length (nm)	Characteristic
	4a 	(58) 0.86	Flexible
	4b (59) 	0.98	Rigid

The same distance measurement criteria were used for each of the designed bivalent ligands. This not only allowed for an assessment of pharmacophore engagement with the key residues within the orthosteric binding pocket but also allowed for a direct comparison between two individual ligands and one bivalent ligand created by joining two with each linker.

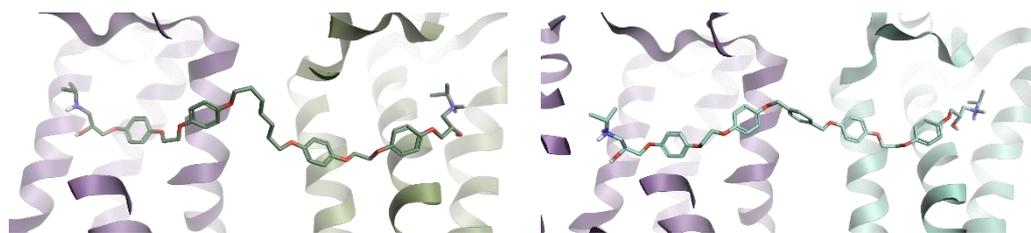


Figure 46 – Molecular representation of bivalent ligands **4a** (left) and **4b** (right) at the final frame of the 100 ns simulations. It is visually evident that both pharmacophores remain within the orthosteric binding pockets to some degree.

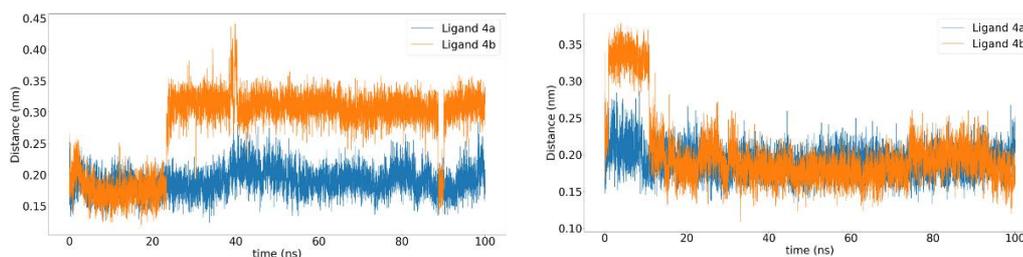


Figure 47 - Distance between COM of ethanolamine backbone atoms of the bivalent ligands and COM of central carbons of the key orthosteric binding pocket residues (Asn 363 and Asp 138). Left: Chain A. Right: Chain B.

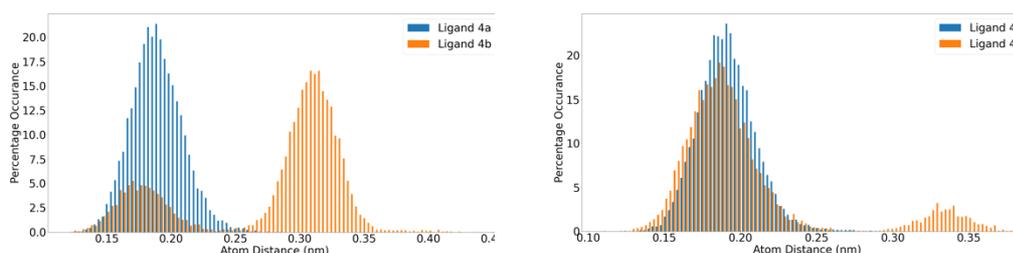


Figure 48 - Distance between COM of ethanolamine backbone atoms of the bivalent ligands and COM of central carbons of the key orthosteric binding pocket residues (Asn 363 and Asp 138) represented as a histogram. The percentage occurrence is measured along the y-axis. Left: Chain A. Right: Chain B.

Table 12 - Mean distance and standard deviation between COM of ethanolamine backbone atoms of the bivalent ligands and COM of central carbons of the key orthosteric binding pocket residues (Asn 363 and Asp 138).

Ligand	Chain	Mean distance (nm)
<i>4a</i>	A	0.18 ± 0.02
	B	0.19 ± 0.02
<i>4b</i>	A	0.28 ± 0.06
	B	0.20 ± 0.05

Both pharmacophores of the two bivalent ligands' ethanolamine backbones maintain residence within their respective chain's binding pocket with distances all below 0.3 nm from the COM of the key residues. Ligand **4a** maintained residence within the binding pocket with less deviation, showing more stability

than ligand **4b**, which demonstrated greater deviation in distance in both chains. Ligand **47** and ligand **4a** share the same mean distance and standard deviation of distance within chain A, and share a very similar distance within chain B. Ligand **4b** however demonstrates an average further distance from the key residues in chain A, but similar average distance within chain B. Both with larger deviations. Despite the greater changes in distance in ligand **4b**, the pharmacophores still remained in the binding pocket over the course of the simulation, however, it appears as though the ligand 'tugs' to and from the binding pockets. While chain A's distance is ~ 0.2 nm, chain B's distance is ~ 0.35 nm, and this is reversed around the 2000th frame, suggesting that the ligand is unable to occupy both binding pockets simultaneously to the degree that ligand **47** and **4a** do.

4.3.4 Ligand-residue interactions

Ligand residue interactions were measured using the same criteria used in Chapter 2. The interactions between the ligands and the residues within the receptor's orthosteric binding site were measured as distances and classified into the different interaction types. The atom identities are present within the table along with the specific atom numbers; however, the numbering is present not to show the specific atom's interactions, but the type of atoms involved and where they are found. The study was conducted to show the presence of the interactions between residues that the ligand as a whole as a percentage of their occurrence throughout the 100 ns simulation. To consider the interactions that occurred the most frequently, a cut-off of 30% was applied.

Table 13 – Residue interactions between the two **47** monomers and the residues within their respective receptors as a percentage of their occurrence throughout the 100 ns simulation (with a 30% threshold). The first column details the type of interaction taking place between an atom of the ligand and an atom of a residue within the receptor. The second details the ligand atoms (yellow fill: atoms of the ethanolamine backbone, brown: atoms of the ligand within the benzyloxy head end). The third column details the receptor atom involved in the interaction (blue: atoms of the ASP138 and ASN363 key residues, purple: atoms of residues surrounding the keyhole). The final column details the percentage occurrence over the duration of the trajectory (the darker the green, the more frequently the interaction occurs).

Ligand 47

Bond Type	Ligand Atom	Protein Atom	% Frames Present
<i>Salt Bridge</i>	A Ligand 47 N1	A ASP 138 OD1	83.30%
<i>Salt Bridge</i>	B Ligand 47 N1	B ASP 138 O1	63.30%
<i>Salt Bridge</i>	B Ligand 47 N1	B ASP 138 O2	36.70%
<i>Hydrogen bond</i>	A Ligand 47 H67	A ASP 138 OD2	38.90%
<i>Hydrogen bond</i>	A Ligand 47 H15	A ASP 138 OD2	60.30%
<i>Hydrogen bond</i>	A Ligand 47 H4	A ASN 363 OD1	100.00%
<i>Hydrogen bond</i>	A Ligand 47 H67	A ASP 138 OD1	92.80%
<i>Hydrogen bond</i>	A Ligand 47 O1	A ASN 363 HD22	97.90%
<i>Hydrogen bond</i>	B Ligand 47 O5	B SER 232 HG	43.40%
<i>Hydrogen bond</i>	B Ligand 47 H4	B ASN 363 O1	98.90%
<i>Hydrogen bond</i>	B Ligand 47 H67	B ASP 138 O1	82.00%
<i>Hydrogen bond</i>	B Ligand 47 H67	B ASP 138 O2	62.40%
<i>Hydrogen bond</i>	B Ligand 47 O1	B ASN 363 H6	96.50%
<i>Aromatic-Aromatic</i>	A Ligand 47 C12	A PHE 341 HE2	62.60%
<i>Aromatic-Aromatic</i>	B Ligand 47 C12	B PHE 341 H8	55.10%
<i>Aromatic-Aromatic</i>	B Ligand 47 C25	B TYR 224 H5	48.50%

Table 14 – Residue interactions between the two **4a** monomers and the residues within their respective receptors as a percentage of their occurrence throughout the 100 ns simulation (with a 30% threshold). The first column details the type of interaction taking place between an atom of the ligand and an atom of a residue within the receptor. The second details the ligand atoms (yellow fill: atoms of the ethanolamine backbone, brown: atoms of the ligand within the benzyloxy head end). The third column details the receptor atom involved in the interaction (blue: atoms of the ASP138 and ASN363 key residues, purple: atoms of residues surrounding the keyhole). The final column details the percentage occurrence over the duration of the trajectory (the darker the green, the more frequently the interaction occurs).

Ligand 4a

Bond Type	Ligand Atom	Protein Atom	% Frames Present
<i>Salt Bridge</i>	Ligand 4a N1	A ASP 138 OD2	95.60%
<i>Salt Bridge</i>	Ligand 4a N2	B ASP 138 O2	77.70%
<i>Hydrogen bond</i>	Ligand 4a H15	A ASP 138 OD1	92.50%
<i>Hydrogen bond</i>	Ligand 4a H3	A ASP 138 OD2	97.60%
<i>Hydrogen bond</i>	Ligand 4a H4	A ASN 363 OD1	99.30%
<i>Hydrogen bond</i>	Ligand 4a O	A ASN 363 HD22	98.20%
<i>Hydrogen bond</i>	Ligand 4a H56	B ASP 138 O1	77.00%
<i>Hydrogen bond</i>	Ligand 4a H30	B ASP 138 O2	77.90%
<i>Hydrogen bond</i>	Ligand 4a O10	B SER 232 H5	33.60%
<i>Hydrogen bond</i>	Ligand 4a H32	B ASN 363 O1	100.00%
<i>Hydrogen bond</i>	Ligand 4a O6	B ASN 363 H6	95.60%
<i>Cation-pi</i>	Ligand 4a N2	B PHE 218 C8	62.20%
<i>Aromatic-Aromatic</i>	Ligand 4a C10	A PHE 341 HZ	48.10%
<i>Aromatic-Aromatic</i>	Ligand 4a C10	A PHE 341 HE2	35.00%
<i>Aromatic-Aromatic</i>	Ligand 4a C36	B PHE 341 H7	57.00%
<i>Aromatic-Aromatic</i>	Ligand 4a C36	B PHE 341 H8	34.10%

Table 15 – Residue interactions between the two **47** monomers and the residues within their respective receptors as a percentage of their occurrence throughout the 100 ns simulation (with a 30% threshold). The first column details the type of interaction taking place between an atom of the ligand and an atom of a residue within the receptor. The second details the ligand atoms (yellow fill: atoms of the ethanolamine backbone, brown: atoms of the ligand within the benzyloxy head end). The third column details the receptor atom involved in the interaction (blue: atoms of the ASP138 and ASN363 key residues, purple: atoms of residues surrounding the keyhole, blue-grey: atoms of residues present on the outside of the receptor). The final column details the percentage occurrence over the duration of the trajectory (the darker the green, the more frequently the interaction occurs).

Ligand 4b

Bond Type	Ligand Atom	Protein Atom	% Frames Present
<i>Salt Bridge</i>	Ligand 4b N1	A ASP 138 OD1	33.00%
<i>Salt Bridge</i>	Ligand 4b N1	A ASP 138 OD2	67.00%
<i>Salt Bridge</i>	Ligand 4b N2	B ASP 138 O2	86.30%
<i>Hydrogen bond</i>	Ligand 4b H3	A ASP 138 OD1	48.50%
<i>Hydrogen bond</i>	Ligand 4b H3	A ASP 138 OD2	86.90%
<i>Hydrogen bond</i>	Ligand 4b H15	A ASP 138 OD2	40.80%
<i>Hydrogen bond</i>	Ligand 4b H4	A ASN 363 OD1	96.50%
<i>Hydrogen bond</i>	Ligand 4b H48	B ASP 138 O1	81.00%
<i>Hydrogen bond</i>	Ligand 4b H30	B ASP 138 O2	90.80%
<i>Hydrogen bond</i>	Ligand 4b H32	B ASN 363 O1	99.30%
<i>Hydrogen bond</i>	Ligand 4b O6	B ASN 363 H6	88.60%
<i>Aromatic-Aromatic</i>	Ligand 4b H35	B TRP 198 C5	37.70%
<i>Aromatic-Aromatic</i>	Ligand 4b H37	B TRP 198 C5	33.40%
<i>Aromatic-Aromatic</i>	Ligand 4b H55	B TYR 224 C3	30.50%
<i>Aromatic-Aromatic</i>	Ligand 4b C34	B PHE 341 H8	49.60%

The residue interaction investigation yielded results coherent to the result found in section 4.3.4. As expected, the salt bridge between the ethanolamine nitrogen atom and the donor oxygen of the Asp138 residue are present over 30% of the time in all dimer systems. The salt bridge constitutes one of the strongest bonding interactions between the ligand and the receptors' orthosteric binding site. In all systems, the salt bridge was not severed, leaving the ethanolamine backbone present within the binding pocket.

It is also important to note that, in the ligand **4b** dimer system, the hydrogen bonding interaction between the ethanolamine oxygen atom of the pharmacophore within chain A of the dimer and the Asn363 amide hydrogen was

not present within 30% of the trajectory. This corresponds well to the results found in section 4.3.4. It appears the pharmacophore within chain A was unable to maintain this specific interaction frequently. These interactions were maintained however for over 90% of the trajectory in the case of both ligand **47** and **4a**, a result also supported by the COM investigation.

Although the details of the interactions with the key Asp138 and Asn363 residues can be directly corroborated with the previous study, details of the ligands' interactions with residues beyond these residues are difficult to determine from the COM study. All systems demonstrated interactions with residues lining the keyhole within chain B, however, the ligand **47** monomer was able to maintain interactions over 40% of the time, whereas the bivalent ligands were only able to sustain these interactions for just above 30%.

In addition to these findings within chain B, it was shown that the ligand **47** monomer was able to elicit both a hydrogen bond and an aromatic-aromatic interaction with residues surrounding the keyhole. Whereas ligand **4a** was only able to elicit a hydrogen bond (as expected, due to the absence of the final aromatic ring at the head end), and ligand **4b** was only able to elicit an aromatic-aromatic interaction over 30% of the time (despite containing the same structural features as ligand **47**). Not only was this an interesting result, but the residue involved in this interaction is not one defined as lining the keyhole but is instead a residue lining the outside of the receptor between TM4 and ECL2 (Trp198). This was an interaction not seen in the other two systems.

Overall, it appears that interactions involving the ethanolamine backbone of the ligand and the key Asp138 and Asn363 occur more often than all other ligand-residue interactions, regardless of the presence or properties of the ligands investigated.

4.3.5 Investigating potential contact between ligands at the transmembrane space

Due to the highly flexible nature of the benzyloxy head end of ligand **47** within the transmembrane space, it was of interest to investigate whether this component of both ligands engage in interactions with each other during the trajectory. To do this, measurements were taken from the 1 and 4 position carbons of the benzyloxy ring to gather the distances of the closest possible contacts at the two extrema of each ring. The general distance of an aromatic-aromatic interaction would be 0.35 nm. If no distances between each atom pairs are below this measurement, it is unlikely that an aromatic-aromatic interaction is taking place.

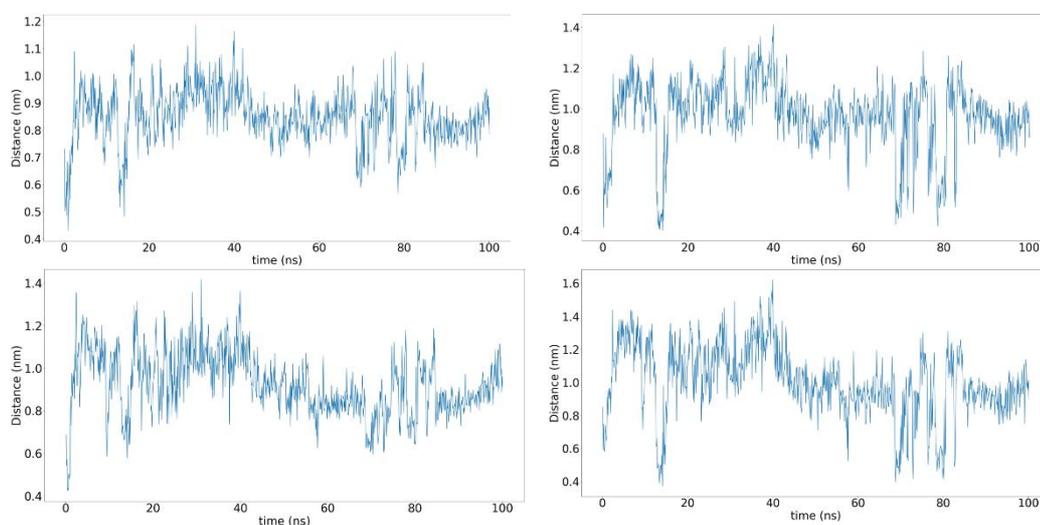


Figure 49 – Distances (nm) between benzyloxy atom pairs of the two **47** ligands bound to each receptor monomer over the course of the 100 ns simulation. Top left: Chain A carbon 1 and Chain B carbon 1. Top right: Chain A carbon 1 and Chain B carbon 4. Bottom left: Chain A carbon 4 and Chain B carbon 1. Bottom right Chain A carbon 4 and Chain B carbon 4.

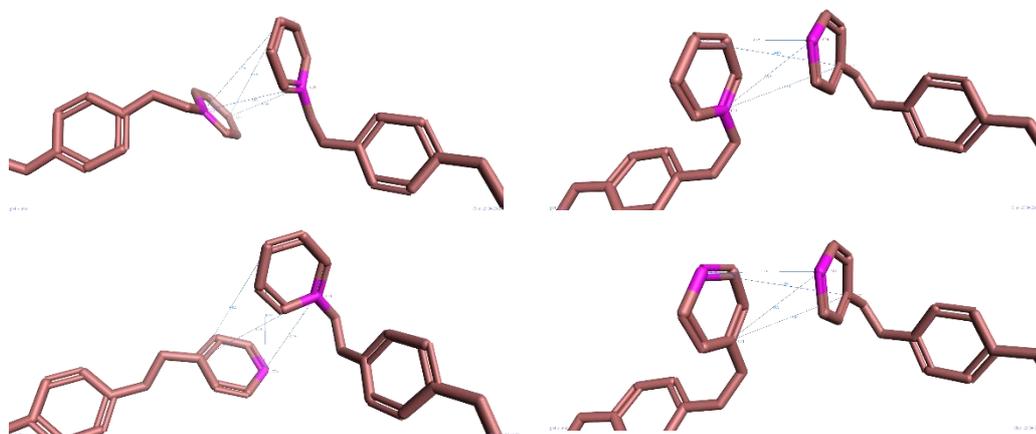


Figure 50 – Molecular representation of ligands’ benzyloxy head end conformations at their closest point for each atom pairing over 100 ns. The ligand is represented in pale pink, and atom pairing is represented in magenta. The ligand of Chain A is presented on the left and the ligand of Chain B is on the right for all diagrams. Top left: Chain A carbon 1 and Chain B carbon 1. Top right: Chain A carbon 1 and Chain B carbon 4. Bottom left: Chain A carbon 4 and Chain B carbon 1. Bottom right Chain A carbon 4 and Chain B carbon 4.

Table 16 – Ligand **47** monomers’ benzyloxy atom pairing and the minimum distances calculated over 100 ns. The final column also shows the frame (snapshot) where the minimum distance occurs.

Chain A benzyloxy atom number	Chain B benzyloxy atom number	Minimum distance (nm)	Frame number
1	1	0.43	10
1	4	0.40	141
4	1	0.42	4
4	4	0.37	141

Over the course of the trajectory the closest distance between any benzyloxy atom pairs of the ligands occurs between atom number 4 of Chain A and atom number 4 of Chain B (0.37 nm). This pairing being the closest in proximity to each other is the most likely pairing to achieve the closest distance due to both atoms being at the farthest atom into the transmembrane space of both ligands. No distances measured over any atom pairs were measured to be below 0.35 nm, the typical cut-off for an aromatic-aromatic interaction. It is interesting to note that the closest distances measured for both chain A atom 1 and chain A atom 4 to chain B atom 4 occurred during the same snapshot of the simulation. This interaction is possibly the closest the ligand comes to an aromatic-aromatic interaction over the duration of the trajectory. That doesn’t rule out that any interactions are occurring between these components of each ligand, but that an aromatic-aromatic interaction is unlikely. This could be due to the ligands’ pharmacophores being held by their strong interactions within the binding pocket, and the presence of lipid between the receptors could also cause an obstruction between ligands.

4.3.6 Contact analysis of residues at the dimer interface

The residues involved at the dimer interface were identified to be Arg148 and Ser151 of ICL2 on both chains. Other less defined interactions may also be involved along the outside of TM4/5; however, these hydrogen bonds constituted the strongest interactions and were therefore chosen for investigation.

Table 17 – Details of the interaction of residues at the dimer interface. Both the bond type and the distance of the interaction in the 5F8U crystal structure are included.

Bond type	Chain A Residue	Chain B Residue	Crystal Structure Distance (nm)
Hydrogen	ARG148	SER151	2.98
Hydrogen	ARG148	ARG148	2.41
Hydrogen	SER151	ARG148	3.10
Hydrogen	ARG148	ARG148	2.52

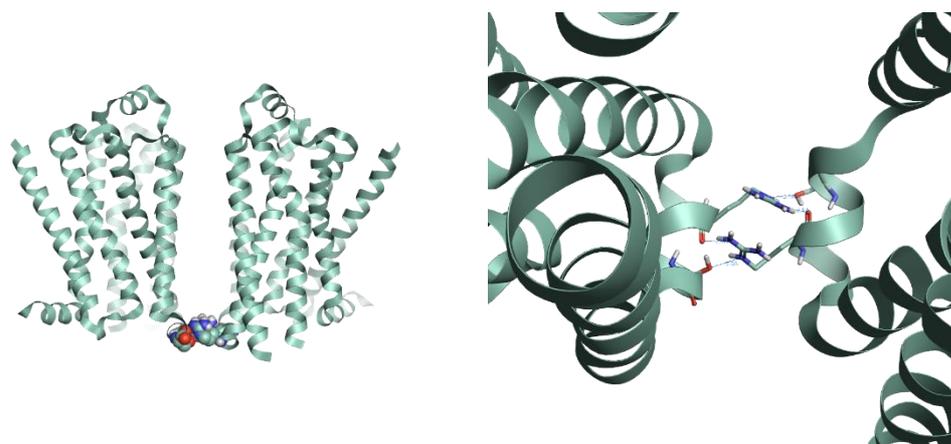


Figure 51 - Molecular representation of the residues at the FG8U TM4/5 dimer interface. Left: The residues are represented by CPK models. Right: Residues are represented by capped stick models.

The average distance of all four interactions were measured over the duration of the trajectory to identify whether the collective distance of the residues at the interface is influenced by the type of ligand present at the orthosteric binding sites.

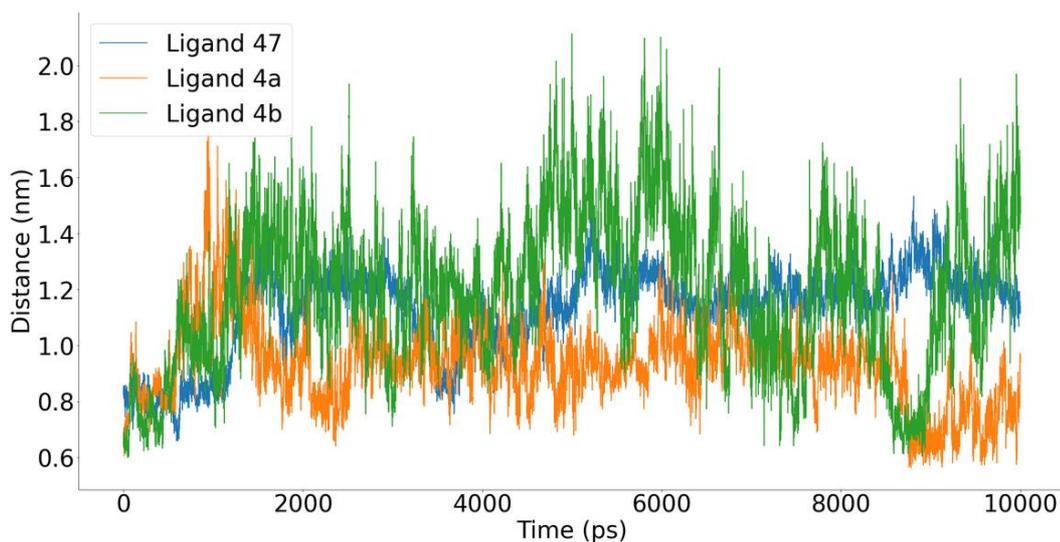


Figure 52 – Average distance (nm) between residues involved in interactions at the TM4/5 dimer interface.

Table 18 - Mean distance and standard deviation of the mean distance of all residue interactions at the TM4/5 dimer interface.

Ligand system	Mean distance (nm)
47	1.13 ± 0.15
4a	0.93 ± 0.15
4b	1.20 ± 0.27

All distances remained within the 2.5 Å cut-off definition for a hydrogen bond within this investigation, however the measured distances and standard deviations varied between systems. The dimer system containing to monomers of ligand **4a** maintained the lowest average distance (0.93 nm) and the smallest standard deviation (0.15 nm), whereas in the systems containing ligands' **47** and **4b**, the average distances were higher (1.13 nm and 1.2 nm respectively). The standard deviation for the dimer system containing ligand **47** maintained a standard deviation of (0.15 nm), matching the deviation seen in the ligand **4a** monomer system. However, the system containing ligand **4b** had a comparatively higher standard deviation (0.27 nm), demonstrating reduced stability in the presence of this bivalent ligand when compared to the other two systems.

Although it appears as though the dimer interface remains more closely associated in the presence of ligand **4a**, the differences between all the ligand systems remains quite small, suggesting that the difference in ligand may only negligibly affect the dimer interface at the timescales measured. However, if the data collected is a true representation of the system at longer timescales, it could be suggested that the presence of a flexible bivalent ligand may be ideal for reducing fluctuation and maintaining close contact between monomers.

4.4 Discussion

The length of time that a GPCR maintains a dimer conformation is likely to differ between receptors, and for GPCRs that likely function as monomers and whose dimeric functions are unclear, it's not certain what may induce or terminate the dimer conformation. In the 5 μ s coarse-grained simulation of the TM4/5 interface dimer conformation, it appears as though the protomers of the conformation remain at a stable distance from each other, in agreement with Altwaijry et al., 2017²²⁴. If the bivalent ligands do not induce dimerization in the receptors, it is possible that the ligands may be able to target an existing dimer formation.

The spacer length investigations revealed that a spacer between two pharmacophores using the keyhole as a route would only require a length of ~ 0.86 nm. Significantly lower than previous attempts at bivalent ligands via the extracellular space¹⁰⁴. As mentioned earlier in this thesis, it has been suggested that ideal bivalent ligands should contain pharmacophores that possess low to medium molecular weight and have existing high binding affinities in the low nanomolar range¹⁰⁴. Ligand **47** (the basis of the bivalent ligand design) displays a binding affinity of 1.2 nM, as recorded in competition assays performed by Prof. Jillian G Baker of the University of Nottingham. The ligand is a small molecule with a molecular weight of 452, which would be considered a molecule of low molecular weight, therefore satisfying both of those criteria.

Altwaijry et al., 2017 found that the β_1 -AR (4GPO) TM4-TM5 interface homodimer conformation displayed two pairs of interacting residues²²⁴; K159^{4.43}-Y231^{5.58} and W166^{4.50}-Y227^{5.62}, with distances of 0.90 ± 0.22 nm and 0.79 ± 0.199 nm respectively. It is important to note that the distance between these residues within the crystal structure was not determined within the study. In addition to this, the model used within this study had been homologised, and therefore the residues referenced here differ. As a result, it was not possible to obtain a direct comparison with respect to the specific residues. Instead, the crystal structure used within this study (PDBID: 5F8U) was analysed for its backbone contacts using PISA²²³. Fortunately, the residues identified using this tool are conserved between the human and avian structures, allowing for direct analysis of these residue interactions. When taking the average measurement of the interactions for each system, the results agree well with the results found by Altwaijry et al., 2017. It appears the distance between interacting atoms remain beneath the 2.5 Å hydrogen bond cut-off as defined within this study. The dimer in the presence of the unlinked and flexibly linked pharmacophores behaved more stable than their rigid counterpart. This could be explained by the ligand's tendency to sever important contacts within the orthosteric binding pocket, seen in both the binding pocket COM and ligand-residue interaction studies. This Perhaps leads to a slightly higher level of instability of the dimer at the TM4/5 interface.

However, the distance between the phenoxy oxygen atoms of the individual ligands within each protomer's binding pocket (~0.86 nm) agrees well with these figures. It is important to note that this distance between the two ligands' phenoxy oxygen is not necessarily an indication of interactions occurring, nor was identifying them the objective of this study.

It was found that the ligands are unlikely to engage in aromatic-aromatic interactions at the transmembrane space during this study. There are three likely explanations for this; the first being that the presence of lipid between the dimer interface could be an obstructing force, preventing contacts from being initiated.

The second explanation could be due to the strong binding pocket interactions with the pharmacophore of the ligands. These several highly favourable interactions between each ligands' pharmacophores and their respective protomers orthosteric binding pocket could prevent the ligands from extending further into the binding pocket and close enough to elicit strong or sustained interactions with each other. The third explanation could be due to the interactions between the benzyloxy component of each ligand and residues on the outer surface of the protein. These interactions were seen relatively frequently in Chain B, which would place the benzyloxy component of the ligand of chain B further from the benzyloxy component of the ligand of Chain A.

The designed spacers for ligands **4a** and **4b**, satisfied the minimum length between the pharmacophores defined within the study, as linkers that are too short to allow both pharmacophores to occupy the orthosteric binding pocket would not perform the role of a bivalent ligand. Linkers that are too long could introduce improved flexibility, at the expense of increased molecular bulk, weight and negatively impact bioavailability. Decker et al., 2009 and Zhang et al., 2011 proposed that hydroxyl, amine and carboxylic groups were the ideal moieties to link two pharmacophores^{105,106}, however the linkers designed in this study were aliphatic and aromatic. The reasoning for the decisions by Decker et al., 2009 and Zhang et al., 2011 is to mitigate the hinderance against the affinity or potency of the pharmacophore's functional groups. Perhaps aliphatic and aromatic chains were not explored due to the potentially long length that would be required to bridge two pharmacophores via the extracellular space. This would significantly affect the ligands lipophilicity, which could be detrimental for some targets.

Several groups have worked towards designing bivalent ligands for opioid receptors and found that a spacer length of between 18 – 25 atoms was optimal. However, it is important to note that this is necessary for spacers to be longer if the route for bivalency is further. In the case of the work in this thesis, linkers designed were around 7 atoms in length. Significantly smaller than proposed linkers in other receptors. This significant reduction in spacer atoms positively

impacts the bioavailability of these ligands and has minimal effect on the pharmacophore's ability to interact with the desired target.

Within this study, we compare a flexible linker with a rigid linker to get an idea of how well each occupies the binding pocket based on the minimum length required for to bind to each pharmacophore. The results show that the more flexible linker was able to maintain a strong prolonged presence within the binding pocket, whereas the rigid linker of a similar length was more likely to deviate away from the pocket. This evaluation was also supported by the ligand-residue contact analysis. It was shown that the systems in the presence of ligand **4a** and **4b** were able to maintain ethanolamine contact with the key residues within the binding pocket (namely the strong salt bridge interaction between the ethanolamine nitrogen and the ASP138 donor oxygen).

The behaviour seen in with ligand **4b** during the binding pocket COM investigation can be explained by the ligand-residue interactions seen in the contact frequency. The low abundance of the important hydrogen bonds between the ethanolamine hydroxyl group and the key residues Asp138 and Asn363 can explain the overall displacement of the ethanolamine core's COM within the chain A binding pocket.

As discussed earlier within this thesis, a highly flexible ligand could increase entropic penalties in the ligand's ability to occupy two protomers, however the results with the increase in flexibility suggests that once both protomers are occupied, it is able to maintain occupancy easier than a rigid linker of similar length. Perhaps if the rigid linker was of an increased length, the problems that the lack of flexibility brings to the ligand once bound could be mitigated. It is important to note that the rigid linker (**4b**), despite deviating more than both the monovalent ligands and **4a**, was still able to maintain residence within the pocket, however the distance a pharmacophore moves at any one time may be enough to sever hydrogen bonds between the ethanolamine backbone and the key residues Asp138 and Asn363. This could be problematic, as it is known from pharmacological studies that these interactions are critical¹⁹².

The overall behaviour of both bound bivalent ligands gives confidence that if the ligand can utilise the keyhole as a route for ligand bivalency, then several caveats seen in bivalent ligands utilising the extracellular space can be avoided, allowing for more intricate spacer design. However, the design of these bivalent ligands is dependent on the ligands entry and exit routes with respect to the receptor. Because of this, it is important to further understand whether entry and exit via the keyhole is a viable binding and unbinding pathway.

4.5 Conclusion

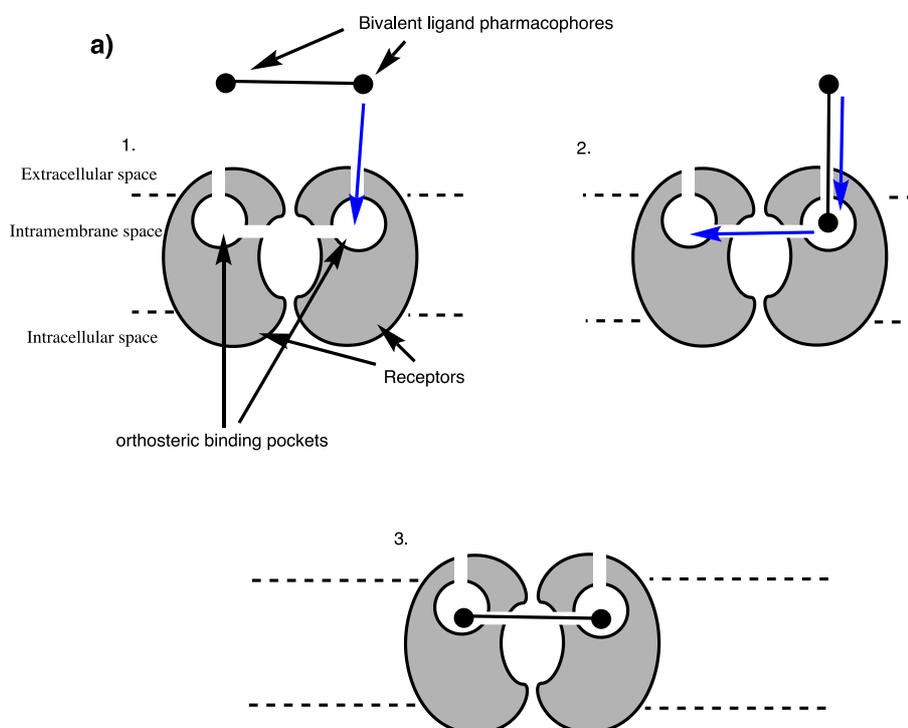
The work shown within this chapter has demonstrated that if a bivalent ligand of a suitable length is capable of binding/unbinding via the keyhole, its pharmacophores would be able to maintain residence within the orthosteric binding pocket over a sustained period irrespective of whether the ligand itself promotes the dimerization process or if dimer has already been formed. The rigid ligand also introduced a higher level of instability of interactions at the dimer interface when compared to the unlinked and flexibly linked pharmacophores. It has also been shown that a flexible bivalent ligand of suitable length can maintain a more consistent residence within the orthosteric binding pocket than a rigid ligand of a similar length. Therefore, suggesting that increasing the length of a rigid linker should be done with caution, as increasing the components of the ligand too much could have other negative implications.

As a result of the findings, it is important that it is assessed whether binding and unbinding is possible via the keyhole.

5 Estimating ligand binding and unbinding kinetics at β_1 -ARs

5.1 Aims and objectives

Understanding the kinetics of a pathway is useful for learning more about the free energy profile of a particular process. In the case of the research within this thesis, the design of selective β_1 -AR ligands by means of bivalent ligands requires understanding of the pathway of the ligand into the protein's binding pocket. The bivalent ligand design investigated here requires at least one of the two recognition units to reach its binding site via the keyhole route. This is not unreasonable, Wang et al., 2009 showed in their investigations that carazolol (**3**) was capable of exiting β_2 -AR via the TM4-TM5 cleft¹⁵¹. The structural similarity between the β -ARs could mean that similar behaviour could be seen in β_1 -AR.



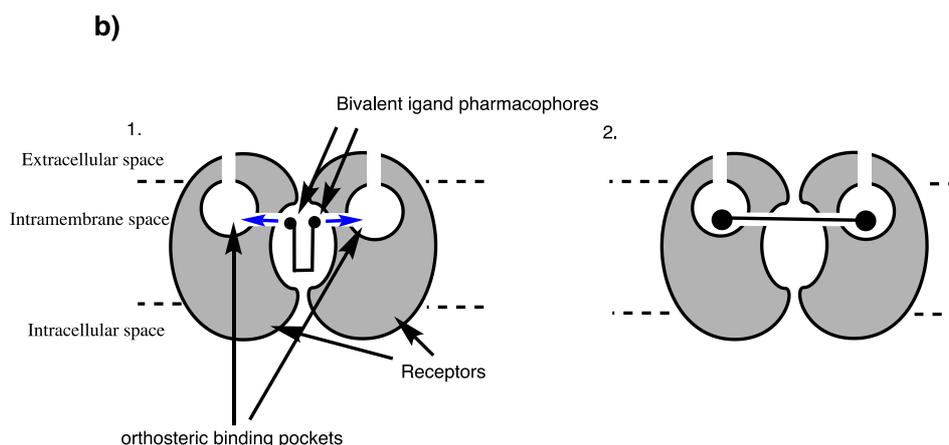


Figure 53 – General schematic of the two possible bivalent ligand pharmacophores entry pathways into the orthosteric binding pockets of a GPCR homodimer. **a)** Describes ligand entry via the extracellular space, where 1) the first pharmacophore enters the orthosteric binding pocket conventionally, then exits the receptor into the intramembrane space. From here, 2) the second pharmacophore enters the first receptor's orthosteric binding site while the first pharmacophore enters the second receptor's orthosteric binding site via the extracellular space. **b)** Describes entry of each pharmacophore into the orthosteric binding pockets of each receptor via the intramembrane space. This pathway assumes that the ligand is already present in the intramembrane space. Dotted lines represent the upper and lower membrane leaflets, and blue arrows represents the bivalent ligand pharmacophores' direction of travel.

This chapter describes the development of a computational method to compare the binding and unbinding rates of ligands from the β_1 -AR via the conventional and keyhole routes. Ligand **47** will be used as a proxy for bivalent ligands, since it shares their key structural features, and we have computational and pharmacological data for it. The computational method will have two steps. First our novel self-avoiding walk MD (SAW-MD) method will be used to generate unbinding pathways from the bound state via both exit routes, and then this will be used to define a non-linear 1D progress coordinate (PC) for a Weighted Ensemble (WE) simulation to obtain predicted binding/unbinding kinetics by each route.

To give confidence in the rate kinetics obtained from the simulations, several steps will be taken. To give confidence in the weighted ensemble (WE) method itself, a system with known rate constants (obtained from pharmacological studies) will be studied. Sykes and Charlton, 2014²²⁵ report the binding and unbinding kinetics of several β -AR ligands to the β_2 -AR, including Bisoprolol which has reasonable

structural similarity to the ligands of interest here. The k_{on} and k_{off} of bisoprolol (**27**) to the β_2 -AR PDBID: 2RH1 (prepared by Emtage, 2013¹⁵⁶) will be predicted using the weighted ensemble method to see whether the results yielded agreed with the pharmacological data.

The protein model used for the β_1 -AR within this study is a humanized version of the turkey β_1 -AR crystal structure PDBID: 4BVN. Unfortunately, as of the time of writing, there were not measured ligand rate kinetics for the β_1 -AR, however, binding affinities for several ligands have been obtained by Prof Jillian G Baker, using competition assays including both Bisoprolol (**27**) and ligand **47**. Therefore, to gain confidence in the β_1 -AR model used here, Bisoprolol (**27**) binding kinetics were predicted using the weighted ensemble method to check whether the rate constants reflected the known binding affinity of this ligand for the two β -ARs.

Lastly, rate kinetics for ligand **47** would be predicted using the WE method for both the conventional and keyhole routes, to compare the rates between the two pathways and determine whether the system preferred a particular route. Although it won't be possible to compare calculated on and off rates for ligand **47** in β_1 -AR with pharmacological data, the relationship between the k_{on} , k_{off} and k_D can be used to provide some level of validation.

5.2 Methods

5.2.1 System preparation

Ligands were docked to the receptors using Glide. The resulting ligand and docking pose were exported and input into the CHARMM-GUI membrane builder^{194,197,198,200} to obtain membrane bilayer systems. The ligand was parameterised using the Antechamber^{176,222} tool. The protein was embedded in a bilayer formed of DPPC molecules oriented in the xy plane. The length of the x and y box axes was set to 58 Å, and the z-axis to 117 Å, and outside the bilayer a

neutralising ion concentration of NaCl was used along with TIP3P as a water model.

5.2.2 Obtaining ligand unbinding pathways using SAWMD

Ligands were docked to the receptors using Glide. The resulting ligand and docking pose were exported and input into CHARMM-gui^{194,197,198,200} to obtain membrane bilayer systems and subsequently into the SAWMD protocol as the starting coordinates. The SAWMD protocol employed used OpenMM²²⁶ and the atoms used to apply targeted MD forces and base the RMSD calculations on were defined as alpha carbons of residues within a 5 Å zone around the ligand within the binding pocket. The force to be applied upon these atoms was 0.225 nm, and the procedure was run for 100 cycles of 10 ps simulations. Each protein-ligand system was simulated 6 times to obtain 24 trajectories in total.

5.2.3 Obtaining rate kinetic estimates using WEMD

The protein-ligand membrane systems used to initialise SAWMD simulations were also used as the starting coordinates (bound state) for the WEMD simulations. The end point (unbound state) was defined as the snapshot from the selected SAWMD trajectory where the ligand was fully submerged within the solvent or membrane, as appropriate. For the study, the unbinding trajectory used to make the definition of the initial and final states was reversed. The unbinding (or binding) pathway was then defined as the shortest route between the initial and final states obtained by hopping from one SAW-MD trajectory snapshot to another (not necessarily in the chronological order they were originally sampled, and not necessarily making use of every snapshot), using RMSD between selected atoms (see below) as the distance metric. Snapshots that contribute to the pathway are referred to as “waymarks”, and the overall pathway consists of the straight-line segments that link these waymarks. When running WE simulations, the 1D

progress coordinate was defined based on the position of a structure along this non-linear pathway as follows. The path segment whose two waymarks had the smallest RMSD from the structure of interest was identified. Next the projection of this structure onto the line segment was calculated. The progress coordinate was then defined as the sum of all the preceding path segments plus the relevant portion of the current segment. The atoms used to calculate the progress coordinate were defined as the OCCN+ atoms of the ligand's ethanolamine backbone, the donor oxygen atoms on the Asp^{3.32} residue R group and the donor oxygen and Nitrogen of the Asn^{7.39} residue R group. These atoms were chosen as they are critical in determining the binding or unbinding of ligands in the β -ARs. The one-dimensional progress coordinator was then used for the adaptive binning, where when a value above the previous highest progress coordinate creates a new bin boundary. To prevent large bin boundaries, a maximum bin width was defined as 0.1 nm. If a bin boundary was beyond this, intermittent bins were created between the new highest boundary and the previous highest boundary. The number of walkers per bin (N) was 20, individual simulations were run for 2 ps, 200 cycles of WEMD were run per simulation, and each simulation was repeated three times with different initial velocity distributions. Simulations were performed using AMBER 2018¹⁷⁰, and the AMBER-ff19SB²²⁷ forcefield was used.

5.2.4 Initial iterations of the proposed methods

The SAWMD method generates 'waymarks' (.pdb snapshots at user defined RMSD intervals based on the user selected atoms), that were initially used with the WEMD technique. The idea was to use the waymarks as the user defined bins and during the resampling phase of the technique, trajectories would be assigned to bins based on the waymark they are closest to, based on RMSD. However, this approach was found not to be suitable for several reasons.

Due to technical limitations of the packages used within the code, a maximum of 31 waymarks could be produced before the method is terminated. It was

therefore important to define waymark intervals (i.e., how large the change in RMSD must be from the previous waymark before a new waymark is produced) that 1) did not produce waymarks too close to each other and therefore failing to reach an unbound state before 31 waymarks are reached and 2) did not produce waymarks too far from each other so that the ligand is expelled too quickly, therefore potentially creating energy barriers that would make it difficult for the WEMD method to reach a steady-state flux.

Initially, (as discussed at the start of section 5.2.4) the waymarks produced by the SAWMD method were going to be used to define Voronoi bin centres, however this proved to be inefficient, as binding events were not witnessed and hardly any progression throughout the bins occurred. It was suspected that energy barriers between waymarks could be too high (despite defining smaller waymark RMSD intervals explained above), and that trajectories with smaller movements from previous bins towards the succeeding bin would be resampled into the same bin, thus creating the overall weighting to build up at these waymarks. This not only prevented trajectories from reaching the target bin and therefore being recycled, but it also would have heavily impacted the recycled flux if a binding event were to occur.

Initially, the WEMD method used the same atom definition as the one used in the SAWMD method (the ligand and the residues within 5 Å zone around the ligand within the binding pocket). This proved problematic, since the portion of the ligand present in the membrane is highly flexible. This caused two major issues; Firstly, the RMSD measured would be heavily overrepresented by flexible head end, resulting in trajectories being assigned to new bins based on RMSD calculations influenced mainly by this component of the ligand. Secondly, the resampling phase would be required to perform a calculation on a large selection of atoms, therefore negatively impacting performance. To circumvent this, the atoms described in section 5.2.3 were selected to represent the ligand and key residues.

Previous iterations of the WEMD investigation used simulations with many atoms (183589). The system was the one produced for use within Chapter 2 of this thesis. However, this proved to be a problem with regards to memory management. The WEMD code used utilises MDTraj²²⁸ during the resampling phase. When several trajectories are created, the protocol would fail citing a 'Memory Error' upon loading several trajectories with substantial amounts of atoms. To remedy this, a 'mini-membrane' was created using CHARMM-GUI^{194,197,198,200}. This resulted in several systems with substantially less atoms (~37000 average for both β_1 - and β_2 -AR systems), therefore reducing the hardware's memory load.

Using a mini-membrane system also meant that a box size could be defined for use with the k_{on} calculations, which relies on a defined volume to calculate a concentration. Previous attempts at optimising WEMD calculations included the use of implicit solvent, which would significantly reduce the number of atoms in the system at the expense of losing an explicit lipid membrane and box dimensions. The former being ideal for best replicating the ligand's membrane interactions (which would be lost to an implicit solvent) and the latter being essential for calculating k_{on} , as the value is concentration dependent (Equation #7) section 1.9.7).

The methods detailed in section 5.2.1, 5.2.2 and 5.2.3 were the methods used to circumvent the issue that arose during these initial iterations.

5.3 Results

5.3.1 Obtaining ligand unbinding pathways using SAWMD

All SAWMD runs resulted in ligands exiting from the orthosteric binding site. Both β_1 - and β_2 -AR systems containing Bisoprolol (**27**) bound to the orthosteric site saw the ligand exit via the conventional route and into the extracellular space. Ligand

47 bound to β_1 -AR exited through the keyhole 50% of the time and exited via the conventional route 50% of the time.

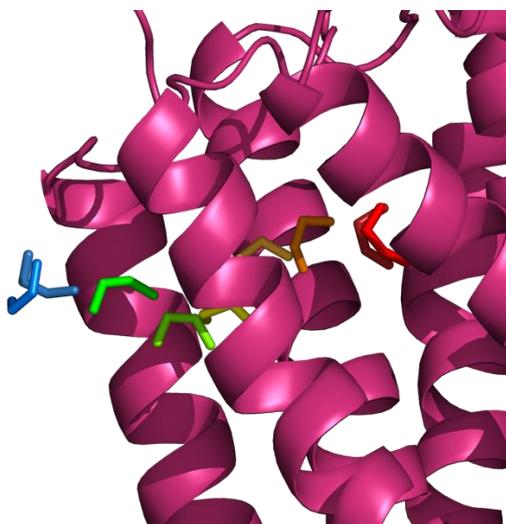


Figure 54 – Example exit pathway via the keyhole into the intramembrane space. The coloured atoms represent the OCCN atoms of the ethanolamine core of the ligand. Warmer colours indicate that the ethanolamine core is closer to the key residues within the orthosteric binding site and cooler colours indicate that the ethanolamine core is further away from the orthosteric binding pocket and closer to the intramembrane space.

Table 19 – Ligand exit path rate occurrence count using SAWMD.

Receptor	Ligand	No. of conventional route exits	No. of keyhole route exits
β_2	Bisoprolol	6	0
β_1	Bisoprolol	6	0
β_1	47	3	3

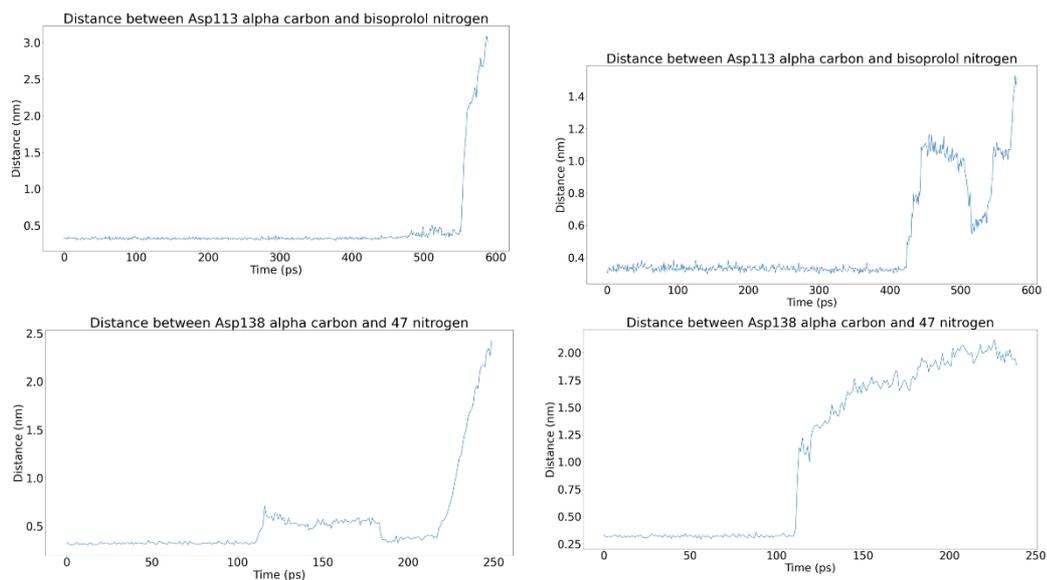


Figure 55 – Distance between ligand ethanolamine nitrogen and receptor D^{3.32} over the duration of the SAWMD trajectory. The graph depicts the distance of the ligand from the key residue as it is expelled out of the orthosteric binding site. Top left: bisoprolol in β_2 via conventional route. Top right: bisoprolol in β_1 via conventional route. Bottom left: Ligand **47** in β_1 via conventional route. Bottom right ligand **47** in β_1 via keyhole route.

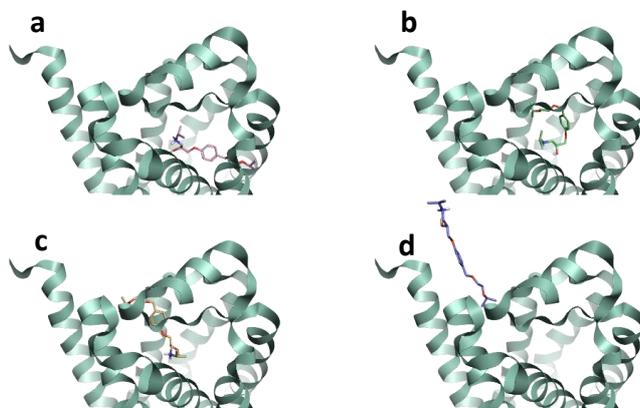


Figure 56 – Schematic of the unbinding trajectory of bisoprolol exiting the β_2 -AR via the conventional route, produced using the SAWMD method. Tiles **a** – **d** show the unbinding event in chronological order, beginning at tile **a** (bound) and ending at tile **d** (unbound).

5.3.2 Obtaining rate kinetic estimates using WEMD

A “mini-membrane” system with a volume of $\sim 300000 \text{ nm}^3$ was used for each system (exact average volumes were obtained from the average volume over the final nanosecond of the system’s production run before SAWMD). The significant reduction in atoms sped up calculations and prevented memory errors. The box dimensions were large enough to host the protein and the exited ligand within the membrane.

Despite extensive method optimisation, it was still not possible to generate an unbinding event within 200 WEMD cycles (which could consume several days of compute time on the available GPUs). This meant that it was not possible to obtain an off rate using the proposed WEMD method, however, it was possible to collect on rates for these systems without much difficulty. Since we have experimental values for k_D for these systems, it allowed for a prediction of the off rates using the relationship between the k_{on} , k_{off} and k_D .

Table 20 – Pharmacologically derived rate kinetics of bisoprolol in β_2 -AR by Sykes and Charlton 2014²²⁵. The calculated predicted koff is calculated using the formula $k_D = k_{off} / k_{on}$.

Ligand	Receptor	calculated k_{off} (min^{-1})	calculated k_{on} ($\text{M}^{-1} \text{min}^{-1}$)	k_D (koff/kon)
Bisoprolol	β_2	6.86 ± 2.09	$4.61 \pm 1.30 \times 10^7$	149 nM

Table 21 – Rate kinetics derived from WEMD. The ‘calculated predicted k_{off} ’ column is calculated using the formula $k_D = k_{off} / k_{on}$. The k_D of bisoprolol reported is pharmacologically derived by Baker 2005³⁴. The k_D of ligand **47** is also pharmacologically derived by Jillian G Baker.

Ligand	Receptor	Route	calculated predicted k_{off} (min^{-1})	calculated k_{on} ($\times 10^7 \text{ M}^{-1} \text{min}^{-1}$)	k_D (Baker (competition assay))
Bisoprolol	β_2	Conventional	9.05 ± 1.04	4.53 ± 0.52	200 nM ³⁴
Bisoprolol	β_1	Conventional	0.59 ± 0.01	3.79 ± 0.09	16 nM
47	β_1	Conventional	0.04 ± 0.007	3.15 ± 0.60	1.2 nM
47	β_1	Keyhole	0.05 ± 0.006	3.87 ± 0.47	1.2 nM

The k_{on} value obtained by WEMD simulation for bisoprolol (**27**) in β_2 -AR was $4.53 \pm 0.52 \times 10^7 \text{ M}^{-1} \text{ min}^{-1}$, which is in excellent agreement with the pharmacologically derived data obtained by Sykes and Charlton 2014²²⁵ as seen in Table 12. The k_{D} value reported in Table 12 is obtained using the relationship between the rate kinetics and the binding affinity obtained within that study, which differs from the binding affinity obtained from competition assays. The k_{off} for the WEMD simulations were calculated using the k_{D} from Jillian G Baker's competition assays as the binding affinity for **47** was calculated using the same method and therefore maintain consistency throughout the data. This difference in binding affinities can explain the slight difference in k_{off} for this system, but irrespective of this minor difference, the value of $9.05 \pm 1.04 \text{ min}^{-1}$ is in excellent agreement with the obtained pharmacological k_{off} .

It was shown by Baker 2005³⁴ that bisoprolol (**27**) had a higher binding affinity for the β_1 -AR than the β_2 -AR, therefore the slower k_{on} in β_1 -AR compared to β_2 -AR could be explained if the k_{off} is also slower, which could be expected from a ligand with a higher binding affinity. In general, the on rates for each of the systems do not differ largely, which puts more emphasis on the differences in off rates. The off rate of bisoprolol (**27**) in β_1 -AR is considerably slower than in β_2 -AR, again, expected from a ligand with a higher binding affinity to the β_1 -AR.

Continuing the trend are the ligand **47** systems in the β_1 -AR. The on rates for both entry routes do not differ largely from each other as well as the bisoprolol (**27**) systems, however the off rates are considerably lower than bisoprolol (**27**) in β_2 -AR and are lower than bisoprolol (**27**) in β_1 -AR. Ligand **47** has an almost 10-fold higher binding affinity to the β_1 -AR than bisoprolol (**27**), reflecting in its ~ 10 -fold slower off rate than bisoprolol (**27**).

5.4 Discussion

Pharmacologically derived binding kinetics obtained by Sykes and Charlton, 2014²²⁵ showed that on rates do not differ largely between ligands within β_2 -AR. There are some exceptions to this, but most of the ligands investigated display similar k_{on} measurements but largely differing in k_{off} measurements. This could be an indication that where both on and off rates are important, the difference in off rates in this series of ligands in β_2 -AR could have a large impact in the ligands' binding affinities. The k_{on} results obtained in the WEMD studies showed largely similar values, in line with the type of similarity seen in the Sykes and Charlton study, with the differing k_{off} rates explaining differences in binding affinities. Although this may indeed be the case generally, a difference in k_{on} between ligands is to be expected, however slight, due to the biomolecular and structural differences between systems and the effects this can have²²⁹.

The on rate obtained for bisoprolol (**27**) in the β_2 -AR agreed with the rate obtained pharmacologically, giving confidence to the WEMD method used within this study. The resulting calculated k_{off} also matched well with the pharmacologically derived data. The binding affinity obtained through competition assays are expected to differ slightly from the binding affinity obtained from kinetic assays, therefore, for the calculated predicted k_{off} in each case, there is a degree of possible variation depending on the binding affinity used.

For bisoprolol (**27**) in β_1 -AR, as mentioned previously, no rate kinetic data was available at the time of writing, however, binding affinity data is available³⁴. The degree of structural similarity between the β_1 - and β_2 -ARs could explain the similarities between the on-rate kinetics of bisoprolol (**27**), but despite the similarity, the on rate is still notably slower. The similar k_{on} value meant that a significantly lower k_{off} value was to be expected. When considering the calculated k_{on} and calculated predicted k_{off} for bisoprolol (**27**) in β_1 -AR, the value is in line with what we'd expect based on the known pharmacological binding affinity and the

rate kinetics of the ligand in β_2 -AR, therefore giving confidence in the β_1 -AR model used within the study.

Ligand **47** has a ~ 10 -fold higher binding affinity for β_1 -AR than bisoprolol (**27**), suggesting that if the on-rates are similar, that a ~ 10 -fold slower off rate is to be expected. The calculated k_{on} for ligand **47** binding via the conventional route was slower than bisoprolol (**27**). This could be due to **47** being a longer and more flexible ligand, which could introduce entropic penalties upon binding, causing it to slow down. The resulting calculated predicted k_{off} is ~ 10 -fold slower than for bisoprolol (**27**).

When comparing the conventional route to the keyhole route for ligand entry and exit for ligand **47** in β_1 -AR, the calculated k_{on} is slightly faster than that of the conventional route and is also higher than the k_{on} of bisoprolol (**27**) in the same receptor. This suggests that there isn't a strong preference for a particular entry route into the receptor but could reflect the absence of any slightly energetically unfavourable conformational changes that would be required for binding via the conventional route. The relatively similar k_{on} values for both routes result in similar k_{off} values.

The initial distribution of exit routes obtained by the SAWMD trajectories for ligand **47** in β_1 -AR, do reflect a lack of preference for the entry or exit of the ligand via a particular route, however it is important to note that the SAWMD method used in this study does not quantify the energetic changes along the unbinding pathway, therefore it isn't possible to directly compare the energetics of the SAWMD trajectories with the rate kinetics obtained from the WEMD. However, it is worth noting this correlation, since the same distribution was not observed in bisoprolol (**27**) or β_2 -AR. It is also important to mention that the exit routes obtained by Wang et al., 2009¹⁵¹ resemble exiting via the keyhole.

Critically and encouragingly, the simulations support the hypothesis that, if ligand **47** can diffuse into the membrane bilayer, then it should be possible for it to go on to bind in the orthosteric binding site via the TM helices.

Bivalent ligands designed in the previous chapter of this thesis would require at least one of their two receptor binding moieties to enter the receptor via the membrane bilayer. Further work would have to be conducted to assess how readily ligands are able to diffuse into the bilayer, but the highly lipophilic nature of the designed bivalent ligands should mean that diffusion into the membrane is possible, therefore supporting the case for a ligand of this nature entry via the transmembrane space.

5.5 Conclusion

The SAWMD method developed as part of this thesis successfully produces ligand exit pathway trajectories that are suitable for use with WEMD. The WEMD simulations were in very good agreement with the pharmacological data for bisoprolol (**27**) binding to β_2 -AR giving confidence in the applicability of this technique to study of GPCR ligand binding kinetics. The reasonable value for k_{on} obtained for bisoprolol (**27**) binding to β_1 -AR gives confidence in the model for the β_1 -AR developed here, and therefore gives confidence in the kinetic data obtained using the WEMD method for ligand **47** interacting with the β_1 -AR.

The WEMD kinetics data suggests that the β_1 -AR has no preference for a ligand's entry or exit route, meaning that it may indeed be energetically possible for ligands to bind to the orthosteric site via the keyhole from the TM space. This outcome gives confidence that ligands such as the bivalent ligands designed in the previous chapter are capable of binding to the receptor via the keyhole.

6 Conclusions and future work

6.1 General Conclusions

Research within this thesis has been conducted with the intention of validating the idea that a transmembrane fissure within the β_1 -AR could be the explanation as to why extended ligands are able to bind more favourably to this receptor than the other β -AR subtypes, as well as the potential for this structural feature to also act as a ligand exit route and ultimately be used as a potential route for ligand bivalency. These research aims all complement each other on the larger search for a method of improving selectivity for the β_1 -AR.

Binding affinity data can only tell a portion of the protein-ligand interaction story. Without crystal structure information, it's not possible to understand what binding conformation results in the binding affinities found pharmacologically. Therefore, molecular modelling and molecular dynamics are powerful tools for investigations in these conditions. Mutagenesis studies can explain which interactions are key for eliciting a response within the ligand target. Fortunately, key residues within the β_1 -AR orthosteric binding pocket are known³⁹ and can aid in the rationalisation of binding poses produced in molecular modelling.

Emtage et al.,^{150,156} conducted investigations into the role of the fissure between TM4/5 (the keyhole) in ligand selectivity between β_1 - and β_2 -AR. Research within the works included using active site pressurisation (ASP) to expand the orthosteric binding pocket with the intention of expanding the keyhole. Such methods were not used within this study, suggesting that the keyhole in the β_1 -AR is indeed existent in nature.

The role of the keyhole as a means of compensating for extended ligands via receptor plasticity in β_1 -AR was investigated within this thesis. Using knowledge of key residues within the binding pocket and molecular modelling and molecular

dynamics, it appears as though a ligand can maintain key interactions within the binding pocket while the head group is exposed to the intramembrane space via the keyhole.

Investigations into the effect of changing the electron density of the head group aryl ring on the binding affinity yielded interesting results. It appears the change in electron density has a negligible effect on the binding affinity. Although some additions improved binding affinity with respect to the base ligand, the improvements were not significant.

The idea of using the keyhole as a route for ligand bivalency was suggested by Emtage in 2013¹⁵⁶. Although no bivalent ligand design was conducted in their study, the potential based on existing ligands was mentioned. The research in this thesis has shown that it may be possible to design bivalent ligands in β_1 -AR that use the intramembrane space that are also capable of mitigating known caveats of traditional extracellular bivalent ligands.

Covalently linked bivalent ligands intended to occupy two β_1 -AR protomers introduce an important question, whether a ligand can enter and exit via the keyhole and into the transmembrane space. Kinetics information can help learn about the binding and unbinding rates and pathways of ligands to their targets. The study within these works has demonstrated that ligand entry and exit via the keyhole shares similar rates to the conventional route, suggesting that a ligand present in the transmembrane space is just as capable as entering the receptor via the keyhole as a ligand in the extracellular space entering conventionally.

6.2 Future work

There are several further investigations that would complement the work in this thesis, one of those would be to investigate different extended ligand head group

molecular properties to assess the effect potential transmembrane interactions has on binding affinity. This could, in turn, help in the design of high affinity transmembrane bivalent ligands. Understanding which molecular characteristics are the most favourable within this space could lead to linkers that can have a positive impact on binding affinity.

Leading on from the last point, linker designs could be vastly explored. Different degrees of flexibility as a result of different functional groups could also be found to have significant effects on stability. These bivalent ligand designs, and those explored within this study could also be synthesised to obtain pharmacological data. This will consolidate the idea that these molecules are indeed capable of occupying two orthosteric binding sites simultaneously. If this does turn out to be the case, then bivalent ligand design could become a focal point in the search for β_1 -AR selective ligands.

Knowing how likely the ligands of the nature explored in this thesis diffuse into the transmembrane space will also be advantageous. This could be investigated using LogP information, but also by using dedicated lipophilicity testing equipment.

Understanding the strength of the force that sustains dimer conformation is also important. Perhaps pulling simulations of the dimer protomers away from each other could explain whether this conformation is readily created or broken. It could also be explored whether the presence of a ligand strengthens or weakens this interaction. Knowing the residues involved in the sustaining dimer interface could also prove to be targets for bivalent ligand linkers.

It would be interesting to investigate the ligand exit routes of other β -AR ligands obtained by self-avoiding walk molecular dynamics (SAWMD). These routes could then be analysed by weighted ensemble molecular dynamics (WEMD) to investigate whether other ligands are able to readily enter the orthosteric site via alternative routes like the ligand investigated in this thesis. As mentioned earlier,

the rate at which the ligands diffuse into the lipid bilayer would be informative. This information could be gathered using the WEMD method used in this thesis.

Appendices

Appendix 1

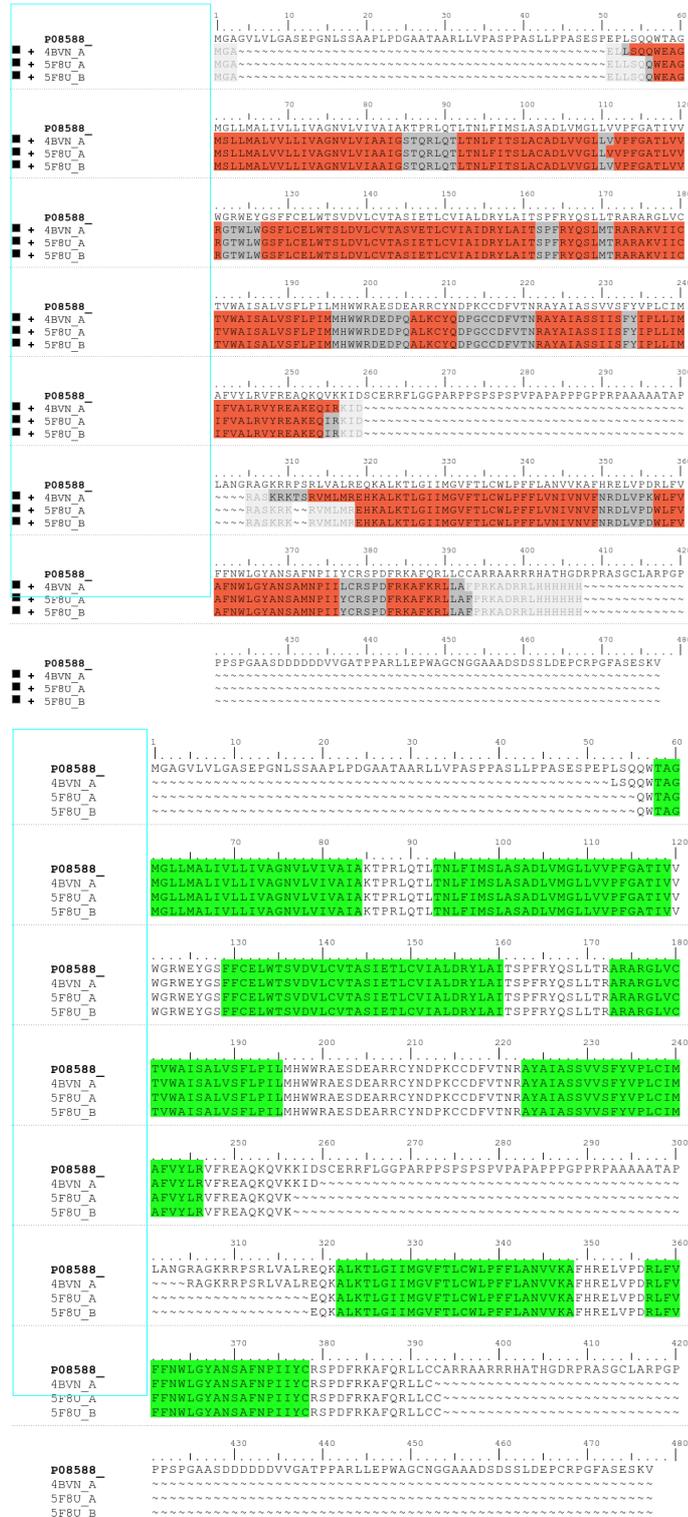


Figure 57 – Top: sequence alignment of β_1 -AR crystallographic templates (PDBID: 4BNV and 5F8U) against the human β_1 -AR wildtype sequence (uniprot ID: P08588) before homology modelling. Orange regions represents

the helices regions, and grey represents loops. Bottom: sequence alignment of β_1 -AR crystallographic templates (PDBID: 4BVN and 5F8U) against the human β_1 -AR wildtype sequence (uniport ID: P08588) after homology modelling. Green represents the TM helices regions.

Appendix 2 – Ligand 47 AMBERff99SB parameters in GROMACS .itp format

; GB4_GMX.top created by acpype (Rev: 0) on Tue Nov 13 14:59:59 2018

[atomtypes]

;name	bond_type	mass	charge	ptype	sigma	epsilon	Amb
c3	c3	0.00000	0.00000	A	3.39771e-01	4.51035e-01	;1.91 0.1078
hc	hc	0.00000	0.00000	A	2.60018e-01	8.70272e-02	;1.46 0.0208
hx	hx	0.00000	0.00000	A	1.88746e-01	8.70272e-02	;1.06 0.0208
n4	n4	0.00000	0.00000	A	2.49951e-01	1.62122e+01	;1.40 3.8748
hn	hn	0.00000	0.00000	A	1.10650e-01	4.18400e-02	;0.62 0.0100
oh	oh	0.00000	0.00000	A	3.24287e-01	3.89112e-01	;1.82 0.0930
ho	ho	0.00000	0.00000	A	5.37925e-02	1.96648e-02	;0.30 0.0047
h1	h1	0.00000	0.00000	A	2.42200e-01	8.70272e-02	;1.36 0.0208
os	os	0.00000	0.00000	A	3.15610e-01	3.03758e-01	;1.77 0.0726
ca	ca	0.00000	0.00000	A	3.31521e-01	4.13379e-01	;1.86 0.0988
ha	ha	0.00000	0.00000	A	2.62548e-01	6.73624e-02	;1.47 0.0161

[moleculetype]

;name	nrexcl
GB4	3

[atoms]

; nr	type	resi	res	atom	cgmr	charge	mass	; qtot	bond_type
1	c3	1	GB4	C3	1	-0.131600	12.01000	; qtot	-0.132
2	hc	1	GB4	H6	2	0.076200	1.00800	; qtot	-0.055
3	hc	1	GB4	H7	3	0.076200	1.00800	; qtot	0.021

4	hc	1	GB4	H8	4	0.076200	1.00800 ; qtot 0.097
5	c3	1	GB4	C2	5	0.132500	12.01000 ; qtot 0.230
6	c3	1	GB4	C4	6	-0.131600	12.01000 ; qtot 0.098
7	hc	1	GB4	H9	7	0.076200	1.00800 ; qtot 0.174
8	hc	1	GB4	H10	8	0.076200	1.00800 ; qtot 0.250
9	hc	1	GB4	H11	9	0.076200	1.00800 ; qtot 0.327
10	hx	1	GB4	H5	10	0.102700	1.00800 ; qtot 0.429
11	n4	1	GB4	N1	11	-0.760002	14.01000 ; qtot -0.331
12	hn	1	GB4	H4	12	0.469800	1.00800 ; qtot 0.139
13	hn	1	GB4	H67	13	0.469800	1.00800 ; qtot 0.609
14	c3	1	GB4	C1	14	0.116800	12.01000 ; qtot 0.726
15	hx	1	GB4	H1	15	0.106700	1.00800 ; qtot 0.832
16	hx	1	GB4	H2	16	0.106700	1.00800 ; qtot 0.939
17	c3	1	GB4	C5	17	0.121100	12.01000 ; qtot 1.060
18	oh	1	GB4	O1	18	-0.615801	16.00000 ; qtot 0.444
19	ho	1	GB4	H15	19	0.447000	1.00800 ; qtot 0.891
20	h1	1	GB4	H12	20	0.091700	1.00800 ; qtot 0.983
21	c3	1	GB4	C6	21	0.085400	12.01000 ; qtot 1.068
22	h1	1	GB4	H13	22	0.069200	1.00800 ; qtot 1.138
23	h1	1	GB4	H14	23	0.069200	1.00800 ; qtot 1.207
24	os	1	GB4	O2	24	-0.400900	16.00000 ; qtot 0.806
25	ca	1	GB4	C7	25	0.012100	12.01000 ; qtot 0.818
26	ca	1	GB4	C8	26	-0.100000	12.01000 ; qtot 0.718
27	ca	1	GB4	C9	27	-0.166500	12.01000 ; qtot 0.551
28	ha	1	GB4	H17	28	0.161500	1.00800 ; qtot 0.713
29	ha	1	GB4	H16	29	0.142000	1.00800 ; qtot 0.855
30	ca	1	GB4	C12	30	-0.100000	12.01000 ; qtot 0.755
31	ha	1	GB4	H19	31	0.142000	1.00800 ; qtot 0.897
32	ca	1	GB4	C11	32	-0.166500	12.01000 ; qtot 0.730
33	ha	1	GB4	H18	33	0.161500	1.00800 ; qtot 0.892
34	ca	1	GB4	C10	34	0.156100	12.01000 ; qtot 1.048
35	os	1	GB4	O3	35	-0.317900	16.00000 ; qtot 0.730

36	c3	1	GB4	C13	36	0.119400	12.01000 ; qtot 0.850
37	h1	1	GB4	H20	37	0.061700	1.00800 ; qtot 0.911
38	h1	1	GB4	H21	38	0.061700	1.00800 ; qtot 0.973
39	c3	1	GB4	C27	39	0.125400	12.01000 ; qtot 1.098
40	h1	1	GB4	H33	40	0.062200	1.00800 ; qtot 1.161
41	h1	1	GB4	H34	41	0.062200	1.00800 ; qtot 1.223
42	os	1	GB4	O5	42	-0.337900	16.00000 ; qtot 0.885
43	ca	1	GB4	C14	43	0.078100	12.01000 ; qtot 0.963
44	ca	1	GB4	C15	44	-0.138500	12.01000 ; qtot 0.824
45	ca	1	GB4	C16	45	-0.151000	12.01000 ; qtot 0.673
46	ha	1	GB4	H23	46	0.149000	1.00800 ; qtot 0.822
47	ha	1	GB4	H22	47	0.142500	1.00800 ; qtot 0.965
48	ca	1	GB4	C19	48	-0.138500	12.01000 ; qtot 0.826
49	ha	1	GB4	H25	49	0.142500	1.00800 ; qtot 0.969
50	ca	1	GB4	C18	50	-0.151000	12.01000 ; qtot 0.818
51	ha	1	GB4	H24	51	0.149000	1.00800 ; qtot 0.967
52	ca	1	GB4	C17	52	0.109100	12.01000 ; qtot 1.076
53	os	1	GB4	O4	53	-0.323900	16.00000 ; qtot 0.752
54	c3	1	GB4	C20	54	0.175700	12.01000 ; qtot 0.928
55	h1	1	GB4	H26	55	0.047200	1.00800 ; qtot 0.975
56	h1	1	GB4	H27	56	0.047200	1.00800 ; qtot 1.022
57	ca	1	GB4	C21	57	-0.098300	12.01000 ; qtot 0.924
58	ca	1	GB4	C26	58	-0.111500	12.01000 ; qtot 0.812
59	ha	1	GB4	H32	59	0.137500	1.00800 ; qtot 0.950
60	ca	1	GB4	C25	60	-0.130000	12.01000 ; qtot 0.820
61	ha	1	GB4	H31	61	0.135000	1.00800 ; qtot 0.955
62	ca	1	GB4	C24	62	-0.121000	12.01000 ; qtot 0.834
63	ha	1	GB4	H30	63	0.135000	1.00800 ; qtot 0.969
64	ca	1	GB4	C23	64	-0.130000	12.01000 ; qtot 0.839
65	ha	1	GB4	H29	65	0.135000	1.00800 ; qtot 0.974
66	ca	1	GB4	C22	66	-0.111500	12.01000 ; qtot 0.862
67	ha	1	GB4	H28	67	0.137500	1.00800 ; qtot 1.000

[bonds]

```
; ai  aj funct  r      k
  1   2   1  1.0970e-01  3.1455e+05 ; C3 - H6
  1   3   1  1.0970e-01  3.1455e+05 ; C3 - H7
  1   4   1  1.0970e-01  3.1455e+05 ; C3 - H8
  1   5   1  1.5380e-01  1.9456e+05 ; C3 - C2
  5   6   1  1.5380e-01  1.9456e+05 ; C2 - C4
  5  10   1  1.0910e-01  3.2342e+05 ; C2 - H5
  5  11   1  1.5110e-01  1.8627e+05 ; C2 - N1
  6   7   1  1.0970e-01  3.1455e+05 ; C4 - H9
  6   8   1  1.0970e-01  3.1455e+05 ; C4 - H10
  6   9   1  1.0970e-01  3.1455e+05 ; C4 - H11
 11  12   1  1.0300e-01  4.0409e+05 ; N1 - H4
 11  13   1  1.0300e-01  4.0409e+05 ; N1 - H67
 11  14   1  1.5110e-01  1.8627e+05 ; N1 - C1
 14  15   1  1.0910e-01  3.2342e+05 ; C1 - H1
 14  16   1  1.0910e-01  3.2342e+05 ; C1 - H2
 14  17   1  1.5380e-01  1.9456e+05 ; C1 - C5
 17  18   1  1.4230e-01  2.4552e+05 ; C5 - O1
 17  20   1  1.0970e-01  3.1455e+05 ; C5 - H12
 17  21   1  1.5380e-01  1.9456e+05 ; C5 - C6
 18  19   1  9.7300e-02  4.7154e+05 ; O1 - H15
 21  22   1  1.0970e-01  3.1455e+05 ; C6 - H13
 21  23   1  1.0970e-01  3.1455e+05 ; C6 - H14
 21  24   1  1.4320e-01  2.3832e+05 ; C6 - O2
 24  25   1  1.3700e-01  2.9916e+05 ; O2 - C7
 25  26   1  1.3980e-01  3.1681e+05 ; C7 - C8
 25  30   1  1.3980e-01  3.1681e+05 ; C7 - C12
 26  27   1  1.3980e-01  3.1681e+05 ; C8 - C9
 26  29   1  1.0860e-01  3.3112e+05 ; C8 - H16
 27  28   1  1.0860e-01  3.3112e+05 ; C9 - H17
```

27 34 1 1.3980e-01 3.1681e+05 ; C9 - C10
30 31 1 1.0860e-01 3.3112e+05 ; C12 - H19
30 32 1 1.3980e-01 3.1681e+05 ; C12 - C11
32 33 1 1.0860e-01 3.3112e+05 ; C11 - H18
32 34 1 1.3980e-01 3.1681e+05 ; C11 - C10
34 35 1 1.3700e-01 2.9916e+05 ; C10 - O3
35 36 1 1.4320e-01 2.3832e+05 ; O3 - C13
36 37 1 1.0970e-01 3.1455e+05 ; C13 - H20
36 38 1 1.0970e-01 3.1455e+05 ; C13 - H21
36 39 1 1.5380e-01 1.9456e+05 ; C13 - C27
39 40 1 1.0970e-01 3.1455e+05 ; C27 - H33
39 41 1 1.0970e-01 3.1455e+05 ; C27 - H34
39 42 1 1.4320e-01 2.3832e+05 ; C27 - O5
42 43 1 1.3700e-01 2.9916e+05 ; O5 - C14
43 44 1 1.3980e-01 3.1681e+05 ; C14 - C15
43 48 1 1.3980e-01 3.1681e+05 ; C14 - C19
44 45 1 1.3980e-01 3.1681e+05 ; C15 - C16
44 47 1 1.0860e-01 3.3112e+05 ; C15 - H22
45 46 1 1.0860e-01 3.3112e+05 ; C16 - H23
45 52 1 1.3980e-01 3.1681e+05 ; C16 - C17
48 49 1 1.0860e-01 3.3112e+05 ; C19 - H25
48 50 1 1.3980e-01 3.1681e+05 ; C19 - C18
50 51 1 1.0860e-01 3.3112e+05 ; C18 - H24
50 52 1 1.3980e-01 3.1681e+05 ; C18 - C17
52 53 1 1.3700e-01 2.9916e+05 ; C17 - O4
53 54 1 1.4320e-01 2.3832e+05 ; O4 - C20
54 55 1 1.0970e-01 3.1455e+05 ; C20 - H26
54 56 1 1.0970e-01 3.1455e+05 ; C20 - H27
54 57 1 1.5160e-01 2.0945e+05 ; C20 - C21
57 58 1 1.3980e-01 3.1681e+05 ; C21 - C26
57 66 1 1.3980e-01 3.1681e+05 ; C21 - C22
58 59 1 1.0860e-01 3.3112e+05 ; C26 - H32

58	60	1	1.3980e-01	3.1681e+05	; C26 - C25
60	61	1	1.0860e-01	3.3112e+05	; C25 - H31
60	62	1	1.3980e-01	3.1681e+05	; C25 - C24
62	63	1	1.0860e-01	3.3112e+05	; C24 - H30
62	64	1	1.3980e-01	3.1681e+05	; C24 - C23
64	65	1	1.0860e-01	3.3112e+05	; C23 - H29
64	66	1	1.3980e-01	3.1681e+05	; C23 - C22
66	67	1	1.0860e-01	3.3112e+05	; C22 - H28

[pairs]

	ai	aj	funct
	1	7	1; C3 - H9
	1	8	1; C3 - H10
	1	9	1; C3 - H11
	1	12	1; C3 - H4
	1	13	1; C3 - H67
	1	14	1; C3 - C1
	2	6	1; H6 - C4
	2	10	1; H6 - H5
	2	11	1; H6 - N1
	3	6	1; H7 - C4
	3	10	1; H7 - H5
	3	11	1; H7 - N1
	4	6	1; H8 - C4
	4	10	1; H8 - H5
	4	11	1; H8 - N1
	5	15	1; C2 - H1
	5	16	1; C2 - H2
	5	17	1; C2 - C5
	6	12	1; C4 - H4
	6	13	1; C4 - H67
	6	14	1; C4 - C1

7 10 1; H9 - H5
7 11 1; H9 - N1
8 10 1; H10 - H5
8 11 1; H10 - N1
9 10 1; H11 - H5
9 11 1; H11 - N1
10 12 1; H5 - H4
10 13 1; H5 - H67
10 14 1; H5 - C1
11 18 1; N1 - O1
11 20 1; N1 - H12
11 21 1; N1 - C6
12 15 1; H4 - H1
12 16 1; H4 - H2
12 17 1; H4 - C5
13 15 1; H67 - H1
13 16 1; H67 - H2
13 17 1; H67 - C5
14 19 1; C1 - H15
14 22 1; C1 - H13
14 23 1; C1 - H14
14 24 1; C1 - O2
15 18 1; H1 - O1
15 20 1; H1 - H12
15 21 1; H1 - C6
16 18 1; H2 - O1
16 20 1; H2 - H12
16 21 1; H2 - C6
17 25 1; C5 - C7
18 22 1; O1 - H13
18 23 1; O1 - H14
18 24 1; O1 - O2

19 20 1; H15 - H12
19 21 1; H15 - C6
20 22 1; H12 - H13
20 23 1; H12 - H14
20 24 1; H12 - O2
21 26 1; C6 - C8
21 30 1; C6 - C12
22 25 1; H13 - C7
23 25 1; H14 - C7
24 27 1; O2 - C9
24 29 1; O2 - H16
24 31 1; O2 - H19
24 32 1; O2 - C11
25 28 1; C7 - H17
25 33 1; C7 - H18
25 34 1; C7 - C10
26 31 1; C8 - H19
26 32 1; C8 - C11
26 35 1; C8 - O3
27 30 1; C9 - C12
27 33 1; C9 - H18
27 36 1; C9 - C13
28 29 1; H17 - H16
28 32 1; H17 - C11
28 35 1; H17 - O3
29 30 1; H16 - C12
29 34 1; H16 - C10
30 35 1; C12 - O3
31 33 1; H19 - H18
31 34 1; H19 - C10
32 36 1; C11 - C13
33 35 1; H18 - O3

34 37 1; C10 - H20
34 38 1; C10 - H21
34 39 1; C10 - C27
35 40 1; O3 - H33
35 41 1; O3 - H34
35 42 1; O3 - O5
36 43 1; C13 - C14
37 40 1; H20 - H33
37 41 1; H20 - H34
37 42 1; H20 - O5
38 40 1; H21 - H33
38 41 1; H21 - H34
38 42 1; H21 - O5
39 44 1; C27 - C15
39 48 1; C27 - C19
40 43 1; H33 - C14
41 43 1; H34 - C14
42 45 1; O5 - C16
42 47 1; O5 - H22
42 49 1; O5 - H25
42 50 1; O5 - C18
43 46 1; C14 - H23
43 51 1; C14 - H24
43 52 1; C14 - C17
44 49 1; C15 - H25
44 50 1; C15 - C18
44 53 1; C15 - O4
45 48 1; C16 - C19
45 51 1; C16 - H24
45 54 1; C16 - C20
46 47 1; H23 - H22
46 50 1; H23 - C18

46 53 1; H23 - O4
47 48 1; H22 - C19
47 52 1; H22 - C17
48 53 1; C19 - O4
49 51 1; H25 - H24
49 52 1; H25 - C17
50 54 1; C18 - C20
51 53 1; H24 - O4
52 55 1; C17 - H26
52 56 1; C17 - H27
52 57 1; C17 - C21
53 58 1; O4 - C26
53 66 1; O4 - C22
54 59 1; C20 - H32
54 60 1; C20 - C25
54 64 1; C20 - C23
54 67 1; C20 - H28
55 58 1; H26 - C26
55 66 1; H26 - C22
56 58 1; H27 - C26
56 66 1; H27 - C22
57 61 1; C21 - H31
57 62 1; C21 - C24
57 65 1; C21 - H29
58 63 1; C26 - H30
58 64 1; C26 - C23
58 67 1; C26 - H28
59 61 1; H32 - H31
59 62 1; H32 - C24
59 66 1; H32 - C22
60 65 1; C25 - H29
60 66 1; C25 - C22

61 63 1; H31 - H30
 61 64 1; H31 - C23
 62 67 1; C24 - H28
 63 65 1; H30 - H29
 63 66 1; H30 - C22
 65 67 1; H29 - H28

[angles]

	ai	aj	ak	funct	theta	cth		
	1	5	6	1	1.1151e+02	5.4308e+02	C3 - C2	- C4
	1	5	10	1	1.1056e+02	3.9079e+02	C3 - C2	- H5
	1	5	11	1	1.1421e+02	6.7781e+02	C3 - C2	- N1
	2	1	3	1	1.0758e+02	3.2635e+02	H6 - C3	- H7
	2	1	4	1	1.0758e+02	3.2635e+02	H6 - C3	- H8
	2	1	5	1	1.0980e+02	3.9162e+02	H6 - C3	- C2
	3	1	4	1	1.0758e+02	3.2635e+02	H7 - C3	- H8
	3	1	5	1	1.0980e+02	3.9162e+02	H7 - C3	- C2
	4	1	5	1	1.0980e+02	3.9162e+02	H8 - C3	- C2
	5	6	7	1	1.0980e+02	3.9162e+02	C2 - C4	- H9
	5	6	8	1	1.0980e+02	3.9162e+02	C2 - C4	- H10
	5	6	9	1	1.0980e+02	3.9162e+02	C2 - C4	- H11
	5	11	12	1	1.1011e+02	3.8660e+02	C2 - N1	- H4
	5	11	13	1	1.1011e+02	3.8660e+02	C2 - N1	- H67
	5	11	14	1	1.0966e+02	5.3974e+02	C2 - N1	- C1
	6	5	10	1	1.1056e+02	3.9079e+02	C4 - C2	- H5
	6	5	11	1	1.1421e+02	6.7781e+02	C4 - C2	- N1
	7	6	8	1	1.0758e+02	3.2635e+02	H9 - C4	- H10
	7	6	9	1	1.0758e+02	3.2635e+02	H9 - C4	- H11
	8	6	9	1	1.0758e+02	3.2635e+02	H10 - C4	- H11
	10	5	11	1	1.0801e+02	5.0292e+02	H5 - C2	- N1
	11	14	15	1	1.0801e+02	5.0292e+02	N1 - C1	- H1
	11	14	16	1	1.0801e+02	5.0292e+02	N1 - C1	- H2

11	14	17	1	1.1421e+02	6.7781e+02 ;	N1 - C1	- C5
12	11	13	1	1.0830e+02	3.3472e+02 ;	H4 - N1	- H67
12	11	14	1	1.1011e+02	3.8660e+02 ;	H4 - N1	- C1
13	11	14	1	1.1011e+02	3.8660e+02 ;	H67 - N1	- C1
14	17	18	1	1.1019e+02	7.0793e+02 ;	C1 - C5	- O1
14	17	20	1	1.0956e+02	3.9246e+02 ;	C1 - C5	- H12
14	17	21	1	1.1151e+02	5.4308e+02 ;	C1 - C5	- C6
15	14	16	1	1.0975e+02	3.2468e+02 ;	H1 - C1	- H2
15	14	17	1	1.1056e+02	3.9079e+02 ;	H1 - C1	- C5
16	14	17	1	1.1056e+02	3.9079e+02 ;	H2 - C1	- C5
17	18	19	1	1.0726e+02	4.1003e+02 ;	C5 - O1	- H15
17	21	22	1	1.0956e+02	3.9246e+02 ;	C5 - C6	- H13
17	21	23	1	1.0956e+02	3.9246e+02 ;	C5 - C6	- H14
17	21	24	1	1.0797e+02	7.1379e+02 ;	C5 - C6	- O2
18	17	20	1	1.1026e+02	5.2300e+02 ;	O1 - C5	- H12
18	17	21	1	1.1019e+02	7.0793e+02 ;	O1 - C5	- C6
20	17	21	1	1.0956e+02	3.9246e+02 ;	H12 - C5	- C6
21	24	25	1	1.1796e+02	5.5312e+02 ;	C6 - O2	- C7
22	21	23	1	1.0846e+02	3.2468e+02 ;	H13 - C6	- H14
22	21	24	1	1.0978e+02	5.2216e+02 ;	H13 - C6	- O2
23	21	24	1	1.0978e+02	5.2216e+02 ;	H14 - C6	- O2
24	25	26	1	1.1920e+02	7.3053e+02 ;	O2 - C7	- C8
24	25	30	1	1.1920e+02	7.3053e+02 ;	O2 - C7	- C12
25	26	27	1	1.2002e+02	5.7572e+02 ;	C7 - C8	- C9
25	26	29	1	1.1988e+02	4.0752e+02 ;	C7 - C8	- H16
25	30	31	1	1.1988e+02	4.0752e+02 ;	C7 - C12	- H19
25	30	32	1	1.2002e+02	5.7572e+02 ;	C7 - C12	- C11
26	25	30	1	1.2002e+02	5.7572e+02 ;	C8 - C7	- C12
26	27	28	1	1.1988e+02	4.0752e+02 ;	C8 - C9	- H17
26	27	34	1	1.2002e+02	5.7572e+02 ;	C8 - C9	- C10
27	26	29	1	1.1988e+02	4.0752e+02 ;	C9 - C8	- H16
27	34	32	1	1.2002e+02	5.7572e+02 ;	C9 - C10	- C11

27	34	35	1	1.1920e+02	7.3053e+02 ;	C9 - C10	- O3
28	27	34	1	1.1988e+02	4.0752e+02 ;	H17 - C9	- C10
30	32	33	1	1.1988e+02	4.0752e+02 ;	C12 - C11	- H18
30	32	34	1	1.2002e+02	5.7572e+02 ;	C12 - C11	- C10
31	30	32	1	1.1988e+02	4.0752e+02 ;	H19 - C12	- C11
32	34	35	1	1.1920e+02	7.3053e+02 ;	C11 - C10	- O3
33	32	34	1	1.1988e+02	4.0752e+02 ;	H18 - C11	- C10
34	35	36	1	1.1796e+02	5.5312e+02 ;	C10 - O3	- C13
35	36	37	1	1.0978e+02	5.2216e+02 ;	O3 - C13	- H20
35	36	38	1	1.0978e+02	5.2216e+02 ;	O3 - C13	- H21
35	36	39	1	1.0797e+02	7.1379e+02 ;	O3 - C13	- C27
36	39	40	1	1.0956e+02	3.9246e+02 ;	C13 - C27	- H33
36	39	41	1	1.0956e+02	3.9246e+02 ;	C13 - C27	- H34
36	39	42	1	1.0797e+02	7.1379e+02 ;	C13 - C27	- O5
37	36	38	1	1.0846e+02	3.2468e+02 ;	H20 - C13	- H21
37	36	39	1	1.0956e+02	3.9246e+02 ;	H20 - C13	- C27
38	36	39	1	1.0956e+02	3.9246e+02 ;	H21 - C13	- C27
39	42	43	1	1.1796e+02	5.5312e+02 ;	C27 - O5	- C14
40	39	41	1	1.0846e+02	3.2468e+02 ;	H33 - C27	- H34
40	39	42	1	1.0978e+02	5.2216e+02 ;	H33 - C27	- O5
41	39	42	1	1.0978e+02	5.2216e+02 ;	H34 - C27	- O5
42	43	44	1	1.1920e+02	7.3053e+02 ;	O5 - C14	- C15
42	43	48	1	1.1920e+02	7.3053e+02 ;	O5 - C14	- C19
43	44	45	1	1.2002e+02	5.7572e+02 ;	C14 - C15	- C16
43	44	47	1	1.1988e+02	4.0752e+02 ;	C14 - C15	- H22
43	48	49	1	1.1988e+02	4.0752e+02 ;	C14 - C19	- H25
43	48	50	1	1.2002e+02	5.7572e+02 ;	C14 - C19	- C18
44	43	48	1	1.2002e+02	5.7572e+02 ;	C15 - C14	- C19
44	45	46	1	1.1988e+02	4.0752e+02 ;	C15 - C16	- H23
44	45	52	1	1.2002e+02	5.7572e+02 ;	C15 - C16	- C17
45	44	47	1	1.1988e+02	4.0752e+02 ;	C16 - C15	- H22
45	52	50	1	1.2002e+02	5.7572e+02 ;	C16 - C17	- C18

45	52	53	1	1.1920e+02	7.3053e+02 ;	C16 - C17	- O4
46	45	52	1	1.1988e+02	4.0752e+02 ;	H23 - C16	- C17
48	50	51	1	1.1988e+02	4.0752e+02 ;	C19 - C18	- H24
48	50	52	1	1.2002e+02	5.7572e+02 ;	C19 - C18	- C17
49	48	50	1	1.1988e+02	4.0752e+02 ;	H25 - C19	- C18
50	52	53	1	1.1920e+02	7.3053e+02 ;	C18 - C17	- O4
51	50	52	1	1.1988e+02	4.0752e+02 ;	H24 - C18	- C17
52	53	54	1	1.1796e+02	5.5312e+02 ;	C17 - O4	- C20
53	54	55	1	1.0978e+02	5.2216e+02 ;	O4 - C20	- H26
53	54	56	1	1.0978e+02	5.2216e+02 ;	O4 - C20	- H27
53	54	57	1	1.0895e+02	7.1630e+02 ;	O4 - C20	- C21
54	57	58	1	1.2077e+02	5.4894e+02 ;	C20 - C21	- C26
54	57	66	1	1.2077e+02	5.4894e+02 ;	C20 - C21	- C22
55	54	56	1	1.0846e+02	3.2468e+02 ;	H26 - C20	- H27
55	54	57	1	1.0956e+02	3.9748e+02 ;	H26 - C20	- C21
56	54	57	1	1.0956e+02	3.9748e+02 ;	H27 - C20	- C21
57	58	59	1	1.1988e+02	4.0752e+02 ;	C21 - C26	- H32
57	58	60	1	1.2002e+02	5.7572e+02 ;	C21 - C26	- C25
57	66	64	1	1.2002e+02	5.7572e+02 ;	C21 - C22	- C23
57	66	67	1	1.1988e+02	4.0752e+02 ;	C21 - C22	- H28
58	57	66	1	1.2002e+02	5.7572e+02 ;	C26 - C21	- C22
58	60	61	1	1.1988e+02	4.0752e+02 ;	C26 - C25	- H31
58	60	62	1	1.2002e+02	5.7572e+02 ;	C26 - C25	- C24
59	58	60	1	1.1988e+02	4.0752e+02 ;	H32 - C26	- C25
60	62	63	1	1.1988e+02	4.0752e+02 ;	C25 - C24	- H30
60	62	64	1	1.2002e+02	5.7572e+02 ;	C25 - C24	- C23
61	60	62	1	1.1988e+02	4.0752e+02 ;	H31 - C25	- C24
62	64	65	1	1.1988e+02	4.0752e+02 ;	C24 - C23	- H29
62	64	66	1	1.2002e+02	5.7572e+02 ;	C24 - C23	- C22
63	62	64	1	1.1988e+02	4.0752e+02 ;	H30 - C24	- C23
64	66	67	1	1.1988e+02	4.0752e+02 ;	C23 - C22	- H28
65	64	66	1	1.1988e+02	4.0752e+02 ;	H29 - C23	- C22

[dihedrals] ; propers

; treated as RBs in GROMACS to use combine multiple AMBER torsions per quartet

```
; i j k l func C0 C1 C2 C3 C4 C5
  1 5 6 7 3 0.33472 1.00416 0.00000 -1.33888 0.00000 0.00000
; C3- C2- C4- H9
  1 5 6 8 3 0.33472 1.00416 0.00000 -1.33888 0.00000 0.00000
; C3- C2- C4- H10
  1 5 6 9 3 0.33472 1.00416 0.00000 -1.33888 0.00000 0.00000
; C3- C2- C4- H11
  1 5 11 12 3 0.65084 1.95253 0.00000 -2.60338 0.00000
0.00000; C3- C2- N1- H4
  1 5 11 13 3 0.65084 1.95253 0.00000 -2.60338 0.00000
0.00000; C3- C2- N1- H67
  1 5 11 14 3 0.65270 1.95811 0.00000 -2.61082 0.00000
0.00000; C3- C2- N1- C1
  2 1 5 6 3 0.33472 1.00416 0.00000 -1.33888 0.00000 0.00000
; H6- C3- C2- C4
  2 1 5 10 3 0.65084 1.95253 0.00000 -2.60338 0.00000
0.00000; H6- C3- C2- H5
  2 1 5 11 3 0.65084 1.95253 0.00000 -2.60338 0.00000
0.00000; H6- C3- C2- N1
  3 1 5 6 3 0.33472 1.00416 0.00000 -1.33888 0.00000 0.00000
; H7- C3- C2- C4
  3 1 5 10 3 0.65084 1.95253 0.00000 -2.60338 0.00000
0.00000; H7- C3- C2- H5
  3 1 5 11 3 0.65084 1.95253 0.00000 -2.60338 0.00000
0.00000; H7- C3- C2- N1
  4 1 5 6 3 0.33472 1.00416 0.00000 -1.33888 0.00000 0.00000
; H8- C3- C2- C4
  4 1 5 10 3 0.65084 1.95253 0.00000 -2.60338 0.00000
0.00000; H8- C3- C2- H5
```

4	1	5	11	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H8- C3- C2- N1									
5	11	14	15	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; C2- N1- C1- H1									
5	11	14	16	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; C2- N1- C1- H2									
5	11	14	17	3	0.65270	1.95811	0.00000	-2.61082	0.00000
0.00000; C2- N1- C1- C5									
6	5	11	12	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; C4- C2- N1- H4									
6	5	11	13	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; C4- C2- N1- H67									
6	5	11	14	3	0.65270	1.95811	0.00000	-2.61082	0.00000
0.00000; C4- C2- N1- C1									
7	6	5	10	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H9- C4- C2- H5									
7	6	5	11	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H9- C4- C2- N1									
8	6	5	10	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H10- C4- C2- H5									
8	6	5	11	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H10- C4- C2- N1									
9	6	5	10	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H11- C4- C2- H5									
9	6	5	11	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H11- C4- C2- N1									
10	5	11	12	3	0.45606	1.36817	0.00000	-1.82422	0.00000
0.00000; H5- C2- N1- H4									
10	5	11	13	3	0.45606	1.36817	0.00000	-1.82422	0.00000
0.00000; H5- C2- N1- H67									
10	5	11	14	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H5- C2- N1- C1									

11	14	17	18	3	0.60250	1.80749	10.87840	-2.40998	0.00000
0.00000; N1- C1- C5- O1									
11	14	17	20	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; N1- C1- C5- H12									
11	14	17	21	3	0.87864	2.63592	0.00000	-3.51456	0.00000
0.00000; N1- C1- C5- C6									
12	11	14	15	3	0.45606	1.36817	0.00000	-1.82422	0.00000
0.00000; H4- N1- C1- H1									
12	11	14	16	3	0.45606	1.36817	0.00000	-1.82422	0.00000
0.00000; H4- N1- C1- H2									
12	11	14	17	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H4- N1- C1- C5									
13	11	14	15	3	0.45606	1.36817	0.00000	-1.82422	0.00000
0.00000; H67- N1- C1- H1									
13	11	14	16	3	0.45606	1.36817	0.00000	-1.82422	0.00000
0.00000; H67- N1- C1- H2									
13	11	14	17	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H67- N1- C1- C5									
14	17	18	19	3	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000; C1- C5- O1- H15									
14	17	21	22	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; C1- C5- C6- H13									
14	17	21	23	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; C1- C5- C6- H14									
14	17	21	24	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; C1- C5- C6- O2									
15	14	17	18	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H1- C1- C5- O1									
15	14	17	20	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H1- C1- C5- H12									
15	14	17	21	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000; H1- C1- C5- C6									

16 14 17 18 3 0.65084 1.95253 0.00000 -2.60338 0.00000
 0.00000; H2- C1- C5- O1
 16 14 17 20 3 0.65084 1.95253 0.00000 -2.60338 0.00000
 0.00000; H2- C1- C5- H12
 16 14 17 21 3 0.65084 1.95253 0.00000 -2.60338 0.00000
 0.00000; H2- C1- C5- C6
 17 21 24 25 3 1.60387 4.81160 0.00000 -6.41547 0.00000
 0.00000; C5- C6- O2- C7
 18 17 21 22 3 1.04600 -1.04600 0.00000 0.00000 0.00000
 0.00000; O1- C5- C6- H13
 18 17 21 23 3 1.04600 -1.04600 0.00000 0.00000 0.00000
 0.00000; O1- C5- C6- H14
 18 17 21 24 3 4.30952 12.76120 0.00000 -16.90336 0.00000
 0.00000; O1- C5- C6- O2
 19 18 17 20 3 0.47279 1.41838 0.00000 -1.89117 0.00000
 0.00000; H15- O1- C5- H12
 19 18 17 21 3 0.00000 0.00000 0.00000 0.00000 0.00000
 0.00000; H15- O1- C5- C6
 20 17 21 22 3 0.65084 1.95253 0.00000 -2.60338 0.00000
 0.00000; H12- C5- C6- H13
 20 17 21 23 3 0.65084 1.95253 0.00000 -2.60338 0.00000
 0.00000; H12- C5- C6- H14
 20 17 21 24 3 1.04600 -1.04600 0.00000 0.00000 0.00000
 0.00000; H12- C5- C6- O2
 21 24 25 26 3 13.47248 0.00000 -13.47248 0.00000 0.00000
 0.00000; C6- O2- C7- C8
 21 24 25 30 3 13.47248 0.00000 -13.47248 0.00000 0.00000
 0.00000; C6- O2- C7- C12
 22 21 24 25 3 1.60387 4.81160 0.00000 -6.41547 0.00000
 0.00000; H13- C6- O2- C7
 23 21 24 25 3 1.60387 4.81160 0.00000 -6.41547 0.00000
 0.00000; H14- C6- O2- C7

24 25 26 27 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; O2- C7- C8- C9
 24 25 26 29 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; O2- C7- C8- H16
 24 25 30 31 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; O2- C7- C12- H19
 24 25 30 32 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; O2- C7- C12- C11
 25 26 27 28 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C7- C8- C9- H17
 25 26 27 34 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C7- C8- C9- C10
 25 30 32 33 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C7- C12- C11- H18
 25 30 32 34 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C7- C12- C11- C10
 26 25 30 31 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C8- C7- C12- H19
 26 25 30 32 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C8- C7- C12- C11
 26 27 34 32 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C8- C9- C10- C11
 26 27 34 35 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C8- C9- C10- O3
 27 26 25 30 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C9- C8- C7- C12
 27 34 32 30 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C9- C10- C11- C12
 27 34 32 33 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C9- C10- C11- H18
 27 34 35 36 3 13.47248 0.00000 -13.47248 0.00000 0.00000
 0.00000; C9- C10- O3- C13

28 27 26 29 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H17- C9- C8- H16
 28 27 34 32 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H17- C9- C10- C11
 28 27 34 35 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H17- C9- C10- O3
 29 26 25 30 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H16- C8- C7- C12
 29 26 27 34 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H16- C8- C9- C10
 30 32 34 35 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C12- C11- C10- O3
 31 30 32 33 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H19- C12- C11- H18
 31 30 32 34 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H19- C12- C11- C10
 32 34 35 36 3 13.47248 0.00000 -13.47248 0.00000 0.00000
 0.00000; C11- C10- O3- C13
 33 32 34 35 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H18- C11- C10- O3
 34 35 36 37 3 1.60387 4.81160 0.00000 -6.41547 0.00000
 0.00000; C10- O3- C13- H20
 34 35 36 38 3 1.60387 4.81160 0.00000 -6.41547 0.00000
 0.00000; C10- O3- C13- H21
 34 35 36 39 3 1.60387 4.81160 0.00000 -6.41547 0.00000
 0.00000; C10- O3- C13- C27
 35 36 39 40 3 1.04600 -1.04600 0.00000 0.00000 0.00000
 0.00000; O3- C13- C27- H33
 35 36 39 41 3 1.04600 -1.04600 0.00000 0.00000 0.00000
 0.00000; O3- C13- C27- H34
 35 36 39 42 3 0.71128 0.71128 0.00000 0.00000 0.00000
 0.00000; O3- C13- C27- O5

36	39	42	43	3	1.60387	4.81160	0.00000	-6.41547	0.00000
0.00000 ; C13- C27- O5- C14									
37	36	39	40	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000 ; H20- C13- C27- H33									
37	36	39	41	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000 ; H20- C13- C27- H34									
37	36	39	42	3	1.04600	-1.04600	0.00000	0.00000	0.00000
0.00000 ; H20- C13- C27- O5									
38	36	39	40	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000 ; H21- C13- C27- H33									
38	36	39	41	3	0.65084	1.95253	0.00000	-2.60338	0.00000
0.00000 ; H21- C13- C27- H34									
38	36	39	42	3	1.04600	-1.04600	0.00000	0.00000	0.00000
0.00000 ; H21- C13- C27- O5									
39	42	43	44	3	13.47248	0.00000	-13.47248	0.00000	0.00000
0.00000 ; C27- O5- C14- C15									
39	42	43	48	3	13.47248	0.00000	-13.47248	0.00000	0.00000
0.00000 ; C27- O5- C14- C19									
40	39	42	43	3	1.60387	4.81160	0.00000	-6.41547	0.00000
0.00000 ; H33- C27- O5- C14									
41	39	42	43	3	1.60387	4.81160	0.00000	-6.41547	0.00000
0.00000 ; H34- C27- O5- C14									
42	43	44	45	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000 ; O5- C14- C15- C16									
42	43	44	47	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000 ; O5- C14- C15- H22									
42	43	48	49	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000 ; O5- C14- C19- H25									
42	43	48	50	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000 ; O5- C14- C19- C18									
43	44	45	46	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000 ; C14- C15- C16- H23									

43 44 45 52 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C14- C15- C16- C17
 43 48 50 51 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C14- C19- C18- H24
 43 48 50 52 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C14- C19- C18- C17
 44 43 48 49 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C15- C14- C19- H25
 44 43 48 50 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C15- C14- C19- C18
 44 45 52 50 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C15- C16- C17- C18
 44 45 52 53 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C15- C16- C17- O4
 45 44 43 48 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C16- C15- C14- C19
 45 52 50 48 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C16- C17- C18- C19
 45 52 50 51 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; C16- C17- C18- H24
 45 52 53 54 3 13.47248 0.00000 -13.47248 0.00000 0.00000
 0.00000 ; C16- C17- O4- C20
 46 45 44 47 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; H23- C16- C15- H22
 46 45 52 50 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; H23- C16- C17- C18
 46 45 52 53 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; H23- C16- C17- O4
 47 44 43 48 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; H22- C15- C14- C19
 47 44 45 52 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000 ; H22- C15- C16- C17

48	50	52	53	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; C19- C18- C17- O4									
49	48	50	51	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; H25- C19- C18- H24									
49	48	50	52	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; H25- C19- C18- C17									
50	52	53	54	3	13.47248	0.00000	-13.47248	0.00000	0.00000
0.00000; C18- C17- O4- C20									
51	50	52	53	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; H24- C18- C17- O4									
52	53	54	55	3	1.60387	4.81160	0.00000	-6.41547	0.00000
0.00000; C17- O4- C20- H26									
52	53	54	56	3	1.60387	4.81160	0.00000	-6.41547	0.00000
0.00000; C17- O4- C20- H27									
52	53	54	57	3	1.60387	4.81160	0.00000	-6.41547	0.00000
0.00000; C17- O4- C20- C21									
53	54	57	58	3	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000; O4- C20- C21- C26									
53	54	57	66	3	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000; O4- C20- C21- C22									
54	57	58	59	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; C20- C21- C26- H32									
54	57	58	60	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; C20- C21- C26- C25									
54	57	66	64	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; C20- C21- C22- C23									
54	57	66	67	3	30.33400	0.00000	-30.33400	0.00000	0.00000
0.00000; C20- C21- C22- H28									
55	54	57	58	3	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000; H26- C20- C21- C26									
55	54	57	66	3	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000; H26- C20- C21- C22									

56 54 57 58 3 0.00000 0.00000 0.00000 0.00000 0.00000
 0.00000; H27- C20- C21- C26

56 54 57 66 3 0.00000 0.00000 0.00000 0.00000 0.00000
 0.00000; H27- C20- C21- C22

57 58 60 61 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C21- C26- C25- H31

57 58 60 62 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C21- C26- C25- C24

57 66 64 62 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C21- C22- C23- C24

57 66 64 65 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C21- C22- C23- H29

58 57 66 64 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C26- C21- C22- C23

58 57 66 67 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C26- C21- C22- H28

58 60 62 63 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C26- C25- C24- H30

58 60 62 64 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C26- C25- C24- C23

59 58 57 66 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H32- C26- C21- C22

59 58 60 61 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H32- C26- C25- H31

59 58 60 62 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; H32- C26- C25- C24

60 58 57 66 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C25- C26- C21- C22

60 62 64 65 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C25- C24- C23- H29

60 62 64 66 3 30.33400 0.00000 -30.33400 0.00000 0.00000
 0.00000; C25- C24- C23- C22

```

61 60 62 63 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000; H31- C25- C24- H30
61 60 62 64 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000; H31- C25- C24- C23
62 64 66 67 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000; C24- C23- C22- H28
63 62 64 65 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000; H30- C24- C23- H29
63 62 64 66 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000; H30- C24- C23- C22
65 64 66 67 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000; H29- C23- C22- H28

```

[dihedrals] ; impropers

; treated as propers in GROMACS to use correct AMBER analytical function

```

; i j k l func phase kd pn
25 27 26 29 1 180.00 4.60240 2; C7- C9- C8- H16
25 32 30 31 1 180.00 4.60240 2; C7- C11- C12- H19
26 30 25 24 1 180.00 4.60240 2; C8- C12- C7- O2
26 34 27 28 1 180.00 4.60240 2; C8- C10- C9- H17
27 32 34 35 1 180.00 4.60240 2; C9- C11- C10- O3
30 34 32 33 1 180.00 4.60240 2; C12- C10- C11- H18
43 45 44 47 1 180.00 4.60240 2; C14- C16- C15- H22
43 50 48 49 1 180.00 4.60240 2; C14- C18- C19- H25
44 48 43 42 1 180.00 4.60240 2; C15- C19- C14- O5
44 52 45 46 1 180.00 4.60240 2; C15- C17- C16- H23
45 50 52 53 1 180.00 4.60240 2; C16- C18- C17- O4
48 52 50 51 1 180.00 4.60240 2; C19- C17- C18- H24
57 60 58 59 1 180.00 4.60240 2; C21- C25- C26- H32
57 64 66 67 1 180.00 4.60240 2; C21- C23- C22- H28
58 62 60 61 1 180.00 4.60240 2; C26- C24- C25- H31
58 66 57 54 1 180.00 4.60240 2; C26- C22- C21- C20

```

60 64 62 63 1 180.00 4.60240 2 ; C25- C23- C24- H30
62 66 64 65 1 180.00 4.60240 2 ; C24- C22- C23- H29

Appendix 3 - Ligand **58** AMBERff99SB parameters in GROMACS
.itp format

; 7c_capped_GMX.itp created by acpype (v: 2019-11-07T23:16:00CET) on Thu Dec
19 11:36:04 2019

[atomtypes]

;name	bond_type	mass	charge	ptype	sigma	epsilon	Amb
c3	c3	0.00000	0.00000	A	3.39967e-01	4.57730e-01	; 1.91 0.1094
n4	n4	0.00000	0.00000	A	3.25000e-01	7.11280e-01	; 1.82 0.1700
oh	oh	0.00000	0.00000	A	3.06647e-01	8.80314e-01	; 1.72 0.2104
os	os	0.00000	0.00000	A	3.00001e-01	7.11280e-01	; 1.68 0.1700
ca	ca	0.00000	0.00000	A	3.39967e-01	3.59824e-01	; 1.91 0.0860
hx	hx	0.00000	0.00000	A	1.95998e-01	6.56888e-02	; 1.10 0.0157
hn	hn	0.00000	0.00000	A	1.06908e-01	6.56888e-02	; 0.60 0.0157
hc	hc	0.00000	0.00000	A	2.64953e-01	6.56888e-02	; 1.49 0.0157
h1	h1	0.00000	0.00000	A	2.47135e-01	6.56888e-02	; 1.39 0.0157
ho	ho	0.00000	0.00000	A	0.00000e+00	0.00000e+00	; 0.00 0.0000
ha	ha	0.00000	0.00000	A	2.59964e-01	6.27600e-02	; 1.46 0.0150

[moleculetype]

```
;name      nrexcl
7c_capped  3
```

[atoms]

; nr	type	resi	res	atom	cgmr	charge	mass	; qtot	bond_type
1	c3	1	UNK	C1	1	0.100800	12.01000	; qtot	0.101
2	n4	1	UNK	N1	2	-0.762504	14.01000	; qtot	-0.662
3	c3	1	UNK	C2	3	0.133000	12.01000	; qtot	-0.529
4	c3	1	UNK	C3	4	-0.134600	12.01000	; qtot	-0.663

5	c3	1	UNK	C4	5	-0.134600	12.01000 ; qtot -0.798
6	c3	1	UNK	C5	6	0.122100	12.01000 ; qtot -0.676
7	c3	1	UNK	C6	7	0.097400	12.01000 ; qtot -0.578
8	oh	1	UNK	O1	8	-0.632801	16.00000 ; qtot -1.211
9	os	1	UNK	O2	9	-0.356400	16.00000 ; qtot -1.568
10	ca	1	UNK	C7	10	0.020600	12.01000 ; qtot -1.547
11	ca	1	UNK	C8	11	-0.097500	12.01000 ; qtot -1.645
12	ca	1	UNK	C9	12	-0.169000	12.01000 ; qtot -1.814
13	ca	1	UNK	C10	13	0.145600	12.01000 ; qtot -1.668
14	ca	1	UNK	C11	14	-0.169000	12.01000 ; qtot -1.837
15	ca	1	UNK	C12	15	-0.097500	12.01000 ; qtot -1.934
16	os	1	UNK	O3	16	-0.315400	16.00000 ; qtot -2.250
17	c3	1	UNK	C13	17	0.119400	12.01000 ; qtot -2.130
18	ca	1	UNK	C14	18	0.096100	12.01000 ; qtot -2.034
19	ca	1	UNK	C15	19	-0.153000	12.01000 ; qtot -2.187
20	ca	1	UNK	C16	20	-0.132750	12.01000 ; qtot -2.320
21	ca	1	UNK	C17	21	0.087100	12.01000 ; qtot -2.233
22	ca	1	UNK	C18	22	-0.132750	12.01000 ; qtot -2.366
23	ca	1	UNK	C19	23	-0.153000	12.01000 ; qtot -2.519
24	os	1	UNK	O4	24	-0.335400	16.00000 ; qtot -2.854
25	c3	1	UNK	C27	25	0.111400	12.01000 ; qtot -2.743
26	os	1	UNK	O5	26	-0.331400	16.00000 ; qtot -3.074
27	c3	1	UNK	C20	27	0.100800	12.01000 ; qtot -2.973
28	n4	1	UNK	N2	28	-0.762504	14.01000 ; qtot -3.736
29	c3	1	UNK	C22	29	0.133000	12.01000 ; qtot -3.603
30	c3	1	UNK	C24	30	-0.134600	12.01000 ; qtot -3.737
31	c3	1	UNK	C26	31	-0.134600	12.01000 ; qtot -3.872
32	c3	1	UNK	C29	32	0.122100	12.01000 ; qtot -3.750
33	c3	1	UNK	C31	33	0.097400	12.01000 ; qtot -3.653
34	oh	1	UNK	O6	34	-0.632801	16.00000 ; qtot -4.285
35	os	1	UNK	O7	35	-0.356400	16.00000 ; qtot -4.642
36	ca	1	UNK	C33	36	0.020600	12.01000 ; qtot -4.621

37	ca	1	UNK	C35	37	-0.097500	12.01000 ; qtot -4.719
38	ca	1	UNK	C36	38	-0.169000	12.01000 ; qtot -4.888
39	ca	1	UNK	C37	39	0.145600	12.01000 ; qtot -4.742
40	ca	1	UNK	C38	40	-0.169000	12.01000 ; qtot -4.911
41	ca	1	UNK	C39	41	-0.097500	12.01000 ; qtot -5.009
42	os	1	UNK	O8	42	-0.315400	16.00000 ; qtot -5.324
43	c3	1	UNK	C40	43	0.119400	12.01000 ; qtot -5.205
44	ca	1	UNK	C41	44	0.096100	12.01000 ; qtot -5.108
45	ca	1	UNK	C42	45	-0.153000	12.01000 ; qtot -5.261
46	ca	1	UNK	C43	46	-0.132750	12.01000 ; qtot -5.394
47	ca	1	UNK	C44	47	0.087100	12.01000 ; qtot -5.307
48	ca	1	UNK	C45	48	-0.132750	12.01000 ; qtot -5.440
49	ca	1	UNK	C46	49	-0.153000	12.01000 ; qtot -5.593
50	os	1	UNK	O9	50	-0.335400	16.00000 ; qtot -5.928
51	c3	1	UNK	C47	51	0.111400	12.01000 ; qtot -5.817
52	os	1	UNK	O10	52	-0.331400	16.00000 ; qtot -6.148
53	c3	1	UNK	C21	53	0.129400	12.01000 ; qtot -6.019
54	c3	1	UNK	C23	54	0.129400	12.01000 ; qtot -5.889
55	c3	1	UNK	C25	55	-0.096400	12.01000 ; qtot -5.986
56	c3	1	UNK	C28	56	-0.076900	12.01000 ; qtot -6.063
57	c3	1	UNK	C30	57	-0.082400	12.01000 ; qtot -6.145
58	c3	1	UNK	C32	58	-0.076900	12.01000 ; qtot -6.222
59	c3	1	UNK	C34	59	-0.096400	12.01000 ; qtot -6.318
60	hx	1	UNK	H1	60	0.115950	1.00800 ; qtot -6.202
61	hx	1	UNK	H2	61	0.115950	1.00800 ; qtot -6.087
62	hn	1	UNK	H3	62	0.457550	1.00800 ; qtot -5.629
63	hn	1	UNK	H4	63	0.457550	1.00800 ; qtot -5.171
64	hx	1	UNK	H5	64	0.106700	1.00800 ; qtot -5.065
65	hc	1	UNK	H6	65	0.078867	1.00800 ; qtot -4.986
66	hc	1	UNK	H7	66	0.078867	1.00800 ; qtot -4.907
67	hc	1	UNK	H8	67	0.078867	1.00800 ; qtot -4.828
68	hc	1	UNK	H9	68	0.078867	1.00800 ; qtot -4.749

69	hc	1	UNK	H10	69	0.078867	1.00800 ; qtot -4.670
70	hc	1	UNK	H11	70	0.078867	1.00800 ; qtot -4.592
71	h1	1	UNK	H12	71	0.073700	1.00800 ; qtot -4.518
72	h1	1	UNK	H13	72	0.061200	1.00800 ; qtot -4.457
73	h1	1	UNK	H14	73	0.061200	1.00800 ; qtot -4.395
74	ho	1	UNK	H15	74	0.460500	1.00800 ; qtot -3.935
75	ha	1	UNK	H16	75	0.143750	1.00800 ; qtot -3.791
76	ha	1	UNK	H17	76	0.157750	1.00800 ; qtot -3.633
77	ha	1	UNK	H18	77	0.157750	1.00800 ; qtot -3.476
78	ha	1	UNK	H19	78	0.143750	1.00800 ; qtot -3.332
79	h1	1	UNK	H20	79	0.072450	1.00800 ; qtot -3.259
80	h1	1	UNK	H21	80	0.072450	1.00800 ; qtot -3.187
81	ha	1	UNK	H22	81	0.143250	1.00800 ; qtot -3.044
82	ha	1	UNK	H23	82	0.147500	1.00800 ; qtot -2.896
83	ha	1	UNK	H24	83	0.147500	1.00800 ; qtot -2.749
84	ha	1	UNK	H25	84	0.143250	1.00800 ; qtot -2.606
85	h1	1	UNK	H33	85	0.056200	1.00800 ; qtot -2.549
86	h1	1	UNK	H34	86	0.056200	1.00800 ; qtot -2.493
87	hx	1	UNK	H26	87	0.115950	1.00800 ; qtot -2.377
88	hx	1	UNK	H28	88	0.115950	1.00800 ; qtot -2.261
89	hn	1	UNK	H30	89	0.457550	1.00800 ; qtot -1.804
90	hn	1	UNK	H32	90	0.457550	1.00800 ; qtot -1.346
91	hx	1	UNK	H36	91	0.106700	1.00800 ; qtot -1.239
92	hc	1	UNK	H38	92	0.078867	1.00800 ; qtot -1.161
93	hc	1	UNK	H40	93	0.078867	1.00800 ; qtot -1.082
94	hc	1	UNK	H42	94	0.078867	1.00800 ; qtot -1.003
95	hc	1	UNK	H44	95	0.078867	1.00800 ; qtot -0.924
96	hc	1	UNK	H46	96	0.078867	1.00800 ; qtot -0.845
97	hc	1	UNK	H48	97	0.078867	1.00800 ; qtot -0.766
98	h1	1	UNK	H50	98	0.073700	1.00800 ; qtot -0.693
99	h1	1	UNK	H52	99	0.061200	1.00800 ; qtot -0.631
100	h1	1	UNK	H54	100	0.061200	1.00800 ; qtot -0.570

101	ho	1	UNK	H56	101	0.460500	1.00800 ; qtot -0.110
102	ha	1	UNK	H57	102	0.143750	1.00800 ; qtot 0.034
103	ha	1	UNK	H58	103	0.157750	1.00800 ; qtot 0.192
104	ha	1	UNK	H59	104	0.157750	1.00800 ; qtot 0.350
105	ha	1	UNK	H60	105	0.143750	1.00800 ; qtot 0.493
106	h1	1	UNK	H61	106	0.072450	1.00800 ; qtot 0.566
107	h1	1	UNK	H62	107	0.072450	1.00800 ; qtot 0.638
108	ha	1	UNK	H63	108	0.143250	1.00800 ; qtot 0.782
109	ha	1	UNK	H64	109	0.147500	1.00800 ; qtot 0.929
110	ha	1	UNK	H65	110	0.147500	1.00800 ; qtot 1.077
111	ha	1	UNK	H66	111	0.143250	1.00800 ; qtot 1.220
112	h1	1	UNK	H67	112	0.056200	1.00800 ; qtot 1.276
113	h1	1	UNK	H68	113	0.056200	1.00800 ; qtot 1.332
114	h1	1	UNK	H27	114	0.045950	1.00800 ; qtot 1.378
115	h1	1	UNK	H29	115	0.045950	1.00800 ; qtot 1.424
116	hc	1	UNK	H31	116	0.053950	1.00800 ; qtot 1.478
117	hc	1	UNK	H35	117	0.053950	1.00800 ; qtot 1.532
118	hc	1	UNK	H37	118	0.043700	1.00800 ; qtot 1.576
119	hc	1	UNK	H39	119	0.043700	1.00800 ; qtot 1.619
120	hc	1	UNK	H41	120	0.046700	1.00800 ; qtot 1.666
121	hc	1	UNK	H43	121	0.046700	1.00800 ; qtot 1.713
122	hc	1	UNK	H45	122	0.043700	1.00800 ; qtot 1.756
123	hc	1	UNK	H47	123	0.043700	1.00800 ; qtot 1.800
124	h1	1	UNK	H49	124	0.045950	1.00800 ; qtot 1.846
125	h1	1	UNK	H51	125	0.045950	1.00800 ; qtot 1.892
126	hc	1	UNK	H53	126	0.053950	1.00800 ; qtot 1.946
127	hc	1	UNK	H55	127	0.053950	1.00800 ; qtot 2.000

[bonds]

; ai	aj	funct	r	k	
1	2	1	1.5110e-01	2.3707e+05	; C1 - N1
1	6	1	1.5375e-01	2.5179e+05	; C1 - C5

1	60	1	1.0910e-01	2.8342e+05 ;	C1 - H1
1	61	1	1.0910e-01	2.8342e+05 ;	C1 - H2
2	3	1	1.5110e-01	2.3707e+05 ;	N1 - C2
2	62	1	1.0304e-01	3.1229e+05 ;	N1 - H3
2	63	1	1.0304e-01	3.1229e+05 ;	N1 - H4
3	4	1	1.5375e-01	2.5179e+05 ;	C2 - C3
3	5	1	1.5375e-01	2.5179e+05 ;	C2 - C4
3	64	1	1.0910e-01	2.8342e+05 ;	C2 - H5
4	65	1	1.0969e-01	2.7665e+05 ;	C3 - H6
4	66	1	1.0969e-01	2.7665e+05 ;	C3 - H7
4	67	1	1.0969e-01	2.7665e+05 ;	C3 - H8
5	68	1	1.0969e-01	2.7665e+05 ;	C4 - H9
5	69	1	1.0969e-01	2.7665e+05 ;	C4 - H10
5	70	1	1.0969e-01	2.7665e+05 ;	C4 - H11
6	7	1	1.5375e-01	2.5179e+05 ;	C5 - C6
6	8	1	1.4233e-01	2.6501e+05 ;	C5 - O1
6	71	1	1.0969e-01	2.7665e+05 ;	C5 - H12
7	9	1	1.4316e-01	2.5824e+05 ;	C6 - O2
7	72	1	1.0969e-01	2.7665e+05 ;	C6 - H13
7	73	1	1.0969e-01	2.7665e+05 ;	C6 - H14
8	74	1	9.7300e-02	3.1079e+05 ;	O1 - H15
9	10	1	1.3696e-01	3.1514e+05 ;	O2 - C7
10	11	1	1.3984e-01	3.8585e+05 ;	C7 - C8
10	15	1	1.3984e-01	3.8585e+05 ;	C7 - C12
11	12	1	1.3984e-01	3.8585e+05 ;	C8 - C9
11	75	1	1.0860e-01	2.8937e+05 ;	C8 - H16
12	13	1	1.3984e-01	3.8585e+05 ;	C9 - C10
12	76	1	1.0860e-01	2.8937e+05 ;	C9 - H17
13	14	1	1.3984e-01	3.8585e+05 ;	C10 - C11
13	16	1	1.3696e-01	3.1514e+05 ;	C10 - O3
14	15	1	1.3984e-01	3.8585e+05 ;	C11 - C12
14	77	1	1.0860e-01	2.8937e+05 ;	C11 - H18

15	78	1	1.0860e-01	2.8937e+05	; C12 - H19
16	17	1	1.4316e-01	2.5824e+05	; O3 - C13
17	25	1	1.5375e-01	2.5179e+05	; C13 - C27
17	79	1	1.0969e-01	2.7665e+05	; C13 - H20
17	80	1	1.0969e-01	2.7665e+05	; C13 - H21
18	19	1	1.3984e-01	3.8585e+05	; C14 - C15
18	23	1	1.3984e-01	3.8585e+05	; C14 - C19
18	26	1	1.3696e-01	3.1514e+05	; C14 - O5
19	20	1	1.3984e-01	3.8585e+05	; C15 - C16
19	81	1	1.0860e-01	2.8937e+05	; C15 - H22
20	21	1	1.3984e-01	3.8585e+05	; C16 - C17
20	82	1	1.0860e-01	2.8937e+05	; C16 - H23
21	22	1	1.3984e-01	3.8585e+05	; C17 - C18
21	24	1	1.3696e-01	3.1514e+05	; C17 - O4
22	23	1	1.3984e-01	3.8585e+05	; C18 - C19
22	83	1	1.0860e-01	2.8937e+05	; C18 - H24
23	84	1	1.0860e-01	2.8937e+05	; C19 - H25
24	54	1	1.4316e-01	2.5824e+05	; O4 - C23
25	26	1	1.4316e-01	2.5824e+05	; C27 - O5
25	85	1	1.0969e-01	2.7665e+05	; C27 - H33
25	86	1	1.0969e-01	2.7665e+05	; C27 - H34
27	28	1	1.5110e-01	2.3707e+05	; C20 - N2
27	32	1	1.5375e-01	2.5179e+05	; C20 - C29
27	87	1	1.0910e-01	2.8342e+05	; C20 - H26
27	88	1	1.0910e-01	2.8342e+05	; C20 - H28
28	29	1	1.5110e-01	2.3707e+05	; N2 - C22
28	89	1	1.0304e-01	3.1229e+05	; N2 - H30
28	90	1	1.0304e-01	3.1229e+05	; N2 - H32
29	30	1	1.5375e-01	2.5179e+05	; C22 - C24
29	31	1	1.5375e-01	2.5179e+05	; C22 - C26
29	91	1	1.0910e-01	2.8342e+05	; C22 - H36
30	92	1	1.0969e-01	2.7665e+05	; C24 - H38

30 93 1 1.0969e-01 2.7665e+05 ; C24 - H40
30 94 1 1.0969e-01 2.7665e+05 ; C24 - H42
31 95 1 1.0969e-01 2.7665e+05 ; C26 - H44
31 96 1 1.0969e-01 2.7665e+05 ; C26 - H46
31 97 1 1.0969e-01 2.7665e+05 ; C26 - H48
32 33 1 1.5375e-01 2.5179e+05 ; C29 - C31
32 34 1 1.4233e-01 2.6501e+05 ; C29 - O6
32 98 1 1.0969e-01 2.7665e+05 ; C29 - H50
33 35 1 1.4316e-01 2.5824e+05 ; C31 - O7
33 99 1 1.0969e-01 2.7665e+05 ; C31 - H52
33 100 1 1.0969e-01 2.7665e+05 ; C31 - H54
34 101 1 9.7300e-02 3.1079e+05 ; O6 - H56
35 36 1 1.3696e-01 3.1514e+05 ; O7 - C33
36 37 1 1.3984e-01 3.8585e+05 ; C33 - C35
36 41 1 1.3984e-01 3.8585e+05 ; C33 - C39
37 38 1 1.3984e-01 3.8585e+05 ; C35 - C36
37 102 1 1.0860e-01 2.8937e+05 ; C35 - H57
38 39 1 1.3984e-01 3.8585e+05 ; C36 - C37
38 103 1 1.0860e-01 2.8937e+05 ; C36 - H58
39 40 1 1.3984e-01 3.8585e+05 ; C37 - C38
39 42 1 1.3696e-01 3.1514e+05 ; C37 - O8
40 41 1 1.3984e-01 3.8585e+05 ; C38 - C39
40 104 1 1.0860e-01 2.8937e+05 ; C38 - H59
41 105 1 1.0860e-01 2.8937e+05 ; C39 - H60
42 43 1 1.4316e-01 2.5824e+05 ; O8 - C40
43 51 1 1.5375e-01 2.5179e+05 ; C40 - C47
43 106 1 1.0969e-01 2.7665e+05 ; C40 - H61
43 107 1 1.0969e-01 2.7665e+05 ; C40 - H62
44 45 1 1.3984e-01 3.8585e+05 ; C41 - C42
44 49 1 1.3984e-01 3.8585e+05 ; C41 - C46
44 52 1 1.3696e-01 3.1514e+05 ; C41 - O10
45 46 1 1.3984e-01 3.8585e+05 ; C42 - C43

45 108 1 1.0860e-01 2.8937e+05 ; C42 - H63
46 47 1 1.3984e-01 3.8585e+05 ; C43 - C44
46 109 1 1.0860e-01 2.8937e+05 ; C43 - H64
47 48 1 1.3984e-01 3.8585e+05 ; C44 - C45
47 50 1 1.3696e-01 3.1514e+05 ; C44 - O9
48 49 1 1.3984e-01 3.8585e+05 ; C45 - C46
48 110 1 1.0860e-01 2.8937e+05 ; C45 - H65
49 111 1 1.0860e-01 2.8937e+05 ; C46 - H66
50 53 1 1.4316e-01 2.5824e+05 ; O9 - C21
51 52 1 1.4316e-01 2.5824e+05 ; C47 - O10
51 112 1 1.0969e-01 2.7665e+05 ; C47 - H67
51 113 1 1.0969e-01 2.7665e+05 ; C47 - H68
53 55 1 1.5375e-01 2.5179e+05 ; C21 - C25
53 114 1 1.0969e-01 2.7665e+05 ; C21 - H27
53 115 1 1.0969e-01 2.7665e+05 ; C21 - H29
54 59 1 1.5375e-01 2.5179e+05 ; C23 - C34
54 124 1 1.0969e-01 2.7665e+05 ; C23 - H49
54 125 1 1.0969e-01 2.7665e+05 ; C23 - H51
55 56 1 1.5375e-01 2.5179e+05 ; C25 - C28
55 116 1 1.0969e-01 2.7665e+05 ; C25 - H31
55 117 1 1.0969e-01 2.7665e+05 ; C25 - H35
56 57 1 1.5375e-01 2.5179e+05 ; C28 - C30
56 118 1 1.0969e-01 2.7665e+05 ; C28 - H37
56 119 1 1.0969e-01 2.7665e+05 ; C28 - H39
57 58 1 1.5375e-01 2.5179e+05 ; C30 - C32
57 120 1 1.0969e-01 2.7665e+05 ; C30 - H41
57 121 1 1.0969e-01 2.7665e+05 ; C30 - H43
58 59 1 1.5375e-01 2.5179e+05 ; C32 - C34
58 122 1 1.0969e-01 2.7665e+05 ; C32 - H45
58 123 1 1.0969e-01 2.7665e+05 ; C32 - H47
59 126 1 1.0969e-01 2.7665e+05 ; C34 - H53
59 127 1 1.0969e-01 2.7665e+05 ; C34 - H55

[pairs]

; ai aj funct

1	4	1;	C1 - C3
1	5	1;	C1 - C4
1	9	1;	C1 - O2
1	64	1;	C1 - H5
1	72	1;	C1 - H13
1	73	1;	C1 - H14
1	74	1;	C1 - H15
2	7	1;	N1 - C6
2	8	1;	N1 - O1
2	65	1;	N1 - H6
2	66	1;	N1 - H7
2	67	1;	N1 - H8
2	68	1;	N1 - H9
2	69	1;	N1 - H10
2	70	1;	N1 - H11
2	71	1;	N1 - H12
4	62	1;	C3 - H3
4	63	1;	C3 - H4
4	68	1;	C3 - H9
4	69	1;	C3 - H10
4	70	1;	C3 - H11
5	62	1;	C4 - H3
5	63	1;	C4 - H4
5	65	1;	C4 - H6
5	66	1;	C4 - H7
5	67	1;	C4 - H8
6	3	1;	C5 - C2
6	10	1;	C5 - C7
6	62	1;	C5 - H3

6 63 1; C5 - H4
7 11 1; C6 - C8
7 15 1; C6 - C12
7 74 1; C6 - H15
8 9 1; O1 - O2
8 72 1; O1 - H13
8 73 1; O1 - H14
9 12 1; O2 - C9
9 14 1; O2 - C11
9 71 1; O2 - H12
9 75 1; O2 - H16
9 78 1; O2 - H19
10 13 1; C7 - C10
10 72 1; C7 - H13
10 73 1; C7 - H14
10 76 1; C7 - H17
10 77 1; C7 - H18
11 14 1; C8 - C11
11 16 1; C8 - O3
11 78 1; C8 - H19
12 15 1; C9 - C12
12 17 1; C9 - C13
12 77 1; C9 - H18
13 25 1; C10 - C27
13 75 1; C10 - H16
13 78 1; C10 - H19
13 79 1; C10 - H20
13 80 1; C10 - H21
14 17 1; C11 - C13
14 76 1; C11 - H17
15 16 1; C12 - O3
15 75 1; C12 - H16

16 26 1; O3 - O5
16 76 1; O3 - H17
16 77 1; O3 - H18
16 85 1; O3 - H33
16 86 1; O3 - H34
17 18 1; C13 - C14
18 21 1; C14 - C17
18 82 1; C14 - H23
18 83 1; C14 - H24
18 85 1; C14 - H33
18 86 1; C14 - H34
19 22 1; C15 - C18
19 24 1; C15 - O4
19 25 1; C15 - C27
19 84 1; C15 - H25
20 23 1; C16 - C19
20 26 1; C16 - O5
20 54 1; C16 - C23
20 83 1; C16 - H24
21 59 1; C17 - C34
21 81 1; C17 - H22
21 84 1; C17 - H25
21 124 1; C17 - H49
21 125 1; C17 - H51
22 26 1; C18 - O5
22 54 1; C18 - C23
22 82 1; C18 - H23
23 24 1; C19 - O4
23 25 1; C19 - C27
23 81 1; C19 - H22
24 58 1; O4 - C32
24 82 1; O4 - H23

24 83 1; O4 - H24
24 126 1; O4 - H53
24 127 1; O4 - H55
26 79 1; O5 - H20
26 80 1; O5 - H21
26 81 1; O5 - H22
26 84 1; O5 - H25
27 30 1; C20 - C24
27 31 1; C20 - C26
27 35 1; C20 - O7
27 91 1; C20 - H36
27 99 1; C20 - H52
27 100 1; C20 - H54
27 101 1; C20 - H56
28 33 1; N2 - C31
28 34 1; N2 - O6
28 92 1; N2 - H38
28 93 1; N2 - H40
28 94 1; N2 - H42
28 95 1; N2 - H44
28 96 1; N2 - H46
28 97 1; N2 - H48
28 98 1; N2 - H50
29 32 1; C22 - C29
29 87 1; C22 - H26
29 88 1; C22 - H28
30 89 1; C24 - H30
30 90 1; C24 - H32
30 95 1; C24 - H44
30 96 1; C24 - H46
30 97 1; C24 - H48
31 89 1; C26 - H30

31 90 1; C26 - H32
31 92 1; C26 - H38
31 93 1; C26 - H40
31 94 1; C26 - H42
32 36 1; C29 - C33
32 89 1; C29 - H30
32 90 1; C29 - H32
33 37 1; C31 - C35
33 41 1; C31 - C39
33 87 1; C31 - H26
33 88 1; C31 - H28
33 101 1; C31 - H56
34 35 1; O6 - O7
34 87 1; O6 - H26
34 88 1; O6 - H28
34 99 1; O6 - H52
34 100 1; O6 - H54
35 38 1; O7 - C36
35 40 1; O7 - C38
35 98 1; O7 - H50
35 102 1; O7 - H57
35 105 1; O7 - H60
36 39 1; C33 - C37
36 99 1; C33 - H52
36 100 1; C33 - H54
36 103 1; C33 - H58
36 104 1; C33 - H59
37 40 1; C35 - C38
37 42 1; C35 - O8
37 105 1; C35 - H60
38 41 1; C36 - C39
38 43 1; C36 - C40

38 104 1; C36 - H59
39 51 1; C37 - C47
39 102 1; C37 - H57
39 105 1; C37 - H60
39 106 1; C37 - H61
39 107 1; C37 - H62
40 43 1; C38 - C40
40 103 1; C38 - H58
41 42 1; C39 - O8
41 102 1; C39 - H57
42 52 1; O8 - O10
42 103 1; O8 - H58
42 104 1; O8 - H59
42 112 1; O8 - H67
42 113 1; O8 - H68
43 44 1; C40 - C41
44 47 1; C41 - C44
44 109 1; C41 - H64
44 110 1; C41 - H65
44 112 1; C41 - H67
44 113 1; C41 - H68
45 48 1; C42 - C45
45 50 1; C42 - O9
45 51 1; C42 - C47
45 111 1; C42 - H66
46 49 1; C43 - C46
46 52 1; C43 - O10
46 53 1; C43 - C21
46 110 1; C43 - H65
47 55 1; C44 - C25
47 108 1; C44 - H63
47 111 1; C44 - H66

47 114 1; C44 - H27
47 115 1; C44 - H29
48 52 1; C45 - O10
48 53 1; C45 - C21
48 109 1; C45 - H64
49 50 1; C46 - O9
49 51 1; C46 - C47
49 108 1; C46 - H63
50 56 1; O9 - C28
50 109 1; O9 - H64
50 110 1; O9 - H65
50 116 1; O9 - H31
50 117 1; O9 - H35
52 106 1; O10 - H61
52 107 1; O10 - H62
52 108 1; O10 - H63
52 111 1; O10 - H66
53 57 1; C21 - C30
53 118 1; C21 - H37
53 119 1; C21 - H39
54 57 1; C23 - C30
54 122 1; C23 - H45
54 123 1; C23 - H47
55 58 1; C25 - C32
55 120 1; C25 - H41
55 121 1; C25 - H43
56 59 1; C28 - C34
56 114 1; C28 - H27
56 115 1; C28 - H29
56 122 1; C28 - H45
56 123 1; C28 - H47
57 116 1; C30 - H31

57 117 1; C30 - H35
57 126 1; C30 - H53
57 127 1; C30 - H55
58 118 1; C32 - H37
58 119 1; C32 - H39
58 124 1; C32 - H49
58 125 1; C32 - H51
59 120 1; C34 - H41
59 121 1; C34 - H43
60 3 1; H1 - C2
60 7 1; H1 - C6
60 8 1; H1 - O1
60 62 1; H1 - H3
60 63 1; H1 - H4
60 71 1; H1 - H12
61 3 1; H2 - C2
61 7 1; H2 - C6
61 8 1; H2 - O1
61 62 1; H2 - H3
61 63 1; H2 - H4
61 71 1; H2 - H12
62 64 1; H3 - H5
63 64 1; H4 - H5
64 65 1; H5 - H6
64 66 1; H5 - H7
64 67 1; H5 - H8
64 68 1; H5 - H9
64 69 1; H5 - H10
64 70 1; H5 - H11
71 72 1; H12 - H13
71 73 1; H12 - H14
71 74 1; H12 - H15

75 76 1; H16 - H17
77 78 1; H18 - H19
79 85 1; H20 - H33
79 86 1; H20 - H34
80 85 1; H21 - H33
80 86 1; H21 - H34
81 82 1; H22 - H23
83 84 1; H24 - H25
87 89 1; H26 - H30
87 90 1; H26 - H32
87 98 1; H26 - H50
88 89 1; H28 - H30
88 90 1; H28 - H32
88 98 1; H28 - H50
89 91 1; H30 - H36
90 91 1; H32 - H36
91 92 1; H36 - H38
91 93 1; H36 - H40
91 94 1; H36 - H42
91 95 1; H36 - H44
91 96 1; H36 - H46
91 97 1; H36 - H48
98 99 1; H50 - H52
98 100 1; H50 - H54
98 101 1; H50 - H56
102 103 1; H57 - H58
104 105 1; H59 - H60
106 112 1; H61 - H67
106 113 1; H61 - H68
107 112 1; H62 - H67
107 113 1; H62 - H68
108 109 1; H63 - H64

110 111 1; H65 - H66
 114 116 1; H27 - H31
 114 117 1; H27 - H35
 115 116 1; H29 - H31
 115 117 1; H29 - H35
 116 118 1; H31 - H37
 116 119 1; H31 - H39
 117 118 1; H35 - H37
 117 119 1; H35 - H39
 118 120 1; H37 - H41
 118 121 1; H37 - H43
 119 120 1; H39 - H41
 119 121 1; H39 - H43
 120 122 1; H41 - H45
 120 123 1; H41 - H47
 121 122 1; H43 - H45
 121 123 1; H43 - H47
 122 126 1; H45 - H53
 122 127 1; H45 - H55
 123 126 1; H47 - H53
 123 127 1; H47 - H55
 124 126 1; H49 - H53
 124 127 1; H49 - H55
 125 126 1; H51 - H53
 125 127 1; H51 - H55

[angles]

	ai	aj	ak	funct	theta	cth		
	1	2	3	1	1.0966e+02	5.2400e+02	;	C1 - N1 - C2
	1	2	62	1	1.1011e+02	3.8367e+02	;	C1 - N1 - H3
	1	2	63	1	1.1011e+02	3.8367e+02	;	C1 - N1 - H4
	1	6	7	1	1.1151e+02	5.2601e+02	;	C1 - C5 - C6

1	6	8	1	1.1019e+02	5.6459e+02 ;	C1 - C5	- O1
1	6	71	1	1.0956e+02	3.8819e+02 ;	C1 - C5	- H12
2	1	6	1	1.1421e+02	5.3706e+02 ;	N1 - C1	- C5
2	1	60	1	1.0801e+02	4.0710e+02 ;	N1 - C1	- H1
2	1	61	1	1.0801e+02	4.0710e+02 ;	N1 - C1	- H2
2	3	4	1	1.1421e+02	5.3706e+02 ;	N1 - C2	- C3
2	3	5	1	1.1421e+02	5.3706e+02 ;	N1 - C2	- C4
2	3	64	1	1.0801e+02	4.0710e+02 ;	N1 - C2	- H5
3	2	62	1	1.1011e+02	3.8367e+02 ;	C2 - N1	- H3
3	2	63	1	1.1011e+02	3.8367e+02 ;	C2 - N1	- H4
3	4	65	1	1.0980e+02	3.8777e+02 ;	C2 - C3	- H6
3	4	66	1	1.0980e+02	3.8777e+02 ;	C2 - C3	- H7
3	4	67	1	1.0980e+02	3.8777e+02 ;	C2 - C3	- H8
3	5	68	1	1.0980e+02	3.8777e+02 ;	C2 - C4	- H9
3	5	69	1	1.0980e+02	3.8777e+02 ;	C2 - C4	- H10
3	5	70	1	1.0980e+02	3.8777e+02 ;	C2 - C4	- H11
4	3	5	1	1.1151e+02	5.2601e+02 ;	C3 - C2	- C4
4	3	64	1	1.1056e+02	3.8660e+02 ;	C3 - C2	- H5
5	3	64	1	1.1056e+02	3.8660e+02 ;	C4 - C2	- H5
6	1	60	1	1.1056e+02	3.8660e+02 ;	C5 - C1	- H1
6	1	61	1	1.1056e+02	3.8660e+02 ;	C5 - C1	- H2
6	7	9	1	1.0797e+02	5.6902e+02 ;	C5 - C6	- O2
6	7	72	1	1.0956e+02	3.8819e+02 ;	C5 - C6	- H13
6	7	73	1	1.0956e+02	3.8819e+02 ;	C5 - C6	- H14
6	8	74	1	1.0726e+02	3.9648e+02 ;	C5 - O1	- H15
7	6	8	1	1.1019e+02	5.6459e+02 ;	C6 - C5	- O1
7	6	71	1	1.0956e+02	3.8819e+02 ;	C6 - C5	- H12
7	9	10	1	1.1796e+02	5.2317e+02 ;	C6 - O2	- C7
8	6	71	1	1.1026e+02	4.2618e+02 ;	O1 - C5	- H12
9	7	72	1	1.0978e+02	4.2509e+02 ;	O2 - C6	- H13
9	7	73	1	1.0978e+02	4.2509e+02 ;	O2 - C6	- H14
9	10	11	1	1.1920e+02	5.8225e+02 ;	O2 - C7	- C8

9	10	15	1	1.1920e+02	5.8225e+02 ;	O2 - C7	- C12
10	11	12	1	1.2002e+02	5.5748e+02 ;	C7 - C8	- C9
10	11	75	1	1.1988e+02	4.0317e+02 ;	C7 - C8	- H16
10	15	14	1	1.2002e+02	5.5748e+02 ;	C7 - C12	- C11
10	15	78	1	1.1988e+02	4.0317e+02 ;	C7 - C12	- H19
11	10	15	1	1.2002e+02	5.5748e+02 ;	C8 - C7	- C12
11	12	13	1	1.2002e+02	5.5748e+02 ;	C8 - C9	- C10
11	12	76	1	1.1988e+02	4.0317e+02 ;	C8 - C9	- H17
12	11	75	1	1.1988e+02	4.0317e+02 ;	C9 - C8	- H16
12	13	14	1	1.2002e+02	5.5748e+02 ;	C9 - C10	- C11
12	13	16	1	1.1920e+02	5.8225e+02 ;	C9 - C10	- O3
13	12	76	1	1.1988e+02	4.0317e+02 ;	C10 - C9	- H17
13	14	15	1	1.2002e+02	5.5748e+02 ;	C10 - C11	- C12
13	14	77	1	1.1988e+02	4.0317e+02 ;	C10 - C11	- H18
13	16	17	1	1.1796e+02	5.2317e+02 ;	C10 - O3	- C13
14	13	16	1	1.1920e+02	5.8225e+02 ;	C11 - C10	- O3
14	15	78	1	1.1988e+02	4.0317e+02 ;	C11 - C12	- H19
15	14	77	1	1.1988e+02	4.0317e+02 ;	C12 - C11	- H18
16	17	25	1	1.0797e+02	5.6902e+02 ;	O3 - C13	- C27
16	17	79	1	1.0978e+02	4.2509e+02 ;	O3 - C13	- H20
16	17	80	1	1.0978e+02	4.2509e+02 ;	O3 - C13	- H21
17	25	26	1	1.0797e+02	5.6902e+02 ;	C13 - C27	- O5
17	25	85	1	1.0956e+02	3.8819e+02 ;	C13 - C27	- H33
17	25	86	1	1.0956e+02	3.8819e+02 ;	C13 - C27	- H34
18	19	20	1	1.2002e+02	5.5748e+02 ;	C14 - C15	- C16
18	19	81	1	1.1988e+02	4.0317e+02 ;	C14 - C15	- H22
18	23	22	1	1.2002e+02	5.5748e+02 ;	C14 - C19	- C18
18	23	84	1	1.1988e+02	4.0317e+02 ;	C14 - C19	- H25
18	26	25	1	1.1796e+02	5.2317e+02 ;	C14 - O5	- C27
19	18	23	1	1.2002e+02	5.5748e+02 ;	C15 - C14	- C19
19	18	26	1	1.1920e+02	5.8225e+02 ;	C15 - C14	- O5
19	20	21	1	1.2002e+02	5.5748e+02 ;	C15 - C16	- C17

19	20	82	1	1.1988e+02	4.0317e+02 ;	C15 - C16	- H23
20	19	81	1	1.1988e+02	4.0317e+02 ;	C16 - C15	- H22
20	21	22	1	1.2002e+02	5.5748e+02 ;	C16 - C17	- C18
20	21	24	1	1.1920e+02	5.8225e+02 ;	C16 - C17	- O4
21	20	82	1	1.1988e+02	4.0317e+02 ;	C17 - C16	- H23
21	22	23	1	1.2002e+02	5.5748e+02 ;	C17 - C18	- C19
21	22	83	1	1.1988e+02	4.0317e+02 ;	C17 - C18	- H24
21	24	54	1	1.1796e+02	5.2317e+02 ;	C17 - O4	- C23
22	21	24	1	1.1920e+02	5.8225e+02 ;	C18 - C17	- O4
22	23	84	1	1.1988e+02	4.0317e+02 ;	C18 - C19	- H25
23	18	26	1	1.1920e+02	5.8225e+02 ;	C19 - C14	- O5
23	22	83	1	1.1988e+02	4.0317e+02 ;	C19 - C18	- H24
24	54	59	1	1.0797e+02	5.6902e+02 ;	O4 - C23	- C34
24	54	124	1	1.0978e+02	4.2509e+02 ;	O4 - C23	- H49
24	54	125	1	1.0978e+02	4.2509e+02 ;	O4 - C23	- H51
25	17	79	1	1.0956e+02	3.8819e+02 ;	C27 - C13	- H20
25	17	80	1	1.0956e+02	3.8819e+02 ;	C27 - C13	- H21
26	25	85	1	1.0978e+02	4.2509e+02 ;	O5 - C27	- H33
26	25	86	1	1.0978e+02	4.2509e+02 ;	O5 - C27	- H34
27	28	29	1	1.0966e+02	5.2400e+02 ;	C20 - N2	- C22
27	28	89	1	1.1011e+02	3.8367e+02 ;	C20 - N2	- H30
27	28	90	1	1.1011e+02	3.8367e+02 ;	C20 - N2	- H32
27	32	33	1	1.1151e+02	5.2601e+02 ;	C20 - C29	- C31
27	32	34	1	1.1019e+02	5.6459e+02 ;	C20 - C29	- O6
27	32	98	1	1.0956e+02	3.8819e+02 ;	C20 - C29	- H50
28	27	32	1	1.1421e+02	5.3706e+02 ;	N2 - C20	- C29
28	27	87	1	1.0801e+02	4.0710e+02 ;	N2 - C20	- H26
28	27	88	1	1.0801e+02	4.0710e+02 ;	N2 - C20	- H28
28	29	30	1	1.1421e+02	5.3706e+02 ;	N2 - C22	- C24
28	29	31	1	1.1421e+02	5.3706e+02 ;	N2 - C22	- C26
28	29	91	1	1.0801e+02	4.0710e+02 ;	N2 - C22	- H36
29	28	89	1	1.1011e+02	3.8367e+02 ;	C22 - N2	- H30

29	28	90	1	1.1011e+02	3.8367e+02 ;	C22 - N2	- H32
29	30	92	1	1.0980e+02	3.8777e+02 ;	C22 - C24	- H38
29	30	93	1	1.0980e+02	3.8777e+02 ;	C22 - C24	- H40
29	30	94	1	1.0980e+02	3.8777e+02 ;	C22 - C24	- H42
29	31	95	1	1.0980e+02	3.8777e+02 ;	C22 - C26	- H44
29	31	96	1	1.0980e+02	3.8777e+02 ;	C22 - C26	- H46
29	31	97	1	1.0980e+02	3.8777e+02 ;	C22 - C26	- H48
30	29	31	1	1.1151e+02	5.2601e+02 ;	C24 - C22	- C26
30	29	91	1	1.1056e+02	3.8660e+02 ;	C24 - C22	- H36
31	29	91	1	1.1056e+02	3.8660e+02 ;	C26 - C22	- H36
32	27	87	1	1.1056e+02	3.8660e+02 ;	C29 - C20	- H26
32	27	88	1	1.1056e+02	3.8660e+02 ;	C29 - C20	- H28
32	33	35	1	1.0797e+02	5.6902e+02 ;	C29 - C31	- O7
32	33	99	1	1.0956e+02	3.8819e+02 ;	C29 - C31	- H52
32	33	100	1	1.0956e+02	3.8819e+02 ;	C29 - C31	- H54
32	34	101	1	1.0726e+02	3.9648e+02 ;	C29 - O6	- H56
33	32	34	1	1.1019e+02	5.6459e+02 ;	C31 - C29	- O6
33	32	98	1	1.0956e+02	3.8819e+02 ;	C31 - C29	- H50
33	35	36	1	1.1796e+02	5.2317e+02 ;	C31 - O7	- C33
34	32	98	1	1.1026e+02	4.2618e+02 ;	O6 - C29	- H50
35	33	99	1	1.0978e+02	4.2509e+02 ;	O7 - C31	- H52
35	33	100	1	1.0978e+02	4.2509e+02 ;	O7 - C31	- H54
35	36	37	1	1.1920e+02	5.8225e+02 ;	O7 - C33	- C35
35	36	41	1	1.1920e+02	5.8225e+02 ;	O7 - C33	- C39
36	37	38	1	1.2002e+02	5.5748e+02 ;	C33 - C35	- C36
36	37	102	1	1.1988e+02	4.0317e+02 ;	C33 - C35	- H57
36	41	40	1	1.2002e+02	5.5748e+02 ;	C33 - C39	- C38
36	41	105	1	1.1988e+02	4.0317e+02 ;	C33 - C39	- H60
37	36	41	1	1.2002e+02	5.5748e+02 ;	C35 - C33	- C39
37	38	39	1	1.2002e+02	5.5748e+02 ;	C35 - C36	- C37
37	38	103	1	1.1988e+02	4.0317e+02 ;	C35 - C36	- H58
38	37	102	1	1.1988e+02	4.0317e+02 ;	C36 - C35	- H57

38	39	40	1	1.2002e+02	5.5748e+02 ;	C36 - C37 - C38
38	39	42	1	1.1920e+02	5.8225e+02 ;	C36 - C37 - O8
39	38	103	1	1.1988e+02	4.0317e+02 ;	C37 - C36 - H58
39	40	41	1	1.2002e+02	5.5748e+02 ;	C37 - C38 - C39
39	40	104	1	1.1988e+02	4.0317e+02 ;	C37 - C38 - H59
39	42	43	1	1.1796e+02	5.2317e+02 ;	C37 - O8 - C40
40	39	42	1	1.1920e+02	5.8225e+02 ;	C38 - C37 - O8
40	41	105	1	1.1988e+02	4.0317e+02 ;	C38 - C39 - H60
41	40	104	1	1.1988e+02	4.0317e+02 ;	C39 - C38 - H59
42	43	51	1	1.0797e+02	5.6902e+02 ;	O8 - C40 - C47
42	43	106	1	1.0978e+02	4.2509e+02 ;	O8 - C40 - H61
42	43	107	1	1.0978e+02	4.2509e+02 ;	O8 - C40 - H62
43	51	52	1	1.0797e+02	5.6902e+02 ;	C40 - C47 - O10
43	51	112	1	1.0956e+02	3.8819e+02 ;	C40 - C47 - H67
43	51	113	1	1.0956e+02	3.8819e+02 ;	C40 - C47 - H68
44	45	46	1	1.2002e+02	5.5748e+02 ;	C41 - C42 - C43
44	45	108	1	1.1988e+02	4.0317e+02 ;	C41 - C42 - H63
44	49	48	1	1.2002e+02	5.5748e+02 ;	C41 - C46 - C45
44	49	111	1	1.1988e+02	4.0317e+02 ;	C41 - C46 - H66
44	52	51	1	1.1796e+02	5.2317e+02 ;	C41 - O10 - C47
45	44	49	1	1.2002e+02	5.5748e+02 ;	C42 - C41 - C46
45	44	52	1	1.1920e+02	5.8225e+02 ;	C42 - C41 - O10
45	46	47	1	1.2002e+02	5.5748e+02 ;	C42 - C43 - C44
45	46	109	1	1.1988e+02	4.0317e+02 ;	C42 - C43 - H64
46	45	108	1	1.1988e+02	4.0317e+02 ;	C43 - C42 - H63
46	47	48	1	1.2002e+02	5.5748e+02 ;	C43 - C44 - C45
46	47	50	1	1.1920e+02	5.8225e+02 ;	C43 - C44 - O9
47	46	109	1	1.1988e+02	4.0317e+02 ;	C44 - C43 - H64
47	48	49	1	1.2002e+02	5.5748e+02 ;	C44 - C45 - C46
47	48	110	1	1.1988e+02	4.0317e+02 ;	C44 - C45 - H65
47	50	53	1	1.1796e+02	5.2317e+02 ;	C44 - O9 - C21
48	47	50	1	1.1920e+02	5.8225e+02 ;	C45 - C44 - O9

48	49	111	1	1.1988e+02	4.0317e+02 ;	C45 - C46	- H66
49	44	52	1	1.1920e+02	5.8225e+02 ;	C46 - C41	- O10
49	48	110	1	1.1988e+02	4.0317e+02 ;	C46 - C45	- H65
50	53	55	1	1.0797e+02	5.6902e+02 ;	O9 - C21	- C25
50	53	114	1	1.0978e+02	4.2509e+02 ;	O9 - C21	- H27
50	53	115	1	1.0978e+02	4.2509e+02 ;	O9 - C21	- H29
51	43	106	1	1.0956e+02	3.8819e+02 ;	C47 - C40	- H61
51	43	107	1	1.0956e+02	3.8819e+02 ;	C47 - C40	- H62
52	51	112	1	1.0978e+02	4.2509e+02 ;	O10 - C47	- H67
52	51	113	1	1.0978e+02	4.2509e+02 ;	O10 - C47	- H68
53	55	56	1	1.1151e+02	5.2601e+02 ;	C21 - C25	- C28
53	55	116	1	1.0980e+02	3.8777e+02 ;	C21 - C25	- H31
53	55	117	1	1.0980e+02	3.8777e+02 ;	C21 - C25	- H35
54	59	58	1	1.1151e+02	5.2601e+02 ;	C23 - C34	- C32
54	59	126	1	1.0980e+02	3.8777e+02 ;	C23 - C34	- H53
54	59	127	1	1.0980e+02	3.8777e+02 ;	C23 - C34	- H55
55	53	114	1	1.0956e+02	3.8819e+02 ;	C25 - C21	- H27
55	53	115	1	1.0956e+02	3.8819e+02 ;	C25 - C21	- H29
55	56	57	1	1.1151e+02	5.2601e+02 ;	C25 - C28	- C30
55	56	118	1	1.0980e+02	3.8777e+02 ;	C25 - C28	- H37
55	56	119	1	1.0980e+02	3.8777e+02 ;	C25 - C28	- H39
56	55	116	1	1.0980e+02	3.8777e+02 ;	C28 - C25	- H31
56	55	117	1	1.0980e+02	3.8777e+02 ;	C28 - C25	- H35
56	57	58	1	1.1151e+02	5.2601e+02 ;	C28 - C30	- C32
56	57	120	1	1.0980e+02	3.8777e+02 ;	C28 - C30	- H41
56	57	121	1	1.0980e+02	3.8777e+02 ;	C28 - C30	- H43
57	56	118	1	1.0980e+02	3.8777e+02 ;	C30 - C28	- H37
57	56	119	1	1.0980e+02	3.8777e+02 ;	C30 - C28	- H39
57	58	59	1	1.1151e+02	5.2601e+02 ;	C30 - C32	- C34
57	58	122	1	1.0980e+02	3.8777e+02 ;	C30 - C32	- H45
57	58	123	1	1.0980e+02	3.8777e+02 ;	C30 - C32	- H47
58	57	120	1	1.0980e+02	3.8777e+02 ;	C32 - C30	- H41

58	57	121	1	1.0980e+02	3.8777e+02 ;	C32 - C30	- H43
58	59	126	1	1.0980e+02	3.8777e+02 ;	C32 - C34	- H53
58	59	127	1	1.0980e+02	3.8777e+02 ;	C32 - C34	- H55
59	54	124	1	1.0956e+02	3.8819e+02 ;	C34 - C23	- H49
59	54	125	1	1.0956e+02	3.8819e+02 ;	C34 - C23	- H51
59	58	122	1	1.0980e+02	3.8777e+02 ;	C34 - C32	- H45
59	58	123	1	1.0980e+02	3.8777e+02 ;	C34 - C32	- H47
60	1	61	1	1.0975e+02	3.2819e+02 ;	H1 - C1	- H2
62	2	63	1	1.0830e+02	3.3957e+02 ;	H3 - N1	- H4
65	4	66	1	1.0758e+02	3.2970e+02 ;	H6 - C3	- H7
65	4	67	1	1.0758e+02	3.2970e+02 ;	H6 - C3	- H8
66	4	67	1	1.0758e+02	3.2970e+02 ;	H7 - C3	- H8
68	5	69	1	1.0758e+02	3.2970e+02 ;	H9 - C4	- H10
68	5	70	1	1.0758e+02	3.2970e+02 ;	H9 - C4	- H11
69	5	70	1	1.0758e+02	3.2970e+02 ;	H10 - C4	- H11
72	7	73	1	1.0846e+02	3.2836e+02 ;	H13 - C6	- H14
79	17	80	1	1.0846e+02	3.2836e+02 ;	H20 - C13	- H21
85	25	86	1	1.0846e+02	3.2836e+02 ;	H33 - C27	- H34
87	27	88	1	1.0975e+02	3.2819e+02 ;	H26 - C20	- H28
89	28	90	1	1.0830e+02	3.3957e+02 ;	H30 - N2	- H32
92	30	93	1	1.0758e+02	3.2970e+02 ;	H38 - C24	- H40
92	30	94	1	1.0758e+02	3.2970e+02 ;	H38 - C24	- H42
93	30	94	1	1.0758e+02	3.2970e+02 ;	H40 - C24	- H42
95	31	96	1	1.0758e+02	3.2970e+02 ;	H44 - C26	- H46
95	31	97	1	1.0758e+02	3.2970e+02 ;	H44 - C26	- H48
96	31	97	1	1.0758e+02	3.2970e+02 ;	H46 - C26	- H48
99	33	100	1	1.0846e+02	3.2836e+02 ;	H52 - C31	- H54
106	43	107	1	1.0846e+02	3.2836e+02 ;	H61 - C40	- H62
112	51	113	1	1.0846e+02	3.2836e+02 ;	H67 - C47	- H68
114	53	115	1	1.0846e+02	3.2836e+02 ;	H27 - C21	- H29
116	55	117	1	1.0758e+02	3.2970e+02 ;	H31 - C25	- H35
118	56	119	1	1.0758e+02	3.2970e+02 ;	H37 - C28	- H39

```

120  57 121  1 1.0758e+02 3.2970e+02 ; H41 - C30 - H43
122  58 123  1 1.0758e+02 3.2970e+02 ; H45 - C32 - H47
124  54 125  1 1.0846e+02 3.2836e+02 ; H49 - C23 - H51
126  59 127  1 1.0758e+02 3.2970e+02 ; H53 - C34 - H55

```

[dihedrals] ; propers

; for gromacs 4.5 or higher, using funct 9

```

; i  j  k  l  func  phase  kd  pn
  1  2  3  4  9  0.00 0.65084 3 ; C1- N1- C2- C3
  1  2  3  5  9  0.00 0.65084 3 ; C1- N1- C2- C4
  1  2  3  64 9  0.00 0.65084 3 ; C1- N1- C2- H5
  1  6  7  9  9  0.00 0.65084 3 ; C1- C5- C6- O2
  1  6  7  72 9  0.00 0.65084 3 ; C1- C5- C6- H13
  1  6  7  73 9  0.00 0.65084 3 ; C1- C5- C6- H14
  1  6  8  74 9  0.00 0.66944 3 ; C1- C5- O1- H15
  1  6  8  74 9  0.00 1.04600 1 ; C1- C5- O1- H15
  2  1  6  7  9  0.00 0.65084 3 ; N1- C1- C5- C6
  2  1  6  8  9  0.00 0.60250 3 ; N1- C1- C5- O1
  2  1  6  8  9  0.00 5.43920 2 ; N1- C1- C5- O1
  2  1  6  71 9  0.00 0.65084 3 ; N1- C1- C5- H12
  2  3  4  65 9  0.00 0.65084 3 ; N1- C2- C3- H6
  2  3  4  66 9  0.00 0.65084 3 ; N1- C2- C3- H7
  2  3  4  67 9  0.00 0.65084 3 ; N1- C2- C3- H8
  2  3  5  68 9  0.00 0.65084 3 ; N1- C2- C4- H9
  2  3  5  69 9  0.00 0.65084 3 ; N1- C2- C4- H10
  2  3  5  70 9  0.00 0.65084 3 ; N1- C2- C4- H11
  4  3  2  62 9  0.00 0.65084 3 ; C3- C2- N1- H3
  4  3  2  63 9  0.00 0.65084 3 ; C3- C2- N1- H4
  4  3  5  68 9  0.00 0.66944 3 ; C3- C2- C4- H9
  4  3  5  69 9  0.00 0.66944 3 ; C3- C2- C4- H10
  4  3  5  70 9  0.00 0.66944 3 ; C3- C2- C4- H11
  5  3  2  62 9  0.00 0.65084 3 ; C4- C2- N1- H3

```

5 3 2 63 9 0.00 0.65084 3 ; C4- C2- N1- H4
5 3 4 65 9 0.00 0.66944 3 ; C4- C2- C3- H6
5 3 4 66 9 0.00 0.66944 3 ; C4- C2- C3- H7
5 3 4 67 9 0.00 0.66944 3 ; C4- C2- C3- H8
6 1 2 3 9 0.00 0.65084 3 ; C5- C1- N1- C2
6 1 2 62 9 0.00 0.65084 3 ; C5- C1- N1- H3
6 1 2 63 9 0.00 0.65084 3 ; C5- C1- N1- H4
6 7 9 10 9 0.00 1.60387 3 ; C5- C6- O2- C7
7 6 8 74 9 0.00 0.66944 3 ; C6- C5- O1- H15
7 6 8 74 9 0.00 1.04600 1 ; C6- C5- O1- H15
7 9 10 11 9 180.00 3.76560 2 ; C6- O2- C7- C8
7 9 10 15 9 180.00 3.76560 2 ; C6- O2- C7- C12
8 6 7 9 9 0.00 0.60250 3 ; O1- C5- C6- O2
8 6 7 9 9 0.00 4.91620 2 ; O1- C5- C6- O2
8 6 7 72 9 0.00 0.00000 0 ; O1- C5- C6- H13
8 6 7 72 9 0.00 1.04600 1 ; O1- C5- C6- H13
8 6 7 73 9 0.00 0.00000 0 ; O1- C5- C6- H14
8 6 7 73 9 0.00 1.04600 1 ; O1- C5- C6- H14
9 7 6 71 9 0.00 0.00000 0 ; O2- C6- C5- H12
9 7 6 71 9 0.00 1.04600 1 ; O2- C6- C5- H12
9 10 11 12 9 180.00 15.16700 2 ; O2- C7- C8- C9
9 10 11 75 9 180.00 15.16700 2 ; O2- C7- C8- H16
9 10 15 14 9 180.00 15.16700 2 ; O2- C7- C12- C11
9 10 15 78 9 180.00 15.16700 2 ; O2- C7- C12- H19
10 9 7 72 9 0.00 1.60387 3 ; C7- O2- C6- H13
10 9 7 73 9 0.00 1.60387 3 ; C7- O2- C6- H14
10 11 12 13 9 180.00 15.16700 2 ; C7- C8- C9- C10
10 11 12 76 9 180.00 15.16700 2 ; C7- C8- C9- H17
10 15 14 13 9 180.00 15.16700 2 ; C7- C12- C11- C10
10 15 14 77 9 180.00 15.16700 2 ; C7- C12- C11- H18
11 10 15 14 9 180.00 15.16700 2 ; C8- C7- C12- C11
11 10 15 78 9 180.00 15.16700 2 ; C8- C7- C12- H19

11 12 13 14 9 180.00 15.16700 2 ; C8- C9- C10- C11
 11 12 13 16 9 180.00 15.16700 2 ; C8- C9- C10- O3
 12 11 10 15 9 180.00 15.16700 2 ; C9- C8- C7- C12
 12 13 14 15 9 180.00 15.16700 2 ; C9- C10- C11- C12
 12 13 14 77 9 180.00 15.16700 2 ; C9- C10- C11- H18
 12 13 16 17 9 180.00 3.76560 2 ; C9- C10- O3- C13
 13 12 11 75 9 180.00 15.16700 2 ; C10- C9- C8- H16
 13 14 15 78 9 180.00 15.16700 2 ; C10- C11- C12- H19
 13 16 17 25 9 0.00 1.60387 3 ; C10- O3- C13- C27
 13 16 17 79 9 0.00 1.60387 3 ; C10- O3- C13- H20
 13 16 17 80 9 0.00 1.60387 3 ; C10- O3- C13- H21
 14 13 12 76 9 180.00 15.16700 2 ; C11- C10- C9- H17
 14 13 16 17 9 180.00 3.76560 2 ; C11- C10- O3- C13
 15 10 11 75 9 180.00 15.16700 2 ; C12- C7- C8- H16
 15 14 13 16 9 180.00 15.16700 2 ; C12- C11- C10- O3
 16 13 12 76 9 180.00 15.16700 2 ; O3- C10- C9- H17
 16 13 14 77 9 180.00 15.16700 2 ; O3- C10- C11- H18
 16 17 25 26 9 0.00 0.60250 3 ; O3- C13- C27- O5
 16 17 25 26 9 0.00 4.91620 2 ; O3- C13- C27- O5
 16 17 25 85 9 0.00 0.00000 0 ; O3- C13- C27- H33
 16 17 25 85 9 0.00 1.04600 1 ; O3- C13- C27- H33
 16 17 25 86 9 0.00 0.00000 0 ; O3- C13- C27- H34
 16 17 25 86 9 0.00 1.04600 1 ; O3- C13- C27- H34
 17 25 26 18 9 0.00 1.60387 3 ; C13- C27- O5- C14
 18 19 20 21 9 180.00 15.16700 2 ; C14- C15- C16- C17
 18 19 20 82 9 180.00 15.16700 2 ; C14- C15- C16- H23
 18 23 22 21 9 180.00 15.16700 2 ; C14- C19- C18- C17
 18 23 22 83 9 180.00 15.16700 2 ; C14- C19- C18- H24
 18 26 25 85 9 0.00 1.60387 3 ; C14- O5- C27- H33
 18 26 25 86 9 0.00 1.60387 3 ; C14- O5- C27- H34
 19 18 23 22 9 180.00 15.16700 2 ; C15- C14- C19- C18
 19 18 23 84 9 180.00 15.16700 2 ; C15- C14- C19- H25

19 18 26 25 9 180.00 3.76560 2 ; C15- C14- O5- C27
 19 20 21 22 9 180.00 15.16700 2 ; C15- C16- C17- C18
 19 20 21 24 9 180.00 15.16700 2 ; C15- C16- C17- O4
 20 19 18 23 9 180.00 15.16700 2 ; C16- C15- C14- C19
 20 19 18 26 9 180.00 15.16700 2 ; C16- C15- C14- O5
 20 21 22 23 9 180.00 15.16700 2 ; C16- C17- C18- C19
 20 21 22 83 9 180.00 15.16700 2 ; C16- C17- C18- H24
 20 21 24 54 9 180.00 3.76560 2 ; C16- C17- O4- C23
 21 20 19 81 9 180.00 15.16700 2 ; C17- C16- C15- H22
 21 22 23 84 9 180.00 15.16700 2 ; C17- C18- C19- H25
 21 24 54 59 9 0.00 1.60387 3 ; C17- O4- C23- C34
 21 24 54 124 9 0.00 1.60387 3 ; C17- O4- C23- H49
 21 24 54 125 9 0.00 1.60387 3 ; C17- O4- C23- H51
 22 21 20 82 9 180.00 15.16700 2 ; C18- C17- C16- H23
 22 21 24 54 9 180.00 3.76560 2 ; C18- C17- O4- C23
 22 23 18 26 9 180.00 15.16700 2 ; C18- C19- C14- O5
 23 18 19 81 9 180.00 15.16700 2 ; C19- C14- C15- H22
 23 18 26 25 9 180.00 3.76560 2 ; C19- C14- O5- C27
 23 22 21 24 9 180.00 15.16700 2 ; C19- C18- C17- O4
 24 21 20 82 9 180.00 15.16700 2 ; O4- C17- C16- H23
 24 21 22 83 9 180.00 15.16700 2 ; O4- C17- C18- H24
 24 54 59 58 9 0.00 0.65084 3 ; O4- C23- C34- C32
 24 54 59 126 9 0.00 0.00000 0 ; O4- C23- C34- H53
 24 54 59 126 9 0.00 1.04600 1 ; O4- C23- C34- H53
 24 54 59 127 9 0.00 0.00000 0 ; O4- C23- C34- H55
 24 54 59 127 9 0.00 1.04600 1 ; O4- C23- C34- H55
 26 18 19 81 9 180.00 15.16700 2 ; O5- C14- C15- H22
 26 18 23 84 9 180.00 15.16700 2 ; O5- C14- C19- H25
 26 25 17 79 9 0.00 0.00000 0 ; O5- C27- C13- H20
 26 25 17 79 9 0.00 1.04600 1 ; O5- C27- C13- H20
 26 25 17 80 9 0.00 0.00000 0 ; O5- C27- C13- H21
 26 25 17 80 9 0.00 1.04600 1 ; O5- C27- C13- H21

27	28	29	30	9	0.00	0.65084	3 ;	C20-	N2-	C22-	C24
27	28	29	31	9	0.00	0.65084	3 ;	C20-	N2-	C22-	C26
27	28	29	91	9	0.00	0.65084	3 ;	C20-	N2-	C22-	H36
27	32	33	35	9	0.00	0.65084	3 ;	C20-	C29-	C31-	O7
27	32	33	99	9	0.00	0.65084	3 ;	C20-	C29-	C31-	H52
27	32	33	100	9	0.00	0.65084	3 ;	C20-	C29-	C31-	H54
27	32	34	101	9	0.00	0.66944	3 ;	C20-	C29-	O6-	H56
27	32	34	101	9	0.00	1.04600	1 ;	C20-	C29-	O6-	H56
28	27	32	33	9	0.00	0.65084	3 ;	N2-	C20-	C29-	C31
28	27	32	34	9	0.00	0.60250	3 ;	N2-	C20-	C29-	O6
28	27	32	34	9	0.00	5.43920	2 ;	N2-	C20-	C29-	O6
28	27	32	98	9	0.00	0.65084	3 ;	N2-	C20-	C29-	H50
28	29	30	92	9	0.00	0.65084	3 ;	N2-	C22-	C24-	H38
28	29	30	93	9	0.00	0.65084	3 ;	N2-	C22-	C24-	H40
28	29	30	94	9	0.00	0.65084	3 ;	N2-	C22-	C24-	H42
28	29	31	95	9	0.00	0.65084	3 ;	N2-	C22-	C26-	H44
28	29	31	96	9	0.00	0.65084	3 ;	N2-	C22-	C26-	H46
28	29	31	97	9	0.00	0.65084	3 ;	N2-	C22-	C26-	H48
29	28	27	32	9	0.00	0.65084	3 ;	C22-	N2-	C20-	C29
29	28	27	87	9	0.00	0.65084	3 ;	C22-	N2-	C20-	H26
29	28	27	88	9	0.00	0.65084	3 ;	C22-	N2-	C20-	H28
30	29	28	89	9	0.00	0.65084	3 ;	C24-	C22-	N2-	H30
30	29	28	90	9	0.00	0.65084	3 ;	C24-	C22-	N2-	H32
30	29	31	95	9	0.00	0.66944	3 ;	C24-	C22-	C26-	H44
30	29	31	96	9	0.00	0.66944	3 ;	C24-	C22-	C26-	H46
30	29	31	97	9	0.00	0.66944	3 ;	C24-	C22-	C26-	H48
31	29	28	89	9	0.00	0.65084	3 ;	C26-	C22-	N2-	H30
31	29	28	90	9	0.00	0.65084	3 ;	C26-	C22-	N2-	H32
31	29	30	92	9	0.00	0.66944	3 ;	C26-	C22-	C24-	H38
31	29	30	93	9	0.00	0.66944	3 ;	C26-	C22-	C24-	H40
31	29	30	94	9	0.00	0.66944	3 ;	C26-	C22-	C24-	H42
32	27	28	89	9	0.00	0.65084	3 ;	C29-	C20-	N2-	H30

32 27 28 90 9 0.00 0.65084 3 ; C29- C20- N2- H32
 32 33 35 36 9 0.00 1.60387 3 ; C29- C31- O7- C33
 33 32 27 87 9 0.00 0.65084 3 ; C31- C29- C20- H26
 33 32 27 88 9 0.00 0.65084 3 ; C31- C29- C20- H28
 33 32 34 101 9 0.00 0.66944 3 ; C31- C29- O6- H56
 33 32 34 101 9 0.00 1.04600 1 ; C31- C29- O6- H56
 33 35 36 37 9 180.00 3.76560 2 ; C31- O7- C33- C35
 33 35 36 41 9 180.00 3.76560 2 ; C31- O7- C33- C39
 34 32 27 87 9 0.00 0.65084 3 ; O6- C29- C20- H26
 34 32 27 88 9 0.00 0.65084 3 ; O6- C29- C20- H28
 34 32 33 35 9 0.00 0.60250 3 ; O6- C29- C31- O7
 34 32 33 35 9 0.00 4.91620 2 ; O6- C29- C31- O7
 34 32 33 99 9 0.00 0.00000 0 ; O6- C29- C31- H52
 34 32 33 99 9 0.00 1.04600 1 ; O6- C29- C31- H52
 34 32 33 100 9 0.00 0.00000 0 ; O6- C29- C31- H54
 34 32 33 100 9 0.00 1.04600 1 ; O6- C29- C31- H54
 35 33 32 98 9 0.00 0.00000 0 ; O7- C31- C29- H50
 35 33 32 98 9 0.00 1.04600 1 ; O7- C31- C29- H50
 35 36 37 38 9 180.00 15.16700 2 ; O7- C33- C35- C36
 35 36 37 102 9 180.00 15.16700 2 ; O7- C33- C35- H57
 35 36 41 40 9 180.00 15.16700 2 ; O7- C33- C39- C38
 35 36 41 105 9 180.00 15.16700 2 ; O7- C33- C39- H60
 36 35 33 99 9 0.00 1.60387 3 ; C33- O7- C31- H52
 36 35 33 100 9 0.00 1.60387 3 ; C33- O7- C31- H54
 36 37 38 39 9 180.00 15.16700 2 ; C33- C35- C36- C37
 36 37 38 103 9 180.00 15.16700 2 ; C33- C35- C36- H58
 36 41 40 39 9 180.00 15.16700 2 ; C33- C39- C38- C37
 36 41 40 104 9 180.00 15.16700 2 ; C33- C39- C38- H59
 37 36 41 40 9 180.00 15.16700 2 ; C35- C33- C39- C38
 37 36 41 105 9 180.00 15.16700 2 ; C35- C33- C39- H60
 37 38 39 40 9 180.00 15.16700 2 ; C35- C36- C37- C38
 37 38 39 42 9 180.00 15.16700 2 ; C35- C36- C37- O8

38 37 36 41 9 180.00 15.16700 2 ; C36- C35- C33- C39
 38 39 40 41 9 180.00 15.16700 2 ; C36- C37- C38- C39
 38 39 40 104 9 180.00 15.16700 2 ; C36- C37- C38- H59
 38 39 42 43 9 180.00 3.76560 2 ; C36- C37- O8- C40
 39 38 37 102 9 180.00 15.16700 2 ; C37- C36- C35- H57
 39 40 41 105 9 180.00 15.16700 2 ; C37- C38- C39- H60
 39 42 43 51 9 0.00 1.60387 3 ; C37- O8- C40- C47
 39 42 43 106 9 0.00 1.60387 3 ; C37- O8- C40- H61
 39 42 43 107 9 0.00 1.60387 3 ; C37- O8- C40- H62
 40 39 38 103 9 180.00 15.16700 2 ; C38- C37- C36- H58
 40 39 42 43 9 180.00 3.76560 2 ; C38- C37- O8- C40
 41 36 37 102 9 180.00 15.16700 2 ; C39- C33- C35- H57
 41 40 39 42 9 180.00 15.16700 2 ; C39- C38- C37- O8
 42 39 38 103 9 180.00 15.16700 2 ; O8- C37- C36- H58
 42 39 40 104 9 180.00 15.16700 2 ; O8- C37- C38- H59
 42 43 51 52 9 0.00 0.60250 3 ; O8- C40- C47- O10
 42 43 51 52 9 0.00 4.91620 2 ; O8- C40- C47- O10
 42 43 51 112 9 0.00 0.00000 0 ; O8- C40- C47- H67
 42 43 51 112 9 0.00 1.04600 1 ; O8- C40- C47- H67
 42 43 51 113 9 0.00 0.00000 0 ; O8- C40- C47- H68
 42 43 51 113 9 0.00 1.04600 1 ; O8- C40- C47- H68
 43 51 52 44 9 0.00 1.60387 3 ; C40- C47- O10- C41
 44 45 46 47 9 180.00 15.16700 2 ; C41- C42- C43- C44
 44 45 46 109 9 180.00 15.16700 2 ; C41- C42- C43- H64
 44 49 48 47 9 180.00 15.16700 2 ; C41- C46- C45- C44
 44 49 48 110 9 180.00 15.16700 2 ; C41- C46- C45- H65
 44 52 51 112 9 0.00 1.60387 3 ; C41- O10- C47- H67
 44 52 51 113 9 0.00 1.60387 3 ; C41- O10- C47- H68
 45 44 49 48 9 180.00 15.16700 2 ; C42- C41- C46- C45
 45 44 49 111 9 180.00 15.16700 2 ; C42- C41- C46- H66
 45 44 52 51 9 180.00 3.76560 2 ; C42- C41- O10- C47
 45 46 47 48 9 180.00 15.16700 2 ; C42- C43- C44- C45

45 46 47 50 9 180.00 15.16700 2 ; C42- C43- C44- O9
 46 45 44 49 9 180.00 15.16700 2 ; C43- C42- C41- C46
 46 45 44 52 9 180.00 15.16700 2 ; C43- C42- C41- O10
 46 47 48 49 9 180.00 15.16700 2 ; C43- C44- C45- C46
 46 47 48 110 9 180.00 15.16700 2 ; C43- C44- C45- H65
 46 47 50 53 9 180.00 3.76560 2 ; C43- C44- O9- C21
 47 46 45 108 9 180.00 15.16700 2 ; C44- C43- C42- H63
 47 48 49 111 9 180.00 15.16700 2 ; C44- C45- C46- H66
 47 50 53 55 9 0.00 1.60387 3 ; C44- O9- C21- C25
 47 50 53 114 9 0.00 1.60387 3 ; C44- O9- C21- H27
 47 50 53 115 9 0.00 1.60387 3 ; C44- O9- C21- H29
 48 47 46 109 9 180.00 15.16700 2 ; C45- C44- C43- H64
 48 47 50 53 9 180.00 3.76560 2 ; C45- C44- O9- C21
 48 49 44 52 9 180.00 15.16700 2 ; C45- C46- C41- O10
 49 44 45 108 9 180.00 15.16700 2 ; C46- C41- C42- H63
 49 44 52 51 9 180.00 3.76560 2 ; C46- C41- O10- C47
 49 48 47 50 9 180.00 15.16700 2 ; C46- C45- C44- O9
 50 47 46 109 9 180.00 15.16700 2 ; O9- C44- C43- H64
 50 47 48 110 9 180.00 15.16700 2 ; O9- C44- C45- H65
 50 53 55 56 9 0.00 0.65084 3 ; O9- C21- C25- C28
 50 53 55 116 9 0.00 0.00000 0 ; O9- C21- C25- H31
 50 53 55 116 9 0.00 1.04600 1 ; O9- C21- C25- H31
 50 53 55 117 9 0.00 0.00000 0 ; O9- C21- C25- H35
 50 53 55 117 9 0.00 1.04600 1 ; O9- C21- C25- H35
 52 44 45 108 9 180.00 15.16700 2 ; O10- C41- C42- H63
 52 44 49 111 9 180.00 15.16700 2 ; O10- C41- C46- H66
 52 51 43 106 9 0.00 0.00000 0 ; O10- C47- C40- H61
 52 51 43 106 9 0.00 1.04600 1 ; O10- C47- C40- H61
 52 51 43 107 9 0.00 0.00000 0 ; O10- C47- C40- H62
 52 51 43 107 9 0.00 1.04600 1 ; O10- C47- C40- H62
 53 55 56 57 9 0.00 0.75312 3 ; C21- C25- C28- C30
 53 55 56 57 9 180.00 0.83680 1 ; C21- C25- C28- C30

53 55 56 57 9 180.00 1.04600 2 ; C21- C25- C28- C30
53 55 56 118 9 0.00 0.66944 3 ; C21- C25- C28- H37
53 55 56 119 9 0.00 0.66944 3 ; C21- C25- C28- H39
54 59 58 57 9 0.00 0.75312 3 ; C23- C34- C32- C30
54 59 58 57 9 180.00 0.83680 1 ; C23- C34- C32- C30
54 59 58 57 9 180.00 1.04600 2 ; C23- C34- C32- C30
54 59 58 122 9 0.00 0.66944 3 ; C23- C34- C32- H45
54 59 58 123 9 0.00 0.66944 3 ; C23- C34- C32- H47
55 56 57 58 9 0.00 0.75312 3 ; C25- C28- C30- C32
55 56 57 58 9 180.00 0.83680 1 ; C25- C28- C30- C32
55 56 57 58 9 180.00 1.04600 2 ; C25- C28- C30- C32
55 56 57 120 9 0.00 0.66944 3 ; C25- C28- C30- H41
55 56 57 121 9 0.00 0.66944 3 ; C25- C28- C30- H43
56 55 53 114 9 0.00 0.65084 3 ; C28- C25- C21- H27
56 55 53 115 9 0.00 0.65084 3 ; C28- C25- C21- H29
56 57 58 59 9 0.00 0.75312 3 ; C28- C30- C32- C34
56 57 58 59 9 180.00 0.83680 1 ; C28- C30- C32- C34
56 57 58 59 9 180.00 1.04600 2 ; C28- C30- C32- C34
56 57 58 122 9 0.00 0.66944 3 ; C28- C30- C32- H45
56 57 58 123 9 0.00 0.66944 3 ; C28- C30- C32- H47
57 56 55 116 9 0.00 0.66944 3 ; C30- C28- C25- H31
57 56 55 117 9 0.00 0.66944 3 ; C30- C28- C25- H35
57 58 59 126 9 0.00 0.66944 3 ; C30- C32- C34- H53
57 58 59 127 9 0.00 0.66944 3 ; C30- C32- C34- H55
58 57 56 118 9 0.00 0.66944 3 ; C32- C30- C28- H37
58 57 56 119 9 0.00 0.66944 3 ; C32- C30- C28- H39
58 59 54 124 9 0.00 0.65084 3 ; C32- C34- C23- H49
58 59 54 125 9 0.00 0.65084 3 ; C32- C34- C23- H51
59 58 57 120 9 0.00 0.66944 3 ; C34- C32- C30- H41
59 58 57 121 9 0.00 0.66944 3 ; C34- C32- C30- H43
60 1 2 3 9 0.00 0.65084 3 ; H1- C1- N1- C2
60 1 2 62 9 0.00 0.65084 3 ; H1- C1- N1- H3

60 1 2 63 9 0.00 0.65084 3 ; H1- C1- N1- H4
60 1 6 7 9 0.00 0.65084 3 ; H1- C1- C5- C6
60 1 6 8 9 0.00 0.65084 3 ; H1- C1- C5- O1
60 1 6 71 9 0.00 0.65084 3 ; H1- C1- C5- H12
61 1 2 3 9 0.00 0.65084 3 ; H2- C1- N1- C2
61 1 2 62 9 0.00 0.65084 3 ; H2- C1- N1- H3
61 1 2 63 9 0.00 0.65084 3 ; H2- C1- N1- H4
61 1 6 7 9 0.00 0.65084 3 ; H2- C1- C5- C6
61 1 6 8 9 0.00 0.65084 3 ; H2- C1- C5- O1
61 1 6 71 9 0.00 0.65084 3 ; H2- C1- C5- H12
62 2 3 64 9 0.00 0.65084 3 ; H3- N1- C2- H5
63 2 3 64 9 0.00 0.65084 3 ; H4- N1- C2- H5
64 3 4 65 9 0.00 0.65084 3 ; H5- C2- C3- H6
64 3 4 66 9 0.00 0.65084 3 ; H5- C2- C3- H7
64 3 4 67 9 0.00 0.65084 3 ; H5- C2- C3- H8
64 3 5 68 9 0.00 0.65084 3 ; H5- C2- C4- H9
64 3 5 69 9 0.00 0.65084 3 ; H5- C2- C4- H10
64 3 5 70 9 0.00 0.65084 3 ; H5- C2- C4- H11
71 6 7 72 9 0.00 0.65084 3 ; H12- C5- C6- H13
71 6 7 73 9 0.00 0.65084 3 ; H12- C5- C6- H14
71 6 8 74 9 0.00 0.69733 3 ; H12- C5- O1- H15
75 11 12 76 9 180.00 15.16700 2 ; H16- C8- C9- H17
77 14 15 78 9 180.00 15.16700 2 ; H18- C11- C12- H19
79 17 25 85 9 0.00 0.65084 3 ; H20- C13- C27- H33
79 17 25 86 9 0.00 0.65084 3 ; H20- C13- C27- H34
80 17 25 85 9 0.00 0.65084 3 ; H21- C13- C27- H33
80 17 25 86 9 0.00 0.65084 3 ; H21- C13- C27- H34
81 19 20 82 9 180.00 15.16700 2 ; H22- C15- C16- H23
83 22 23 84 9 180.00 15.16700 2 ; H24- C18- C19- H25
87 27 28 89 9 0.00 0.65084 3 ; H26- C20- N2- H30
87 27 28 90 9 0.00 0.65084 3 ; H26- C20- N2- H32
87 27 32 98 9 0.00 0.65084 3 ; H26- C20- C29- H50

88	27	28	89	9	0.00	0.65084	3	; H28-	C20-	N2-	H30
88	27	28	90	9	0.00	0.65084	3	; H28-	C20-	N2-	H32
88	27	32	98	9	0.00	0.65084	3	; H28-	C20-	C29-	H50
89	28	29	91	9	0.00	0.65084	3	; H30-	N2-	C22-	H36
90	28	29	91	9	0.00	0.65084	3	; H32-	N2-	C22-	H36
91	29	30	92	9	0.00	0.65084	3	; H36-	C22-	C24-	H38
91	29	30	93	9	0.00	0.65084	3	; H36-	C22-	C24-	H40
91	29	30	94	9	0.00	0.65084	3	; H36-	C22-	C24-	H42
91	29	31	95	9	0.00	0.65084	3	; H36-	C22-	C26-	H44
91	29	31	96	9	0.00	0.65084	3	; H36-	C22-	C26-	H46
91	29	31	97	9	0.00	0.65084	3	; H36-	C22-	C26-	H48
98	32	33	99	9	0.00	0.65084	3	; H50-	C29-	C31-	H52
98	32	33	100	9	0.00	0.65084	3	; H50-	C29-	C31-	H54
98	32	34	101	9	0.00	0.69733	3	; H50-	C29-	O6-	H56
102	37	38	103	9	180.00	15.16700	2	; H57-	C35-	C36-	H58
104	40	41	105	9	180.00	15.16700	2	; H59-	C38-	C39-	H60
106	43	51	112	9	0.00	0.65084	3	; H61-	C40-	C47-	H67
106	43	51	113	9	0.00	0.65084	3	; H61-	C40-	C47-	H68
107	43	51	112	9	0.00	0.65084	3	; H62-	C40-	C47-	H67
107	43	51	113	9	0.00	0.65084	3	; H62-	C40-	C47-	H68
108	45	46	109	9	180.00	15.16700	2	; H63-	C42-	C43-	H64
110	48	49	111	9	180.00	15.16700	2	; H65-	C45-	C46-	H66
114	53	55	116	9	0.00	0.65084	3	; H27-	C21-	C25-	H31
114	53	55	117	9	0.00	0.65084	3	; H27-	C21-	C25-	H35
115	53	55	116	9	0.00	0.65084	3	; H29-	C21-	C25-	H31
115	53	55	117	9	0.00	0.65084	3	; H29-	C21-	C25-	H35
116	55	56	118	9	0.00	0.62760	3	; H31-	C25-	C28-	H37
116	55	56	119	9	0.00	0.62760	3	; H31-	C25-	C28-	H39
117	55	56	118	9	0.00	0.62760	3	; H35-	C25-	C28-	H37
117	55	56	119	9	0.00	0.62760	3	; H35-	C25-	C28-	H39
118	56	57	120	9	0.00	0.62760	3	; H37-	C28-	C30-	H41
118	56	57	121	9	0.00	0.62760	3	; H37-	C28-	C30-	H43

119	56	57	120	9	0.00	0.62760	3	; H39- C28- C30- H41
119	56	57	121	9	0.00	0.62760	3	; H39- C28- C30- H43
120	57	58	122	9	0.00	0.62760	3	; H41- C30- C32- H45
120	57	58	123	9	0.00	0.62760	3	; H41- C30- C32- H47
121	57	58	122	9	0.00	0.62760	3	; H43- C30- C32- H45
121	57	58	123	9	0.00	0.62760	3	; H43- C30- C32- H47
122	58	59	126	9	0.00	0.62760	3	; H45- C32- C34- H53
122	58	59	127	9	0.00	0.62760	3	; H45- C32- C34- H55
123	58	59	126	9	0.00	0.62760	3	; H47- C32- C34- H53
123	58	59	127	9	0.00	0.62760	3	; H47- C32- C34- H55
124	54	59	126	9	0.00	0.65084	3	; H49- C23- C34- H53
124	54	59	127	9	0.00	0.65084	3	; H49- C23- C34- H55
125	54	59	126	9	0.00	0.65084	3	; H51- C23- C34- H53
125	54	59	127	9	0.00	0.65084	3	; H51- C23- C34- H55

[dihedrals] ; impropers

; treated as propers in GROMACS to use correct AMBER analytical function

; i j k l func phase kd pn

10	12	11	75	4	180.00	4.60240	2	; C7- C9- C8- H16
10	14	15	78	4	180.00	4.60240	2	; C7- C11- C12- H19
11	13	12	76	4	180.00	4.60240	2	; C8- C10- C9- H17
11	15	10	9	4	180.00	4.60240	2	; C8- C12- C7- O2
12	14	13	16	4	180.00	4.60240	2	; C9- C11- C10- O3
13	15	14	77	4	180.00	4.60240	2	; C10- C12- C11- H18
18	20	19	81	4	180.00	4.60240	2	; C14- C16- C15- H22
18	22	23	84	4	180.00	4.60240	2	; C14- C18- C19- H25
19	21	20	82	4	180.00	4.60240	2	; C15- C17- C16- H23
19	23	18	26	4	180.00	4.60240	2	; C15- C19- C14- O5
20	22	21	24	4	180.00	4.60240	2	; C16- C18- C17- O4
21	23	22	83	4	180.00	4.60240	2	; C17- C19- C18- H24
36	38	37	102	4	180.00	4.60240	2	; C33- C36- C35- H57
36	40	41	105	4	180.00	4.60240	2	; C33- C38- C39- H60

37 39 38 103 4 180.00 4.60240 2 ; C35- C37- C36- H58
37 41 36 35 4 180.00 4.60240 2 ; C35- C39- C33- O7
38 40 39 42 4 180.00 4.60240 2 ; C36- C38- C37- O8
39 41 40 104 4 180.00 4.60240 2 ; C37- C39- C38- H59
44 46 45 108 4 180.00 4.60240 2 ; C41- C43- C42- H63
44 48 49 111 4 180.00 4.60240 2 ; C41- C45- C46- H66
45 47 46 109 4 180.00 4.60240 2 ; C42- C44- C43- H64
45 49 44 52 4 180.00 4.60240 2 ; C42- C46- C41- O10
46 48 47 50 4 180.00 4.60240 2 ; C43- C45- C44- O9
47 49 48 110 4 180.00 4.60240 2 ; C44- C46- C45- H65

Appendix 4 - Ligand **59** AMBERff99SB parameters in GROMACS
.itp format

; GBS004_linker_GMX.itp created by acpype (v: 2019-11-07T23:16:00CET) on Thu
Dec 19 11:42:24 2019

[atomtypes]

;name	bond_type	mass	charge	ptype	sigma	epsilon	Amb
c3	c3	0.00000	0.00000	A	3.39967e-01	4.57730e-01 ; 1.91	0.1094
n4	n4	0.00000	0.00000	A	3.25000e-01	7.11280e-01 ; 1.82	0.1700
oh	oh	0.00000	0.00000	A	3.06647e-01	8.80314e-01 ; 1.72	0.2104
os	os	0.00000	0.00000	A	3.00001e-01	7.11280e-01 ; 1.68	0.1700
ca	ca	0.00000	0.00000	A	3.39967e-01	3.59824e-01 ; 1.91	0.0860
hx	hx	0.00000	0.00000	A	1.95998e-01	6.56888e-02 ; 1.10	0.0157
hn	hn	0.00000	0.00000	A	1.06908e-01	6.56888e-02 ; 0.60	0.0157
hc	hc	0.00000	0.00000	A	2.64953e-01	6.56888e-02 ; 1.49	0.0157
h1	h1	0.00000	0.00000	A	2.47135e-01	6.56888e-02 ; 1.39	0.0157
ho	ho	0.00000	0.00000	A	0.00000e+00	0.00000e+00 ; 0.00	0.0000
ha	ha	0.00000	0.00000	A	2.59964e-01	6.27600e-02 ; 1.46	0.0150

[moleculetype]

;name nrexcl
GBS004_linker 3

[atoms]

; nr	type	resi	res	atom	cgmr	charge	mass	; qtot	bond_type
1	c3	1	UNK	C1	1	0.100300	12.01000	; qtot	0.100
2	n4	1	UNK	N1	2	-0.763004	14.01000	; qtot	-0.663
3	c3	1	UNK	C2	3	0.133000	12.01000	; qtot	-0.530
4	c3	1	UNK	C3	4	-0.134600	12.01000	; qtot	-0.664

5	c3	1	UNK	C4	5	-0.134600	12.01000 ; qtot -0.799
6	c3	1	UNK	C5	6	0.122600	12.01000 ; qtot -0.676
7	c3	1	UNK	C6	7	0.097400	12.01000 ; qtot -0.579
8	oh	1	UNK	O1	8	-0.633301	16.00000 ; qtot -1.212
9	os	1	UNK	O2	9	-0.357900	16.00000 ; qtot -1.570
10	ca	1	UNK	C7	10	0.021100	12.01000 ; qtot -1.549
11	ca	1	UNK	C8	11	-0.100000	12.01000 ; qtot -1.649
12	ca	1	UNK	C9	12	-0.166250	12.01000 ; qtot -1.815
13	ca	1	UNK	C10	13	0.143600	12.01000 ; qtot -1.672
14	ca	1	UNK	C11	14	-0.166250	12.01000 ; qtot -1.838
15	ca	1	UNK	C12	15	-0.100000	12.01000 ; qtot -1.938
16	os	1	UNK	O3	16	-0.323400	16.00000 ; qtot -2.261
17	c3	1	UNK	C13	17	0.121400	12.01000 ; qtot -2.140
18	ca	1	UNK	C14	18	0.071100	12.01000 ; qtot -2.069
19	ca	1	UNK	C15	19	-0.129250	12.01000 ; qtot -2.198
20	ca	1	UNK	C16	20	-0.154000	12.01000 ; qtot -2.352
21	ca	1	UNK	C17	21	0.110100	12.01000 ; qtot -2.242
22	ca	1	UNK	C18	22	-0.154000	12.01000 ; qtot -2.396
23	ca	1	UNK	C19	23	-0.129250	12.01000 ; qtot -2.525
24	os	1	UNK	O4	24	-0.325900	16.00000 ; qtot -2.851
25	c3	1	UNK	C20	25	0.173200	12.01000 ; qtot -2.678
26	c3	1	UNK	C27	26	0.111900	12.01000 ; qtot -2.566
27	os	1	UNK	O5	27	-0.347900	16.00000 ; qtot -2.914
28	c3	1	UNK	C21	28	0.100300	12.01000 ; qtot -2.814
29	n4	1	UNK	N2	29	-0.763004	14.01000 ; qtot -3.577
30	c3	1	UNK	C23	30	0.133000	12.01000 ; qtot -3.444
31	c3	1	UNK	C25	31	-0.134600	12.01000 ; qtot -3.578
32	c3	1	UNK	C28	32	-0.134600	12.01000 ; qtot -3.713
33	c3	1	UNK	C30	33	0.122600	12.01000 ; qtot -3.590
34	c3	1	UNK	C32	34	0.097400	12.01000 ; qtot -3.493
35	oh	1	UNK	O6	35	-0.633301	16.00000 ; qtot -4.126
36	os	1	UNK	O7	36	-0.357900	16.00000 ; qtot -4.484

37	ca	1	UNK	C34	37	0.021100	12.01000 ; qtot -4.463
38	ca	1	UNK	C36	38	-0.100000	12.01000 ; qtot -4.563
39	ca	1	UNK	C37	39	-0.166250	12.01000 ; qtot -4.729
40	ca	1	UNK	C38	40	0.143600	12.01000 ; qtot -4.586
41	ca	1	UNK	C39	41	-0.166250	12.01000 ; qtot -4.752
42	ca	1	UNK	C40	42	-0.100000	12.01000 ; qtot -4.852
43	os	1	UNK	O8	43	-0.323400	16.00000 ; qtot -5.175
44	c3	1	UNK	C41	44	0.121400	12.01000 ; qtot -5.054
45	ca	1	UNK	C42	45	0.071100	12.01000 ; qtot -4.983
46	ca	1	UNK	C43	46	-0.129250	12.01000 ; qtot -5.112
47	ca	1	UNK	C44	47	-0.154000	12.01000 ; qtot -5.266
48	ca	1	UNK	C45	48	0.110100	12.01000 ; qtot -5.156
49	ca	1	UNK	C46	49	-0.154000	12.01000 ; qtot -5.310
50	ca	1	UNK	C47	50	-0.129250	12.01000 ; qtot -5.439
51	os	1	UNK	O9	51	-0.325900	16.00000 ; qtot -5.765
52	c3	1	UNK	C48	52	0.111900	12.01000 ; qtot -5.653
53	os	1	UNK	O10	53	-0.347900	16.00000 ; qtot -6.001
54	c3	1	UNK	C22	54	0.173200	12.01000 ; qtot -5.828
55	ca	1	UNK	C24	55	-0.105300	12.01000 ; qtot -5.933
56	ca	1	UNK	C26	56	-0.113500	12.01000 ; qtot -6.047
57	ca	1	UNK	C29	57	-0.113500	12.01000 ; qtot -6.160
58	ca	1	UNK	C31	58	-0.105300	12.01000 ; qtot -6.265
59	ca	1	UNK	C33	59	-0.113500	12.01000 ; qtot -6.379
60	ca	1	UNK	C35	60	-0.113500	12.01000 ; qtot -6.492
61	hx	1	UNK	H1	61	0.116200	1.00800 ; qtot -6.376
62	hx	1	UNK	H2	62	0.116200	1.00800 ; qtot -6.260
63	hn	1	UNK	H3	63	0.457300	1.00800 ; qtot -5.803
64	hn	1	UNK	H4	64	0.457300	1.00800 ; qtot -5.345
65	hx	1	UNK	H5	65	0.107200	1.00800 ; qtot -5.238
66	hc	1	UNK	H6	66	0.078617	1.00800 ; qtot -5.160
67	hc	1	UNK	H7	67	0.078617	1.00800 ; qtot -5.081
68	hc	1	UNK	H8	68	0.078617	1.00800 ; qtot -5.002

69	hc	1	UNK	H9	69	0.078617	1.00800 ; qtot -4.924
70	hc	1	UNK	H10	70	0.078617	1.00800 ; qtot -4.845
71	hc	1	UNK	H11	71	0.078617	1.00800 ; qtot -4.767
72	h1	1	UNK	H12	72	0.073700	1.00800 ; qtot -4.693
73	h1	1	UNK	H13	73	0.061450	1.00800 ; qtot -4.631
74	h1	1	UNK	H14	74	0.061450	1.00800 ; qtot -4.570
75	ho	1	UNK	H15	75	0.460000	1.00800 ; qtot -4.110
76	ha	1	UNK	H16	76	0.143250	1.00800 ; qtot -3.967
77	ha	1	UNK	H17	77	0.163500	1.00800 ; qtot -3.803
78	ha	1	UNK	H18	78	0.163500	1.00800 ; qtot -3.640
79	ha	1	UNK	H19	79	0.143250	1.00800 ; qtot -3.496
80	h1	1	UNK	H20	80	0.070450	1.00800 ; qtot -3.426
81	h1	1	UNK	H21	81	0.070450	1.00800 ; qtot -3.356
82	ha	1	UNK	H22	82	0.144750	1.00800 ; qtot -3.211
83	ha	1	UNK	H23	83	0.149750	1.00800 ; qtot -3.061
84	ha	1	UNK	H24	84	0.149750	1.00800 ; qtot -2.911
85	ha	1	UNK	H25	85	0.144750	1.00800 ; qtot -2.767
86	h1	1	UNK	H33	86	0.058950	1.00800 ; qtot -2.708
87	h1	1	UNK	H34	87	0.058950	1.00800 ; qtot -2.649
88	hx	1	UNK	H26	88	0.116200	1.00800 ; qtot -2.532
89	hx	1	UNK	H28	89	0.116200	1.00800 ; qtot -2.416
90	hn	1	UNK	H30	90	0.457300	1.00800 ; qtot -1.959
91	hn	1	UNK	H32	91	0.457300	1.00800 ; qtot -1.502
92	hx	1	UNK	H36	92	0.107200	1.00800 ; qtot -1.394
93	hc	1	UNK	H38	93	0.078617	1.00800 ; qtot -1.316
94	hc	1	UNK	H40	94	0.078617	1.00800 ; qtot -1.237
95	hc	1	UNK	H41	95	0.078617	1.00800 ; qtot -1.159
96	hc	1	UNK	H42	96	0.078617	1.00800 ; qtot -1.080
97	hc	1	UNK	H43	97	0.078617	1.00800 ; qtot -1.001
98	hc	1	UNK	H44	98	0.078617	1.00800 ; qtot -0.923
99	h1	1	UNK	H45	99	0.073700	1.00800 ; qtot -0.849
100	h1	1	UNK	H46	100	0.061450	1.00800 ; qtot -0.788

101	h1	1	UNK	H47	101	0.061450	1.00800 ; qtot -0.726
102	ho	1	UNK	H48	102	0.460000	1.00800 ; qtot -0.266
103	ha	1	UNK	H49	103	0.143250	1.00800 ; qtot -0.123
104	ha	1	UNK	H50	104	0.163500	1.00800 ; qtot 0.041
105	ha	1	UNK	H51	105	0.163500	1.00800 ; qtot 0.204
106	ha	1	UNK	H52	106	0.143250	1.00800 ; qtot 0.347
107	h1	1	UNK	H53	107	0.070450	1.00800 ; qtot 0.418
108	h1	1	UNK	H54	108	0.070450	1.00800 ; qtot 0.488
109	ha	1	UNK	H55	109	0.144750	1.00800 ; qtot 0.633
110	ha	1	UNK	H56	110	0.149750	1.00800 ; qtot 0.783
111	ha	1	UNK	H57	111	0.149750	1.00800 ; qtot 0.933
112	ha	1	UNK	H58	112	0.144750	1.00800 ; qtot 1.077
113	h1	1	UNK	H59	113	0.058950	1.00800 ; qtot 1.136
114	h1	1	UNK	H60	114	0.058950	1.00800 ; qtot 1.195
115	h1	1	UNK	H27	115	0.058950	1.00800 ; qtot 1.254
116	h1	1	UNK	H29	116	0.058950	1.00800 ; qtot 1.313
117	ha	1	UNK	H31	117	0.142250	1.00800 ; qtot 1.455
118	ha	1	UNK	H35	118	0.142250	1.00800 ; qtot 1.598
119	ha	1	UNK	H37	119	0.142250	1.00800 ; qtot 1.740
120	ha	1	UNK	H39	120	0.142250	1.00800 ; qtot 1.882
121	h1	1	UNK	H201	121	0.058950	1.00800 ; qtot 1.941
122	h1	1	UNK	H202	122	0.058950	1.00800 ; qtot 2.000

[bonds]

	ai	aj	funct	r	k	
	1	2	1	1.5110e-01	2.3707e+05	C1 - N1
	1	6	1	1.5375e-01	2.5179e+05	C1 - C5
	1	61	1	1.0910e-01	2.8342e+05	C1 - H1
	1	62	1	1.0910e-01	2.8342e+05	C1 - H2
	2	3	1	1.5110e-01	2.3707e+05	N1 - C2
	2	63	1	1.0304e-01	3.1229e+05	N1 - H3
	2	64	1	1.0304e-01	3.1229e+05	N1 - H4

3	4	1	1.5375e-01	2.5179e+05 ;	C2 - C3
3	5	1	1.5375e-01	2.5179e+05 ;	C2 - C4
3	65	1	1.0910e-01	2.8342e+05 ;	C2 - H5
4	66	1	1.0969e-01	2.7665e+05 ;	C3 - H6
4	67	1	1.0969e-01	2.7665e+05 ;	C3 - H7
4	68	1	1.0969e-01	2.7665e+05 ;	C3 - H8
5	69	1	1.0969e-01	2.7665e+05 ;	C4 - H9
5	70	1	1.0969e-01	2.7665e+05 ;	C4 - H10
5	71	1	1.0969e-01	2.7665e+05 ;	C4 - H11
6	7	1	1.5375e-01	2.5179e+05 ;	C5 - C6
6	8	1	1.4233e-01	2.6501e+05 ;	C5 - O1
6	72	1	1.0969e-01	2.7665e+05 ;	C5 - H12
7	9	1	1.4316e-01	2.5824e+05 ;	C6 - O2
7	73	1	1.0969e-01	2.7665e+05 ;	C6 - H13
7	74	1	1.0969e-01	2.7665e+05 ;	C6 - H14
8	75	1	9.7300e-02	3.1079e+05 ;	O1 - H15
9	10	1	1.3696e-01	3.1514e+05 ;	O2 - C7
10	11	1	1.3984e-01	3.8585e+05 ;	C7 - C8
10	15	1	1.3984e-01	3.8585e+05 ;	C7 - C12
11	12	1	1.3984e-01	3.8585e+05 ;	C8 - C9
11	76	1	1.0860e-01	2.8937e+05 ;	C8 - H16
12	13	1	1.3984e-01	3.8585e+05 ;	C9 - C10
12	77	1	1.0860e-01	2.8937e+05 ;	C9 - H17
13	14	1	1.3984e-01	3.8585e+05 ;	C10 - C11
13	16	1	1.3696e-01	3.1514e+05 ;	C10 - O3
14	15	1	1.3984e-01	3.8585e+05 ;	C11 - C12
14	78	1	1.0860e-01	2.8937e+05 ;	C11 - H18
15	79	1	1.0860e-01	2.8937e+05 ;	C12 - H19
16	17	1	1.4316e-01	2.5824e+05 ;	O3 - C13
17	26	1	1.5375e-01	2.5179e+05 ;	C13 - C27
17	80	1	1.0969e-01	2.7665e+05 ;	C13 - H20
17	81	1	1.0969e-01	2.7665e+05 ;	C13 - H21

18	19	1	1.3984e-01	3.8585e+05 ;	C14 - C15
18	23	1	1.3984e-01	3.8585e+05 ;	C14 - C19
18	27	1	1.3696e-01	3.1514e+05 ;	C14 - O5
19	20	1	1.3984e-01	3.8585e+05 ;	C15 - C16
19	82	1	1.0860e-01	2.8937e+05 ;	C15 - H22
20	21	1	1.3984e-01	3.8585e+05 ;	C16 - C17
20	83	1	1.0860e-01	2.8937e+05 ;	C16 - H23
21	22	1	1.3984e-01	3.8585e+05 ;	C17 - C18
21	24	1	1.3696e-01	3.1514e+05 ;	C17 - O4
22	23	1	1.3984e-01	3.8585e+05 ;	C18 - C19
22	84	1	1.0860e-01	2.8937e+05 ;	C18 - H24
23	85	1	1.0860e-01	2.8937e+05 ;	C19 - H25
24	25	1	1.4316e-01	2.5824e+05 ;	O4 - C20
25	55	1	1.5156e-01	2.6861e+05 ;	C20 - C24
25	121	1	1.0969e-01	2.7665e+05 ;	C20 - H201
25	122	1	1.0969e-01	2.7665e+05 ;	C20 - H202
26	27	1	1.4316e-01	2.5824e+05 ;	C27 - O5
26	86	1	1.0969e-01	2.7665e+05 ;	C27 - H33
26	87	1	1.0969e-01	2.7665e+05 ;	C27 - H34
28	29	1	1.5110e-01	2.3707e+05 ;	C21 - N2
28	33	1	1.5375e-01	2.5179e+05 ;	C21 - C30
28	88	1	1.0910e-01	2.8342e+05 ;	C21 - H26
28	89	1	1.0910e-01	2.8342e+05 ;	C21 - H28
29	30	1	1.5110e-01	2.3707e+05 ;	N2 - C23
29	90	1	1.0304e-01	3.1229e+05 ;	N2 - H30
29	91	1	1.0304e-01	3.1229e+05 ;	N2 - H32
30	31	1	1.5375e-01	2.5179e+05 ;	C23 - C25
30	32	1	1.5375e-01	2.5179e+05 ;	C23 - C28
30	92	1	1.0910e-01	2.8342e+05 ;	C23 - H36
31	93	1	1.0969e-01	2.7665e+05 ;	C25 - H38
31	94	1	1.0969e-01	2.7665e+05 ;	C25 - H40
31	95	1	1.0969e-01	2.7665e+05 ;	C25 - H41

32 96 1 1.0969e-01 2.7665e+05 ; C28 - H42
32 97 1 1.0969e-01 2.7665e+05 ; C28 - H43
32 98 1 1.0969e-01 2.7665e+05 ; C28 - H44
33 34 1 1.5375e-01 2.5179e+05 ; C30 - C32
33 35 1 1.4233e-01 2.6501e+05 ; C30 - O6
33 99 1 1.0969e-01 2.7665e+05 ; C30 - H45
34 36 1 1.4316e-01 2.5824e+05 ; C32 - O7
34 100 1 1.0969e-01 2.7665e+05 ; C32 - H46
34 101 1 1.0969e-01 2.7665e+05 ; C32 - H47
35 102 1 9.7300e-02 3.1079e+05 ; O6 - H48
36 37 1 1.3696e-01 3.1514e+05 ; O7 - C34
37 38 1 1.3984e-01 3.8585e+05 ; C34 - C36
37 42 1 1.3984e-01 3.8585e+05 ; C34 - C40
38 39 1 1.3984e-01 3.8585e+05 ; C36 - C37
38 103 1 1.0860e-01 2.8937e+05 ; C36 - H49
39 40 1 1.3984e-01 3.8585e+05 ; C37 - C38
39 104 1 1.0860e-01 2.8937e+05 ; C37 - H50
40 41 1 1.3984e-01 3.8585e+05 ; C38 - C39
40 43 1 1.3696e-01 3.1514e+05 ; C38 - O8
41 42 1 1.3984e-01 3.8585e+05 ; C39 - C40
41 105 1 1.0860e-01 2.8937e+05 ; C39 - H51
42 106 1 1.0860e-01 2.8937e+05 ; C40 - H52
43 44 1 1.4316e-01 2.5824e+05 ; O8 - C41
44 52 1 1.5375e-01 2.5179e+05 ; C41 - C48
44 107 1 1.0969e-01 2.7665e+05 ; C41 - H53
44 108 1 1.0969e-01 2.7665e+05 ; C41 - H54
45 46 1 1.3984e-01 3.8585e+05 ; C42 - C43
45 50 1 1.3984e-01 3.8585e+05 ; C42 - C47
45 53 1 1.3696e-01 3.1514e+05 ; C42 - O10
46 47 1 1.3984e-01 3.8585e+05 ; C43 - C44
46 109 1 1.0860e-01 2.8937e+05 ; C43 - H55
47 48 1 1.3984e-01 3.8585e+05 ; C44 - C45

47 110 1 1.0860e-01 2.8937e+05 ; C44 - H56
 48 49 1 1.3984e-01 3.8585e+05 ; C45 - C46
 48 51 1 1.3696e-01 3.1514e+05 ; C45 - O9
 49 50 1 1.3984e-01 3.8585e+05 ; C46 - C47
 49 111 1 1.0860e-01 2.8937e+05 ; C46 - H57
 50 112 1 1.0860e-01 2.8937e+05 ; C47 - H58
 51 54 1 1.4316e-01 2.5824e+05 ; O9 - C22
 52 53 1 1.4316e-01 2.5824e+05 ; C48 - O10
 52 113 1 1.0969e-01 2.7665e+05 ; C48 - H59
 52 114 1 1.0969e-01 2.7665e+05 ; C48 - H60
 54 58 1 1.5156e-01 2.6861e+05 ; C22 - C31
 54 115 1 1.0969e-01 2.7665e+05 ; C22 - H27
 54 116 1 1.0969e-01 2.7665e+05 ; C22 - H29
 55 56 1 1.3984e-01 3.8585e+05 ; C24 - C26
 55 60 1 1.3984e-01 3.8585e+05 ; C24 - C35
 56 57 1 1.3984e-01 3.8585e+05 ; C26 - C29
 56 117 1 1.0860e-01 2.8937e+05 ; C26 - H31
 57 58 1 1.3984e-01 3.8585e+05 ; C29 - C31
 57 118 1 1.0860e-01 2.8937e+05 ; C29 - H35
 58 59 1 1.3984e-01 3.8585e+05 ; C31 - C33
 59 60 1 1.3984e-01 3.8585e+05 ; C33 - C35
 59 119 1 1.0860e-01 2.8937e+05 ; C33 - H37
 60 120 1 1.0860e-01 2.8937e+05 ; C35 - H39

[pairs]

; ai aj funct

1 4 1 ; C1 - C3
 1 5 1 ; C1 - C4
 1 9 1 ; C1 - O2
 1 65 1 ; C1 - H5
 1 73 1 ; C1 - H13
 1 74 1 ; C1 - H14

1 75 1; C1 - H15
2 7 1; N1 - C6
2 8 1; N1 - O1
2 66 1; N1 - H6
2 67 1; N1 - H7
2 68 1; N1 - H8
2 69 1; N1 - H9
2 70 1; N1 - H10
2 71 1; N1 - H11
2 72 1; N1 - H12
4 63 1; C3 - H3
4 64 1; C3 - H4
4 69 1; C3 - H9
4 70 1; C3 - H10
4 71 1; C3 - H11
5 63 1; C4 - H3
5 64 1; C4 - H4
5 66 1; C4 - H6
5 67 1; C4 - H7
5 68 1; C4 - H8
6 3 1; C5 - C2
6 10 1; C5 - C7
6 63 1; C5 - H3
6 64 1; C5 - H4
7 11 1; C6 - C8
7 15 1; C6 - C12
7 75 1; C6 - H15
8 9 1; O1 - O2
8 73 1; O1 - H13
8 74 1; O1 - H14
9 12 1; O2 - C9
9 14 1; O2 - C11

9 72 1; O2 - H12
9 76 1; O2 - H16
9 79 1; O2 - H19
10 13 1; C7 - C10
10 73 1; C7 - H13
10 74 1; C7 - H14
10 77 1; C7 - H17
10 78 1; C7 - H18
11 14 1; C8 - C11
11 16 1; C8 - O3
11 79 1; C8 - H19
12 15 1; C9 - C12
12 17 1; C9 - C13
12 78 1; C9 - H18
13 26 1; C10 - C27
13 76 1; C10 - H16
13 79 1; C10 - H19
13 80 1; C10 - H20
13 81 1; C10 - H21
14 17 1; C11 - C13
14 77 1; C11 - H17
15 16 1; C12 - O3
15 76 1; C12 - H16
16 27 1; O3 - O5
16 77 1; O3 - H17
16 78 1; O3 - H18
16 86 1; O3 - H33
16 87 1; O3 - H34
17 18 1; C13 - C14
18 21 1; C14 - C17
18 83 1; C14 - H23
18 84 1; C14 - H24

18 86 1; C14 - H33
18 87 1; C14 - H34
19 22 1; C15 - C18
19 24 1; C15 - O4
19 26 1; C15 - C27
19 85 1; C15 - H25
20 23 1; C16 - C19
20 25 1; C16 - C20
20 27 1; C16 - O5
20 84 1; C16 - H24
21 55 1; C17 - C24
21 82 1; C17 - H22
21 85 1; C17 - H25
21 121 1; C17 - H201
21 122 1; C17 - H202
22 25 1; C18 - C20
22 27 1; C18 - O5
22 83 1; C18 - H23
23 24 1; C19 - O4
23 26 1; C19 - C27
23 82 1; C19 - H22
24 56 1; O4 - C26
24 60 1; O4 - C35
24 83 1; O4 - H23
24 84 1; O4 - H24
25 57 1; C20 - C29
25 59 1; C20 - C33
25 117 1; C20 - H31
25 120 1; C20 - H39
27 80 1; O5 - H20
27 81 1; O5 - H21
27 82 1; O5 - H22

27 85 1; O5 - H25
28 31 1; C21 - C25
28 32 1; C21 - C28
28 36 1; C21 - O7
28 92 1; C21 - H36
28 100 1; C21 - H46
28 101 1; C21 - H47
28 102 1; C21 - H48
29 34 1; N2 - C32
29 35 1; N2 - O6
29 93 1; N2 - H38
29 94 1; N2 - H40
29 95 1; N2 - H41
29 96 1; N2 - H42
29 97 1; N2 - H43
29 98 1; N2 - H44
29 99 1; N2 - H45
30 33 1; C23 - C30
30 88 1; C23 - H26
30 89 1; C23 - H28
31 90 1; C25 - H30
31 91 1; C25 - H32
31 96 1; C25 - H42
31 97 1; C25 - H43
31 98 1; C25 - H44
32 90 1; C28 - H30
32 91 1; C28 - H32
32 93 1; C28 - H38
32 94 1; C28 - H40
32 95 1; C28 - H41
33 37 1; C30 - C34
33 90 1; C30 - H30

33 91 1; C30 - H32
34 38 1; C32 - C36
34 42 1; C32 - C40
34 88 1; C32 - H26
34 89 1; C32 - H28
34 102 1; C32 - H48
35 36 1; O6 - O7
35 88 1; O6 - H26
35 89 1; O6 - H28
35 100 1; O6 - H46
35 101 1; O6 - H47
36 39 1; O7 - C37
36 41 1; O7 - C39
36 99 1; O7 - H45
36 103 1; O7 - H49
36 106 1; O7 - H52
37 40 1; C34 - C38
37 100 1; C34 - H46
37 101 1; C34 - H47
37 104 1; C34 - H50
37 105 1; C34 - H51
38 41 1; C36 - C39
38 43 1; C36 - O8
38 106 1; C36 - H52
39 42 1; C37 - C40
39 44 1; C37 - C41
39 105 1; C37 - H51
40 52 1; C38 - C48
40 103 1; C38 - H49
40 106 1; C38 - H52
40 107 1; C38 - H53
40 108 1; C38 - H54

41 44 1; C39 - C41
41 104 1; C39 - H50
42 43 1; C40 - O8
42 103 1; C40 - H49
43 53 1; O8 - O10
43 104 1; O8 - H50
43 105 1; O8 - H51
43 113 1; O8 - H59
43 114 1; O8 - H60
44 45 1; C41 - C42
45 48 1; C42 - C45
45 110 1; C42 - H56
45 111 1; C42 - H57
45 113 1; C42 - H59
45 114 1; C42 - H60
46 49 1; C43 - C46
46 51 1; C43 - O9
46 52 1; C43 - C48
46 112 1; C43 - H58
47 50 1; C44 - C47
47 53 1; C44 - O10
47 54 1; C44 - C22
47 111 1; C44 - H57
48 58 1; C45 - C31
48 109 1; C45 - H55
48 112 1; C45 - H58
48 115 1; C45 - H27
48 116 1; C45 - H29
49 53 1; C46 - O10
49 54 1; C46 - C22
49 110 1; C46 - H56
50 51 1; C47 - O9

50 52 1; C47 - C48
50 109 1; C47 - H55
51 57 1; O9 - C29
51 59 1; O9 - C33
51 110 1; O9 - H56
51 111 1; O9 - H57
53 107 1; O10 - H53
53 108 1; O10 - H54
53 109 1; O10 - H55
53 112 1; O10 - H58
54 56 1; C22 - C26
54 60 1; C22 - C35
54 118 1; C22 - H35
54 119 1; C22 - H37
55 58 1; C24 - C31
55 118 1; C24 - H35
55 119 1; C24 - H37
56 59 1; C26 - C33
56 120 1; C26 - H39
56 121 1; C26 - H201
56 122 1; C26 - H202
57 60 1; C29 - C35
57 115 1; C29 - H27
57 116 1; C29 - H29
57 119 1; C29 - H37
58 117 1; C31 - H31
58 120 1; C31 - H39
59 115 1; C33 - H27
59 116 1; C33 - H29
59 118 1; C33 - H35
60 117 1; C35 - H31
60 121 1; C35 - H201

60 122 1; C35 - H202
61 3 1; H1 - C2
61 7 1; H1 - C6
61 8 1; H1 - O1
61 63 1; H1 - H3
61 64 1; H1 - H4
61 72 1; H1 - H12
62 3 1; H2 - C2
62 7 1; H2 - C6
62 8 1; H2 - O1
62 63 1; H2 - H3
62 64 1; H2 - H4
62 72 1; H2 - H12
63 65 1; H3 - H5
64 65 1; H4 - H5
65 66 1; H5 - H6
65 67 1; H5 - H7
65 68 1; H5 - H8
65 69 1; H5 - H9
65 70 1; H5 - H10
65 71 1; H5 - H11
72 73 1; H12 - H13
72 74 1; H12 - H14
72 75 1; H12 - H15
76 77 1; H16 - H17
78 79 1; H18 - H19
80 86 1; H20 - H33
80 87 1; H20 - H34
81 86 1; H21 - H33
81 87 1; H21 - H34
82 83 1; H22 - H23
84 85 1; H24 - H25

88 90 1; H26 - H30
 88 91 1; H26 - H32
 88 99 1; H26 - H45
 89 90 1; H28 - H30
 89 91 1; H28 - H32
 89 99 1; H28 - H45
 90 92 1; H30 - H36
 91 92 1; H32 - H36
 92 93 1; H36 - H38
 92 94 1; H36 - H40
 92 95 1; H36 - H41
 92 96 1; H36 - H42
 92 97 1; H36 - H43
 92 98 1; H36 - H44
 99 100 1; H45 - H46
 99 101 1; H45 - H47
 99 102 1; H45 - H48
 103 104 1; H49 - H50
 105 106 1; H51 - H52
 107 113 1; H53 - H59
 107 114 1; H53 - H60
 108 113 1; H54 - H59
 108 114 1; H54 - H60
 109 110 1; H55 - H56
 111 112 1; H57 - H58
 117 118 1; H31 - H35
 119 120 1; H37 - H39

[angles]

	ai	aj	ak	funct	theta	cth		
	1	2	3	1	1.0966e+02	5.2400e+02	; C1 - N1	- C2
	1	2	63	1	1.1011e+02	3.8367e+02	; C1 - N1	- H3

1	2	64	1	1.1011e+02	3.8367e+02 ;	C1 - N1	- H4
1	6	7	1	1.1151e+02	5.2601e+02 ;	C1 - C5	- C6
1	6	8	1	1.1019e+02	5.6459e+02 ;	C1 - C5	- O1
1	6	72	1	1.0956e+02	3.8819e+02 ;	C1 - C5	- H12
2	1	6	1	1.1421e+02	5.3706e+02 ;	N1 - C1	- C5
2	1	61	1	1.0801e+02	4.0710e+02 ;	N1 - C1	- H1
2	1	62	1	1.0801e+02	4.0710e+02 ;	N1 - C1	- H2
2	3	4	1	1.1421e+02	5.3706e+02 ;	N1 - C2	- C3
2	3	5	1	1.1421e+02	5.3706e+02 ;	N1 - C2	- C4
2	3	65	1	1.0801e+02	4.0710e+02 ;	N1 - C2	- H5
3	2	63	1	1.1011e+02	3.8367e+02 ;	C2 - N1	- H3
3	2	64	1	1.1011e+02	3.8367e+02 ;	C2 - N1	- H4
3	4	66	1	1.0980e+02	3.8777e+02 ;	C2 - C3	- H6
3	4	67	1	1.0980e+02	3.8777e+02 ;	C2 - C3	- H7
3	4	68	1	1.0980e+02	3.8777e+02 ;	C2 - C3	- H8
3	5	69	1	1.0980e+02	3.8777e+02 ;	C2 - C4	- H9
3	5	70	1	1.0980e+02	3.8777e+02 ;	C2 - C4	- H10
3	5	71	1	1.0980e+02	3.8777e+02 ;	C2 - C4	- H11
4	3	5	1	1.1151e+02	5.2601e+02 ;	C3 - C2	- C4
4	3	65	1	1.1056e+02	3.8660e+02 ;	C3 - C2	- H5
5	3	65	1	1.1056e+02	3.8660e+02 ;	C4 - C2	- H5
6	1	61	1	1.1056e+02	3.8660e+02 ;	C5 - C1	- H1
6	1	62	1	1.1056e+02	3.8660e+02 ;	C5 - C1	- H2
6	7	9	1	1.0797e+02	5.6902e+02 ;	C5 - C6	- O2
6	7	73	1	1.0956e+02	3.8819e+02 ;	C5 - C6	- H13
6	7	74	1	1.0956e+02	3.8819e+02 ;	C5 - C6	- H14
6	8	75	1	1.0726e+02	3.9648e+02 ;	C5 - O1	- H15
7	6	8	1	1.1019e+02	5.6459e+02 ;	C6 - C5	- O1
7	6	72	1	1.0956e+02	3.8819e+02 ;	C6 - C5	- H12
7	9	10	1	1.1796e+02	5.2317e+02 ;	C6 - O2	- C7
8	6	72	1	1.1026e+02	4.2618e+02 ;	O1 - C5	- H12
9	7	73	1	1.0978e+02	4.2509e+02 ;	O2 - C6	- H13

9	7	74	1	1.0978e+02	4.2509e+02 ;	O2 - C6	- H14
9	10	11	1	1.1920e+02	5.8225e+02 ;	O2 - C7	- C8
9	10	15	1	1.1920e+02	5.8225e+02 ;	O2 - C7	- C12
10	11	12	1	1.2002e+02	5.5748e+02 ;	C7 - C8	- C9
10	11	76	1	1.1988e+02	4.0317e+02 ;	C7 - C8	- H16
10	15	14	1	1.2002e+02	5.5748e+02 ;	C7 - C12	- C11
10	15	79	1	1.1988e+02	4.0317e+02 ;	C7 - C12	- H19
11	10	15	1	1.2002e+02	5.5748e+02 ;	C8 - C7	- C12
11	12	13	1	1.2002e+02	5.5748e+02 ;	C8 - C9	- C10
11	12	77	1	1.1988e+02	4.0317e+02 ;	C8 - C9	- H17
12	11	76	1	1.1988e+02	4.0317e+02 ;	C9 - C8	- H16
12	13	14	1	1.2002e+02	5.5748e+02 ;	C9 - C10	- C11
12	13	16	1	1.1920e+02	5.8225e+02 ;	C9 - C10	- O3
13	12	77	1	1.1988e+02	4.0317e+02 ;	C10 - C9	- H17
13	14	15	1	1.2002e+02	5.5748e+02 ;	C10 - C11	- C12
13	14	78	1	1.1988e+02	4.0317e+02 ;	C10 - C11	- H18
13	16	17	1	1.1796e+02	5.2317e+02 ;	C10 - O3	- C13
14	13	16	1	1.1920e+02	5.8225e+02 ;	C11 - C10	- O3
14	15	79	1	1.1988e+02	4.0317e+02 ;	C11 - C12	- H19
15	14	78	1	1.1988e+02	4.0317e+02 ;	C12 - C11	- H18
16	17	26	1	1.0797e+02	5.6902e+02 ;	O3 - C13	- C27
16	17	80	1	1.0978e+02	4.2509e+02 ;	O3 - C13	- H20
16	17	81	1	1.0978e+02	4.2509e+02 ;	O3 - C13	- H21
17	26	27	1	1.0797e+02	5.6902e+02 ;	C13 - C27	- O5
17	26	86	1	1.0956e+02	3.8819e+02 ;	C13 - C27	- H33
17	26	87	1	1.0956e+02	3.8819e+02 ;	C13 - C27	- H34
18	19	20	1	1.2002e+02	5.5748e+02 ;	C14 - C15	- C16
18	19	82	1	1.1988e+02	4.0317e+02 ;	C14 - C15	- H22
18	23	22	1	1.2002e+02	5.5748e+02 ;	C14 - C19	- C18
18	23	85	1	1.1988e+02	4.0317e+02 ;	C14 - C19	- H25
18	27	26	1	1.1796e+02	5.2317e+02 ;	C14 - O5	- C27
19	18	23	1	1.2002e+02	5.5748e+02 ;	C15 - C14	- C19

19	18	27	1	1.1920e+02	5.8225e+02 ;	C15 - C14	- O5
19	20	21	1	1.2002e+02	5.5748e+02 ;	C15 - C16	- C17
19	20	83	1	1.1988e+02	4.0317e+02 ;	C15 - C16	- H23
20	19	82	1	1.1988e+02	4.0317e+02 ;	C16 - C15	- H22
20	21	22	1	1.2002e+02	5.5748e+02 ;	C16 - C17	- C18
20	21	24	1	1.1920e+02	5.8225e+02 ;	C16 - C17	- O4
21	20	83	1	1.1988e+02	4.0317e+02 ;	C17 - C16	- H23
21	22	23	1	1.2002e+02	5.5748e+02 ;	C17 - C18	- C19
21	22	84	1	1.1988e+02	4.0317e+02 ;	C17 - C18	- H24
21	24	25	1	1.1796e+02	5.2317e+02 ;	C17 - O4	- C20
22	21	24	1	1.1920e+02	5.8225e+02 ;	C18 - C17	- O4
22	23	85	1	1.1988e+02	4.0317e+02 ;	C18 - C19	- H25
23	18	27	1	1.1920e+02	5.8225e+02 ;	C19 - C14	- O5
23	22	84	1	1.1988e+02	4.0317e+02 ;	C19 - C18	- H24
24	25	55	1	1.0895e+02	5.7120e+02 ;	O4 - C20	- C24
24	25	121	1	1.0978e+02	4.2509e+02 ;	O4 - C20	- H201
24	25	122	1	1.0978e+02	4.2509e+02 ;	O4 - C20	- H202
25	55	56	1	1.2077e+02	5.3162e+02 ;	C20 - C24	- C26
25	55	60	1	1.2077e+02	5.3162e+02 ;	C20 - C24	- C35
26	17	80	1	1.0956e+02	3.8819e+02 ;	C27 - C13	- H20
26	17	81	1	1.0956e+02	3.8819e+02 ;	C27 - C13	- H21
27	26	86	1	1.0978e+02	4.2509e+02 ;	O5 - C27	- H33
27	26	87	1	1.0978e+02	4.2509e+02 ;	O5 - C27	- H34
28	29	30	1	1.0966e+02	5.2400e+02 ;	C21 - N2	- C23
28	29	90	1	1.1011e+02	3.8367e+02 ;	C21 - N2	- H30
28	29	91	1	1.1011e+02	3.8367e+02 ;	C21 - N2	- H32
28	33	34	1	1.1151e+02	5.2601e+02 ;	C21 - C30	- C32
28	33	35	1	1.1019e+02	5.6459e+02 ;	C21 - C30	- O6
28	33	99	1	1.0956e+02	3.8819e+02 ;	C21 - C30	- H45
29	28	33	1	1.1421e+02	5.3706e+02 ;	N2 - C21	- C30
29	28	88	1	1.0801e+02	4.0710e+02 ;	N2 - C21	- H26
29	28	89	1	1.0801e+02	4.0710e+02 ;	N2 - C21	- H28

29	30	31	1	1.1421e+02	5.3706e+02 ;	N2 - C23	- C25
29	30	32	1	1.1421e+02	5.3706e+02 ;	N2 - C23	- C28
29	30	92	1	1.0801e+02	4.0710e+02 ;	N2 - C23	- H36
30	29	90	1	1.1011e+02	3.8367e+02 ;	C23 - N2	- H30
30	29	91	1	1.1011e+02	3.8367e+02 ;	C23 - N2	- H32
30	31	93	1	1.0980e+02	3.8777e+02 ;	C23 - C25	- H38
30	31	94	1	1.0980e+02	3.8777e+02 ;	C23 - C25	- H40
30	31	95	1	1.0980e+02	3.8777e+02 ;	C23 - C25	- H41
30	32	96	1	1.0980e+02	3.8777e+02 ;	C23 - C28	- H42
30	32	97	1	1.0980e+02	3.8777e+02 ;	C23 - C28	- H43
30	32	98	1	1.0980e+02	3.8777e+02 ;	C23 - C28	- H44
31	30	32	1	1.1151e+02	5.2601e+02 ;	C25 - C23	- C28
31	30	92	1	1.1056e+02	3.8660e+02 ;	C25 - C23	- H36
32	30	92	1	1.1056e+02	3.8660e+02 ;	C28 - C23	- H36
33	28	88	1	1.1056e+02	3.8660e+02 ;	C30 - C21	- H26
33	28	89	1	1.1056e+02	3.8660e+02 ;	C30 - C21	- H28
33	34	36	1	1.0797e+02	5.6902e+02 ;	C30 - C32	- O7
33	34	100	1	1.0956e+02	3.8819e+02 ;	C30 - C32	- H46
33	34	101	1	1.0956e+02	3.8819e+02 ;	C30 - C32	- H47
33	35	102	1	1.0726e+02	3.9648e+02 ;	C30 - O6	- H48
34	33	35	1	1.1019e+02	5.6459e+02 ;	C32 - C30	- O6
34	33	99	1	1.0956e+02	3.8819e+02 ;	C32 - C30	- H45
34	36	37	1	1.1796e+02	5.2317e+02 ;	C32 - O7	- C34
35	33	99	1	1.1026e+02	4.2618e+02 ;	O6 - C30	- H45
36	34	100	1	1.0978e+02	4.2509e+02 ;	O7 - C32	- H46
36	34	101	1	1.0978e+02	4.2509e+02 ;	O7 - C32	- H47
36	37	38	1	1.1920e+02	5.8225e+02 ;	O7 - C34	- C36
36	37	42	1	1.1920e+02	5.8225e+02 ;	O7 - C34	- C40
37	38	39	1	1.2002e+02	5.5748e+02 ;	C34 - C36	- C37
37	38	103	1	1.1988e+02	4.0317e+02 ;	C34 - C36	- H49
37	42	41	1	1.2002e+02	5.5748e+02 ;	C34 - C40	- C39
37	42	106	1	1.1988e+02	4.0317e+02 ;	C34 - C40	- H52

38	37	42	1	1.2002e+02	5.5748e+02 ;	C36 - C34	- C40
38	39	40	1	1.2002e+02	5.5748e+02 ;	C36 - C37	- C38
38	39	104	1	1.1988e+02	4.0317e+02 ;	C36 - C37	- H50
39	38	103	1	1.1988e+02	4.0317e+02 ;	C37 - C36	- H49
39	40	41	1	1.2002e+02	5.5748e+02 ;	C37 - C38	- C39
39	40	43	1	1.1920e+02	5.8225e+02 ;	C37 - C38	- O8
40	39	104	1	1.1988e+02	4.0317e+02 ;	C38 - C37	- H50
40	41	42	1	1.2002e+02	5.5748e+02 ;	C38 - C39	- C40
40	41	105	1	1.1988e+02	4.0317e+02 ;	C38 - C39	- H51
40	43	44	1	1.1796e+02	5.2317e+02 ;	C38 - O8	- C41
41	40	43	1	1.1920e+02	5.8225e+02 ;	C39 - C38	- O8
41	42	106	1	1.1988e+02	4.0317e+02 ;	C39 - C40	- H52
42	41	105	1	1.1988e+02	4.0317e+02 ;	C40 - C39	- H51
43	44	52	1	1.0797e+02	5.6902e+02 ;	O8 - C41	- C48
43	44	107	1	1.0978e+02	4.2509e+02 ;	O8 - C41	- H53
43	44	108	1	1.0978e+02	4.2509e+02 ;	O8 - C41	- H54
44	52	53	1	1.0797e+02	5.6902e+02 ;	C41 - C48	- O10
44	52	113	1	1.0956e+02	3.8819e+02 ;	C41 - C48	- H59
44	52	114	1	1.0956e+02	3.8819e+02 ;	C41 - C48	- H60
45	46	47	1	1.2002e+02	5.5748e+02 ;	C42 - C43	- C44
45	46	109	1	1.1988e+02	4.0317e+02 ;	C42 - C43	- H55
45	50	49	1	1.2002e+02	5.5748e+02 ;	C42 - C47	- C46
45	50	112	1	1.1988e+02	4.0317e+02 ;	C42 - C47	- H58
45	53	52	1	1.1796e+02	5.2317e+02 ;	C42 - O10	- C48
46	45	50	1	1.2002e+02	5.5748e+02 ;	C43 - C42	- C47
46	45	53	1	1.1920e+02	5.8225e+02 ;	C43 - C42	- O10
46	47	48	1	1.2002e+02	5.5748e+02 ;	C43 - C44	- C45
46	47	110	1	1.1988e+02	4.0317e+02 ;	C43 - C44	- H56
47	46	109	1	1.1988e+02	4.0317e+02 ;	C44 - C43	- H55
47	48	49	1	1.2002e+02	5.5748e+02 ;	C44 - C45	- C46
47	48	51	1	1.1920e+02	5.8225e+02 ;	C44 - C45	- O9
48	47	110	1	1.1988e+02	4.0317e+02 ;	C45 - C44	- H56

48	49	50	1	1.2002e+02	5.5748e+02 ;	C45 - C46 - C47
48	49	111	1	1.1988e+02	4.0317e+02 ;	C45 - C46 - H57
48	51	54	1	1.1796e+02	5.2317e+02 ;	C45 - O9 - C22
49	48	51	1	1.1920e+02	5.8225e+02 ;	C46 - C45 - O9
49	50	112	1	1.1988e+02	4.0317e+02 ;	C46 - C47 - H58
50	45	53	1	1.1920e+02	5.8225e+02 ;	C47 - C42 - O10
50	49	111	1	1.1988e+02	4.0317e+02 ;	C47 - C46 - H57
51	54	58	1	1.0895e+02	5.7120e+02 ;	O9 - C22 - C31
51	54	115	1	1.0978e+02	4.2509e+02 ;	O9 - C22 - H27
51	54	116	1	1.0978e+02	4.2509e+02 ;	O9 - C22 - H29
52	44	107	1	1.0956e+02	3.8819e+02 ;	C48 - C41 - H53
52	44	108	1	1.0956e+02	3.8819e+02 ;	C48 - C41 - H54
53	52	113	1	1.0978e+02	4.2509e+02 ;	O10 - C48 - H59
53	52	114	1	1.0978e+02	4.2509e+02 ;	O10 - C48 - H60
54	58	57	1	1.2077e+02	5.3162e+02 ;	C22 - C31 - C29
54	58	59	1	1.2077e+02	5.3162e+02 ;	C22 - C31 - C33
55	25	121	1	1.0956e+02	3.9321e+02 ;	C24 - C20 - H201
55	25	122	1	1.0956e+02	3.9321e+02 ;	C24 - C20 - H202
55	56	57	1	1.2002e+02	5.5748e+02 ;	C24 - C26 - C29
55	56	117	1	1.1988e+02	4.0317e+02 ;	C24 - C26 - H31
55	60	59	1	1.2002e+02	5.5748e+02 ;	C24 - C35 - C33
55	60	120	1	1.1988e+02	4.0317e+02 ;	C24 - C35 - H39
56	55	60	1	1.2002e+02	5.5748e+02 ;	C26 - C24 - C35
56	57	58	1	1.2002e+02	5.5748e+02 ;	C26 - C29 - C31
56	57	118	1	1.1988e+02	4.0317e+02 ;	C26 - C29 - H35
57	56	117	1	1.1988e+02	4.0317e+02 ;	C29 - C26 - H31
57	58	59	1	1.2002e+02	5.5748e+02 ;	C29 - C31 - C33
58	54	115	1	1.0956e+02	3.9321e+02 ;	C31 - C22 - H27
58	54	116	1	1.0956e+02	3.9321e+02 ;	C31 - C22 - H29
58	57	118	1	1.1988e+02	4.0317e+02 ;	C31 - C29 - H35
58	59	60	1	1.2002e+02	5.5748e+02 ;	C31 - C33 - C35
58	59	119	1	1.1988e+02	4.0317e+02 ;	C31 - C33 - H37

```

59 60 120 1 1.1988e+02 4.0317e+02 ; C33 - C35 - H39
60 59 119 1 1.1988e+02 4.0317e+02 ; C35 - C33 - H37
61 1 62 1 1.0975e+02 3.2819e+02 ; H1 - C1 - H2
63 2 64 1 1.0830e+02 3.3957e+02 ; H3 - N1 - H4
66 4 67 1 1.0758e+02 3.2970e+02 ; H6 - C3 - H7
66 4 68 1 1.0758e+02 3.2970e+02 ; H6 - C3 - H8
67 4 68 1 1.0758e+02 3.2970e+02 ; H7 - C3 - H8
69 5 70 1 1.0758e+02 3.2970e+02 ; H9 - C4 - H10
69 5 71 1 1.0758e+02 3.2970e+02 ; H9 - C4 - H11
70 5 71 1 1.0758e+02 3.2970e+02 ; H10 - C4 - H11
73 7 74 1 1.0846e+02 3.2836e+02 ; H13 - C6 - H14
80 17 81 1 1.0846e+02 3.2836e+02 ; H20 - C13 - H21
86 26 87 1 1.0846e+02 3.2836e+02 ; H33 - C27 - H34
88 28 89 1 1.0975e+02 3.2819e+02 ; H26 - C21 - H28
90 29 91 1 1.0830e+02 3.3957e+02 ; H30 - N2 - H32
93 31 94 1 1.0758e+02 3.2970e+02 ; H38 - C25 - H40
93 31 95 1 1.0758e+02 3.2970e+02 ; H38 - C25 - H41
94 31 95 1 1.0758e+02 3.2970e+02 ; H40 - C25 - H41
96 32 97 1 1.0758e+02 3.2970e+02 ; H42 - C28 - H43
96 32 98 1 1.0758e+02 3.2970e+02 ; H42 - C28 - H44
97 32 98 1 1.0758e+02 3.2970e+02 ; H43 - C28 - H44
100 34 101 1 1.0846e+02 3.2836e+02 ; H46 - C32 - H47
107 44 108 1 1.0846e+02 3.2836e+02 ; H53 - C41 - H54
113 52 114 1 1.0846e+02 3.2836e+02 ; H59 - C48 - H60
115 54 116 1 1.0846e+02 3.2836e+02 ; H27 - C22 - H29
121 25 122 1 1.0846e+02 3.2836e+02 ; H201 - C20 - H202

```

[dihedrals] ; propers

; for gromacs 4.5 or higher, using funct 9

; i j k l func phase kd pn

```
1 2 3 4 9 0.00 0.65084 3; C1- N1- C2- C3
```

```
1 2 3 5 9 0.00 0.65084 3; C1- N1- C2- C4
```

1	2	3	65	9	0.00	0.65084	3	;	C1-	N1-	C2-	H5
1	6	7	9	9	0.00	0.65084	3	;	C1-	C5-	C6-	O2
1	6	7	73	9	0.00	0.65084	3	;	C1-	C5-	C6-	H13
1	6	7	74	9	0.00	0.65084	3	;	C1-	C5-	C6-	H14
1	6	8	75	9	0.00	0.66944	3	;	C1-	C5-	O1-	H15
1	6	8	75	9	0.00	1.04600	1	;	C1-	C5-	O1-	H15
2	1	6	7	9	0.00	0.65084	3	;	N1-	C1-	C5-	C6
2	1	6	8	9	0.00	0.60250	3	;	N1-	C1-	C5-	O1
2	1	6	8	9	0.00	5.43920	2	;	N1-	C1-	C5-	O1
2	1	6	72	9	0.00	0.65084	3	;	N1-	C1-	C5-	H12
2	3	4	66	9	0.00	0.65084	3	;	N1-	C2-	C3-	H6
2	3	4	67	9	0.00	0.65084	3	;	N1-	C2-	C3-	H7
2	3	4	68	9	0.00	0.65084	3	;	N1-	C2-	C3-	H8
2	3	5	69	9	0.00	0.65084	3	;	N1-	C2-	C4-	H9
2	3	5	70	9	0.00	0.65084	3	;	N1-	C2-	C4-	H10
2	3	5	71	9	0.00	0.65084	3	;	N1-	C2-	C4-	H11
4	3	2	63	9	0.00	0.65084	3	;	C3-	C2-	N1-	H3
4	3	2	64	9	0.00	0.65084	3	;	C3-	C2-	N1-	H4
4	3	5	69	9	0.00	0.66944	3	;	C3-	C2-	C4-	H9
4	3	5	70	9	0.00	0.66944	3	;	C3-	C2-	C4-	H10
4	3	5	71	9	0.00	0.66944	3	;	C3-	C2-	C4-	H11
5	3	2	63	9	0.00	0.65084	3	;	C4-	C2-	N1-	H3
5	3	2	64	9	0.00	0.65084	3	;	C4-	C2-	N1-	H4
5	3	4	66	9	0.00	0.66944	3	;	C4-	C2-	C3-	H6
5	3	4	67	9	0.00	0.66944	3	;	C4-	C2-	C3-	H7
5	3	4	68	9	0.00	0.66944	3	;	C4-	C2-	C3-	H8
6	1	2	3	9	0.00	0.65084	3	;	C5-	C1-	N1-	C2
6	1	2	63	9	0.00	0.65084	3	;	C5-	C1-	N1-	H3
6	1	2	64	9	0.00	0.65084	3	;	C5-	C1-	N1-	H4
6	7	9	10	9	0.00	1.60387	3	;	C5-	C6-	O2-	C7
7	6	8	75	9	0.00	0.66944	3	;	C6-	C5-	O1-	H15
7	6	8	75	9	0.00	1.04600	1	;	C6-	C5-	O1-	H15

7 9 10 11 9 180.00 3.76560 2; C6- O2- C7- C8
7 9 10 15 9 180.00 3.76560 2; C6- O2- C7- C12
8 6 7 9 9 0.00 0.60250 3; O1- C5- C6- O2
8 6 7 9 9 0.00 4.91620 2; O1- C5- C6- O2
8 6 7 73 9 0.00 0.00000 0; O1- C5- C6- H13
8 6 7 73 9 0.00 1.04600 1; O1- C5- C6- H13
8 6 7 74 9 0.00 0.00000 0; O1- C5- C6- H14
8 6 7 74 9 0.00 1.04600 1; O1- C5- C6- H14
9 7 6 72 9 0.00 0.00000 0; O2- C6- C5- H12
9 7 6 72 9 0.00 1.04600 1; O2- C6- C5- H12
9 10 11 12 9 180.00 15.16700 2; O2- C7- C8- C9
9 10 11 76 9 180.00 15.16700 2; O2- C7- C8- H16
9 10 15 14 9 180.00 15.16700 2; O2- C7- C12- C11
9 10 15 79 9 180.00 15.16700 2; O2- C7- C12- H19
10 9 7 73 9 0.00 1.60387 3; C7- O2- C6- H13
10 9 7 74 9 0.00 1.60387 3; C7- O2- C6- H14
10 11 12 13 9 180.00 15.16700 2; C7- C8- C9- C10
10 11 12 77 9 180.00 15.16700 2; C7- C8- C9- H17
10 15 14 13 9 180.00 15.16700 2; C7- C12- C11- C10
10 15 14 78 9 180.00 15.16700 2; C7- C12- C11- H18
11 10 15 14 9 180.00 15.16700 2; C8- C7- C12- C11
11 10 15 79 9 180.00 15.16700 2; C8- C7- C12- H19
11 12 13 14 9 180.00 15.16700 2; C8- C9- C10- C11
11 12 13 16 9 180.00 15.16700 2; C8- C9- C10- O3
12 11 10 15 9 180.00 15.16700 2; C9- C8- C7- C12
12 13 14 15 9 180.00 15.16700 2; C9- C10- C11- C12
12 13 14 78 9 180.00 15.16700 2; C9- C10- C11- H18
12 13 16 17 9 180.00 3.76560 2; C9- C10- O3- C13
13 12 11 76 9 180.00 15.16700 2; C10- C9- C8- H16
13 14 15 79 9 180.00 15.16700 2; C10- C11- C12- H19
13 16 17 26 9 0.00 1.60387 3; C10- O3- C13- C27
13 16 17 80 9 0.00 1.60387 3; C10- O3- C13- H20

13 16 17 81 9 0.00 1.60387 3 ; C10- O3- C13- H21
14 13 12 77 9 180.00 15.16700 2 ; C11- C10- C9- H17
14 13 16 17 9 180.00 3.76560 2 ; C11- C10- O3- C13
15 10 11 76 9 180.00 15.16700 2 ; C12- C7- C8- H16
15 14 13 16 9 180.00 15.16700 2 ; C12- C11- C10- O3
16 13 12 77 9 180.00 15.16700 2 ; O3- C10- C9- H17
16 13 14 78 9 180.00 15.16700 2 ; O3- C10- C11- H18
16 17 26 27 9 0.00 0.60250 3 ; O3- C13- C27- O5
16 17 26 27 9 0.00 4.91620 2 ; O3- C13- C27- O5
16 17 26 86 9 0.00 0.00000 0 ; O3- C13- C27- H33
16 17 26 86 9 0.00 1.04600 1 ; O3- C13- C27- H33
16 17 26 87 9 0.00 0.00000 0 ; O3- C13- C27- H34
16 17 26 87 9 0.00 1.04600 1 ; O3- C13- C27- H34
17 26 27 18 9 0.00 1.60387 3 ; C13- C27- O5- C14
18 19 20 21 9 180.00 15.16700 2 ; C14- C15- C16- C17
18 19 20 83 9 180.00 15.16700 2 ; C14- C15- C16- H23
18 23 22 21 9 180.00 15.16700 2 ; C14- C19- C18- C17
18 23 22 84 9 180.00 15.16700 2 ; C14- C19- C18- H24
18 27 26 86 9 0.00 1.60387 3 ; C14- O5- C27- H33
18 27 26 87 9 0.00 1.60387 3 ; C14- O5- C27- H34
19 18 23 22 9 180.00 15.16700 2 ; C15- C14- C19- C18
19 18 23 85 9 180.00 15.16700 2 ; C15- C14- C19- H25
19 18 27 26 9 180.00 3.76560 2 ; C15- C14- O5- C27
19 20 21 22 9 180.00 15.16700 2 ; C15- C16- C17- C18
19 20 21 24 9 180.00 15.16700 2 ; C15- C16- C17- O4
20 19 18 23 9 180.00 15.16700 2 ; C16- C15- C14- C19
20 19 18 27 9 180.00 15.16700 2 ; C16- C15- C14- O5
20 21 22 23 9 180.00 15.16700 2 ; C16- C17- C18- C19
20 21 22 84 9 180.00 15.16700 2 ; C16- C17- C18- H24
20 21 24 25 9 180.00 3.76560 2 ; C16- C17- O4- C20
21 20 19 82 9 180.00 15.16700 2 ; C17- C16- C15- H22
21 22 23 85 9 180.00 15.16700 2 ; C17- C18- C19- H25

21 24 25 55 9 0.00 1.60387 3 ; C17- O4- C20- C24
 21 24 25 121 9 0.00 1.60387 3 ; C17- O4- C20- H201
 21 24 25 122 9 0.00 1.60387 3 ; C17- O4- C20- H202
 22 21 20 83 9 180.00 15.16700 2 ; C18- C17- C16- H23
 22 21 24 25 9 180.00 3.76560 2 ; C18- C17- O4- C20
 22 23 18 27 9 180.00 15.16700 2 ; C18- C19- C14- O5
 23 18 19 82 9 180.00 15.16700 2 ; C19- C14- C15- H22
 23 18 27 26 9 180.00 3.76560 2 ; C19- C14- O5- C27
 23 22 21 24 9 180.00 15.16700 2 ; C19- C18- C17- O4
 24 21 20 83 9 180.00 15.16700 2 ; O4- C17- C16- H23
 24 21 22 84 9 180.00 15.16700 2 ; O4- C17- C18- H24
 24 25 55 56 9 0.00 0.00000 0 ; O4- C20- C24- C26
 24 25 55 60 9 0.00 0.00000 0 ; O4- C20- C24- C35
 25 55 56 57 9 180.00 15.16700 2 ; C20- C24- C26- C29
 25 55 56 117 9 180.00 15.16700 2 ; C20- C24- C26- H31
 25 55 60 59 9 180.00 15.16700 2 ; C20- C24- C35- C33
 25 55 60 120 9 180.00 15.16700 2 ; C20- C24- C35- H39
 27 18 19 82 9 180.00 15.16700 2 ; O5- C14- C15- H22
 27 18 23 85 9 180.00 15.16700 2 ; O5- C14- C19- H25
 27 26 17 80 9 0.00 0.00000 0 ; O5- C27- C13- H20
 27 26 17 80 9 0.00 1.04600 1 ; O5- C27- C13- H20
 27 26 17 81 9 0.00 0.00000 0 ; O5- C27- C13- H21
 27 26 17 81 9 0.00 1.04600 1 ; O5- C27- C13- H21
 28 29 30 31 9 0.00 0.65084 3 ; C21- N2- C23- C25
 28 29 30 32 9 0.00 0.65084 3 ; C21- N2- C23- C28
 28 29 30 92 9 0.00 0.65084 3 ; C21- N2- C23- H36
 28 33 34 36 9 0.00 0.65084 3 ; C21- C30- C32- O7
 28 33 34 100 9 0.00 0.65084 3 ; C21- C30- C32- H46
 28 33 34 101 9 0.00 0.65084 3 ; C21- C30- C32- H47
 28 33 35 102 9 0.00 0.66944 3 ; C21- C30- O6- H48
 28 33 35 102 9 0.00 1.04600 1 ; C21- C30- O6- H48
 29 28 33 34 9 0.00 0.65084 3 ; N2- C21- C30- C32

29	28	33	35	9	0.00	0.60250	3 ;	N2-	C21-	C30-	O6
29	28	33	35	9	0.00	5.43920	2 ;	N2-	C21-	C30-	O6
29	28	33	99	9	0.00	0.65084	3 ;	N2-	C21-	C30-	H45
29	30	31	93	9	0.00	0.65084	3 ;	N2-	C23-	C25-	H38
29	30	31	94	9	0.00	0.65084	3 ;	N2-	C23-	C25-	H40
29	30	31	95	9	0.00	0.65084	3 ;	N2-	C23-	C25-	H41
29	30	32	96	9	0.00	0.65084	3 ;	N2-	C23-	C28-	H42
29	30	32	97	9	0.00	0.65084	3 ;	N2-	C23-	C28-	H43
29	30	32	98	9	0.00	0.65084	3 ;	N2-	C23-	C28-	H44
30	29	28	33	9	0.00	0.65084	3 ;	C23-	N2-	C21-	C30
30	29	28	88	9	0.00	0.65084	3 ;	C23-	N2-	C21-	H26
30	29	28	89	9	0.00	0.65084	3 ;	C23-	N2-	C21-	H28
31	30	29	90	9	0.00	0.65084	3 ;	C25-	C23-	N2-	H30
31	30	29	91	9	0.00	0.65084	3 ;	C25-	C23-	N2-	H32
31	30	32	96	9	0.00	0.66944	3 ;	C25-	C23-	C28-	H42
31	30	32	97	9	0.00	0.66944	3 ;	C25-	C23-	C28-	H43
31	30	32	98	9	0.00	0.66944	3 ;	C25-	C23-	C28-	H44
32	30	29	90	9	0.00	0.65084	3 ;	C28-	C23-	N2-	H30
32	30	29	91	9	0.00	0.65084	3 ;	C28-	C23-	N2-	H32
32	30	31	93	9	0.00	0.66944	3 ;	C28-	C23-	C25-	H38
32	30	31	94	9	0.00	0.66944	3 ;	C28-	C23-	C25-	H40
32	30	31	95	9	0.00	0.66944	3 ;	C28-	C23-	C25-	H41
33	28	29	90	9	0.00	0.65084	3 ;	C30-	C21-	N2-	H30
33	28	29	91	9	0.00	0.65084	3 ;	C30-	C21-	N2-	H32
33	34	36	37	9	0.00	1.60387	3 ;	C30-	C32-	O7-	C34
34	33	28	88	9	0.00	0.65084	3 ;	C32-	C30-	C21-	H26
34	33	28	89	9	0.00	0.65084	3 ;	C32-	C30-	C21-	H28
34	33	35	102	9	0.00	0.66944	3 ;	C32-	C30-	O6-	H48
34	33	35	102	9	0.00	1.04600	1 ;	C32-	C30-	O6-	H48
34	36	37	38	9	180.00	3.76560	2 ;	C32-	O7-	C34-	C36
34	36	37	42	9	180.00	3.76560	2 ;	C32-	O7-	C34-	C40
35	33	28	88	9	0.00	0.65084	3 ;	O6-	C30-	C21-	H26

35 33 28 89 9 0.00 0.65084 3 ; O6- C30- C21- H28
 35 33 34 36 9 0.00 0.60250 3 ; O6- C30- C32- O7
 35 33 34 36 9 0.00 4.91620 2 ; O6- C30- C32- O7
 35 33 34 100 9 0.00 0.00000 0 ; O6- C30- C32- H46
 35 33 34 100 9 0.00 1.04600 1 ; O6- C30- C32- H46
 35 33 34 101 9 0.00 0.00000 0 ; O6- C30- C32- H47
 35 33 34 101 9 0.00 1.04600 1 ; O6- C30- C32- H47
 36 34 33 99 9 0.00 0.00000 0 ; O7- C32- C30- H45
 36 34 33 99 9 0.00 1.04600 1 ; O7- C32- C30- H45
 36 37 38 39 9 180.00 15.16700 2 ; O7- C34- C36- C37
 36 37 38 103 9 180.00 15.16700 2 ; O7- C34- C36- H49
 36 37 42 41 9 180.00 15.16700 2 ; O7- C34- C40- C39
 36 37 42 106 9 180.00 15.16700 2 ; O7- C34- C40- H52
 37 36 34 100 9 0.00 1.60387 3 ; C34- O7- C32- H46
 37 36 34 101 9 0.00 1.60387 3 ; C34- O7- C32- H47
 37 38 39 40 9 180.00 15.16700 2 ; C34- C36- C37- C38
 37 38 39 104 9 180.00 15.16700 2 ; C34- C36- C37- H50
 37 42 41 40 9 180.00 15.16700 2 ; C34- C40- C39- C38
 37 42 41 105 9 180.00 15.16700 2 ; C34- C40- C39- H51
 38 37 42 41 9 180.00 15.16700 2 ; C36- C34- C40- C39
 38 37 42 106 9 180.00 15.16700 2 ; C36- C34- C40- H52
 38 39 40 41 9 180.00 15.16700 2 ; C36- C37- C38- C39
 38 39 40 43 9 180.00 15.16700 2 ; C36- C37- C38- O8
 39 38 37 42 9 180.00 15.16700 2 ; C37- C36- C34- C40
 39 40 41 42 9 180.00 15.16700 2 ; C37- C38- C39- C40
 39 40 41 105 9 180.00 15.16700 2 ; C37- C38- C39- H51
 39 40 43 44 9 180.00 3.76560 2 ; C37- C38- O8- C41
 40 39 38 103 9 180.00 15.16700 2 ; C38- C37- C36- H49
 40 41 42 106 9 180.00 15.16700 2 ; C38- C39- C40- H52
 40 43 44 52 9 0.00 1.60387 3 ; C38- O8- C41- C48
 40 43 44 107 9 0.00 1.60387 3 ; C38- O8- C41- H53
 40 43 44 108 9 0.00 1.60387 3 ; C38- O8- C41- H54

41 40 39 104 9 180.00 15.16700 2; C39- C38- C37- H50
 41 40 43 44 9 180.00 3.76560 2; C39- C38- O8- C41
 42 37 38 103 9 180.00 15.16700 2; C40- C34- C36- H49
 42 41 40 43 9 180.00 15.16700 2; C40- C39- C38- O8
 43 40 39 104 9 180.00 15.16700 2; O8- C38- C37- H50
 43 40 41 105 9 180.00 15.16700 2; O8- C38- C39- H51
 43 44 52 53 9 0.00 0.60250 3; O8- C41- C48- O10
 43 44 52 53 9 0.00 4.91620 2; O8- C41- C48- O10
 43 44 52 113 9 0.00 0.00000 0; O8- C41- C48- H59
 43 44 52 113 9 0.00 1.04600 1; O8- C41- C48- H59
 43 44 52 114 9 0.00 0.00000 0; O8- C41- C48- H60
 43 44 52 114 9 0.00 1.04600 1; O8- C41- C48- H60
 44 52 53 45 9 0.00 1.60387 3; C41- C48- O10- C42
 45 46 47 48 9 180.00 15.16700 2; C42- C43- C44- C45
 45 46 47 110 9 180.00 15.16700 2; C42- C43- C44- H56
 45 50 49 48 9 180.00 15.16700 2; C42- C47- C46- C45
 45 50 49 111 9 180.00 15.16700 2; C42- C47- C46- H57
 45 53 52 113 9 0.00 1.60387 3; C42- O10- C48- H59
 45 53 52 114 9 0.00 1.60387 3; C42- O10- C48- H60
 46 45 50 49 9 180.00 15.16700 2; C43- C42- C47- C46
 46 45 50 112 9 180.00 15.16700 2; C43- C42- C47- H58
 46 45 53 52 9 180.00 3.76560 2; C43- C42- O10- C48
 46 47 48 49 9 180.00 15.16700 2; C43- C44- C45- C46
 46 47 48 51 9 180.00 15.16700 2; C43- C44- C45- O9
 47 46 45 50 9 180.00 15.16700 2; C44- C43- C42- C47
 47 46 45 53 9 180.00 15.16700 2; C44- C43- C42- O10
 47 48 49 50 9 180.00 15.16700 2; C44- C45- C46- C47
 47 48 49 111 9 180.00 15.16700 2; C44- C45- C46- H57
 47 48 51 54 9 180.00 3.76560 2; C44- C45- O9- C22
 48 47 46 109 9 180.00 15.16700 2; C45- C44- C43- H55
 48 49 50 112 9 180.00 15.16700 2; C45- C46- C47- H58
 48 51 54 58 9 0.00 1.60387 3; C45- O9- C22- C31

48 51 54 115 9 0.00 1.60387 3 ; C45- O9- C22- H27
 48 51 54 116 9 0.00 1.60387 3 ; C45- O9- C22- H29
 49 48 47 110 9 180.00 15.16700 2 ; C46- C45- C44- H56
 49 48 51 54 9 180.00 3.76560 2 ; C46- C45- O9- C22
 49 50 45 53 9 180.00 15.16700 2 ; C46- C47- C42- O10
 50 45 46 109 9 180.00 15.16700 2 ; C47- C42- C43- H55
 50 45 53 52 9 180.00 3.76560 2 ; C47- C42- O10- C48
 50 49 48 51 9 180.00 15.16700 2 ; C47- C46- C45- O9
 51 48 47 110 9 180.00 15.16700 2 ; O9- C45- C44- H56
 51 48 49 111 9 180.00 15.16700 2 ; O9- C45- C46- H57
 51 54 58 57 9 0.00 0.00000 0 ; O9- C22- C31- C29
 51 54 58 59 9 0.00 0.00000 0 ; O9- C22- C31- C33
 53 45 46 109 9 180.00 15.16700 2 ; O10- C42- C43- H55
 53 45 50 112 9 180.00 15.16700 2 ; O10- C42- C47- H58
 53 52 44 107 9 0.00 0.00000 0 ; O10- C48- C41- H53
 53 52 44 107 9 0.00 1.04600 1 ; O10- C48- C41- H53
 53 52 44 108 9 0.00 0.00000 0 ; O10- C48- C41- H54
 53 52 44 108 9 0.00 1.04600 1 ; O10- C48- C41- H54
 54 58 57 56 9 180.00 15.16700 2 ; C22- C31- C29- C26
 54 58 57 118 9 180.00 15.16700 2 ; C22- C31- C29- H35
 54 58 59 60 9 180.00 15.16700 2 ; C22- C31- C33- C35
 54 58 59 119 9 180.00 15.16700 2 ; C22- C31- C33- H37
 55 56 57 58 9 180.00 15.16700 2 ; C24- C26- C29- C31
 55 56 57 118 9 180.00 15.16700 2 ; C24- C26- C29- H35
 55 60 59 58 9 180.00 15.16700 2 ; C24- C35- C33- C31
 55 60 59 119 9 180.00 15.16700 2 ; C24- C35- C33- H37
 56 55 25 121 9 0.00 0.00000 0 ; C26- C24- C20- H201
 56 55 25 122 9 0.00 0.00000 0 ; C26- C24- C20- H202
 56 55 60 59 9 180.00 15.16700 2 ; C26- C24- C35- C33
 56 55 60 120 9 180.00 15.16700 2 ; C26- C24- C35- H39
 56 57 58 59 9 180.00 15.16700 2 ; C26- C29- C31- C33
 57 56 55 60 9 180.00 15.16700 2 ; C29- C26- C24- C35

57 58 54 115 9 0.00 0.00000 0 ; C29- C31- C22- H27
57 58 54 116 9 0.00 0.00000 0 ; C29- C31- C22- H29
57 58 59 60 9 180.00 15.16700 2 ; C29- C31- C33- C35
57 58 59 119 9 180.00 15.16700 2 ; C29- C31- C33- H37
58 57 56 117 9 180.00 15.16700 2 ; C31- C29- C26- H31
58 59 60 120 9 180.00 15.16700 2 ; C31- C33- C35- H39
59 58 54 115 9 0.00 0.00000 0 ; C33- C31- C22- H27
59 58 54 116 9 0.00 0.00000 0 ; C33- C31- C22- H29
59 58 57 118 9 180.00 15.16700 2 ; C33- C31- C29- H35
60 55 25 121 9 0.00 0.00000 0 ; C35- C24- C20- H201
60 55 25 122 9 0.00 0.00000 0 ; C35- C24- C20- H202
60 55 56 117 9 180.00 15.16700 2 ; C35- C24- C26- H31
61 1 2 3 9 0.00 0.65084 3 ; H1- C1- N1- C2
61 1 2 63 9 0.00 0.65084 3 ; H1- C1- N1- H3
61 1 2 64 9 0.00 0.65084 3 ; H1- C1- N1- H4
61 1 6 7 9 0.00 0.65084 3 ; H1- C1- C5- C6
61 1 6 8 9 0.00 0.65084 3 ; H1- C1- C5- O1
61 1 6 72 9 0.00 0.65084 3 ; H1- C1- C5- H12
62 1 2 3 9 0.00 0.65084 3 ; H2- C1- N1- C2
62 1 2 63 9 0.00 0.65084 3 ; H2- C1- N1- H3
62 1 2 64 9 0.00 0.65084 3 ; H2- C1- N1- H4
62 1 6 7 9 0.00 0.65084 3 ; H2- C1- C5- C6
62 1 6 8 9 0.00 0.65084 3 ; H2- C1- C5- O1
62 1 6 72 9 0.00 0.65084 3 ; H2- C1- C5- H12
63 2 3 65 9 0.00 0.65084 3 ; H3- N1- C2- H5
64 2 3 65 9 0.00 0.65084 3 ; H4- N1- C2- H5
65 3 4 66 9 0.00 0.65084 3 ; H5- C2- C3- H6
65 3 4 67 9 0.00 0.65084 3 ; H5- C2- C3- H7
65 3 4 68 9 0.00 0.65084 3 ; H5- C2- C3- H8
65 3 5 69 9 0.00 0.65084 3 ; H5- C2- C4- H9
65 3 5 70 9 0.00 0.65084 3 ; H5- C2- C4- H10
65 3 5 71 9 0.00 0.65084 3 ; H5- C2- C4- H11

72 6 7 73 9 0.00 0.65084 3 ; H12- C5- C6- H13
72 6 7 74 9 0.00 0.65084 3 ; H12- C5- C6- H14
72 6 8 75 9 0.00 0.69733 3 ; H12- C5- O1- H15
76 11 12 77 9 180.00 15.16700 2 ; H16- C8- C9- H17
78 14 15 79 9 180.00 15.16700 2 ; H18- C11- C12- H19
80 17 26 86 9 0.00 0.65084 3 ; H20- C13- C27- H33
80 17 26 87 9 0.00 0.65084 3 ; H20- C13- C27- H34
81 17 26 86 9 0.00 0.65084 3 ; H21- C13- C27- H33
81 17 26 87 9 0.00 0.65084 3 ; H21- C13- C27- H34
82 19 20 83 9 180.00 15.16700 2 ; H22- C15- C16- H23
84 22 23 85 9 180.00 15.16700 2 ; H24- C18- C19- H25
88 28 29 90 9 0.00 0.65084 3 ; H26- C21- N2- H30
88 28 29 91 9 0.00 0.65084 3 ; H26- C21- N2- H32
88 28 33 99 9 0.00 0.65084 3 ; H26- C21- C30- H45
89 28 29 90 9 0.00 0.65084 3 ; H28- C21- N2- H30
89 28 29 91 9 0.00 0.65084 3 ; H28- C21- N2- H32
89 28 33 99 9 0.00 0.65084 3 ; H28- C21- C30- H45
90 29 30 92 9 0.00 0.65084 3 ; H30- N2- C23- H36
91 29 30 92 9 0.00 0.65084 3 ; H32- N2- C23- H36
92 30 31 93 9 0.00 0.65084 3 ; H36- C23- C25- H38
92 30 31 94 9 0.00 0.65084 3 ; H36- C23- C25- H40
92 30 31 95 9 0.00 0.65084 3 ; H36- C23- C25- H41
92 30 32 96 9 0.00 0.65084 3 ; H36- C23- C28- H42
92 30 32 97 9 0.00 0.65084 3 ; H36- C23- C28- H43
92 30 32 98 9 0.00 0.65084 3 ; H36- C23- C28- H44
99 33 34 100 9 0.00 0.65084 3 ; H45- C30- C32- H46
99 33 34 101 9 0.00 0.65084 3 ; H45- C30- C32- H47
99 33 35 102 9 0.00 0.69733 3 ; H45- C30- O6- H48
103 38 39 104 9 180.00 15.16700 2 ; H49- C36- C37- H50
105 41 42 106 9 180.00 15.16700 2 ; H51- C39- C40- H52
107 44 52 113 9 0.00 0.65084 3 ; H53- C41- C48- H59
107 44 52 114 9 0.00 0.65084 3 ; H53- C41- C48- H60

108 44 52 113 9 0.00 0.65084 3 ; H54- C41- C48- H59
 108 44 52 114 9 0.00 0.65084 3 ; H54- C41- C48- H60
 109 46 47 110 9 180.00 15.16700 2 ; H55- C43- C44- H56
 111 49 50 112 9 180.00 15.16700 2 ; H57- C46- C47- H58
 117 56 57 118 9 180.00 15.16700 2 ; H31- C26- C29- H35
 119 59 60 120 9 180.00 15.16700 2 ; H37- C33- C35- H39

[dihedrals] ; impropers

; treated as propers in GROMACS to use correct AMBER analytical function

i	j	k	l	func	phase	kd	pn
10	12	11	76	4	180.00	4.60240	2 ; C7- C9- C8- H16
10	14	15	79	4	180.00	4.60240	2 ; C7- C11- C12- H19
11	13	12	77	4	180.00	4.60240	2 ; C8- C10- C9- H17
11	15	10	9	4	180.00	4.60240	2 ; C8- C12- C7- O2
12	14	13	16	4	180.00	4.60240	2 ; C9- C11- C10- O3
13	15	14	78	4	180.00	4.60240	2 ; C10- C12- C11- H18
18	20	19	82	4	180.00	4.60240	2 ; C14- C16- C15- H22
18	22	23	85	4	180.00	4.60240	2 ; C14- C18- C19- H25
19	21	20	83	4	180.00	4.60240	2 ; C15- C17- C16- H23
19	23	18	27	4	180.00	4.60240	2 ; C15- C19- C14- O5
20	22	21	24	4	180.00	4.60240	2 ; C16- C18- C17- O4
21	23	22	84	4	180.00	4.60240	2 ; C17- C19- C18- H24
37	39	38	103	4	180.00	4.60240	2 ; C34- C37- C36- H49
37	41	42	106	4	180.00	4.60240	2 ; C34- C39- C40- H52
38	40	39	104	4	180.00	4.60240	2 ; C36- C38- C37- H50
38	42	37	36	4	180.00	4.60240	2 ; C36- C40- C34- O7
39	41	40	43	4	180.00	4.60240	2 ; C37- C39- C38- O8
40	42	41	105	4	180.00	4.60240	2 ; C38- C40- C39- H51
45	47	46	109	4	180.00	4.60240	2 ; C42- C44- C43- H55
45	49	50	112	4	180.00	4.60240	2 ; C42- C46- C47- H58
46	48	47	110	4	180.00	4.60240	2 ; C43- C45- C44- H56
46	50	45	53	4	180.00	4.60240	2 ; C43- C47- C42- O10

47 49 48 51 4 180.00 4.60240 2 ; C44- C46- C45- O9
48 50 49 111 4 180.00 4.60240 2 ; C45- C47- C46- H57
54 57 58 59 4 180.00 4.60240 2 ; C22- C29- C31- C33
55 57 56 117 4 180.00 4.60240 2 ; C24- C29- C26- H31
55 59 60 120 4 180.00 4.60240 2 ; C24- C33- C35- H39
56 58 57 118 4 180.00 4.60240 2 ; C26- C31- C29- H35
56 60 55 25 4 180.00 4.60240 2 ; C26- C35- C24- C20
58 60 59 119 4 180.00 4.60240 2 ; C31- C35- C33- H37

Appendix 5

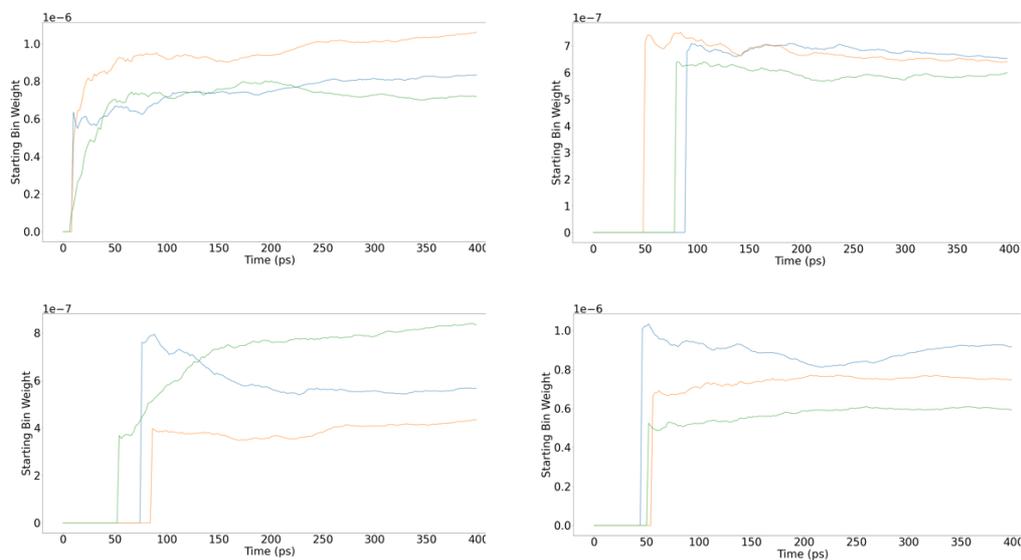


Figure 58 – Starting bin weight for on-rate WEMD simulations against time (ps) showing convergence. Top left: bisoprolol in β_2 -AR using the conventional route. Top right: bisoprolol in β_1 -AR using the conventional route. Bottom left: Ligand **47** in β_1 -AR using the conventional route. Bottom right: Ligand **47** in β_1 -AR using the keyhole route. Each system was run in triplicate.

Appendix 6 – Self-avoiding walk script

```
#!/usr/bin/env python3
from simtk.openmm.app import *
from simtk.openmm import *
from simtk.unit import *
from sys import stdout
import mdtraj as mdt
from mdtraj.reporters import NetCDFReporter
import numpy as np
import argparse
import yaml
from parmed import load_file, unit as u
#from parmed.openmm import StateDataReporter, NetCDFReporter

def waymark(lt, rmsd, atom_indices):
    """
    Identify waymark structures
    """
    r = mdt.rmsd(lt, lt, frame=0, atom_indices=atom_indices)
    j = np.argmax(r > rmsd)
    rlist = np.atleast_2d(r)
    jlist = [0]
    while j > 0:
        jlist.append(j)
        r = mdt.rmsd(lt, lt, frame=j, atom_indices=atom_indices)
        rlist = np.append(rlist, np.atleast_2d(r), axis=0)
        j = np.argmax(rlist.min(axis=0) > rmsd)

    waymarks = jlist
    assignments = rlist.argmax(axis=0)
    counts = [(assignments==i).sum() for i in range(len(jlist))]
    return waymarks, counts

def run(args):
    """
    Executes the self-avoiding walk molecular dynamics cycles
    based on user-defined arguments within a yaml file.
    """
    prmtopfile = args['prmtop']
    inpcrdfile = args['inpcrd']
    selection = args['selection']
    tgtrmsd = args['marksep']
    tgtmdfrc = args['markforce']
    n_cycles = args['ncycles']
    start = args['start']

    cycle_length = 10 # in ps
    save_interval = 1 # in ps
    n_steps = int(cycle_length / 0.002)
    n_save = int(save_interval / 0.002)

    if start > 0:
        trajfile = 'cycle_{:03d}.nc'.format(start - 1)
        tmptraj = mdt.load(trajfile, top=prmtopfile)
        inpcrdfile = 'cycle_{:03d}.ncrst'.format(start - 1)
```

```

    tmptraj[-1].save(inpcrdfile)

    #prmtop = AmberPrmtopFile(prmtopfile)
    inpcrd = AmberInpcrdFile(inpcrdfile)

    prmtop = load_file(prmtopfile)
    #inpcrd = load_file(inpcrdfile)

    tmptraj = mdt.load(inpcrdfile, top=prmtopfile)
    atom_indices = [int(i) for i in
tmptraj.topology.select(selection)]
    seltraj = tmptraj.atom_slice(atom_indices)
    seltraj.save('selection.pdb')
    print('selected atoms saved to file "selection.pdb"')
    ntgtatms = len(atom_indices)
    sigma2 = tgtrmsd * tgtrmsd
    cv_indices = []
    n_cvs = 0
    k_rmsd = tgtmdfrc * ntgtatms
    saw_function = "scalefac_{} * k_rmsd * exp(-(rmsd_{} *
rmsd_{})/(2.0 * sigma2))"

    system = prmtop.createSystem(nonbondedMethod=PME,
        nonbondedCutoff=1*nanometer, constraints=HBonds)
    integrator = LangevinIntegrator(300*kelvin, 1/picosecond,
0.002*picoseconds)
    simulation = Simulation(prmtop.topology, system, integrator)
    platform = simulation.context.getPlatform().getName()
    print('OpenMM is using the {} platform'.format(platform))
    simulation.context.setPositions(inpcrd.positions)
    if inpcrd.boxVectors is not None:

simulation.context.setPeriodicBoxVectors(*inpcrd.boxVectors)
    if start == 0:
        print('minimizing energy...')
        simulation.minimizeEnergy()
    print('Beginning SAW-MD cycles...')
    for icycle in range(start, n_cycles):
        if icycle > 0:
            trajfiles = ['cycle_{:03d}.nc'.format(i) for i in
range(icycle)]
            m_traj = mdt.load(trajfiles, top=prmtopfile)
            print('trajectory = ', m_traj)
            rmsd = tgtrmsd / 2
            waymarks, counts = waymark(m_traj, rmsd, atom_indices)

            n_cvs = len(cv_indices)
            for iref in range(len(waymarks)):
                scalefac = float(counts[iref])
                if iref < n_cvs:

simulation.context.setParameter("scalefac_{}".format(iref),
scalefac)
                else:
                    positions =
Quantity(m_traj.xyz[waymarks[iref]],
nanometers)
                    waymarkfile =
'waymark_{:03d}.pdb'.format(iref)
                    m_traj[waymarks[iref]].save(waymarkfile)

```

```

cv_force=CustomCVForce(saw_function.format(iref, iref, iref))

cv_force.addGlobalParameter("scalefac_{}".format(iref),
                             scalefac)

cv_force.addCollectiveVariable("rmsd_{}".format(iref),
RMSDForce(positions,
atom_indices))

cv_force.addGlobalParameter("k_rmsd".format(iref),
                             k_rmsd)
                             cv_force.addGlobalParameter("sigma2", sigma2)
                             cv_force.setForceGroup(len(cv_indices) + 1)
                             cv_indices.append(system.addForce(cv_force))
state = simulation.context.getState(getPositions=True)
integrator = LangevinIntegrator(300*kelvin,
                                1/picosecond,
                                0.002*picoseconds)
simulation = Simulation(prmtop.topology, system,
integrator)
simulation.context.setPositions(state.getPositions())
boxVectors = state.getPeriodicBoxVectors()
if boxVectors is not None:

simulation.context.setPeriodicBoxVectors(*boxVectors)
    trajfile = 'cycle_{:03d}.nc'.format(icycle)
    simulation.reporters.append(NetCDFReporter(trajfile,
n_save))

    print('running md...')
    simulation.step(n_steps)

    n_cvs = len(cv_indices)
    if n_cvs > 0:
        print('\nwaymarks at end of cycle {}'.format(icycle))
        template = '{:12s}: ' + '{:6d} ' * n_cvs
        print(template.format('snapshot #', *waymarks))
        print(template.format('counts', *counts))
        template = '{:12s}: ' + '{:6.2f} ' * n_cvs + '\n'
        rmsds =
[system.getForce(i).getCollectiveVariableValues(simulation.context
)[0] for i in cv_indices]
        print(template.format('current rmsd', *rmsds))
        cves = [simulation.context.getState(getEnergy=True,
groups=2*i).getPotentialEnergy().value_in_unit(kilojoules/mole)
for i in range(1, len(cv_indices) + 1)]
        print(template.format('current Erep', *cves))

if __name__ == '__main__':

    parser = argparse.ArgumentParser()
    parser.add_argument('yamlfile', help='Configuration file in
YAML format')

    args = parser.parse_args()
    with open(args.yamlfile) as f:
        config = yaml.load(f, Loader=yaml.SafeLoader)

```

```
print (config)
run (config)
```

Appendix 4 – weighted ensemble with pragmatic adaptive binning scheme script

```
# The adaptive approach is as follows:
#
# To begin with there is just the target bin.
# We run a set of MDs.
# If the target bin is reached (4nm Z-coordinate), no more
adaptivity is required.
# If not then:
#     identify the structure with the largest value of the
progress coordinate
#     set a bin edge a fraction behind this.
#     call this "front edge"
#     apply WE reweighting, run simulations.
#     If target bin is reached, no more adaptivity
#     identify structure with the largest value of the progress
coordinate
#     propose a bin edge a fraction behind this.
#     if this bin is in front of "front edge"
#         add this bin edge
#         call it "front edge"
#     apply WE reweighting, run simulations
#
# -----
#
# Import required packages:
#
# In[1]:

from crossflow.kernels import SubprocessKernel
from crossflow.filehandling import FileHandler
from crossflow.clients import Client
import mdtraj as mdt
from mdplus import pca
import matplotlib.pyplot as plt
import networkx as nx
import numpy as np
import pickle
from distributed import Future

# Load the required starting files:

# In[2]:

fh = FileHandler()
mdin = fh.load('md_2ps.in')

# In[3]:
```

```

trajfile = 'reversed_traj.nc'
prmtopfile = 'step5_input.prmtop'
mini_prmtopfile = 'step5_input.prmtop'
selection = '((name CG or name OD1 or name OD2) and residue 85) or
((name CG or name OD1 or name ND2) and residue 265) or ((name N1
or name C1 or name C5 or name O1) and resname GB4)'

```

```
# In[4]:
```

```

traj = mdt.load(trajfile, top=prmtopfile)
sel = traj.topology.select(selection)

```

```
# In[5]:
```

```

crds = fh.load('start_crd.ncrst')
prmtop = fh.load(prmtopfile)
mini_prmtop = fh.load(mini_prmtopfile)

```

```
# In[6]:
```

```

class ProgressCoordinator(object):
    """
    A progress coordinator is trained to generate a 1D progress
    coordinate for a trajectory between start and end points.
    """
    import mdtraj as mdt
    import networkx as nx
    def __init__(self, traj, selection = None, start=None,
end=None):
        """
        Train the pc predictor.

        Args:
            traj: MDTraj trajectory
            selection: MDTraj style atom selection expression
            start: index of the snapshot that defines the start
point
            end: index of the snapshot that defines the end point

        Attributes:
            pathpoints: list of indices of snapshots in traj that
form the
                    shortest pathway from start to end.
            pc: progress coordinate value for each point in the
shortest path. This
                    is the sum of the rmsds between all steps up to
this one.
            path_length: the number of steps in the path.
        """
        if selection is None:
            self._atom_indices = range(len(traj))
        else:
            self._atom_indices = traj.topology.select(selection)
        if start is None:
            start = 0

```

```

    if end is None:
        end = len(traj) - 1
    rmsd2 = []
    for i in range(traj.n_frames):
        rmsd2.append(mdt.rmsd(traj, traj, frame=i,
atom_indices=self._atom_indices))
    rmsd2 = np.array(rmsd2)
    G = nx.Graph()
    for i in range(traj.n_frames - 1):
        rthresh = rmsd2[i, i+1]
        for j in range(i+1, traj.n_frames):
            if rmsd2[i, j] <= rthresh:
                G.add_edge(i, j, weight=rmsd2[i, j])
    self.pathpoints = nx.shortest_path(G, start, end)
    self._ptraj = traj[self.pathpoints]
    prmsd2 = rmsd2[self.pathpoints][:, self.pathpoints]
    self.pc = [0]
    l = 0
    for i in range(1, len(prmsd2)):
        l += prmsd2[i-1, i]
        self.pc.append(l)
    self.path_length = len(self.pc)

def pcoord(self, traj):
    """
    Calculate the progress coordinate for each frame in traj.

    Args:
        traj: an MDTraj trajectory

    Returns:
        pcs: numpy array of predicted progress coordinates
        r2: numpy array of squares of perpendicular distance
of each point from
        closest line segment of shortest path.
    """
    pcs = []
    r2 = []
    for k in range(traj.n_frames):
        r = mdt.rmsd(self._ptraj, traj, frame=k,
atom_indices=self._atom_indices)
        i = np.argmin(r) # closest path point
        rik = r[i]
        # test both path segments: i-1 -> i, and i -> i+1:
        l1 = self.pc[-1] + 1.0
        l2 = self.pc[-1] + 1.0
        if i > 0:
            j = i - 1
            rjk = r[j]
            rij = self.pc[i] - self.pc[j]
            p1 = (rik * rik + rij * rij - rjk * rjk) / (2 *
rij)

            l1 = (rik * rik) - (p1 * p1)
        if i < self.path_length - 1:
            j = i + 1
            rjk = r[j]
            rij = self.pc[j] - self.pc[i]
            p2 = (rik * rik + rij * rij - rjk * rjk) / (2 *
rij)

            l2 = (rik * rik) - (p2 * p2)

```

```

        if l2 > l1: # path i-1 -> is closest
            pck = self.pc[i] - p1
            l = l1
        else: # path i -> i+1 is closest:
            pck = self.pc[i] + p2
            l = l2
        pcs.append(pck)
        r2.append(l)
    return np.array(pcs), np.array(r2)

# Create a SubprocessKernel that runs pmemd. All we need back are
the final coordinates.

# In[8]:

pcoorder = ProgressCoordinator(traj, selection=selection,
start=101)

# In[9]:

pmemd = SubprocessKernel('pmemd.cuda -i mdin -c x.inpcrd -p
x.prmtop -r x.ncrst -o x.log -ref y.ncrst')
pmemd.set_inputs(['mdin', 'x.inpcrd', 'x.parm7', 'y.rst7'])
pmemd.set_outputs(['x.ncrst'])
pmemd.set_constant('mdin', mdin)
pmemd.set_constant('x.prmtop', mini_prmtop)
pmemd.set_constant('y.ncrst', crds)

# Create a crossflow client that talks to our dask cluster:

# In[10]:

client = Client(scheduler_file='dask/dask.dat')

# We will have a target of 4 simulations per WE bin, and the bins
are defined by RMSD from the starting structure, between 0.05 nm
(limit of what is considered 'native' structure) and 0.40 nm
(above which is 'unfolded' target state), in 0.02 nm increments.
We will run the WE simulation for 50 cycles.

# In[11]:

n_cycles = 200
n_reps = 20
bin_edges = [max(pcoorder.pc)]
target_bin = len(bin_edges)
print(bin_edges, flush=True)

# Here is the function that takes a set of coordinates and returns
the bin ids:

```

```

# In[12]:

def get_bins(t, bin_edges):
    pcs, r2 = pcoorder.pcoord(t)
    bin_ids = np.digitize(pcs, bin_edges)
    return bin_ids, pcs

# Here is the workflow:

# In[13]:

class RestartData(object):
    """
    A class to hold everything needed to restart a simulation
    """
    def __init__(self, cycle, coordinates, weights, old_bin_ids,
target_bin, bin_edges, recycled_flux, starting_bin_wt,
sum_leaving_flux):
        self.cycle = cycle # the cycle number
        self.coordinates = [c.result() if isinstance(c, Future)
else c for c in coordinates] # converting the future back to real
data
        self.weights = weights
        self.old_bin_ids = old_bin_ids
        self.target_bin = target_bin
        self.bin_edges = bin_edges
        self.recycled_flux = recycled_flux # this is a list that
grows each cycle
        self.starting_bin_wt = starting_bin_wt # this is a list
that grows each cycle
        self.sum_leaving_flux = sum_leaving_flux # this is a list
that grows each cycle

# In[14]:

restart = True
adapt = False
max_bin_width = 0.2

if restart:
    with open('restart_2ps_20w.pkl', 'rb') as f:
        restart_data = pickle.load(f)
    start_cycle = restart_data.cycle + 1
    starting_coordinates = restart_data.coordinates
    weights = restart_data.weights
    old_bin_ids = restart_data.old_bin_ids
    target_bin = restart_data.target_bin
    bin_edges = restart_data.bin_edges
    recycled_flux = restart_data.recycled_flux
    starting_bin_wt = restart_data.starting_bin_wt
    sum_leaving_flux = restart_data.sum_leaving_flux
    print('Continuing from beginning of cycle
{}'.format(start_cycle), flush=True)
    print('bins=', old_bin_ids, flush=True)
    print('bin edges = ', bin_edges, flush=True)
    print('desired total cycles=', n_cycles, flush=True)

```

```

    print(flush=True)
else:
    # Begin with n_rep copies of the starting structure, with
    equal weights:
    start_cycle = 0
    starting_coordinates = [crds] * n_reps
    weights = [1.0 / n_reps] * n_reps
    # Record what gets recycled each cycle:
    recycled_flux = []
    starting_bin_wt = []
    sum_leaving_flux = []
    old_bin_ids = [0] * n_reps
    print('bins=', old_bin_ids, flush=True)
    print('desired total cycles=', n_cycles, flush=True)
    print(flush=True)
    bin_edges = [max(pcoorder.pc)]
    target_bin = len(bin_edges)

# Run a simulation step on each:
restarts = client.map(pmemd, starting_coordinates)
#print('trajfiles = ', [str(r.result()) for r in restarts])
#print('topfile=', str(mini_prmtop))
# Main loop:
for c in range(start_cycle, n_cycles):
    # Calculate which bin each restart structure falls in to:
    t = mdt.load([str(r.result()) for r in restarts], top =
str(mini_prmtop))
    new_bin_ids, d = get_bins(t, bin_edges)
    print('bins=', new_bin_ids, flush=True)
    '''
    Adaptive binning protocol. A new bin edge is created if the
    progress coordinate
    exceeds the previous highest bin edge until an arbitrary
    distance from the target.
    When the target bin is achieved, adaptive binning terminates.
    '''
    if adapt:
        if target_bin in new_bin_ids:
            adapt = False
            print('target bin reached, adaption turned off',
flush=True)
        else:
            dmax = d.max()
            front_edge = dmax - 0.01
            print('front edge=', front_edge, flush=True)
            front_bin = target_bin - 2
            if front_bin == -1:
                bin_edges.insert(front_bin + 1, front_edge)
            elif bin_edges[front_bin] < front_edge:
                last_front = bin_edges[front_bin]
                new_edge = front_edge
                while new_edge > last_front:
                    bin_edges.insert(front_bin + 1, new_edge)
                    new_edge -= max_bin_width
            target_bin = len(bin_edges)
            new_bin_ids, d = get_bins(t, bin_edges)
            print('adapted bins=', new_bin_ids, flush=True)
            print('bin edges = ', bin_edges, flush=True)

```

```

    # Build the complete flux matrix. Because the number of bins
    can change over the simulation we
    # don't try to use an array for this, but a dictionary
    instead:
    fluxdict = {}
    for n, w in enumerate(new_bin_ids):
        key = (old_bin_ids[w], new_bin_ids[w])
        if not key in fluxdict:
            fluxdict[(old_bin_ids[w], new_bin_ids[w])] =
weights[w]
        else:
            fluxdict[(old_bin_ids[w], new_bin_ids[w])] +=
weights[w]

    # Calculate the weight which has reached the final target bin:
    recycled_weight = np.where(new_bin_ids == target_bin, weights,
0.0).sum()
    print('Cycle {}: recycled weight={}'.format(c,
recycled_weight), flush=True)
    recycled_flux.append(recycled_weight)
    # Recycle any simulations that have reached the target state
    back to the starting state:
    restarts = np.where(new_bin_ids == target_bin, crds, restarts)
    new_bin_ids = np.where(new_bin_ids == target_bin, 0,
new_bin_ids)
    # Assign restart structures to bins and calculate the total
    weight in each bin:
    bins = {}
    bin_wts = {}
    for i, r in enumerate(restarts):
        if new_bin_ids[i] in bins:
            bins[new_bin_ids[i]].append(r)
            bin_wts[new_bin_ids[i]] += weights[i]
        else:
            bins[new_bin_ids[i]] = [r]
            bin_wts[new_bin_ids[i]] = weights[i]
    # Print out the flux "matrix" and bin weights, for post-
    processing. Sorry it's a bit
    # horrible that these just appear for now in the logging, you
    will need to write a
    # new python script to read through this and extract the
    required data...
    print('Bin weights this cycle:', flush=True)
    starting_bin_wt_val = 0
    for index, weight in bin_wts.items():
        print('Bin id: {} weight: {}'.format(index, weight),
flush=True)
        if index == 0:
            starting_bin_wt_val += weight
    starting_bin_wt.append(starting_bin_wt_val)
    sum_leaving_flux_val = 0
    print('Transitions this cycle:', flush=True)
    for key, flux in fluxdict.items():
        print('Transition: {} {}, flux: {}'.format(key[0], key[1],
flux), flush=True)
        if key[0] == 0 and key[1] != 0:
            sum_leaving_flux_val += flux
    sum_leaving_flux.append(sum_leaving_flux_val)
    print("Total flux leaving initial bin:
{}".format(sum_leaving_flux_val), flush=True)
    print(flush=True)

```

```

    # Replicate or cull simulations from each bin to leave n_reps
in each,
    # then reallocate the total bin weight evenly amongst the
structures:
    starting_coordinates = []
    weights = []
    for bin in bins:
        starting_coordinates +=
list(np.random.choice(bins[bin], n_reps))
        weights += [bin_wts[bin] / n_reps] * n_reps
    # Run the next round of simulations:
    old_bin_ids = new_bin_ids
    restart_data = RestartData(c, starting_coordinates, weights,
old_bin_ids, target_bin, bin_edges, recycled_flux,
starting_bin_wt, sum_leaving_flux)
    with open('restart_2ps_20w.pkl', 'wb') as f:
        pickle.dump(restart_data, f, pickle.HIGHEST_PROTOCOL)
    restarts = client.map(pmemd, starting_coordinates)

```

Appendix 7 – example weighted ensemble post-processing script

Test of post-processing methods for WE simulations

```
import numpy as np
from matplotlib import pyplot as plt
%matplotlib inline
```

First section: working out how to scrape the transitions information (which bin each walker started in, and ended up in, each cycle) out of the .out file.

The process is not pretty...

In early cycles, when adaptive/pragmatic binning is active, the initially-allocated destination bins for each walker are reported, but then updated and reported again, before the next cycle runs. So in this case the assignments against "adapted bins" are the ones we want. But later on, once adaption is switched off (because successful recycling has started), its the data from the "bins=" lines that we want.

It's messy because if there are lots of walkers, the numbers for them are spread over multiple lines; these need to be stiched together.

```
# stage 1 = load up everything that looks interesting
resultsfile1 = 'job_2ps_20w-Copy1.out'
collect = False
parsed_data1 = []
with open(resultsfile1) as f:
    for line in f.readlines():
        if 'bins=' in line or 'adapted bins=' in line:
            collect = True
            data= []
            if collect:
                data.append(line[:-1])
            if ']' in line and collect:
                collect = False
                parsed_data1.append(data)
# step 2 - overwrite "bins=" lines with "adapted bins" lines where they exist:
parsed_data2 = []
for d in parsed_data1:
    if '"bins=' in d[0]:
        parsed_data2.append(d)
    if 'adapted bins' in d[0]:
        parsed_data2[-1] = d
# step 3, clean it all up, producing one list on walker final bin ids for each cycle.
# Manually create the first entry, as it's not present in the input file:
parsed_data3 = [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]
for c in parsed_data2[1:]:
    l = ""
    for d in c:
        l += d[:]
    i = l.index('[') + 1
    j = l.index(']')
    parsed_data3.append([int(i) for i in l[i:j].split()])
```

Now we know where each walker ended up each cycle, but we also need to know where they started. We can get this from knowing where they ended up the cycle before.

Example: if at the end of the previous cycle the bin assignments for the 5 walkers were [1,2,0,1,1], then there are now three occupied bins (0, 1, and 2), and there will be `n_reps` walkers starting from each. They are created in the order they first appear in the list, so 1, then 2, then 0: `starts = [1,1,1,1,1,2,2,2,2,2,0,0,0,0,0]`
Also, if a walker has ended up in the target bin, then it will be recycled to bin 0 for the start of the next cycle, so we need to allow for that.

Here's the code:

```
target_bin = 22
n_reps = 20
starts = []
for d in parsed_data3:
    start_bins = []
    for i in d:
        if i == target_bin:
            i = 0
        if not i in start_bins:
            for j in range(n_reps):
                start_bins.append(i)
    starts.append(start_bins)
count = 0
for listElem in starts:
    count += len(listElem)
print('Total Number of independent simulations : ', count)
print('Total elapsed simulation time : ', (count*2)/1000, 'ns')
Total Number of independent simulations : 83660
Total elapsed simulation time : 167.32 ns
# Match up "starts" with "ends". "Ends will be parsed_data3 minus
its first entry,
# and the last entry in "starts" needs to be dropped too:
starts = starts[:-1]
ends = parsed_data3[1:]
```

Now we know where every walker started and ended every cycle, so we can build the transition counts matrix. `tc[i, j]` records how many times a transition was observed from bin `i` to bin `j`:

```
tc = np.zeros((target_bin + 1, target_bin + 1))
for s, e in zip(starts, ends):
    for i in range(len(s)):
        tc[s[i], e[i]] += 1
# There is one missing value in this matrix: we have not so far
captured the recycling
# process. We manually add a value to take care of this:
tc[target_bin, 0] = 1.0
```

Now we need to convert the transition counts matrix `tc` into a proper transition matrix `tmat`. In `tmat`, `tmat[i, j]` is the probability of a walker that starts in bin `i` ending up in bin `j`. We do this by scaling the values in each row of `tc`, so that they always add up to 1 (as the probability that a walker that starts in bin `i` ends up in some bin somewhere, must be 1).

The code below is nice and compact, sorry for the multiple matrix transposes required...

```
tmat = (tc.T / tc.sum(axis=1)).T
print(tmat.sum(axis=1)) # should all be ones
```


Appendix 8 - example weighted ensemble post-processing script for rate equilibration graph

Test of post-processing methods for WE simulations

```
import numpy as np
from matplotlib import pyplot as plt
%matplotlib inline
```

First section: working out how to scrape the transitions information (which bin each walker started in, and ended up in, each cycle) out of the .out file.

The process is not pretty...

In early cycles, when adaptive/pragmatic binning is active, the initially-allocated destination bins for each walker are reported, but then updated and reported again, before the next cycle runs. So in this case the assignments against "adapted bins" are the ones we want. But later on, once adaption is switched off (because succesful recycling has started), its the data from the "bins=" lines that we want.

It's messy because if there are lots of walkers, the numbers for them are spread over multiple lines; these need to be stitched together.

```
# stage 1 = load up everything that looks interesting
resultsfile1 = 'job_2ps_20w-Copy1.out'
collect = False
parsed_data1 = []
with open(resultsfile1) as f:
    for line in f.readlines():
        if 'bins=' in line or 'adapted bins=' in line:
            collect = True
            data= []
            if collect:
                data.append(line[:-1])
            if ']' in line and collect:
                collect = False
                parsed_data1.append(data)
# step 2 - overwrite "bins=" lines with "adapted bins" lines where
they exist:
parsed_data2 = []
for d in parsed_data1:
    if '"bins=' in d[0]:
        parsed_data2.append(d)
    if 'adapted bins' in d[0]:
        parsed_data2[-1] = d
# step 3, clean it all up, producing one list on walker final bin
ids for each cycle.
# Manually create the first entry, as it's not present in the
input file:
parsed_data3 = [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0]]
for c in parsed_data2[1:]:
    l = ""
    for d in c:
        l += d[:]
    i = l.index('[') + 1
    j = l.index(']')
    parsed_data3.append([int(i) for i in l[i:j].split()])
```

Now we know where each walker ended up each cycle, but we also need to know where they started. We can get this from knowing where they ended up the cycle before.

Example: if at the end of the previous cycle the bin assignments for the 5 walkers were [1,2,0,1,1], then there are now three occupied bins (0, 1, and 2), and there will be `n_reps` walkers starting from each. They are created in the order they first appear in the list, so 1, then 2, then 0: `starts = [1,1,1,1,1,2,2,2,2,2,0,0,0,0,0]`
 Also, if a walker has ended up in the target bin, then it will be recycled to bin 0 for the start of the next cycle, so we need to allow for that.

Here's the code:

```
target_bin = 22
n_reps = 20
starts = []
for d in parsed_data3:
    start_bins = []
    for i in d:
        if i == target_bin:
            i = 0
        if not i in start_bins:
            for j in range(n_reps):
                start_bins.append(i)
    starts.append(start_bins)
count = 0
for listElem in starts:
    count += len(listElem)
print('Total Number of independent simulations : ', count)
print('Total elapsed simulation time : ', (count*2)/1000, 'ns')
Total Number of independent simulations : 82240
Total elapsed simulation time : 164.48 ns
# Match up "starts" with "ends". "Ends will be parsed_data3 minus
its first entry,
# and the last entry in "starts" needs to be dropped too:
starts = starts[:-1]
ends = parsed_data3[1:]
```

Now we know where every walker started and ended every cycle, so we can build the transition counts matrix. `tc[i, j]` records how many times a transition was observed from bin `i` to bin `j`:

```
def step(w_old, tm):
    """
    Given a transition matrix tm and current set of bin weights,
    calculate new bin weights"""
    n_bins = len(w_old)
    w_new = np.zeros(n_bins)
    for i in range(n_bins):
        for j in range(n_bins):
            w_new[j] += w_old[i] * tm[i, j]
    w_new = w_new / w_new.sum() # ensures sum of weights is
exactly 1.
    flux = (tmat.T * w_new).T
    on_flux = flux[0,1:].sum()
    rate = flux[0,1:].sum() / (2 * w_new[0] * 3.132940635e-19 *
6.0221409e+23)
    return w_new, on_flux, rate
runs = list(map(list, zip(starts, ends)))
import itertools
of_list = []
r_list = []
```

```
for c in range(len(runs)):
```

Here we produce a list containing the calculated on-rate after each step.

```
tc = np.zeros((target_bin + 1, target_bin + 1))
for s, e in itertools.islice(zip(starts, ends), 0, c):
    for i in range(len(s)):
        tc[s[i], e[i]] += 1

tc[target_bin, 0] = 1.0
tmat = (tc.T / tc.sum(axis=1)).T

old_wts = np.zeros(target_bin + 1)
old_wts[0] = 1.0

for i in range(1000):
    new_wts, on_flux, rate = step(old_wts, tmat)
    old_wts = new_wts

of_list.append(on_flux)
r_list.append(rate)
```

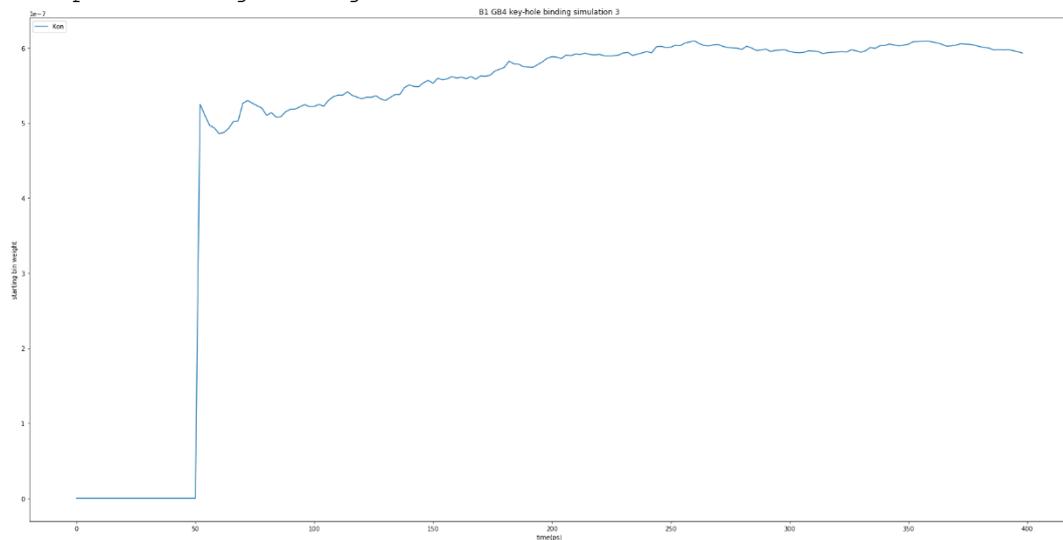
Here we plot the list containing the calculated on-rate after each step.

```
of_list = np.nan_to_num(of_list)
r_list = np.nan_to_num(r_list)
from matplotlib import pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = [30, 15]
flux_out = np.array(of_list)
raterr = np.array(r_list)

time_ps = [i * 2 for i in range(200)]

plt.plot(time_ps, raterr, label='Kon')
plt.title('B1 GB4 key-hole binding simulation 3')
plt.xlabel('time (ps)')
plt.ylabel('starting bin weight')

plt.legend(loc="upper left")
<matplotlib.legend.Legend at 0x7f135b771208>
```



```
import pickle as pkl
with open('rate.pkl', 'wb') as f:
```

```
pkl.dump(r_list, f)
```

References

- (1) Louison, K. A.; Dryden, I. L.; Laughton, C. A. GLIMPS: A Machine Learning Approach to Resolution Transformation for Multiscale Modeling. *Cite This: J. Chem. Theory Comput* **2021**, *17*, 7930–7937. <https://doi.org/10.1021/acs.jctc.1c00735>.
- (2) Insel, P. A.; Sriram, K.; Gorr, M. W.; Wiley, S. Z.; Michkov, A.; Salmerón, C.; Chinn, A. M. GPCRomics: An Approach to Discover GPCR Drug Targets. *Trends in Pharmacological Sciences* **2019**, *40* (6), 378–387. <https://doi.org/10.1016/j.tips.2019.04.001>.
- (3) Sriram, K.; Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Molecular Pharmacology* **2018**, *93* (4), 251–258. <https://doi.org/10.1124/mol.117.111062>.
- (4) Wootten, D.; Christopoulos, A.; Marti-Solano, M.; Babu, M. M.; Sexton, P. M. Mechanisms of Signalling and Biased Agonism in G Protein-Coupled Receptors. *Nature Reviews Molecular Cell Biology* **2018**, *19* (10), 638–653. <https://doi.org/10.1038/s41580-018-0049-3>.
- (5) Hargrave, P. A.; McDowell, J. H.; Curtis, D. R.; Wang, J. K.; Juszczak, E.; Fong, S. L.; Rao, J. K.; Argos, P. The Structure of Bovine Rhodopsin. *Biophys Struct Mech* **1983**, *9* (4), 235–244. <https://doi.org/10.1007/BF00535659>.
- (6) Lagerström, M. C.; Schiöth, H. B. Structural Diversity of G Protein-Coupled Receptors and Significance for Drug Discovery. *Nature Reviews Drug Discovery* **2008**, *7* (4), 339–357. <https://doi.org/10.1038/nrd2518>.
- (7) Lefkowitz, R. J. Historical Review: A Brief History and Personal Retrospective of Seven-Transmembrane Receptors. *Trends in Pharmacological Sciences* **2004**, *25* (8), 413–422. <https://doi.org/10.1016/j.tips.2004.06.006>.
- (8) Li, J.; Edwards, P. C.; Burghammer, M.; Villa, C.; Schertler, G. F. X. Structure of Bovine Rhodopsin in a Trigonal Crystal Form. *Journal of Molecular Biology* **2004**, *343* (5), 1409–1438. <https://doi.org/10.1016/j.jmb.2004.08.090>.
- (9) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.; Yao, X.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K.

- GPCR Engineering Yields High-Resolution Structural Insights into Beta2-Adrenergic Receptor Function. *Science* **2007**, *318* (November), 1266–1273.
- (10) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. a; Søren, G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.; Kuhn, P.; Weis, I.; Kobilka, B. K.; Stevens, R. C. High Resolution Crystal Structure of an Engineered Human B2-Adrenergic G Protein-Coupled Receptor. *Science* **2007**, *318* (5854), 1258–1265. <https://doi.org/10.1126/science.1150577.High>.
- (11) Rasmussen, S. G. F.; Choi, H.-J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R. P.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F. X.; Weis, W. I.; Kobilka, B. K. Crystal Structure of the Human B2 Adrenergic G-Protein-Coupled Receptor. *Nature* **2007**, *450* (7168), 383–387. <https://doi.org/10.1038/nature06325>.
- (12) Rasmussen, S. G. F.; Devree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal Structure of the B2adrenergic Receptor-Gs Protein Complex. *Nature* **2011**, *477* (7366), 549–557. <https://doi.org/10.1038/nature10361>.
- (13) Clark, R. B. Profile of Brian K. Kobilka and Robert J. Lefkowitz, 2012 Nobel Laureates in Chemistry. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (14), 5274–5275. <https://doi.org/10.1073/pnas.1221820110>.
- (14) Caffrey, M. Crystallizing Membrane Proteins for Structure-Function Studies Using Lipidic Mesophases. *NATO Science for Peace and Security Series A: Chemistry and Biology* **2013**, 33–46. <https://doi.org/10.1007/978-94-007-6232-9-4>.
- (15) Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. The Structure and Function of G-Protein-Coupled Receptors. *Nature* **2014**, *459* (7245), 356–363. <https://doi.org/10.1038/nature08144.The>.
- (16) Gurevich, V. v.; Gurevich, E. v. GPCR Signaling Regulation: The Role of GRKs and Arrestins. *Frontiers in Pharmacology* **2019**, *10* (FEB), 1–11. <https://doi.org/10.3389/fphar.2019.00125>.

- (17) Weis, W. I.; Kobilka, B. K. The Molecular Basis of G Protein–Coupled Receptor Activation. *Annual Review of Biochemistry* **2014**, *87* (1), 897–919. <https://doi.org/10.1146/annurev-biochem-060614-033910>.
- (18) Ma, L.; Pei, G. B-Arrestin Signaling and Regulation of Transcription. *Journal of Cell Science* **2007**, *120* (2), 213–218. <https://doi.org/10.1242/jcs.03338>.
- (19) Hoffman, B. B.; Lefkowitz, R. J. Adrenergic Receptors in the Heart. *Annual Review of Physiology* **1982**, *44* (March 1982), 475–482. <https://doi.org/https://doi.org/10.1146/annurev.ph.44.030182.002355>.
- (20) Jones, A. M.; Assmann, S. M. Plants: The Latest Model System for G-Protein Research. *EMBO Reports* **2004**, *5* (6), 572–578. <https://doi.org/10.1038/sj.embor.7400174>.
- (21) Dzimiri, N. Regulation of B-Adrenoceptor Signaling in Cardiac Function and Disease. *Pharmacol.Rev.* **1999**, *51* (3), 465–501.
- (22) Lelias, J. M.; Kaghad, M.; Rodriguez, M.; Chalon, P.; Bonnin, J.; Dupre, I.; Delpech, B.; Bensaid, M.; LeFur, G.; Ferrara, P.; Caput, D. Molecular Cloning of a Human B3-Adrenergic Receptor CDNA. *FEBS Letters* **1993**, *324* (2), 127–130. [https://doi.org/10.1016/0014-5793\(93\)81377-C](https://doi.org/10.1016/0014-5793(93)81377-C).
- (23) Venkatakrisnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Madan Babu, M. Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **2013**, *494* (7436), 185–194. <https://doi.org/10.1038/nature11896>.
- (24) Lee, Y.; Basith, S.; Choi, S. Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. *Journal of Medicinal Chemistry* **2018**, *61* (1), 1–46. <https://doi.org/10.1021/acs.jmedchem.6b01453>.
- (25) Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. The Retinal Conformation and Its Environment in Rhodopsin in Light of a New 2.2 Å Crystal Structure. *Journal of Molecular Biology* **2004**, *342* (2), 571–583. <https://doi.org/10.1016/j.jmb.2004.07.044>.
- (26) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G. W. W.; Tate, C. G.; Schertler, G. F. X. X. Structure of a B1-Adrenergic G-Protein-Coupled Receptor. *Nature* **2008**, *454* (7203), 486–491. <https://doi.org/10.1038/nature07101>.

- (27) Jaakola, V.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Ellen, Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 Å Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322* (5905), 1211–1217. <https://doi.org/10.1126/science.1164772>.The.
- (28) Ballesteros, J. A.; Weinstein, H. Integrated Methods for the Construction of Three Dimensional Models and Computational Probing of Structure-Function Relations in G-Protein Coupled Receptors. *Methods Neurosci.* **1995**, *25*, 366–428. [https://doi.org/10.1016/S1043-9471\(05\)80049-7](https://doi.org/10.1016/S1043-9471(05)80049-7).
- (29) Shi, L.; Liapakis, G.; Xu, R.; Guarnieri, F.; Ballesteros, J. A.; Javitch, J. A. B2 Adrenergic Receptor Activation: Modulation of the Proline Kink in Transmembrane 6 by a Rotamer Toggle Switch. *Journal of Biological Chemistry* **2002**, *277* (43), 40989–40996. <https://doi.org/10.1074/jbc.M206801200>.
- (30) Padayatti, P. S.; Wang, L.; Gupta, S.; Orban, T.; Sun, W.; Salom, D.; Jordan, S. R.; Palczewski, K.; Chance, M. R. A Hybrid Structural Approach to Analyze Ligand Binding by the Serotonin Type 4 Receptor (5-HT4). *Molecular and Cellular Proteomics* **2013**, *12* (5), 1259–1271. <https://doi.org/10.1074/mcp.M112.025536>.
- (31) Swaminath, G.; Deupi, X.; Lee, T. W.; Zhu, W.; Thian, F. S.; Kobilka, T. S.; Kobilka, B. Probing the B2 Adrenoceptor Binding Site with Catechol Reveals Differences in Binding and Activation by Agonists and Partial Agonists. *Journal of Biological Chemistry* **2005**, *280* (23), 22165–22171. <https://doi.org/10.1074/jbc.M502352200>.
- (32) Horn, F. GPCRDB Information System for G Protein-Coupled Receptors. *Nucleic Acids Research* **2003**, *31* (1), 294–297. <https://doi.org/10.1093/nar/gkg103>.
- (33) Conn, P. J.; Christopoulos, A.; Lindsley, C. W. Allosteric Modulators of GPCRs: A Novel Approach for the Treatment of CNS Disorders. *Nature reviews drug discovery* **2010**, *8* (1), 41–54. <https://doi.org/10.1038/nrd2760>.Allosteric.

- (34) Baker, J. G. The Selectivity of β -Adrenoceptor Antagonists at the Human B1, B2 and B3 Adrenoceptors. *British Journal of Pharmacology* **2005**, *144* (3), 317–322. <https://doi.org/10.1038/sj.bjp.0706048>.
- (35) Sugimoto, Y.; Fujisawa, R.; Tanimura, R.; Lattion, A. L.; Cotecchia, S. B1-Selective Agonist (-)-1-(3,4-Dimethoxyphenethylamino)-3-(3,4-Dihydroxy)-2-Propanol [(-)-RO363] Differentially Interacts with Key Amino Acids Responsible for B1-Selective Binding in Resting and Active States. *Pharmacology* **2002**, *301* (1), 51–58.
- (36) Xu, X.; Kaindl, J.; Clark, M. J.; Hübner, H.; Hirata, K.; Sunahara, R. K.; Gmeiner, P.; Kobilka, B. K.; Liu, X. Binding Pathway Determines Norepinephrine Selectivity for the Human B1AR over B2AR. *Cell Research* **2020**, No. September. <https://doi.org/10.1038/s41422-020-00424-2>.
- (37) Kobilka, B. K. G Protein Coupled Receptor Structure and Activation. *Biochimica et biophysica acta* **2007**, *1768* (4), 794–807. <https://doi.org/10.1016/j.bbamem.2006.10.021>.
- (38) Yao, X.; Parnot, C.; Deupi, X.; Ratnala, V. R. P.; Swaminath, G.; Farrens, D.; Kobilka, B. Coupling Ligand Structure to Specific Conformational Switches in the B2-Adrenoceptor. *Nature Chemical Biology* **2006**, *2*, 417.
- (39) Baker, J. G.; Proudman, R. G. W.; Hawley, N. C.; Fischer, P. M.; Hill, S. J. Role of Key Transmembrane Residues in Agonist and Antagonist Actions at the Two Conformations of the Human Beta1-Adrenoceptor. *Molecular pharmacology* **2008**, *74* (5), 1246–1260. <https://doi.org/10.1124/mol.108.048371>.
- (40) Kenakin, T. Being Mindful of Seven-Transmembrane Receptor “guests” When Assessing Agonist Selectivity. *British Journal of Pharmacology* **2010**, *160* (5), 1045–1047. <https://doi.org/10.1111/j.1476-5381.2010.00764.x>.
- (41) Tate, C. G. A Crystal Clear Solution for Determining G-Protein-Coupled Receptor Structures. *Trends in Biochemical Sciences* **2012**, *37* (9), 343–352. <https://doi.org/10.1016/j.tibs.2012.06.003>.
- (42) de Deurwaerdère, P.; Navailles, S.; Berg, K. A.; Clarke, W. P.; Spampinato, U. Constitutive Activity of the Serotonin2C Receptor Inhibits In Vivo Dopamine Release in the Rat Striatum and Nucleus Accumbens. *Journal of*

- Neuroscience* **2004**, *24* (13), 3235–3241.
<https://doi.org/10.1523/JNEUROSCI.0112-04.2004>.
- (43) Baker, J. G.; Hill, S. J.; Summers, R. J. Evolution of β -Blockers: From Anti-Anginal Drugs to Ligand-Directed Signalling. *Trends in Pharmacological Sciences* **2011**, *32* (4), 227–234. <https://doi.org/10.1016/j.tips.2011.02.010>.
- (44) Christopoulos, A. G Protein-Coupled Receptor Allostereism and Complexing. *Pharmacological Reviews* **2002**, *54* (2), 323–374. <https://doi.org/10.1124/pr.54.2.323>.
- (45) Congreve, M.; Langmead, C. J.; Mason, J. S.; Marshall, F. H. Progress in Structure Based Drug Design for G Protein-Coupled Receptors. *Journal of Medicinal Chemistry* **2011**, *54* (13), 4283–4311. <https://doi.org/10.1021/jm200371q>.
- (46) Gregory, K. J.; Nguyen, E. D.; Reiff, S. D.; Squire, E. F.; Stauffer, S. R.; Lindsley, C. W.; Meiler, J.; Conn, P. J. Probing the Metabotropic Glutamate Receptor 5 (mGlu5) Positive Allosteric Modulator (PAM) Binding Pocket: Discovery of Point Mutations That Engender a “Molecular Switch” in PAM Pharmacology. *Molecular Pharmacology* **2013**, *83* (5), 991–1006. <https://doi.org/10.1124/mol.112.083949>.
- (47) Foster, D. J.; Conn, P. J. Allosteric Modulation of GPCRs: New Insights and Potential Utility for Treatment of Schizophrenia and Other CNS Disorders. *Neuron* **2017**, *94* (3), 431–446. <https://doi.org/10.1016/j.neuron.2017.03.016>.
- (48) Tikhonova, I. G.; Selvam, B.; Ivetac, A.; Wereszczynski, J.; McCammon, J. A. Simulations of Biased Agonists in the Beta(2) Adrenergic Receptor with Accelerated Molecular Dynamics. *Biochemistry* **2013**, *52* (33), 5593–5603. <https://doi.org/10.1021/bi400499n>.
- (49) Liu, J.; Horst, R.; Katritch, V.; Stevens, R.; Wuthrich, K. Biased Signalling Pathways in Beta2-Adrenergic Receptor Characterized by 19F-NMR. *Science* **2013**, *335* (6072), 1106–1110. <https://doi.org/10.1126/science.1215802.Biased>.
- (50) Warne, T.; Edwards, P. C.; Leslie, A. G. W.; Tate, C. G. Crystal Structures of a Stabilized B1-Adrenoceptor Bound to the Biased Agonists Bucindolol and

- Carvedilol. *Structure* **2012**, *20* (5), 841–849.
<https://doi.org/10.1016/j.str.2012.03.014>.
- (51) Lympereopoulos, A.; Negussie, S. β -Arrestins in Cardiac G Protein-Coupled Receptor Signaling and Function: Partners in Crime or “Good Cop, Bad Cop”? *International Journal of Molecular Sciences* **2013**, *14* (12), 24726–24741. <https://doi.org/10.3390/ijms141224726>.
- (52) Swaminath, G.; Steenhuis, J.; Kobilka, B.; Lee, T. a E. W.; Hughes, H. Allosteric Modulation of β_2 -Adrenergic Receptor by Zn²⁺. **2002**, *61* (1), 65–72.
- (53) Depardieu, F.; Bonora, M. G.; Reynolds, P. E.; Courvalin, P. The VanG Glycopeptide Resistance Operon from *Enterococcus Faecalis* Revisited. *Molecular Microbiology* **2003**, *50* (3), 931–948.
<https://doi.org/10.1046/j.1365-2958.2003.03737.x>.
- (54) Liu, X.; Masoudi, A.; Kahsai, A. W.; Huang, L. Y.; Pani, B.; Staus, D. P.; Shim, P. J.; Hirata, K.; Simhal, R. K.; Schwalb, A. M.; Rambarat, P. K.; Ahn, S.; Lefkowitz, R. J.; Kobilka, B. Mechanism of β_2 AR Regulation by an Intracellular Positive Allosteric Modulator. *Science* **2019**, *364* (6447), 1283–1287. <https://doi.org/10.1126/science.aaw8981>.
- (55) Ahn, S.; Pani, B.; Kahsai, A. W.; Olsen, E. K.; Husemoen, G.; Vestergaard, M.; Jin, L.; Zhao, S.; Wingler, L. M.; Rambarat, P. K.; Simhal, R. K.; Xu, T. T.; Sun, L. D.; Shim, P. J.; Staus, D. P.; Huang, L. Y.; Franch, T.; Chen, X.; Lefkowitz, R. J. Small-Molecule Positive Allosteric Modulators of the β_2 -Adrenoceptor Isolated from DNA-Encoded Libraries. *Molecular Pharmacology* **2018**, *94* (2), 850–861. <https://doi.org/10.1124/mol.118.111948>.
- (56) Baker, J. G.; Hall, I. P.; Hill, S. J. Agonist Actions of “Beta-Blockers” Provide Evidence for Two Agonist Activation Sites or Conformations of the Human β_1 -Adrenoceptor. *Molecular pharmacology* **2003**, *63* (6), 1312–1321.
<https://doi.org/10.1124/mol.63.6.1312>.
- (57) Baker, J. G. Evidence for a Secondary State of the Human β_3 - Adrenoceptor. *Molecular Pharmacology* **2005**, *68* (6), 1645–1655.
<https://doi.org/10.1124/mol.105.015461.described>.

- (58) Soriano-Ursúa, M. A.; Trujillo-Ferrara, J. G.; Correa-Basurto, J.; Vilar, S. Recent Structural Advances of α_1 and α_2 Adrenoceptors Yield Keys for Ligand Recognition and Drug Design. *Journal of Medicinal Chemistry* **2013**, *56* (21), 8207–8223. <https://doi.org/10.1021/jm400471z>.
- (59) Arnatt, C. K.; Zhang, Y. Bivalent Ligands Targeting Chemokine Receptor Dimerization: Molecular Design and Functional Studies. *Current topics in medicinal chemistry* **2014**, *14* (13), 1606–1618. <https://doi.org/10.2174/1568026614666140827144752>.
- (60) Gouldson, P. R.; Ph, D.; Higgs, C.; Ph, D.; Smith, R. E.; Sc, B.; Dean, M. K.; Sc, B.; Gkoutos, G. v; Sc, M.; Reynolds, C. A.; Ph, D. Dimerization and Domain Swapping in G-Protein-Coupled Receptors : A Computational Study. **1999**, No. 00.
- (61) Karellas, P.; McNaughton, M.; Baker, S. P.; Scammells, P. J. Synthesis of Bivalent Beta2-Adrenergic and Adenosine A1 Receptor Ligands. *Journal of medicinal chemistry* **2008**, *51* (19), 6128–6137. <https://doi.org/10.1021/jm800613s>.
- (62) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular Determinants of Drug-Receptor Binding Kinetics. *Drug Discovery Today* **2013**, *18* (13–14), 667–673. <https://doi.org/10.1016/j.drudis.2013.02.007>.
- (63) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-Target Residence Time and Its Implications for Lead Optimization. *Nature Reviews Drug Discovery* **2006**, *5* (9), 730–739. <https://doi.org/10.1038/nrd2082>.
- (64) Yin, N.; Pei, J.; Lai, L. A Comprehensive Analysis of the Influence of Drug Binding Kinetics on Drug Action at Molecular and Systems Levels. *Molecular BioSystems* **2013**, *9* (6), 1381–1389. <https://doi.org/10.1039/c3mb25471b>.
- (65) Tummino, P. J.; Copeland, R. A. Residence Time of Receptor-Ligand Complexes and Its Effect on Biological Function (Biochemistry (2008) 47, 20, (5481-5492)). *Biochemistry* **2008**, *47* (32), 8465. <https://doi.org/10.1021/bi801226m>.
- (66) Lu, H.; Tonge, P. J. Drug-Target Residence Time: Critical Information for Lead Optimization. *Current Opinion in Chemical Biology* **2010**, *14* (4), 467–474. <https://doi.org/10.1016/j.cbpa.2010.06.176>.

- (67) Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. *Annual Review of Physical Chemistry* **2019**, *70*, 143–171. <https://doi.org/10.1146/annurev-physchem-042018-052340>.
- (68) van't Hoff, M. J. H. Etudes de Dynamique Chimique. *Recueil des Travaux Chimiques des Pays-Bas* **1884**, *3* (10), 333–336. <https://doi.org/10.1002/recl.18840031003>.
- (69) Arrhenius, S. Über Die Reaktionsgeschwindigkeit Bei Der Inversion von Rohrzucker Durch Säuren. *Zeitschrift für Physikalische Chemie* **1889**, *4U* (1), 226–248. <https://doi.org/10.1515/zpch-1889-0416>.
- (70) Farkas, L. Keimbildungsgeschwindigkeit in Übersättigten Dämpfen. *Zeitschrift für Physikalische Chemie* **1927**, *125U* (1), 236–242. <https://doi.org/10.1515/zpch-1927-12513>.
- (71) Ferruz, N.; de Fabritiis, G. Binding Kinetics in Drug Discovery. *Molecular Informatics* **2016**, *35* (6–7), 216–226. <https://doi.org/10.1002/minf.201501018>.
- (72) Hulme, E. C.; Trevethick, M. A. Ligand Binding Assays at Equilibrium: Validation and Interpretation. *British Journal of Pharmacology* **2010**, *161* (6), 1219–1237. <https://doi.org/10.1111/j.1476-5381.2009.00604.x>.
- (73) Ilien, B.; Franchet, C.; Bernard, P.; Morisset, S.; Weill, C. O.; Bourguignon, J. J.; Hibert, M.; Galzi, J. L. Fluorescence Resonance Energy Transfer to Probe Human M1 Muscarinic Receptor Structure and Drug Binding Properties. *Journal of Neurochemistry* **2003**, *85* (3), 768–778. <https://doi.org/10.1046/j.1471-4159.2003.01717.x>.
- (74) Meyners, C.; Baud, M. G. J.; Fuchter, M. J.; Meyer-Almes, F. J. Kinetic Method for the Large-Scale Analysis of the Binding Mechanism of Histone Deacetylase Inhibitors. *Analytical Biochemistry* **2014**, *460*, 39–46. <https://doi.org/10.1016/j.ab.2014.05.014>.
- (75) Hiller, C.; Kühhorn, J.; Gmeiner, P. Class A G-Protein-Coupled Receptor (GPCR) Dimers and Bivalent Ligands. *Journal of Medicinal Chemistry* **2013**, *56* (17), 6542–6559. <https://doi.org/10.1021/jm4004335>.
- (76) Felce, J. H.; Davis, S. J.; Klenerman, D. Single-Molecule Analysis of G Protein-Coupled Receptor Stoichiometry: Approaches and Limitations. *Trends in*

- Pharmacological Sciences* **2018**, 39 (2), 96–108.
<https://doi.org/10.1016/j.tips.2017.10.005>.
- (77) Milligan, G.; Ward, R. J.; Marsango, S. GPCR Homo-Oligomerization. *Current Opinion in Cell Biology* **2019**, 57, 40–47.
<https://doi.org/10.1016/j.ceb.2018.10.007>.
- (78) Dijkman, P. M.; Castell, O. K.; Goddard, A. D.; Munoz-Garcia, J. C.; de Graaf, C.; Wallace, M. I.; Watts, A. Dynamic Tuneable G Protein-Coupled Receptor Monomer-Dimer Populations. *Nature Communications* **2018**, 9 (1).
<https://doi.org/10.1038/s41467-018-03727-6>.
- (79) Pediani, J. D.; Ward, R. J.; Marsango, S.; Milligan, G. Spatial Intensity Distribution Analysis: Studies of G Protein-Coupled Receptor Oligomerisation. *Trends in Pharmacological Sciences* **2018**, 39 (2), 175–186.
<https://doi.org/10.1016/j.tips.2017.09.001>.
- (80) Nenasheva, T. A.; Neary, M.; Mashanov, G. I.; Birdsall, N. J. M.; Breckenridge, R. A.; Molloy, J. E. Abundance, Distribution, Mobility and Oligomeric State of M2 Muscarinic Acetylcholine Receptors in Live Cardiac Muscle. *Journal of Molecular and Cellular Cardiology* **2013**, 57 (1), 129–136.
<https://doi.org/10.1016/j.yjmcc.2013.01.009>.
- (81) Kroeger, K. M.; Pflieger, K. D. G.; Eidne, K. A. G-Protein Coupled Receptor Oligomerization in Neuroendocrine Pathways. *Frontiers in Neuroendocrinology* **2003**, 24 (4), 254–278.
<https://doi.org/10.1016/j.yfrne.2003.10.002>.
- (82) Szidonya, L.; Cserzo, M.; Hunyady, L. Dimerization and Oligomerization of G-Protein-Coupled Receptors: Debated Structures with Established and Emerging Functions. *Journal of Endocrinology* **2008**, 196 (3), 435–453.
<https://doi.org/10.1677/JOE-07-0573>.
- (83) Hebert, T. E.; Moffett, S.; Morello, J. P.; Loisel, T. P.; Bichet, D. G.; Barret, C.; Bouvier, M. A Peptide Derived from a β 2-Adrenergic Receptor Transmembrane Domain Inhibits Both Receptor Dimerization and Activation. *Journal of Biological Chemistry* **1996**, 271 (27), 16384–16392.
<https://doi.org/10.1074/jbc.271.27.16384>.

- (84) Pflieger, K. D. G.; Eidne, K. A. Monitoring the Formation of Dynamic G-Protein-Coupled Receptor-Protein Complexes in Living Cells. *Biochemical Journal* **2005**, *385* (3), 625–637. <https://doi.org/10.1042/BJ20041361>.
- (85) Förster, Th. Zwischenmolekulare Energiewanderung Und Fluoreszenz. *Annalen der Physik* **1948**, *437* (1–2), 55–75. <https://doi.org/10.1002/andp.19484370105>.
- (86) Terrillon, S.; Bouvier, M. Roles of G-Protein-Coupled Receptor Dimerization. *EMBO Reports* **2004**, *5* (1), 30–34. <https://doi.org/10.1038/sj.embor.7400052>.
- (87) Petäjä-Repo, U. E.; Hogue, M.; Laperrière, A.; Walker, P.; Bouvier, M. Export from the Endoplasmic Reticulum Represents the Limiting Step in the Maturation and Cell Surface Expression of the Human δ Opioid Receptor. *Journal of Biological Chemistry* **2000**, *275* (18), 13727–13736. <https://doi.org/10.1074/jbc.275.18.13727>.
- (88) Petäjä-Repo, U. E.; Hogue, M.; Laperrière, A.; Bhalla, S.; Walker, P.; Bouvier, M. Newly Synthesized Human δ Opioid Receptors Retained in the Endoplasmic Reticulum Are Retrotranslocated to the Cytosol, Deglycosylated, Ubiquitinated, and Degraded by the Proteasome. *Journal of Biological Chemistry* **2001**, *276* (6), 4416–4423. <https://doi.org/10.1074/jbc.M007151200>.
- (89) Marshall, F. H.; Jones, K. A.; Kaupmann, K.; Bettler, B. GABA(B) Receptors - The First 7TM Heterodimers. *Trends in Pharmacological Sciences* **1999**, *20* (10), 396–399. [https://doi.org/10.1016/S0165-6147\(99\)01383-8](https://doi.org/10.1016/S0165-6147(99)01383-8).
- (90) Roess, D. A.; Smith, S. M. L. Self-Association and Raft Localization of Functional Luteinizing Hormone Receptors. *Biology of Reproduction* **2003**, *69* (6), 1765–1770. <https://doi.org/10.1095/biolreprod.103.018846>.
- (91) Cheng, Z. J.; Miller, L. J. Agonist-Dependent Dissociation of Oligomeric Complexes of G Protein-Coupled Cholecystokinin Receptors Demonstrated in Living Cells Using Bioluminescence Resonance Energy Transfer. *Journal of Biological Chemistry* **2001**, *276* (51), 48040–48047. <https://doi.org/10.1074/jbc.M105668200>.

- (92) Terrillon, S.; Durroux, T.; Mouillac, B.; Breit, A.; Ayoub, M. A.; Taulan, M.; Jockers, R.; Barberis, C.; Bouvier, M. Oxytocin and Vasopressin V1a and V2 Receptors Form Constitutive Homo- and Heterodimers during Biosynthesis. *Molecular Endocrinology* **2003**, *17* (4), 677–691. <https://doi.org/10.1210/me.2002-0222>.
- (93) Jordan, B. A.; Devi, L. A. G-Protein-Coupled Receptor Heterodimerization Modulates Receptor Function. *Nature* **1999**, *399* (6737), 697–700. <https://doi.org/10.1038/21441>.
- (94) Galvez, T.; Duthey, B.; Kniazeff, J.; Blahos, J.; Rovelli, G.; Bettler, B.; Prézeau, L.; Pin, J. P. Allosteric Interactions between GB1 and GB2 Subunits Are Required for Optimal GABAB Receptor Function. *EMBO Journal* **2001**, *20* (9), 2152–2159. <https://doi.org/10.1093/emboj/20.9.2152>.
- (95) Rocheville, M.; Lange, D. C.; Kumar, U.; Sasi, R.; Patel, R. C.; Patel, Y. C. Subtypes of the Somatostatin Receptor Assemble as Functional Homo- and Heterodimers. *Journal of Biological Chemistry* **2000**, *275* (11), 7862–7869. <https://doi.org/10.1074/jbc.275.11.7862>.
- (96) Franco, R.; Martínez-Pinilla, E.; Lanciego, J. L.; Navarro, G. Basic Pharmacological and Structural Evidence for Class A G-Protein-Coupled Receptor Heteromerization. *Frontiers in Pharmacology* **2016**, *7*, 76. <https://doi.org/10.3389/fphar.2016.00076>.
- (97) Margeta-Mitrovic, M.; Jan, Y. N.; Jan, L. Y. A Trafficking Checkpoint Controls GABA(B) Receptor Heterodimerization. *Neuron* **2000**, *27* (1), 97–106. [https://doi.org/10.1016/S0896-6273\(00\)00012-X](https://doi.org/10.1016/S0896-6273(00)00012-X).
- (98) Hillion, J.; Canals, M.; Torvinen, M.; Casadó, V.; Scott, R.; Terasmaa, A.; Hansson, A.; Watson, S.; Olah, M. E.; Mallol, J.; Canela, E. I.; Zoli, M.; Agnati, L. F.; Ibáñez, C. F.; Lluís, C.; Franco, R.; Ferré, S.; Fuxe, K. Coaggregation, Cointernalization, and Codesensitization of Adenosine A2A Receptors and Dopamine D2 Receptors. *Journal of Biological Chemistry* **2002**, *277* (20), 18091–18097. <https://doi.org/10.1074/jbc.M107731200>.
- (99) Pfeiffer, M.; Koch, T.; Schröder, H.; Laugsch, M.; Höllt, V.; Schulz, S. Heterodimerization of Somatostatin and Opioid Receptors Cross-Modulates Phosphorylation, Internalization, and Desensitization. *Journal of Biological*

- Chemistry* **2002**, 277 (22), 19762–19772.
<https://doi.org/10.1074/jbc.M110373200>.
- (100) Xu, J.; He, J.; Castleberry, A. M.; Balasubramanian, S.; Lau, A. G.; Halls, R. A. Heterodimerization of A2A- and B1-Adrenergic Receptors. *Journal of Biological Chemistry* **2003**, 278 (12), 10770–10777.
<https://doi.org/10.1074/jbc.M207968200>.
- (101) Stanton, B. Z.; Chory, E. J.; Crabtree, G. R. Chemically Induced Proximity in Biology and Medicine. *Science* **2018**, 359 (6380).
<https://doi.org/10.1126/science.aao5902>.
- (102) Valant, C.; Robert Lane, J.; Sexton, P. M.; Christopoulos, A. The Best of Both Worlds? Bitopic Orthosteric/Allosteric Ligands of G Protein–Coupled Receptors. *Annual Review of Pharmacology and Toxicology* **2012**, 52 (1), 153–178. <https://doi.org/10.1146/annurev-pharmtox-010611-134514>.
- (103) Lane, J. R.; Sexton, P. M.; Christopoulos, A. Bridging the Gap: Bitopic Ligands of G-Protein-Coupled Receptors. *Trends in Pharmacological Sciences* **2013**, 34 (1), 59–66. <https://doi.org/10.1016/j.tips.2012.10.003>.
- (104) Shonberg, J.; Scammells, P. J.; Capuano, B. Design Strategies for Bivalent Ligands Targeting GPCRs. *ChemMedChem* **2011**, 6 (6), 963–974.
<https://doi.org/10.1002/cmdc.201100101>.
- (105) Decker, M.; Fulton, B. S.; Zhang, B.; Knapp, B. I.; Bidlack, J. M.; Neumeyer, J. L. Univalent and Bivalent Ligands of Butorphan: Characteristics of the Linking Chain Determine the Affinity and Potency of Such Opioid Ligands. *Journal of Medicinal Chemistry* **2009**, 52 (23), 7389–7396.
<https://doi.org/10.1021/jm900379p>.
- (106) Zhang, B.; Zhang, T.; Sromek, A. W.; Scrimale, T.; Bidlack, J. M.; Neumeyer, J. L. Synthesis and Binding Affinity of Novel Mono- and Bivalent Morphinan Ligands for κ , μ , and δ Opioid Receptors. *Bioorganic and Medicinal Chemistry* **2011**, 19 (9), 2808–2816.
<https://doi.org/10.1016/j.bmc.2011.03.052>.
- (107) Zheng, Y.; Akgu, E.; Miller, L. J.; Portoghese, P. S. Induced Association of μ Opioid (MOP) and Type 2 Cholecystinin (CCK 2) Receptors by Novel Bivalent Ligands. **2009**, No. Figure 1, 247–258.

- (108) Daniels, D. J.; Lenard, N. R.; Etienne, C. L.; Law, P.; Roerig, S. C.; Portoghese, P. S. Opioid-Induced Tolerance and Dependence in Mice Is Modulated by the Distance between Pharmacophores in a Bivalent Ligand Series. **2005**, *102* (52).
- (109) Zhang, S.; Yekkirala, A.; Tang, Y.; Portoghese, P. S. A Bivalent Ligand (KMN-21) Antagonist for μ/κ Heterodimeric Opioid Receptors. *Bioorganic and Medicinal Chemistry Letters* **2009**, *19* (24), 6978–6980. <https://doi.org/10.1016/j.bmcl.2009.10.045>.
- (110) Daniels, D. J.; Kulkarni, A.; Xie, Z.; Bhushan, R. G.; Portoghese, P. S. A Bivalent Ligand (KDAN-18) Containing δ -Antagonist and K -Agonist Pharmacophores Bridges δ_2 and K_1 Opioid Receptor Phenotypes †. **2005**, 1713–1716. <https://doi.org/10.1021/jm034234f>.
- (111) Bhushan, R. G.; Sharma, S. K.; Xie, Z.; Daniels, D. J.; Portoghese, P. S. A Bivalent Ligand (KDN-21) Reveals Spinal δ and K Opioid Receptors Are Organized as Heterodimers That Give Rise to δ_1 and K_2 Phenotypes. Selective Targeting of δ - K Heterodimers. **2004**, *20*, 2969–2972. <https://doi.org/10.1021/jm0342358>.
- (112) Portoghese, P. S.; Larson, D. L.; Sayre, L. M.; Yim, C. B.; Ronsisvalle, G.; Tam, S. W.; Takemori, A. E. Opioid Agonist and Antagonist Bivalent Ligands. The Relationship between Spacer Length and Selectivity at Multiple Opioid Receptors. *Journal of medicinal chemistry* **1986**, *29* (10), 1855–1861.
- (113) Rajeswaran, W. G.; Cao, Y.; Huang, X.; Wroblewski, M. E.; Colclough, T.; Lee, S.; Liu, F.; Nagy, P. I.; Ellis, J.; Levine, B. A.; Nocka, K. H.; Messer, W. S. Design, Synthesis, and Biological Characterization of Bivalent 1-Methyl-1, 2, 5, 6-Tetrahydropyridyl-1, 2, 5-Thiadiazole Derivatives as Selective. **2001**, 4563–4576. <https://doi.org/10.1021/jm0102405>.
- (114) Ghosh, E.; Kumari, P.; Jaiman, D.; Shukla, A. K. Methodological Advances: The Unsung Heroes of the GPCR Structural Revolution. *Nature Reviews Molecular Cell Biology* **2015**, *16* (2), 69–81. <https://doi.org/10.1038/nrm3933>.
- (115) Xiang, J.; Chun, E.; Liu, C.; Jing, L.; Al-Sahouri, Z.; Zhu, L.; Liu, W. Successful Strategies to Determine High-Resolution Structures of GPCRs. *Trends in*

- Pharmacological Sciences* **2016**, *37* (12), 1055–1069.
<https://doi.org/10.1016/j.tips.2016.09.009>.
- (116) Robertson, N.; Jazayeri, A.; Errey, J.; Baig, A.; Hurrell, E.; Zhukov, A.; Langmead, C. J.; Weir, M.; Marshall, F. H. The Properties of Thermostabilised G Protein-Coupled Receptors (StaRs) and Their Use in Drug Discovery. *Neuropharmacology* **2011**, *60* (1), 36–44.
<https://doi.org/10.1016/j.neuropharm.2010.07.001>.
- (117) Christopher, J. A.; Brown, J.; Doré, A. S.; Errey, J. C.; Koglin, M.; Marshall, F. H.; Myszka, D. G.; Rich, R. L.; Tate, C. G.; Tehan, B.; Warne, T.; Congreve, M. Biophysical Fragment Screening of the B1-Adrenergic Receptor: Identification of High Affinity Arylpiperazine Leads Using Structure-Based Drug Design. *Journal of Medicinal Chemistry* **2013**, *56* (9), 3446–3455.
<https://doi.org/10.1021/jm400140q>.
- (118) Sato, T.; Baker, J.; Warne, T.; Brown, G. A.; Leslie, A. G. W.; Congreve, M.; Tate, C. G. Pharmacological Analysis and Structure Determination of 7-Methylcyanopindolol-Bound β_1 -Adrenergic Receptor. *Molecular pharmacology* **2015**, *88* (6), 1024–1034.
<https://doi.org/10.1124/mol.115.101030>.
- (119) Warne, T.; Moukhametzianov, R.; Baker, J. G.; Nehmé, R.; Edwards, P. C.; Leslie, A. G. W.; Schertler, G. F. X.; Tate, C. G. The Structural Basis for Agonist and Partial Agonist Action on a B1-Adrenergic Receptor. *Nature* **2011**, *469* (7329), 241–245. <https://doi.org/10.1038/nature09746>.
- (120) Huang, J.; Chen, S.; Zhang, J. J.; Huang, X. Crystal Structure of Oligomeric B1-Adrenergic G Protein- Coupled Receptors in Ligand-Free Basal State. *Nature Structural and Molecular Biology* **2013**, *20* (4), 419–425.
<https://doi.org/10.1038/nsmb.2504.Crystal>.
- (121) Leslie, A. G. W.; Warne, T.; Tate, C. G. Ligand Occupancy in Crystal Structure of β_1 -Adrenergic G Protein Coupled Receptor. *Nature Structural and Molecular Biology* **2015**, *22* (12), 941–942.
<https://doi.org/10.1038/nsmb.3130>.

- (122) Shonberg, J.; Kling, R. C.; Gmeiner, P.; Löber, S. GPCR Crystal Structures: Medicinal Chemistry in the Pocket. *Bioorganic and Medicinal Chemistry* **2015**, *23* (14), 3880–3906. <https://doi.org/10.1016/j.bmc.2014.12.034>.
- (123) Baker, J. G.; Proudman, R. G. W.; Hill, S. J. Identification of Key Residues in Transmembrane 4 Responsible for the Secondary, Low-Affinity Conformation of the Human B1-Adrenoceptor. *Molecular pharmacology* **2014**, *85* (5), 811–829. <https://doi.org/10.1124/mol.114.091587>.
- (124) Cvicek, V.; Goddard, W. A.; Abrol, R. Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLOS Computational Biology* **2016**, *12* (3), e1004805. <https://doi.org/10.1371/journal.pcbi.1004805>.
- (125) Violin, J. D.; Crombie, A. L.; Soergel, D. G.; Lark, M. W. Biased Ligands at G-Protein-Coupled Receptors: Promise and Progress. *Trends in Pharmacological Sciences* **2014**, *35* (7), 308–316. <https://doi.org/10.1016/j.tips.2014.04.007>.
- (126) Wisler, J. W.; Xiao, K.; Thomsen, A. R. B.; Lefkowitz, R. J. Recent Developments in Biased Agonism. *Current Opinion in Cell Biology* **2014**, *27* (1), 18–24. <https://doi.org/10.1016/j.ceb.2013.10.008>.
- (127) Quoyer, J.; Janz, J. M.; Luo, J.; Ren, Y.; Armando, S.; Lukashova, V.; Benovic, J. L.; Carlson, K. E.; Hunt, S. W.; Bouvier, M. Pepducin Targeting the C-X-C Chemokine Receptor Type 4 Acts as a Biased Agonist Favoring Activation of the Inhibitory G Protein. *Proceedings of the National Academy of Sciences* **2013**, *110* (52), E5088–E5097. <https://doi.org/10.1073/pnas.1312515110>.
- (128) Staus, D. P.; Wingler, L. M.; Strachan, R. T.; Rasmussen, S. G. F.; Pardon, E.; Ahn, S.; Steyaert, J.; Kobilka, B. K.; Lefkowitz, R. J. Regulation of 2-Adrenergic Receptor Function by Conformationally Selective Single-Domain Intrabodies. *Molecular Pharmacology* **2014**, *85* (3), 472–481. <https://doi.org/10.1124/mol.113.089516>.
- (129) Committee, J. F. *British National Formulary (Online)*, 75th ed.; British National Formulary, 2018.
- (130) Kotlyar, E.; Keogh, A. M.; Macdonald, P. S.; Arnold, R. H.; McCaffrey, D. J.; Glanville, A. R. Tolerability of Carvedilol in Patients with Heart Failure and

- Concomitant Chronic Obstructive Pulmonary Disease or Asthma. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation* **2002**, *21* (12), 1290–1295. [https://doi.org/10.1016/S1053-2498\(02\)00459-X](https://doi.org/10.1016/S1053-2498(02)00459-X).
- (131) Alwan, A.; MacLean, D. R.; Riley, L. M.; D'Espaignet, E. T.; Mathers, C. D.; Stevens, G. A.; Bettcher, D. Monitoring and Surveillance of Chronic Non-Communicable Diseases: Progress and Capacity in High-Burden Countries. *The Lancet* **2010**, *376* (9755), 1861–1868. [https://doi.org/10.1016/S0140-6736\(10\)61853-3](https://doi.org/10.1016/S0140-6736(10)61853-3).
- (132) Hoffmann, C.; Leitz, M. R.; Oberdorf-Maass, S.; Lohse, M. J.; Klotz, K. N. Comparative Pharmacology of Human β -Adrenergic Receptor Subtypes - Characterization of Stably Transfected Receptors in CHO Cells. *Naunyn-Schmiedeberg's Archives of Pharmacology* **2004**, *369* (2), 151–159. <https://doi.org/10.1007/s00210-003-0860-y>.
- (133) Baker, J. G. A Full Pharmacological Analysis of the Three Turkey β -Adrenoceptors and Comparison with the Human β -Adrenoceptors. *PLoS ONE* **2010**, *5* (11). <https://doi.org/10.1371/journal.pone.0015487>.
- (134) Baker, J. G.; Proudman, R. G. W.; Tate, C. G. The Pharmacological Effects of the Thermostabilising (M23) Mutations and Intra and Extracellular (B36) Deletions Essential for Crystallisation of the Turkey β -Adrenoceptor. *Naunyn-Schmiedeberg's Archives of Pharmacology* **2011**, *384* (1), 71–91. <https://doi.org/10.1007/s00210-011-0648-4>.
- (135) Baker, J. G.; Hill, S. J. Multiple GPCR Conformations and Signalling Pathways: Implications for Antagonist Affinity Estimates. *Trends in Pharmacological Sciences* **2007**, *28* (8), 374–381. <https://doi.org/10.1016/j.tips.2007.06.011>.
- (136) Baker, J. G.; Hall, I. P.; Hill, S. J. Agonist and Inverse Agonist Actions of Beta-Blockers at the Human Beta 2-Adrenoceptor Provide Evidence for Agonist-Directed Signaling. *Molecular pharmacology* **2003**, *64* (6), 1357–1369. <https://doi.org/10.1124/mol.64.6.1357>.
- (137) Mistry, S. N.; Baker, J. G.; Fischer, P. M.; Hill, S. J.; Gardiner, S. M.; Kellam, B. Synthesis and in Vitro and in Vivo Characterization of Highly B1-Selective β -

- Adrenoceptor Partial Agonists. *Journal of Medicinal Chemistry* **2013**, *56* (10), 3852–3865. <https://doi.org/10.1021/jm400348g>.
- (138) Rasmussen, S. G. F.; Choi, H. J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; Devree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; Schnapp, A.; Konetzki, I.; Sunahara, R. K.; Gellman, S. H.; Pautsch, A.; Steyaert, J.; Weis, W. I.; Kobilka, B. K. Structure of a Nanobody-Stabilized Active State of the B2adrenoceptor. *Nature* **2011**, *469* (7329), 175–181. <https://doi.org/10.1038/nature09648>.
- (139) Strader, C. D.; Sigal, I. S.; Candelore, M. R.; Rands, E.; Hill, W. S.; Dixon, R. A. Conserved Aspartic Acid Residues 79 and 113 of the Beta-Adrenergic Receptor Have Different Roles in Receptor Function. *Journal of Biological Chemistry* **1988**, *263* (21), 10267–10271.
- (140) Strosberg, A. D. Structure and Function of the B3 Adrenoreceptor. *Advances in Pharmacology* **1997**, *42*, 511–513. [https://doi.org/10.1016/S1054-3589\(08\)60801-7](https://doi.org/10.1016/S1054-3589(08)60801-7).
- (141) González, A.; Perez-Acle, T.; Pardo, L.; Deupi, X. Molecular Basis of Ligand Dissociation in β -Adrenergic Receptors. *PLoS ONE* **2011**, *6* (9), e23815. <https://doi.org/10.1371/journal.pone.0023815>.
- (142) Soriano-Ursúa, M. A.; Trujillo-Ferrara, J. G.; Correa-Basurto, J.; Vilar, S. Recent Structural Advances of B1 and B2 Adrenoceptors Yield Keys for Ligand Recognition and Drug Design. *Journal of Medicinal Chemistry* **2013**, *56* (21), 8207–8223. <https://doi.org/10.1021/jm400471z>.
- (143) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. Activation Mechanism of the B2 -Adrenergic Receptor. *Proceedings of the National academy of sciences of the USA* **2011**, *108* (46), 18684–18689. <https://doi.org/10.1073/pnas.1110499108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1110499108>.
- (144) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-Resolution Crystal Structure of an Engineered Human b 2 - Adrenergic. *Science* **2007**, *318* (November), 1258–1266. <https://doi.org/10.1126/science.1150577>.

- (145) Strader, C. D.; Candelore, M. R.; Hill, W. S.; Sigal, I. S.; Dixon, R. A. F. Identification of Two Serine Residues Involved in Agonist Activation of the β -Adrenergic Receptor. *Journal of Biological Chemistry* **1989**, *264* (23), 13572–13578.
- (146) Wieland, K.; Zuurmond, H. M.; Krasel, C.; Ijzerman, A. P.; Lohse, M. J. Involvement of Asn-293 in Stereospecific Agonist Recognition and in Activation of the Beta 2-Adrenergic Receptor. *Proceedings of the National Academy of Sciences* **1996**, *93* (17), 9276–9281. <https://doi.org/10.1073/pnas.93.17.9276>.
- (147) Vanni, S.; Neri, M.; Tavernelli, I.; Rothlisberger, U. Predicting Novel Binding Modes of Agonists to β Adrenergic Receptors Using All-Atom Molecular Dynamics Simulations. *PLoS Computational Biology* **2011**, *7* (1). <https://doi.org/10.1371/journal.pcbi.1001053>.
- (148) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. *Proceedings of the National Academy of Sciences* **2011**, *108* (32), 13118–13123. <https://doi.org/10.1073/pnas.1104614108>.
- (149) Chelikani, P.; Hornak, V.; Eilers, M.; Reeves, P. J.; Smith, S. O.; RajBhandary, U. L.; Khorana, H. G. Role of Group-Conserved Residues in the Helical Core of Beta2-Adrenergic Receptor. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104* (17), 7027–7032. <https://doi.org/10.1073/pnas.0702024104>.
- (150) Emtage, A. L.; Mistry, S. N.; Fischer, P. M.; Kellam, B.; Laughton, C. A. GPCRs through the Keyhole: The Role of Protein Flexibility in Ligand Binding to β -Adrenoceptors. *Journal of Biomolecular Structure and Dynamics* **2017**, *35* (12), 2604–2619. <https://doi.org/10.1080/07391102.2016.1226197>.
- (151) Wang, T.; Duan, Y. Ligand Entry and Exit Pathways in the B2-Adrenergic Receptor. *Journal of Molecular Biology* **2009**, *392* (4), 1102–1115. <https://doi.org/10.1016/j.jmb.2009.07.093>.
- (152) Wang, T.; Duan, Y. Chromophore Channeling in the G-Protein Coupled Receptor Rhodopsin. *Journal of the American Chemical Society* **2007**, *129* (22), 6970–6971. <https://doi.org/10.1021/ja0691977>.

- (153) Hildebrand, P. W.; Scheerer, P.; Park, J. H.; Choe, H. W.; Piechnick, R.; Ernst, O. P.; Hofmann, K. P.; Heck, M. A Ligand Channel through the G Protein Coupled Receptor Opsin. *PLoS ONE* **2009**, *4* (2). <https://doi.org/10.1371/journal.pone.0004382>.
- (154) Christopher, J. A.; Brown, J.; Doré, A. S.; Errey, J. C.; Koglin, M.; Marshall, F. H.; Myszka, D. G.; Rich, R. L.; Tate, C. G.; Tehan, B.; Warne, T.; Congreve, M. Biophysical Fragment Screening of the B1-Adrenergic Receptor: Identification of High Affinity Arylpiperazine Leads Using Structure-Based Drug Design. *Journal of Medicinal Chemistry* **2013**, *56* (9), 3446–3455. <https://doi.org/10.1021/jm400140q>.
- (155) Miller-Gallacher, J. L.; Nehmé, R.; Warne, T.; Edwards, P. C.; Schertler, G. F. X.; Leslie, A. G. W.; Tate, C. G. The 2.1 Å Resolution Structure of Cyanopindolol-Bound B1-Adrenoceptor Identifies an Intramembrane Na⁺ Ion That Stabilises the Ligand-Free Receptor. *PLoS ONE* **2014**, *9* (3). <https://doi.org/10.1371/journal.pone.0092727>.
- (156) Emtage, A. L. Structural Exploration of β 1-Adrenoceptor Selectivity. **2013**, No. September.
- (157) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *Journal of molecular biology* **2002**, *320* (3), 597–608. [https://doi.org/10.1016/S0022-2836\(02\)00470-9](https://doi.org/10.1016/S0022-2836(02)00470-9).
- (158) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins: Structure, Function and Genetics* **2004**, *55* (2), 351–367. <https://doi.org/10.1002/prot.10613>.
- (159) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry* **2004**, *47* (7), 1750–1759. <https://doi.org/10.1021/jm030644s>.
- (160) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate

- Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004**, *47* (7), 1739–1749. <https://doi.org/10.1021/jm0306430>.
- (161) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry* **2006**, *49* (21), 6177–6196. <https://doi.org/10.1021/jm051256o>.
- (162) Scharf, M. M.; Bünemann, M.; Baker, J. G.; Kolb, P. Comparative Docking to Distinct G Protein-Coupled Receptor Conformations Exclusively Yields Ligands with Agonist Efficacy. *Molecular Pharmacology* **2019**, *96* (6), 851. <https://doi.org/10.1124/mol.119.117515>.
- (163) Costanzi, S.; Cohen, A.; Danfora, A.; Dolatmoradi, M. Influence of the Structural Accuracy of Homology Models on Their Applicability to Docking-Based Virtual Screening: The B2 Adrenergic Receptor as a Case Study. *Journal of Chemical Information and Modeling* **2019**, *59* (7), 3177–3190. <https://doi.org/10.1021/acs.jcim.9b00380>.
- (164) Yakar, R.; Akten, E. D. Discovery of High Affinity Ligands for B2-Adrenergic Receptor through Pharmacophore-Based High-Throughput Virtual Screening and Docking. *Journal of Molecular Graphics and Modelling* **2014**, *53*, 148–160. <https://doi.org/10.1016/j.jmglm.2014.07.007>.
- (165) Vilar, S.; Sobarzo-Sanchez, E.; Santana, L.; Uriarte, E. Molecular Docking and Drug Discovery in β -Adrenergic Receptors. *Current Medicinal Chemistry* **2017**, *24* (39). <https://doi.org/10.2174/0929867324666170724101448>.
- (166) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. v.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *Journal of Computer-Aided Molecular Design* **1997**, *11* (5), 425–445. <https://doi.org/10.1023/A:1007996124545>.
- (167) Karplus, M.; McCammon, J. A.; Gelin, B. R. Dynamics of Folded Proteins. *Nature* **1977**, *267* (June), 585–590.

- (168) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577–8593. <https://doi.org/10.1063/1.470117>.
- (169) Leach, A. R. *Molecular Modelling : Principles and Applications*; Prentice Hall: Harlow, England; New York, 2001.
- (170) Case, D. A.; Walker, R. C.; Cheatham, T. E.; Simmerling, C.; Roitberg, A.; Merz, K. M.; Luo, R.; Darden, T. Amber 2018. *University of California, San Francisco. 2018* **2018**, 1–923.
- (171) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry* **2005**, *26* (16), 1668–1688. <https://doi.org/10.1002/jcc.20290>.
- (172) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GRGMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4* (3), 435–447. <https://doi.org/10.1021/ct700301q>.
- (173) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>.
- (174) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (175) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (176) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of*

- Computational Chemistry* **2004**, 25 (9), 1157–1174.
<https://doi.org/10.1002/jcc.20035>.
- (177) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *Journal of Physical Chemistry B* **2007**, 111 (27), 7812–7824.
<https://doi.org/10.1021/jp071097f>.
- (178) González, M. A. Force Fields and Molecular Dynamics Simulations. *Collection SFN* **2011**, 12, 169–200. <https://doi.org/10.1051/sfn/201112009>.
- (179) Urey, H. C.; Bradley, C. A. The Vibrations of Pentatonic Tetrahedral Molecules. *Physical Review* **1931**, 38 (11), 1969.
<https://doi.org/10.1103/PhysRev.38.1969>.
- (180) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How Do Substrates Enter and Products Exit the Buried Active Site of Cytochrome P450cam? 1. Random Expulsion Molecular Dynamics Investigation of Ligand Access Channels and Mechanisms. *Journal of Molecular Biology* **2000**, 303 (5), 797–811.
<https://doi.org/10.1006/jmbi.2000.4154>.
- (181) Cheng, X.; Wang, H.; Grant, B.; Sine, S. M.; McCammon, J. A. Targeted Molecular Dynamics Study of C-Loop Closure and Channel Gating in Nicotinic Receptors. *PLoS Computational Biology* **2006**, 2 (9), 1173–1184.
<https://doi.org/10.1371/journal.pcbi.0020134>.
- (182) Compoin, M.; Picaud, F.; Ramseyer, C.; Girardet, C. Targeted Molecular Dynamics of an Open-State KcsA Channel. *Journal of Chemical Physics* **2005**, 122 (13). <https://doi.org/10.1063/1.1869413>.
- (183) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. **2002**.
- (184) Kahn, H.; Harris, T. E. Estimation of Particle Transmission by Random Sampling. *National Bureau of Standards Applied Math Series*. 1951, pp 27–30.
- (185) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophysical Journal* **1996**, 70 (1), 97–110.
[https://doi.org/10.1016/S0006-3495\(96\)79552-8](https://doi.org/10.1016/S0006-3495(96)79552-8).
- (186) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annual Review of Biophysics*

- 2017**, *46* (1), 43–57. <https://doi.org/10.1146/annurev-biophys-070816-033834>.
- (187) Ray, D.; Gokey, T.; Mobley, D. L.; Andricioaei, I. Kinetics and Free Energy of Ligand Dissociation Using Weighted Ensemble Milestoning. *Journal of Chemical Physics* **2020**, *153* (15). <https://doi.org/10.1063/5.0021953>.
- (188) Pan, A. C.; Xu, H.; Palpant, T.; Shaw, D. E. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2017**, *13* (7), 3372–3377. <https://doi.org/10.1021/acs.jctc.7b00172>.
- (189) Suárez, E.; Zuckerman, D.; Chong, L. Extensive Evaluation of Weighted Ensemble Strategies for Calculating Rate Constants and Binding Affinities of Molecular Association/Dissociation Processes. *Extensive Evaluation of Weighted Ensemble Strategies for Calculating Rate Constants and Binding Affinities of Molecular Association/Dissociation Processes* **2019**, 671172. <https://doi.org/10.1101/671172>.
- (190) Zwier, M. C.; Pratt, A. J.; Adelman, J. L.; Kaus, J. W.; Zuckerman, D. M.; Chong, L. T. Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein-Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered P53 Peptide. *Journal of Physical Chemistry Letters* **2016**, *7* (17), 3440–3445. <https://doi.org/10.1021/acs.jpcllett.6b01502>.
- (191) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophysical Journal* **1997**, *72* (3), 1047–1069. [https://doi.org/10.1016/S0006-3495\(97\)78756-3](https://doi.org/10.1016/S0006-3495(97)78756-3).
- (192) Baker, J. G.; Proudman, R. G. W.; Hawley, N. C.; Fischer, P. M.; Hill, S. J. Role of Key Transmembrane Residues in Agonist and Antagonist Actions at the Two Conformations of the Human Beta1-Adrenoceptor. *Molecular pharmacology* **2008**, *74* (5), 1246–1260. <https://doi.org/10.1124/mol.108.048371>.
- (193) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson,

- J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23* (21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
- (194) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29* (11), 1859–1865. <https://doi.org/10.1002/jcc.20945>.
- (195) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward Realistic Biological Membrane Simulations. *Journal of Computational Chemistry* **2014**, *35* (27), 1997–2004. <https://doi.org/https://doi.org/10.1002/jcc.23702>.
- (196) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal* **2009**, *97* (1), 50–58. <https://doi.org/10.1016/j.bpj.2009.04.013>.
- (197) Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE* **2007**, *2* (9). <https://doi.org/10.1371/journal.pone.0000880>.
- (198) Lee, J.; Patel, D. S.; Stähle, J.; Park, S. J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O.; Knirel, Y. A.; Qi, Y.; Jo, S.; Klauda, J. B.; Widmalm, G.; Im, W. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *Journal of Chemical Theory and Computation* **2019**, *15* (1), 775–786. <https://doi.org/10.1021/acs.jctc.8b01066>.
- (199) Lee, J.; Hitzenberger, M.; Rieger, M.; Kern, N. R.; Zacharias, M.; Im, W. CHARMM-GUI Supports the Amber Force Fields. *Journal of Chemical Physics* **2020**, *153* (3). <https://doi.org/10.1063/5.0012280>.
- (200) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical*

- Theory and Computation* **2016**, *12* (1), 405–413.
<https://doi.org/10.1021/acs.jctc.5b00935>.
- (201) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **2009**, *30* (10), 1545–1614. <https://doi.org/10.1002/jcc.21287>.
- (202) Cresset. Flare Version 5. <http://www.cresset-group.com/flare/>: Litlington, Cambridgeshire, UK.
- (203) Mistry, N, S. Structure Activity Relationships of Novel and Selective P1-Adrenoceptor Ligands. **2009**, No. September.
- (204) Baker, J. G. The Selectivity of β -Adrenoceptor Agonists at Human B1-, B2- and B3-Adrenoceptors: RESEARCH PAPER. *British Journal of Pharmacology* **2010**, *160* (5), 1048–1061. <https://doi.org/10.1111/j.1476-5381.2010.00754.x>.
- (205) Bourke, D. G.; Collins, D. J. Conversion of 7-Methoxy-3,4-Dihydro-2H-1-Benzopyran-2-One into the Corresponding Dimethyl Ortho Ester. *Tetrahedron* **1997**, *53* (11), 3863–3878. [https://doi.org/10.1016/S0040-4020\(97\)00007-0](https://doi.org/10.1016/S0040-4020(97)00007-0).
- (206) Howard, K. T.; Chisholm, J. D. Preparation and Applications of 4-Methoxybenzyl Esters in Organic Synthesis. *Organic Preparations and Procedures International* **2016**, *48* (1), 1–36. <https://doi.org/10.1080/00304948.2016.1127096>.
- (207) Jarowicki, K.; Kocienski, P. Protecting Groups. *Journal of the Chemical Society. Perkin Transactions 1* **2001**, *1* (18), 2109–2135. <https://doi.org/10.1039/b103282h>.
- (208) Fleming, J. J.; Bois, J. du. Communication A Synthesis of (+)-Saxitoxin A Synthesis of (+)-Saxitoxin. *Society* **2006**, *128* (March), 3926–3927.

- (209) Fleming, J. J.; McReynolds, M. D.; du Bois, J. (+)-Saxitoxin: A First and Second Generation Stereoselective Synthesis. *Journal of the American Chemical Society* **2007**, *129* (32), 9964–9975. <https://doi.org/10.1021/ja071501o>.
- (210) Kostenko, N.; Gottfriedsen, J.; Hilfert, L.; Edelman, F. T. A Synthetic Route to Quaternary Pyridinium Salt-Functionalized Silsesquioxanes. *International Journal of Polymer Science* **2012**, *2012*. <https://doi.org/10.1155/2012/586594>.
- (211) Chu, L.; Hutchins, J. E.; Weber, A. E.; Lo, J. L.; Yang, Y. T.; Cheng, K.; Smith, R. G.; Fisher, M. H.; Wyvrat, M. J.; Goulet, M. T. Initial Structure-Activity Relationship of a Novel Class of Nonpeptidyl GnRH Receptor Antagonists: 2-Arylindoles. *Bioorganic and Medicinal Chemistry Letters* **2001**, *11* (4), 509–513. [https://doi.org/10.1016/S0960-894X\(00\)00707-1](https://doi.org/10.1016/S0960-894X(00)00707-1).
- (212) Yi, B.; Jahangir, A.; Evans, A. K.; Briggs, D.; Ravina, K.; Ernest, J.; Farimani, A. B.; Sun, W.; Rajadas, J.; Green, M.; Feinberg, E. N.; Pande, V. S.; Shamloo, M. Discovery of Novel Brain Permeable and G Protein-Biased Beta-1 Adrenergic Receptor Partial Agonists for the Treatment of Neurocognitive Disorders. *PLoS ONE* **2017**, *12* (7), 1–30. <https://doi.org/10.1371/journal.pone.0180319>.
- (213) Kierstead, R. W.; Faraone, A.; Mennona, F.; Mullin, J.; Guthrie, R. W.; Crowley, H.; Simko, B.; Blaber, L. C. ..Beta.1-Selective Adrenoceptor Antagonists. 1. Synthesis and .Beta.-Adrenergic Blocking Activity of a Series of Binary (Aryloxy)Propanolamines. *Journal of Medicinal Chemistry* **1983**, *26* (11), 1561–1569. <https://doi.org/10.1021/jm00365a004>.
- (214) Hubner, H.; Schellhorn, T.; Gienger, M.; Schaab, C.; Kaindl, J.; Leeb, L.; Clark, T.; Moller, D.; Gmeiner, P. Structure-Guided Development of Heterodimer-Selective GPCR Ligands. *Nature Communications* **2016**, *7*, 1–12. <https://doi.org/10.1038/ncomms12298>.
- (215) Birnkammer, T.; Spickenreither, A.; Brunskole, I.; Lopuch, M.; Kagermeier, N.; Bernhardt, G.; Dove, S.; Seifert, R.; Elz, S.; Buschauer, A. The Bivalent Ligand Approach Leads to Highly Potent and Selective Acylguanidine-Type Histamine H₂ Receptor Agonists. *Journal of Medicinal Chemistry* **2012**, *55* (3), 1147–1160. <https://doi.org/10.1021/jm201128q>.

- (216) Mondal, S.; Johnston, J. M.; Wang, H.; Khelashvili, G.; Filizola, M.; Weinstein, H. Membrane Driven Spatial Organization of GPCRs. **2013**. <https://doi.org/10.1038/srep02909>.
- (217) Calebiro, D.; Rieken, F.; Wagner, J.; Sungkaworn, T.; Zabel, U.; Borzi, A.; Cocucci, E.; Zürn, A.; Lohse, M. J. Single-Molecule Analysis of Fluorescently Labeled G-Protein-Coupled Receptors Reveals Complexes with Distinct Dynamics and Organization. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (2), 743–748. <https://doi.org/10.1073/pnas.1205798110>.
- (218) Hu, C. D.; Kerppola, T. K. Simultaneous Visualization of Multiple Protein Interactions in Living Cells Using Multicolor Fluorescence Complementation Analysis. *Nature Biotechnology* **2003**, *21* (5), 539–545. <https://doi.org/10.1038/nbt816>.
- (219) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation* **2008**, *4* (5), 819–834. <https://doi.org/10.1021/ct700324x>.
- (220) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. v.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation* **2013**, *9* (1), 687–697. <https://doi.org/10.1021/ct300646g>.
- (221) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational Lipidomics with Insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation* **2015**, *11* (5), 2144–2155. <https://doi.org/10.1021/acs.jctc.5b00209>.
- (222) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25* (2), 247–260. <https://doi.org/10.1016/j.jmglm.2005.12.005>.

- (223) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **2007**, *372* (3), 774–797. <https://doi.org/10.1016/J.JMB.2007.05.022>.
- (224) Altwaijry, N. A.; Baron, M.; Wright, D. W.; Coveney, P. v.; Townsend-Nicholson, A. An Ensemble-Based Protocol for the Computational Prediction of Helix-Helix Interactions in G Protein-Coupled Receptors Using Coarse-Grained Molecular Dynamics. *Journal of Chemical Theory and Computation* **2017**, *13* (5), 2254–2270. <https://doi.org/10.1021/acs.jctc.6b01246>.
- (225) Sykes, D. A.; Parry, C.; Reilly, J.; Wright, P.; Fairhurst, R. A.; Charlton, S. J. Observed Drug-Receptor Association Rates Are Governed by Membrane Affinity: The Importance of Establishing “Micro-Pharmacokinetic/Pharmacodynamic Relationships” at the B2-Adrenoceptor. *Molecular Pharmacology* **2014**, *85* (4), 608–617. <https://doi.org/10.1124/mol.113.090209>.
- (226) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Computational Biology* **2017**, *13* (7), e1005659.
- (227) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Miguez, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16* (1), 528–552. <https://doi.org/10.1021/acs.jctc.9b00591>.
- (228) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. Computational Tools MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysj* **2015**, *109*, 1528–1532. <https://doi.org/10.1016/j.bpj.2015.08.015>.

- (229) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular Determinants of Drug-Receptor Binding Kinetics. *Drug Discovery Today* **2013**, *18* (13–14), 667–673. <https://doi.org/10.1016/j.drudis.2013.02.007>.