Investigations toward the rational modulation of G protein-coupled receptor signalling pathways using *in silico* methods

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Pharmazie

der Philipps-Universität Marburg

vorgelegt von

Master of Science

Magdalena Martina Scharf

aus

Eberbach

Marburg/Lahn, 2021

Erstgutachter: Prof. Dr. Peter Kolb

Zweitgutachter: Prof. Dr. Moritz Bünemann

Eingereicht am 11.12.2020

Tag der mündlichen Prüfung am 22.01.2021

Hochschulkennziffer: 1180

Man muss an seine Berufung glauben
und alles daran setzen, sein Ziel zu erreichen.
-Marie Curie

Abstract

G protein-coupled receptors (GPCRs) are one of the most important protein families and function as signal transducers located in the cell membrane. Currently, about one third of the marketed drugs target a GPCR, reflecting its importance in therapy and disease. Thus, it is not surprising that GPCRs and their signalling are of major interest for researchers. In this thesis, *in silico* methods were used to investigate the modulation of GPCR signalling pathways.

The modulation of GPCR signalling can take place on different levels, e.g. at the level of GPCR ligands or in the downstream signalling pathways. Interactions of the receptor with small molecules can result either in its inactivation or activation. The latter can lead to the intracellular recruitment of various effector proteins to the receptor which can then induce different signalling pathways inside the cell. Certain ligands can induce a stronger recruitment of one effector protein compared to other effector proteins. On a structural basis it is still unclear why and how these ligands induce such bias. Furthermore, there are many different proteins involved in the downstream signalling of GPCRs. One protein family are the Regulators for G protein Signalling (RGS) which are involved in the deactivation of the G protein and, hence, GPCR signalling. Although the members of this protein family are known to be involved in a variety of processes and diseases –many of which are also related to GPCR signalling– they are still not well understood. GPCR signalling needs to be comprehended better on all of these levels to be able to modulate them rationally. In this thesis, two GPCRs –the β_2-adrenergic receptor (β_2AR) and the Cannabinoid receptor 2 (CB2)– and one member of the RGS protein family –the RGS7– were targeted with *in silico* techniques in five studies to investigate their signalling and its modulation.

Two of the studies described in this thesis targeted the β_2AR. More than 30 structures of this class A GPCR in different activation states are available, allowing for more exhaustive structural investigations. This fact was used and three different structures of the β_2AR in different activation states were targeted with a molecular library using a comparative docking approach. The aim was to predict novel agonists for this receptor based on the assumption that these should rather result from docking calculations against active conformations of the receptor. The selected molecules were then characterised pharmacologically, showing that this approach was very successful. Furthermore, a retrospective analysis of the docking approach showed up the optimal way to increase the chances to discover novel agonists for this receptor or other class A GPCRs.

The aim of the second study targeting the β_2AR was to predict antagonists with novel structural scaffolds for this receptor using docking calculations. The project was conducted in collaboration with *InterAx Biotech AG* who also characterised the selected ligands pharmacologically. An an-

tagonist for the β_2AR with a previously undescribed structural scaffold was successfully predicted in this study and a structure-activity relationship investigation showed the general affinity of this structural scaffold for this receptor.

The second studied GPCR was the CB2. In one study, molecular docking was applied to find structurally novel ligands for the CB2. For that, the docking setups were first optimised using a set known reference ligands. The prediction of water positions in the orthosteric binding pocket was shown to be a useful tool to achieve optimised docking results. These docking setups were then targeted by a large molecular library docking screen and several re-ranking and filtering steps were used to achieve better enrichment, similar to one of the approaches targeting the β_2AR. The selected molecules were then tested by collaboration partners from the Veprintsev lab at the University of Nottingham and preliminary results suggest that this screen was successful.
In the second study, Molecular Dynamics simulations were applied to the CB2 to investigate the structural basis of ligands inducing a certain recruitment bias. The results showed that it might be difficult to track recruitment bias with this method, however, indicators for receptor activation and deactivation could be observed.

In the last study, the RGS7-$G_{\beta5}$ complex was targeted using docking calculations. The overall goal is to find small molecules that can bind to this complex, thereby modulating its conformation and possibly its function. However, no binding sites of small molecules on this complex are known. Therefore, the main part of the study consisted of the prediction and evaluation of possible binding sites. Promising cavities were identified and will be targeted in docking screens to investigate whether they can serve the proposed function. This project was conducted in collaboration with the Martemyanov lab at the Scripps Research Institute in Florida.

Overall, the described studies were able to (1) show up ideas on how to best employ *in silico* tools to obtain the desired results, (2) find potential small molecule binding sites for a quite unexplored but therapeutically interesting target, (3) give insights on dynamic processes and structural rearrangements of receptor-ligand interactions leading to (biased) signalling and (4) successfully predict several novel ligands with different properties for two different GPCR targets with hit rates of up to 37%.

Kurzfassung

G-Protein-gekoppelte Rezeptoren (GPCRs) sind eine der wichtigsten Protein Familien und wirken als Signaltransduktoren in der Zellmembran. Etwa ein Drittel der heutzutage vermarkteten Wirkstoffe adressiert GPCRs, was ihre Bedeutung für Therapie und Krankheit verdeutlicht. Es ist daher nicht überraschend, dass GPCRs und ihre Signalwege von großem Interesse in der Forschung sind. In dieser Arbeit wurden *in silico* Methoden angewendet, um die Modulation der GPCR-Signalwege zu untersuchen.

Die Modulation der GPCR-Signalwege kann auf verschiedenen Niveaus ansetzen, z.B. bei GPCR-Liganden oder bei den weiterführenden Signalwegen. Interaktionen des Rezeptors mit kleinen Molekülen können zu dessen Inaktivierung oder Aktivierung führen. Letzteres kann zur intrazellulären Rekrutierung verschiedener Effektorproteine an den Rezeptor führen, was verschiedene Signalwege in der Zelle auslösen kann. Bestimmte Liganden können dabei eine stärkere Rekrutierung eines Effektorproteins im Vergleich zu anderen bewirken. Auf der strukturellen Ebene ist noch unklar, wie und warum diese Liganden einen solchen Bias bewirken. Überdies sind noch eine Vielzahl weiterer Proteine in die weiterführenden Signalwege von GPCRs verwickelt. Eine dieser Proteinfamilien sind die *Regulators for G protein Signalling* (RGS), die eine Deaktivierung von G-Proteinen und damit von GPCR-Signalwegen bewirken. Auch wenn bekannt ist, dass Mitglieder dieser Proteinfamilie eine Rolle bei vielen Prozessen und Krankheiten spielen –von denen viele mit GPCR-Signalwegen zusammenhängen–, sind sie noch nicht vollständig verstanden. GPCR-Signalwege müssen auf allen Niveaus besser verstanden werden, um in der Lage zu sein, sie rational zu beeinflussen. In dieser Arbeit wurden zwei GPCRs –der β_2-adrenerge Rezeptor (β_2AR) und der Cannabinoid Rezeptor 2 (CB2)– sowie ein Mitglied der RGS-Proteinfamilie –RGS7– in fünf verschiedenen Studien mit *in silico* Methoden addressiert, um ihre Signalwege und deren Modulation zu untersuchen.

Zwei der Studien zielten auf den β_2AR. Mehr als 30 Strukturen dieses *class A*-Rezeptors in verschiedenen Aktivierungszuständen sind verfügbar, was umfangreiche strukturelle Untersuchungen ermöglicht. Basierend auf dieser Tatsache wurden drei verschiedene Strukturen des β_2AR in unterschiedlichen Aktivierungszuständen unter Anwendung eines komparativen Docking-Ansatzes mit einer Molekülbibliothek addressiert. Ziel dabei war die Vorhersage neuartiger Agonisten dieses Rezeptors, basierend auf der Annahme, dass diese eher aus Docking-Berechnungen gegen eine aktive Konformation des Rezeptors resultieren sollten. Die ausgewählten Moleküle wurden pharmakologisch charakterisiert, was den Erfolg dieser Herangehensweise zeigte. Außerdem zeigte eine retrospektive Analyse dieses Docking-Ansatzes optimale Möglichkeiten auf, um die Chancen der

Entdeckung eines neuartigen Agonisten für diesen oder auch andere *class A*-Rezeptoren zu erhöhen.
Das Ziel der zweiten Studie zum β_2AR war es, mittels Docking-Berechnungen Antagonisten dieses Rezeptors mit neuartigen Strukturen vorherzusagen. Dieses Projekt wurde in Kollaboration mit *InterAx Biotech AG* bearbeitet, die auch die pharmakologische Charakterisierung der ausgewählten Moleküle durchführten. Dabei konnte erfolgreich ein Antagonist des β_2AR mit bisher nicht beschriebenem Strukturmotiv vorhergesagt werden und eine Untersuchung der Struktur-Wirkungs-Beziehung zeigte die generelle Affinität von Molekülen mit diesem Strukturmotiv für diesen Rezeptor.

Der zweite untersuchte GPCR war der CB2. In einer Studie wurde molekulares Docking angewendet, um strukturell neuartige Liganden für den CB2 zu finden. Dafür wurde zunächst mit Hilfe von bekannten Referenzliganden die Docking-Struktur optimiert. Die Vorhersage von Wasserpositionen in der orthosterischen Bindetasche stellte sich als nützliches Werkzeug zur Optimierung der Docking-Ergebnisse heraus. Die optimierten Docking-Strukturen wurden dann in einem Docking-Screen einer Molekülbibliothek addressiert und verschiedene *re-ranking*- und Filter-Schritte angewendet, um das *enrichment* zu verbessern, ähnlich zu der Herangehensweise an den β_2AR. Die ausgewählten Moleküle wurden dann von Kollaborationspartnern der Veprintsev-Gruppe der University of Nottingham getestet und die vorläufigen Ergebnisse legen den Erfolg des Screens nahe.
In der zweiten Studie wurden Molekulardynamische Simulationen auf den CB2 angewendet, um die strukturelle Basis von Liganden, die einen Rekrutierungs-Bias hervorrufen, zu untersuchen. Die Ergebnisse zeigen, dass es schwierig sein könnte, einen Rekrutierungs-Bias mit dieser Methode nachzuverfolgen, allerdings konnten Anzeichen für Rezeptor-Aktivierung und -Deaktivierung beobachtet werden.

In der letzten Studie wurde der RGS7-$G_{\beta5}$-Komplex durch Docking-Berechnungen addressiert. Das Ziel dabei ist, kleine Moleküle zu finden, die an den Komplex binden und dadurch seine Konformation und möglicherweise Funktion modulieren können. Allerdings sind keine Bindungsstellen für kleine Moleküle an dem Komplex bekannt. Deshalb bestand die Hauptaufgabe der Studie in der Vorhersage und Evaluation möglicher Bindungsstellen. Dabei konnten vielversprechende Taschen identifiziert werden und werden dann in Docking-Screens addressiert und bezüglich ihrer Eignung untersucht. Dieses Projekt wurde in Zusammenarbeit mit der Martemyanov-Gruppe vom Scripps Research Institute in Florida bearbeitet.

Die beschriebenen Studien konnten (1) Ideen zur besseren Anwendung von *in silico* Methoden zum Erreichen der gewünschten Ergebnisse aufzeigen, (2) potentielle Bindestellen kleiner Moleküle an einem wenig untersuchten, aber therapeutisch interessanten, Zielprotein finden, (3) Einsichten in dynamische Prozesse und strukturelle Umordnungen von Rezeptor-Ligand-Interaktionen, die zu

einem Signal-Bias führen, liefern und (4) mit Hit-Raten bis zu 37% erfolgreich neuartige Liganden mit unterschiedlichen Eigenschaften für zwei verschiedene Ziel-GPCRs vorhersagen.

Contents

List of Figures

List of Tables

Abbreviations

β_1AR	β_1-adrenergic receptor
β_2AR	β_2-adrenergic receptor
β_2AR^{active}	β_2-adrenergic receptor in an active conformation
$\beta_2AR^{inactive}$	β_2-adrenergic receptor in an inactive conformation
βAR	β-adrenergic receptor
2D	two dimentional
Å	Ångström
AUC	Area under the curve
BRET	Bioluminescence Resonance Energy Transfer
cAMP	3',5'-cyclic adenosine-monophosphate
CB1	Cannabinoid receptor 1
CB2	Cannabinoid receptor 2
CGenFF	CHARMM General Force Field
CHO	Chinese Hamster Ovary
CNS	Central Nervous System
COPD	Chronic obstructuive pulmonary disease
CRE	cAMP response element
Cryo-EM	Cryo Electron Microscopy
Da	Dalton
DEP	Disheveled, Egl-10, Pleckstrin
DHEX	DEP helical extension
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DUD-E	Directory of Useful Decoys-Enhanced
E_{max}	Maximum response induced by an agonist
EC_{50}	Half maximal effective concentration
ECFP4	Extended-Connectivity Fingerprint 4
ECL	Extracellular loop
eq.	equation
ERK1/2	extracellular signal-regulated kinase 1/2
FRET	Fluorescence Resonance Energy Transfer

G protein	guanine nucleotide-binding protein
$G_{\beta 5}$	type 5 G_{β} protein
GAP	GTPase activating protein
GDP	Guanosine-5'-diphosphate
GGL	G_{γ}-like protein
GIRK	G protein-coupled inwardly-rectifying potassium channel
GPCR	G protein-coupled receptor
GRK	GPCR kinase
GTP	Guanosine-5'-triphosphate
HEK	Human Embryonic Kidney
HID	δ-protonated histidine
HIE	ε-protonated histidine
HIP	δ- and ε-protonated histidine
HTRF	Homogenous Time-Resolved Fluorescence
IBMX	3-isobutyl-1-methylxanthine
IC_{50}	Half maximal inhibitory concentration
ICL	Intracellular loop
K_D	Equilibrium dissociation constant
k_{off}	Dissociation rate constant
k_{on}	Association rate constant
logP	Partition coefficient
MAP	Mitogen-activated protein
MD	Molecular Dynamics
NAM	Negative allosteric modulators
NPT	constant particle number, pressure and temperature
NVT	constant particle number, volume and temperature
PAM	Positive allosteric modulators
PDB ID	Protein Data Bank identifier
PDE	Phosphodiesterase
K_i	Inhibition constant
PME	Particle Mesh Ewald
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
R7BP	R7 binding protein
R9AP	RGS9 anchoring protein

RGS	Regulator of G protein signalling
RMSD	Root-mean-square deviation
ROC	Receiver operating characteristic
S.E.M	Standard error of the mean
SAR	Structure-activity relationship
SAS	Solvent accessible surface
SBDD	Structure-based Drug Design
SD	Standard deviation
SEED	Solvation energy for exhaustive docking
SMILES	Simplified molecular-input line-entry system
SPAP	Secreted placental alkaline phosphatase
TM	Trans-membrane helix
TR-FRET	Time Resolved Fluorescence Resonance Energy Transfer
ZINC	ZINC is not Commercial

1 | Introduction

G protein-coupled receptors (GPCRs) are an important protein family encoded by more than 800 genes in the human body and targeted by about a third of today's marketed drugs.[1,2] Hence, their signalling and modulation is of great interest for researchers and pharmaceutical industry all over the world. This thesis describes five studies aiming to modulate GPCR signalling on different levels and in various ways, investigating three different protein targets.

In all studies, *in silico* methods were used to make predictions on a structural level and the main focus was on the interaction of small molecules with the target proteins. In four of the five described studies, molecular docking was applied to find such small molecules that would interact with the target protein and, thereby, modulate its function. In one study, Molecular Dynamics simulations were used to evaluate receptor-ligand interactions in a more dynamic way, studying structural changes during a time span of 1 µs. Additionally, the molecules that were predicted as small binders by the docking calculations in three of the studies were pharmacologically characterised, two times by collaborators and once by the author of this thesis.

The first chapter 2 aims to introduce the target proteins and the computational methods to the reader. The GPCR family and GPCR signalling are described, followed by a more detailed presentation of the three proteins that were targeted in this thesis: the β_2-adrenergic receptor, the Cannabinoid receptor 2 and the Regulator for G protein Signalling 7. Subsequently, the theoretical background of Molecular docking and Molecular Dynamics simulations is explained briefly.

In chapters 3 and 4 two docking studies targeting the β_2-adrenergic receptor are described. While the goal of the first was to find antagonists with novel structural scaffolds, the latter aimed to predict agonists using comparative dockings against three different structures of the β_2-adrenergic receptor. The pharmacological characterisation of the ligands discovered in the study described in chapter 4 was conducted by the author of this thesis.

Chapter 5 describes another docking study, this time targeting the Cannabinoid receptor 2. The docking results were first optimised by predicting possible water positions in the orthosteric binding pocket of the receptor. The evaluation of the optimisation steps was based on the docking poses of a set of known reference ligands of the Cannabinoid receptor 2. This was followed by a docking screen aiming to find structurally novel ligands of this receptor.

The study described in chapter 6 was also investigating the Cannabinoid receptor 2. Here, the receptor was simulated in Molecular Dynamics simulations to retrospectively analyse the causes and mechanisms based on which certain ligands induce a certain recruitment bias. The ligands included in these simulations were selected based on experimental recruitment data provided by

a collaboration partner. The receptor was then simulated in complex with these four ligands and different effector proteins.

In the last chapter 7, the Regulator for G protein Signalling 7 protein was investigated further. The aim was to find potential binding sites in the protein complex at which small molecules could bind and modulate the conformation of the complex to potentially modulate its function. Such small molecule binders could be used as tool compounds to investigate and better understand the functions of this quite unexplored protein. Two different docking programs were used for this approach.

2 | Basic concepts

This chapter aims to give an introduction to the main targets and computational methods that were used in the studies described in this thesis. First, the G protein-coupled receptor (GPCR) family as well as their signalling and structural features are introduced followed by a more detailed description of the GPCR targets studied in this thesis: the β_2-adrenergic receptor and the Cannabinoid receptor 2. Subsequently, the Regulator of G protein signalling protein family is introduced and its family member RGS7 which was also studied in this thesis will be described in more detail. Following the description of the targets are explanations about the computational methods that were used most during the studies. Molecular docking as well as the docking programs DOCK and SEED played an important role throughout all projects described in this thesis and the background theory will be outlined. Further, the theoretical background of Molecular dynamics simulations will be explained.

2.1 G protein-coupled receptors and their signalling

G protein-coupled receptors (GPCRs) are one of the most extensively researched protein families. They are deeply involved in a plethora of physiological processes in organisms, e.g. vision, taste and many more, and linked to several physiological and psychological diseases. This is reflected in the fact that roughly one third of the drugs approved by the US Food and Drug Administration (FDA) target members of this protein family.[2] It is the largest protein family in humans with approximately 800 members which attests its importance even further.[1]

GPCRs are cell membrane spanning proteins and involved in various signalling pathways in different cell types. Upon recognition of a stimulus such as a small molecule, peptide, protein or even a photon on the extracellular side of the receptor, conformational changes of the receptor are induced, leading to the recruitment of effector proteins such as G proteins at the intracellular side of the receptor which initialises several signalling cascades. To put it in different words, GPCRs are the mediators connecting a wide variety of stimuli to a limited number of effector proteins and signalling pathways inside the cells, showing how efficient nature solved the problem of signal transduction.

This section discusses different aspects of this fascinating protein family.

2.1.1 GPCR structures and their determination

GPCRs consist of seven trans-membrane α-helices (TM) that are connected by three intracellular loops (ICL) and three extracellular loops (ECL), an extracellular N-terminus and an intracellular C-terminus (see Figure 2.1). The first structure of a GPCR was solved in 2000 for Rhodopsin,

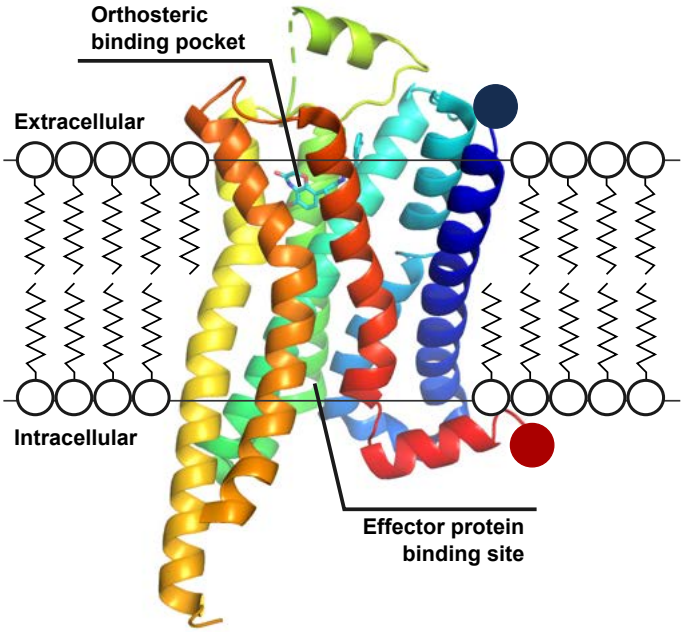


Figure 2.1: Schematic representation of a class A GPCR in the lipid bilayer of a membrane. The connection of the seven TMs and helix VIII is visualised in rainbow colouring. The N-terminus is marked by a blue circle and the C-terminus by a red circle (termini are truncated). Extracellular and intracellular side as well as the orthosteric ligand binding pocket and the effector protein binding site are indicated. The GPCR structure is the β_2-adrenergic receptor structure with PDB ID 3SN6.[8]

confirming earlier observations of the overall GPCR topology and the arrangement of the seven TMs in an anticlockwise manner when observed from the extracellular site.[3–5] Although this was the initial step for further developments and helped to gain a better understanding of GPCRs, it was only of limited use for further Structure-Based Drug Design (SBDD) studies, since Rhodopsin is not a drug target and also covalently bound with its ligand. However, in 2007 the structure of another GPCR, the β_2-adrenergic receptor (β_2AR), could be solved, therefore offering new possibilities for SBDD and a broader understanding of GPCRs.[6,7]

While obtaining the crystal structure of Rhodopsin was comparably easy due to its high expression levels in natural tissues, its relative thermostability and the non-diffusable, covalently bound ligand, the crystallisation of other GPCRs proved a bigger challenge.[9] Firstly, the purification of the receptor in high amounts had to be realised. Secondly, GPCRs are membrane receptors which means they only fold naturally in a membrane-like environment, unless certain thermostabilising mutations are introduced.[10] Lastly, to obtain crystals that are usable for X-ray crystallography certain other tricks were necessary to gain properly sized and well-ordered crystals. Methods to stabilise the overall receptor structure include the mutation of certain residues and the removal of

long, flexible loops and termini. Introduction of specific proteins, which can be covalently but also non-covalently bound, can stabilise the overall receptor structure but also specific conformations of the receptor. A method to obtain more well-ordered crystals is the fusion of N-terminus or ICL3 with small proteins to mediate contacts between single protein molecules.[11]

However, after the first β_2AR structure was solved in 2007 the number of available GPCR structures grew rapidly with 488 structures from 88 unique GPCRs available in November 2020 (numbers from gpcrdb.org;[12] see Figure 2.2). Although techniques and methods to crystallise GPCRs became more handy, another challenge became apparent: Since GPCRs transduce signals from the outside of a cell to the inside upon the influence of a stimulus there has to be a conformational difference between the activated and the inactivated receptor. The activated form of a receptor has to undergo structural changes to bind to an intracellular effector protein. Crystal structures of GPCRs can only represent snapshots of the spectrum of different receptor states. Therefore, it is necessary to obtain snapshots in various activation states of the receptor to gain a more realistic picture and a better starting point for e.g. SBDD. Obtaining crystal structures of receptors in certain active conformations has proven to be difficult due to the higher structural stability and, thus, better crystallographical accessibility of inactive conformations of receptors.[11] Certain active conformations of GPCRs could only be stabilised in the presence of a nanobody which served as a substitute for G proteins. Finally, in 2011 the first structure of the β_2AR in complex with the G_s protein was published.[8]

In spite of that, this first GPCR structure in complex with an effector protein has not kicked off a flood of similar structures from other GPCRs since this complex is continuously difficult to stabilise and crystallise for X-ray crystallography. With recent developments in another structure determination technology, Cryo Electron Microscopy (Cryo-EM) emerged as another option of determining structures of bigger protein complexes.[13] These new developments in Cryo-EM technology opened up more possibilities to determine GPCR-effector protein complexes. The advantage of this technique over classical x-ray crystallography is that lower amounts of protein are needed and that the protein does not need to crystallise in big, well-ordered crystals which resulted in a rising number of GPCR-effector protein complex structures determined by this method.[14,15] However, Cryo-EM has its limitations e.g. concerning high resolution of protein (complexes) smaller than 100 kDa or lower resolutions in more flexible regions which often includes the extracellular, for SBDD interesting binding pocket regions. Thus, X-ray crystallography and the so far developed methods remain important for e.g. determination of the GPCR structures in inactive conformation due to the lower molecular weight.[14,16]

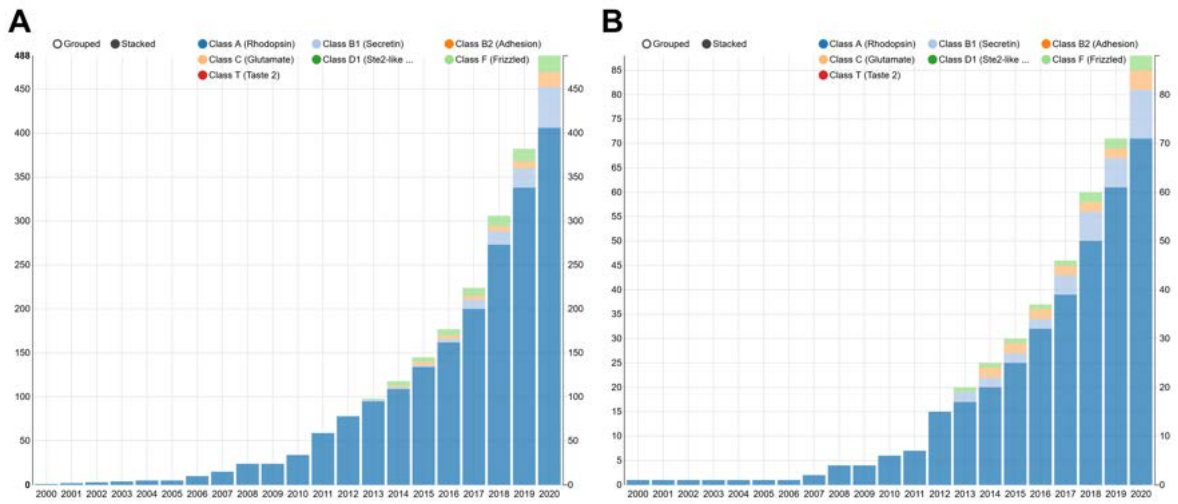


Figure 2.2: Growth of available crystal structures after the release of the first Rhodopsin structure in 2000. A) Overall number of available crystal structures. As of November 2020, a total of 488 structures was available, with 406 class A, 46 class B1 (Secretin), 17 class C and 19 class F structures. B) Number of unique GPCRs with crystal structures. As of November 2020, structures for 88 different GPCRs have been solved, with 71 class A, ten class B1 (Secretin), four class C and three class F GPCRs. Plots were taken from gpcrdb.org[12]

2.1.2 Classification of GPCRs

The GPCR superfamily consists of an enormous number of members with diverse primary structures and mechanisms. A phylogenetic classification approach was applied to group GPCRs in certain classes or clans and, thereby, structure this huge protein family. Although several different attempts were made, the best-known and most used approach was suggested by Kolakowski in 1994.[17] Here, GPCRs are classified into classes A-F which are then divided into subclans. This classification system includes also GPCRs which do not exist in humans, e.g. all GPCRs in classes D and E. In 2003, a new classification system was introduced, the so-called GRAFS classification system.[1] This system, although similar to the one introduced by Kolakowski, utilised the more advanced knowledge about GPCRs encoded in the human genome and multiple different phylogenetic analyses to group GPCRs into the different classes. According to their criteria, most GPCRs could be classified in one of five main groups: *Glutamate*, *Rhodopsin*, *Adhesion*, *Frizzled/Tas2* and *Secretin* (see Figure 2.3). Of these main groups the *Rhodopsin* group is by far the biggest with more than 700 members and was, therefore, subdivided into subgroups α, β, γ and δ.[1] In general, these two classification systems overlap, i.e. class A and the *Rhodopsin* family are mainly the same, while class B consists of the *Adhesion* and *Secretin* family.[4]

Since the receptors that are discussed in more detail in this work, the β_2-adrenergic receptor (β_2AR) and the Cannabinoid receptor 2 (CB2), are both classified as members of the Class A/*Rhodopsin* family (α subgroup), only this subfamily will be discussed in more detail in the following.

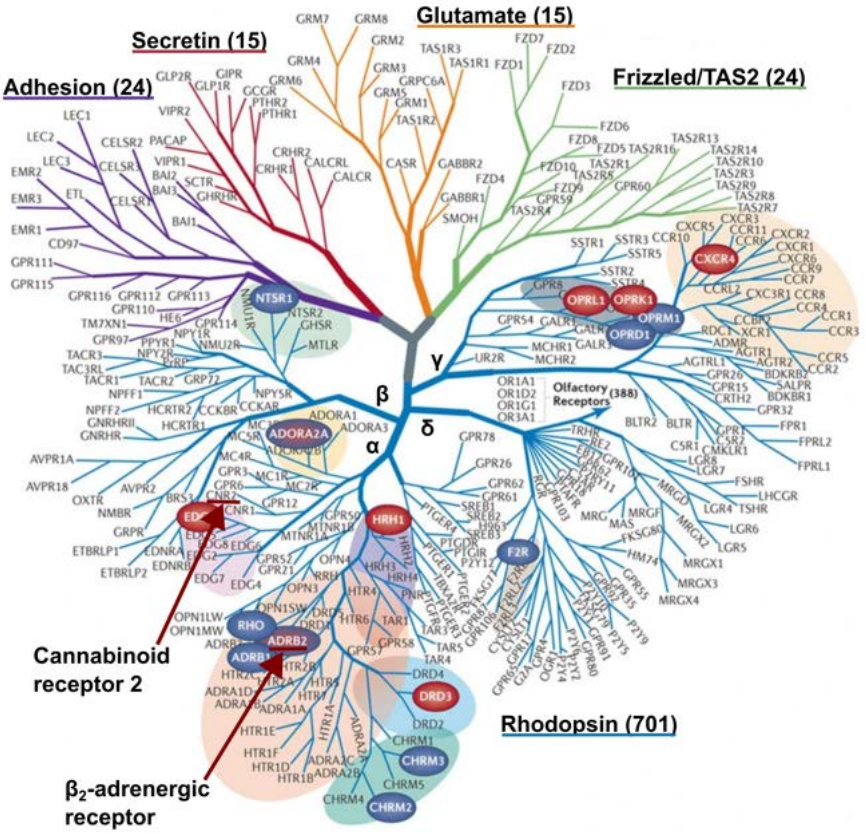


Figure 2.3: Phylogenetic tree of the GPCR family as defined by *Fredriksson et al.*[1] The positions of the two receptors targeted in this thesis –the β_2-adrenergic receptor and the Cannabinoid receptor 2– are indicated in their branches in the α subfamily. Figure adapted from Stevens et al.[18]

Class A GPCRs/*Rhodopsin* family. The class A/*Rhodopsin* GPCR subfamily consists of almost 700 members, most of which are olfactory receptors.[1] Among the remaining 240 non-olfactory receptors are many important drug targets, making this GPCR family an interesting target for GPCR drug discovery. Most members of the *Rhodopsin* family have short N-termini compared to other GPCR families which are mostly not involved in ligand recognition and binding.[4] Furthermore, several residue patterns are conserved between most of the members of the *Rhodopsin* family, including the β_2AR and the CB2. These are (with numbers in superscript according to the Ballesteros-Weinstein enumeration scheme[19]) the D(E)$^{3.49}$R$^{3.50}$ Y$^{3.51}$ motif in the G protein binding site, the N$^{7.49}$ P$^{7.50}$xxY$^{7.53}$ motif which is also located close to the G protein binding site and the C$^{6.47}$W$^{6.48}$ xP$^{6.50}$ motif. All these conserved structural motifs seem to be involved in receptor

activation and deactivation processes.[4,11,20] Additionally, a conserved Na^+ ion spot which is located within the receptor helices between the extracellular and intracellular site plays an important role in the activation/deactivation of most class A GPCRs. In the inactive state of the receptor the residues in this area open up a small binding pocket in which the Na^+ ion and water molecules are located. Upon activation of the receptor, the pocket collapses leaving no space for the Na^+ ion which then leaves the receptor by a not entirely proven mechanism. Interestingly, several residues of this Na^+ binding pocket are conserved between most of the class A GPCRs, most importantly $D^{2.50}$, which is the residue forming ionic interactions to the Na^+ ion in the inactive state while probably remaining uncharged in active state.[21] Opposed to these conserved features in structure and sequence between the members of the *Rhodopsin* family, the ligand binding sites (i.e. orthosteric binding pockets) can be quite diverse and can bind structurally quite different ligands.[4]

2.1.3 Activation and signalling of class A GPCRs

GPCRs are connected to several different signalling pathways in cells and can be activated by binding of ligands to the receptor. The binding mode and the binding location of the ligand, influences the receptor conformation leading to different responses with different intensities inside the cells. These responses are mainly related to the coupling of effector proteins from three major protein classes to the receptor: G proteins, β-arrestins and GPCR kinases (GRKs).[22]

GPCR ligands. Molecules binding to GPCRs can affect signalling pathways in different ways. Ligands can be classified into different types based on the response induced upon their binding to a receptor relative to the basal signalling of that receptor if no ligand is bound (see Figure 2.4). If a ligand induces a signal higher than the basal response it is classified as an agonist. Depending on whether it induces the maximum possible response for a system or a lower than maximum response it can be sub-categorised as full agonist or partial agonist, respectively (Figure 2.4, red and orange curves). If a ligand does not change the response from the basal response upon binding it is categorised as a neutral antagonist (Figure 2.4, blue curve) and if it lowers the response below the basal response it is categorised as an inverse agonist (Figure 2.4, violet curve).[23] Additionally, the binding location can have an effect on the resulting response. Ligands can bind in the orthosteric binding pocket, which is located within the seven TMs on the extracellular side of the receptor, or they can bind to allosteric binding sites, which might be located anywhere on the receptor. These allosteric binding molecules can modulate responses induced by orthosteric ligands by either increasing or lowering this response and are then called positive or negative allosteric modulators (PAM/NAM), respectively.

The different responses induced by different ligands stem from small variations in interaction pat-

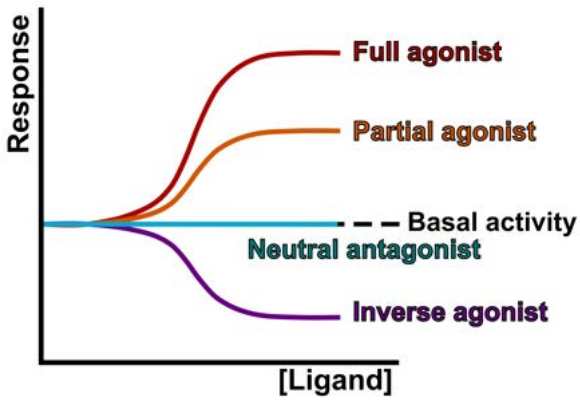


Figure 2.4: Exemplary dose-response curves for different types of GPCR ligands. Compared to the basal activity (black dotted line) a ligand can either induce a maximal increased response (full agonist, red curve), a partial increased response (partial agonist, orange curve), a lowered response (inverse agonist, violet curve) or not change the response at all (neutral antagonist, blue curve). The response of the assay readout is plotted against the ligand concentration used for stimulation.

terns between ligand and receptor. These small differences inside the binding pocket are loosely coupled to structural variations of the overall receptor conformation and, therefore, the stabilisation of different activation states of the receptor.[24] It can be assumed that it is easier to find a novel neutral antagonist for a receptor, i.e. a molecule that does not need to shift the conformational equilibrium of the receptor to a certain conformation (much like in the basal state), as opposed to an agonist, inverse agonist or allosteric modulator that needs to rather stabilise a certain activated conformation.

Conformational changes of GPCRs. In the early days of GPCR research it was assumed that the receptor could either adopt an active conformation that is able to interact with G proteins or an inactive conformation that cannot interact with G proteins.[22] Over time it became clear, that there is more to the receptor activation process than these 'on' or 'off' states. Different studies showed that the receptor can adopt a plethora of different conformations and, therefore, activation states.[22,25] Even in the basal, ligand-unbound state the receptor can dynamically take different conformations, some of which can interact with effector proteins which explains constitutive activity of receptors. This equilibrium between different receptor conformations can then be shifted towards a certain conformation upon binding of ligands or effector proteins, however, it will still remain a dynamic process.[24]

Keeping this in mind, crystal structures of certain activation states, e.g. in complex with an inverse agonist ('inactive') or a G protein ('active'), should be considered snapshots of one part of the

actual ongoing processes since they can only represent one thermodynamically stable state of the complex.[25] However, they can serve as a useful starting point to understand the conformational diversity of GPCRs and can be complemented with dynamic studies to get a more complete picture of the ongoing processes.

G protein mediated signalling. G proteins are heterotrimeric complexes build up of α, β and γ subunits. In the inactive state of the G protein, the G_α subunit binds Guanosine-5'-diphosphate (GDP) and interacts with the $G_{\beta\gamma}$ subunit. Upon interaction with the activated receptor, the GDP is released from the G_α subunit. Due to the high concentration of Guanosine-5'-triphosphate (GTP) inside the cells, the G_α subunit binds GTP causing it to dissociate from receptor and $G_{\beta\gamma}$ subunits. The GTP-bound G_α subunit as well as the $G_{\beta\gamma}$ subunits are then involved in different downstream signalling pathways. Since the G_α subunit contains GTPase, the bound GTP will be hydrolysed to GDP at some point, resulting in the inactivation of the G_α subunit and the reassociation with the $G_{\beta\gamma}$ subunits. The GTP hydrolysis can be accelerated by certain proteins such as the Regulators for G protein Signalling (RGS) by increasing the GTPase activity of the G_α subunit.[26]
This basic mechanism takes place for all main types of G proteins which can be classified in four main subfamilies of the G_α subunit: G_s, $G_{i/o}$ $G_{q/11}$ and $G_{12/13}$.[27] Each of these subfamilies shows different affinities for different receptors and induces different downstream signalling pathways. It is remarkable that each G protein can form interactions and is activated by a variety of different receptors, and vice versa one receptor can interact with various G proteins.[23] The receptors investigated in this work couple mainly to G_s (β_2AR) or G_i (CB2) proteins. Both G_s and G_i act on the adenylyl cyclase by either activating (G_s) or inhibiting (G_i) 3',5'-cyclic adenosine-monophosphate (cAMP) production which influences downstream cAMP dependent processes. The $G_{\beta\gamma}$ subunits can interact with e.g. ion channels influencing the ion influx into the cells.[8,26]

With the first crystal structure of a ternary complex of G protein, receptor and ligand, it was possible to evaluate the interactions between receptor and G protein in more detail.[8] The α5 helix of the G_α subunit is inserted into a cavity between the helices on the intracellular side of the receptor and polar and apolar interactions between various residues of both proteins are formed. Recently and mostly due to improvements in Cryo-EM techniques, several new structures of ternary complexes with different receptors and G proteins were published, which allows to compare interactions between different receptors and G proteins and might help to explain selectivity of certain G proteins for certain receptors.
It has been shown that binding of an agonist increases the binding affinity of the G protein to the receptor, and vice versa.[28,29] The thermodynamically most stable form of the ternary complex seems to be between the ligand bound receptor and the GDP- and GTP-free G protein. Structural

differences can be observed when comparing the structures of receptor-unbound and GDP containing G protein with that of receptor-bound, GDP-unbound G protein. For the final activation of the G protein, resulting in release of the GDP, the G_α subunit and especially the α5 helix have to undergo large conformational changes to allow residues that are involved in interactions within the G_α subunit in the GDP-bound state to interact with the receptor.[30]

Receptor deactivation and the role of β-arrestins. After activation of the G protein and the dissociation of G_α and $G_{\beta\gamma}$, the activated receptor could interact with another G protein and, hence, keep signalling. However, since it is unfavourable to continue signalling inside the cell once the extracellular signal has been received, an intracellular deactivation mechanism is necessary. To deactivate the receptor and stop the signalling process, the receptor C-terminus is phosphorylated by GRKs which leads to binding of β-arrestin to the receptor. There are seven different GRKs and only two different (non-visual) β-arrestins that can bind to a plethora of different receptors, demonstrating once more the efficiency of GPCR mediated signalling.

Although G proteins can still interact with the phosphorylated receptor, β-arrestins have a much higher affinity for the phosphorylated receptor and, therefore, outcompete them. Binding of β-arrestin to the receptor does not only block binding of G proteins but can also lead to internalisation and/or recycling of the receptor and might induce β-arrestin-mediated signalling pathways. The formation of megacomplexes of receptor interacting simultaneously with G protein and β-arrestin is also possible for certain receptor types as can be seen from continued signalling of internalised, arrestin-bound receptors and structural studies with mutated receptors.[31,32]

Recruitment of β-arrestin to the activated receptor might induce the mitogen-activated protein (MAP) kinase pathway and the activation of extracellular signal-regulated kinase 1/2 (ERK1/2), however, this and the involvement of G proteins in these processes are still a case of debate.[33] Initially, the β-arrestin dependent ERK1/2 activation was found to be G protein independent which led to the conclusion that ligands inducing either only G protein or only β-arrestin dependent signalling pathways could be found, resulting in interesting new therapeutical approaches.[34] However, recent studies with "zero functional G" cells indicate that β-arrestin dependent ERK1/2 activation does not happen in absence of G proteins.[35] In the reverse study with cells lacking β-arrestins it was shown, that depending on the knockout method the signalling seemed to be compensated by enhanced G protein signalling. When using a permanent knockout of the gene (e.g. by using CRISPR/Cas) only those cells might survive that are able to compensate the missing function.[36] These results indicate that the true mechanism of a β-arrestin dependent signalling is not yet clear.[37] A signalling bias towards one of both G protein dependent or β-arrestin dependent signalling pathways might be possible, although probably not with zero signalling towards the other pathway.

Biased signalling. In the recent years biased signalling became a topic that attracted more and more interest in GPCR research. It refers to the possibility that several different signalling pathways can be induced with different intensities depending on the bound ligand. In this context, a biased ligand would induce a stronger response for one signalling pathway compared to another signalling pathway upon binding and activating a receptor. Therapeutically, biased ligands might be advantageous because they could lead to drugs with less side effects due to the stronger activation of the favoured signalling pathway over the pathway causing the side effects.[38,39] For this reason, GPCR research into this topic intensified during the last years, trying to describe biased signalling, understanding how it is caused and applying this knowledge to eventually find novel biased ligands. Usually, the bias between G protein and β-arrestin is described in this context but it can also be other types of bias, for example the comparison of signalling induced by different members of the G protein family.

However, biased signalling is more difficult and diverse to understand than initially assumed. Bias can take place at different stages of the signalling process which are not necessarily connected to bias in other stages. Hence, the same ligand can be defined as biased and unbiased depending on which part of the system was observed. The bias can occur at the level of recruitment of effector proteins to the receptor or in downstream signalling responses or at both levels. Additionally, the system itself can inherit bias, e.g. due to different affinities of different effector proteins or even members of one effector protein family towards the receptor.[40] Furthermore, quantification of bias can be difficult if results from different assays and different systems are compared, since especially downstream responses are dependent on many factors such as e.g. receptor expression levels. Bias of a ligand, therefore, usually needs to be defined with respect to a reference ligand with an unknown bias.[41–43] When discussing signalling and ligand bias, all these factors should be kept in mind to make sure that the label 'biased' or 'unbiased' is only used and given within these boundaries. However, the true impact of a biased ligand *in vivo* might still differ from expectations based on *in vitro* results since many additional factors could influence the response *in vivo*.[40,43]

Since the research done in this work is conducted on a molecular and structural level, it is best comparable to bias at the level of effector protein recruitment as opposed to bias in downstream signalling.

2.1.4 The β_2-adrenergic receptor (β_2AR)

The β_2-adrenergic receptor (β_2AR) is a class A GPCR that belongs to the aminergic receptor subfamily of adrenergic receptors. Next to Rhodopsin, this subfamily played an important role during the first discovery and functional description of GPCRs. The first concepts of receptors

were already formed more than a century ago, and in 1948 Raymond Ahlquist classified two types of adrenaline receptors (named after their endogenous ligand adrenaline and the reason for their discovery) which he called α and β.[44–46] Although Ahlquist himself as well as other scientists in the field later on doubted the existence of the receptors, the basis for GPCR research was set with this discovery. [47,48] Later, the adrenergic receptor family was divided into three subfamilies with distinct physiological functions, the α_1-, α_2- and β-adrenergic receptor subfamilies, each of which consists of three members.[49,50] The closest relative to the β_2AR is the β_1-adrenergic receptor (β_1AR) with 41% identity and 51% similarity in their sequences (calculated with gpcrdb.org[12]) and a very high similarity within the orthosteric binding site.

The adrenergic receptors and especially the β_2AR played an important role in the discovery of functions, mechanisms and signalling of GPCRs, with the first human β_2AR cDNA isolated and sequenced in 1987.[51] For a more detailed history on first GPCR discoveries see Lefkowitz[52], Lohse[53], Kobilka[54]. Due to its role during the initial discoveries on GPCRs, the β_2AR is probably one of the best investigated and described GPCRs in functional as well as in structural studies and is, therefore, an optimal model system for the investigation of underlying concepts of class A GPCRs. The β_2AR couples to the G_s signalling pathway and is, thus, involved in the activation of the adenylyl cyclase and also other signalling pathways mediated by the $G_{\beta\gamma}$ subunit, GRKs and β-arrestins. Additionally, it has been shown that the β_2AR can also couple to G_i proteins.[55]

The β_2AR in therapy and disease. The β_2AR is mostly expressed in smooth muscle tissue, especially in airways.[56–58] Agonists of the β_2AR induce a relaxation of the smooth muscle tissue and are, therefore, mainly used to treat respiratory diseases such as Asthma and chronic obstructuive pulmonary disease (COPD).[57–59] Other uses of agonists targeting the β_2AR include for example uterine smooth muscle tissue to prevent preterm labour.[60,61] Antagonists of the β-adrenergic receptors (so called beta-blockers) are usually targeting the β_1AR in the treatment of cardiovascular diseases.[62] Due to the high similarity between β_1AR and β_2AR they can also block the β_2AR causing shortness of breath as a side effect of these drugs.[63] Recently, it has also been reported that the β_2AR antagonist Propranolol seems to be effective in the treatment of infantile hemangioma when applied to the skin.[64] However, after concerns about the effect of this compound on the central nervous system of infantile hemanginoma patients have been raised, the discovery of novel β_2AR antagonists with less side effects is of interest.[65] Polymorphisms of the β_2AR are involved in several diseases and are known to cause changed responses to therapeutical treatment.[66,67]

Structural features of the β_2AR. Since the release of the first crystal structures of the β_2AR in 2007, the number of β_2AR structures has increased continuously.[6,7] In August 2020 a total of 35 structures of the β_2AR has been released in the Protein Data Bank (PDB; *rcsb.org*).[68] Different

techniques were used to obtain these crystal structures and they represent different activation states of the receptor, from a presumably fully inactive conformation in complex with an inverse agonist (e.g. PDB ID 3NY9[69]) to a presumably fully activated conformation in complex with a G_s protein (PDB ID 3SN6[8]). Most notable at the current state of knowledge are probably the crystal structure in complex with the G_s protein,[8] structures in complex with various allosteric modulators,[70–72] the structure of a megacomplex of a β_2V_2 mutant receptor with G_s protein and β-arrestin[32] and a study using serial femtosecond crystallography (SFX) and an X-ray free-electron laser (XFEL).[73]
Structurally, the β_2AR forms the for GPCRs typical bundle of seven TMs. The orthosteric binding site is buried within the seven TM core and accessible from the extracellular side of the receptor. The biggest differences between inactive, inverse agonist bound conformations and active, agonist/G protein bound conformations is visible on the cytoplasmic side of the receptor (see Figure 2.5). To make space for binding of the α5 helix of the G_α subunit, TM VI shifts outwards by approximately 14.8 Å (measured between E268$^{6.30}$ C_α in the active conformation with PDB ID 3SN6 vs. the inactive conformation with PDB ID 3NY9).[8,69] This shift is accomplished by a kink around residue P288$^{6.50}$ in the active conformation compared to the inactive conformation. Aside from the structural changes in TM VI, TM V forms two additional helical turns in the active conformation. Placement of all other helices is comparable and only slight shifts of the TMs can be observed between active and inactive conformations, especially on the extracellular side of the receptor. In the orthosteric binding pocket slight positional changes of the side chains can be seen. These small changes and changed interactions within the binding pocket contribute to the significant changes on the intracellular side of the receptor.

Ligands and their interactions with the β_2AR. First efforts to describe the binding pocket and how ligands interact with the β_2AR were already made long before the first GPCR structure was solved. With several mutational studies the most important interaction partners of known ligands of the β_2AR were investigated, both towards affinity and function (i.e. G_s mediated signalling). In 1987 the general binding location of the ligand was identified using several deletion mutants.[74] Around the same time, the most important residue involved in binding of ligands to the β_2AR was identified, D113$^{3.32}$.[75,76] Residue N293$^{6.55}$ was found to play a role in the enantiomer selectivity of several catecholamine β_2AR agonists and its involvement in interactions with e.g. adrenaline was later confirmed in crystal structures.[77,78] An important role was also found for S203$^{5.42}$, S204$^{5.43}$ and S207$^{5.46}$ regarding agonist binding as well as receptor activation.[79,80] However, although interaction with these residues might be of importance for catecholamines, there are also agonists known that do not interact with any of these three residues which means they are not of general importance for receptor activation by β_2AR agonists (e.g. the novel agonists described in chapter 4[81]).

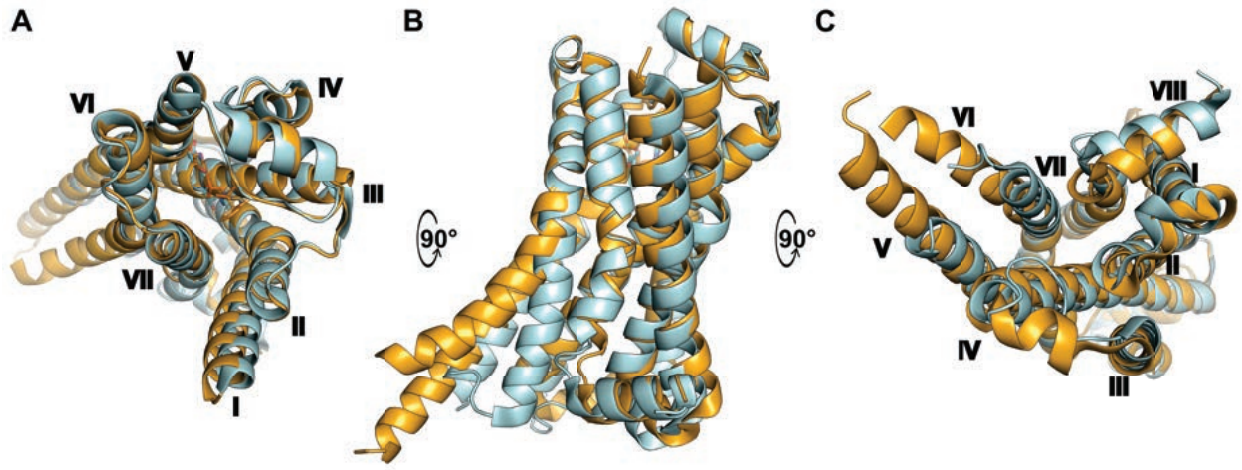


Figure 2.5: Overlay of the β_2AR in an inactive conformation (aquamarine; PDB ID 3NY9[69]) and in an active conformation (orange; PDB ID 3SN6[8]). A) Extracellular view, turned by 90° to the back to B) side view, turned by 90° to the back to C) intracellular view. TMs are labelled in A) and C). On the intracellular side the noticeable outward shift of TM VI as well as the prolongation of TM V upon activation are visible. On the extracellular side only slight TM shifts can be observed.

Many ligands of the β_2AR have a very similar structure to the endogenous ligands adrenaline and noradrenaline. Similar to these catecholamines, ligands usually contain an amine and often a hydroxyl-group in β-position. These functional groups interact with D113$^{3.32}$ and N312$^{7.39}$ via ionic interactions and hydrogen bonds. Connected to this β-hydroxyl-amine is usually an aromatic moiety which is placed within TM III, (IV,) V and VI and stabilised by hydrophobic interactions to several apolar residues in proximity to the ligand (see Figure 2.6). Hydrophilic groups attached to this aromatic moiety can then form hydrogen bonds to e.g. S203$^{5.42}$, S204$^{5.43}$, S207$^{5.46}$ or N293$^{6.55}$. The linker length between β-hydroxyl-amine and aromatic moiety is usually prolonged for antagonists compared to agonists with an additional C atom and ether group. Substituents attached to the amine on the opposite site of the β-hydroxyl-group can be of varying length, from mere hydrogen atoms in the case of noradrenaline to long alkylic chains that almost reach outside of the binding pocket as it is the case for salmeterol.[82] Although many ligands of the β_2AR display these structural features, only the amine group seems essential for ligand binding. Aside from that, β_2AR ligands can vary in their structural features, as can be seen for e.g. the novel ligands discovered in the study described in chapter 3.[83]

Differences between active and inactive structure are relatively small within the binding pocket with only slight changes in placement and angles of side chains. However, these changes overall lead to a smaller and more contracted binding pocket in the active conformation compared to the

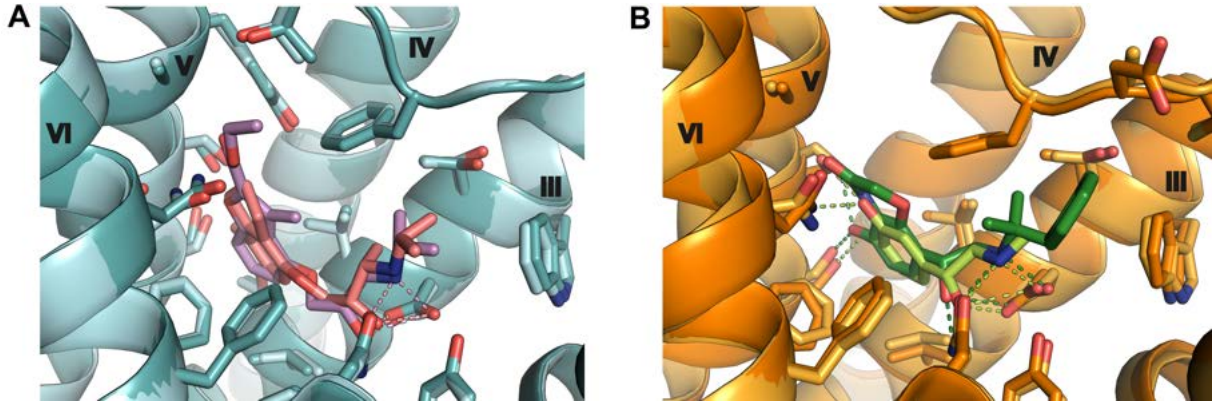


Figure 2.6: Orthosteric binding pocket of the β_2AR in A) inactive and B) active conformations. A) Overlay the β_2AR structures with PDB ID 3NY8[69] (darkteal) and 3NY9[69] (lightblue) with their respective ligands, the inverse agonists ICI 118551 (salmon) and Compound 1 (lilac). Both ligands form polar interactions to D113$^{3.32}$ and N312$^{7.39}$. B) Overlay of the β_2AR structures with PDB ID 4LDO[78] (orange) and 3SN6[8] (yellow) with their respective ligands, the agonists Adrenaline (light green) and BI167107 (dark green). These agonists form polar interactions to D113$^{3.32}$ and N312$^{7.39}$ as well as to S203$^{5.42}$, S207$^{5.46}$ and N293$^{6.55}$. Agonists as well as antagonists are additionally stabilised by apolar interactions between the aromatic moiety placed within TM III, (IV,) V and VI and apolar residues of the receptor. The binding site and ligand binding poses are similar comparing the inactive conformations as well as the active conformations.

inactive conformation. Antagonist molecules with the longer linker between β-hydroxyl-amine and aromatic moiety can, therefore, rather be accommodated in the bigger binding pocket of the inactive conformation than in the more contracted binding pocket of the active conformation and, vice versa, the smaller size of agonist molecules contributes to the contraction of the receptor binding pocket as found in an activated conformation of the β_2AR. As has been shown in different studies, these factors influence each other and work both ways, i.e. the bound ligand has an influence on the receptor conformation and the receptor conformation influences the affinity of ligands with certain functions to the receptor.[29,81,84]

2.1.5 The Cannabinoid receptor 2 (CB2)

The Cannabinoid receptor 2 (CB2) is a class A GPCR and belongs to the lipid receptor subfamily of cannabinoid receptors. Its closest relative is the Cannabinoid receptor 1 (CB1) with 44% amino acid identity.[85] Both receptors interact with lipophilic compounds such as cannabinoids which represent a main class of Cannabinoid receptor ligands and can be found endogenously, expressed in plants or

derived synthetically. The first research that at least indirectly regarded the Cannabinoid receptors was already conducted towards the end of the 19th century, when the first phytocannabinoid, Cannabinol, was isolated from *Cannabis* plants (see reviews by Pertwee[86], Mechoulam and Hanuš[87] for a brief history on Cannabinoid research). In 1975 it was still theorised that the effects of Cannabinoids were not actually induced by binding to receptors but rather by disordering the membrane lipids of cells.[88] Several years later, in 1988, the existence of these receptors was proven by *in vitro* binding experiments with radiolabelled CP55940, a ligand with a strong affinity to both Cannabinoid receptors.[89] Finally, both receptors could be cloned and their sequence elucidated, for the CB1 in 1990[90] and for the CB2 in 1993.[85]

In recent years, it has also been discussed whether there might be additional members of the Cannabinoid receptor subfamily. Based on the ability of certain endo- and phytocannabinoids to bind to and activate GPCRs GPR18 and GPR55 it was suggested that these two receptors might in fact be Cannabinoid receptors.[91,92] However, the IUPHAR committee working on the classification of Cannabinoid receptors came to the conclusion that these two receptors cannot be classified as members of the Cannabinoid receptor family, although these receptors fulfil most of the rules the committee proposed for the classification.[93,94]

The CB2 couples mainly to $G_{i/o}$ proteins and the associated signalling pathways but has also been shown to be involved in other processes such as MAP kinase signalling, transient increases of Ca^{2+}, G_s mediated signalling, G protein-coupled inwardly-rectifying potassium channel (GIRK) effects and also β-arrestin recruitment. [95–99]

The CB2 in therapy and disease. While the CB1 is mostly expressed in the central nervous system (CNS), the CB2 is found in immune tissue, especially B cells and NK cells.[100,101] However, increased expression levels of the CB2 were also found within the CNS in microglia, especially during neuroinflammations or nerve injuries and could, therefore, play an important role in various neurodegenerative diseases.[102–106] Hence, the CB2 has been proposed as target for various medical conditions such as treatment of pain and inflammatory conditions[107] but also in the treatment of neurodegenerative diseases connected to neuroinflammation such as Alzheimer disease or Multiple Sclerosis.[105] Aside from that, the CB2 is also involved in inflammatory processes in various organs and might be targeted to treat these conditions as well.[108] Recently, it has been hypothesised that targeting the CB2 might have positive effects in patients with Corona Virus Disease (COVID-19) caused by the Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) due to its involvement in the modulation of immune and inflammatory responses.[109]

Especially desirable are drugs targeting the CB2 selectively over the CB1 since it has been shown that psychotropic side effects are caused via the CB1 but not the CB2.[110,111]

Structural features of the CB2. In 2019, the first complete crystal structure of the CB2 in complex with antagonist AM10257 was released (PDB ID 5ZTY).[112] Like other class A GPCRs, the CB2 consists of seven TMs and an amphipathic helix VIII on the intracellular side of the receptor.
Only recently, in 2020, the first structures of the CB2 in active conformations were released: an X-ray structure in complex with agonist AM12033 (PDB ID 6KPC[113]) and two Cryo-EM structures in complex with a G_i heterotrimer and two different agonists, AM12033 (PDB ID 6KPF[113]) and WIN 55,212-2 (PDB ID 6PT0[114]). When comparing the structure in an inactive conformation to these structures in active conformations, several differences similar to the ones observed for the β_2AR can be noticed. The biggest differences occur on the intracellular side with an outward shift of approximately 13 Å of TM VI (measured between D240$^{6.30}$ C_α of inactive PDB ID 5ZTY and the G_i-bound structures PDB ID 6KPF and 6PT0) with a kink at roughly P260$^{6.50}$ and two additional helical turns in TM V (see Figure 2.7). These structural differences are only visible for the receptor structure in complex with the G_i protein but not for the structure which was crystallised with an agonist only, which suggests that the insertion of the α5 helix of the G protein into the seven TM core of the receptor is necessary for these major structural changes. On the extracellular side of the receptor, the differences between active and inactive conformations are very slight and mainly visible in changes of site chain positions within the orthosteric binding pocket, much like for the β_2AR.

Ligands and their interactions with the CB2. As described above, cannabinoids are compounds that act on the endocannabinoid system and can be classified into phyto-, endo- and synthetic cannabinoids. A shared feature of all cannabinoids is their lipophilicity while aside from that they can be structurally very diverse (e.g. comparing Anandamide, THC and CP55940). Cannabinoids can be more selective for one of the CBs over the other one and were shown to also interact with other GPCRs such as GPR55. The first closer investigated and described cannabinoids were phytocannabinoids such as Δ^9-tetrahydrocannabinol (Δ^9-THC) and Cannabinol which were extracted from *Cannabis* plants. The search for endocannabinoids began only after the existence of the first Cannabinoid receptor was proven and when it became important to explain the reason for its existence. Soon after, two endocannabinoid ligands were identified, Anandamide and 2-Arachidonoylglycerol (2-AG), showing that the CBs play an internal regulatory role in organisms.[115,116]
The development of synthetic cannabinoids was already pursued before the existence of the receptor was proven, resulting in several widely used compounds with often higher potency than the phytocannabinoids. These synthetic cannabinoids can resemble a cannabinol-like structure or be structurally different from the phyto- or endocannabinoids.[86,87]

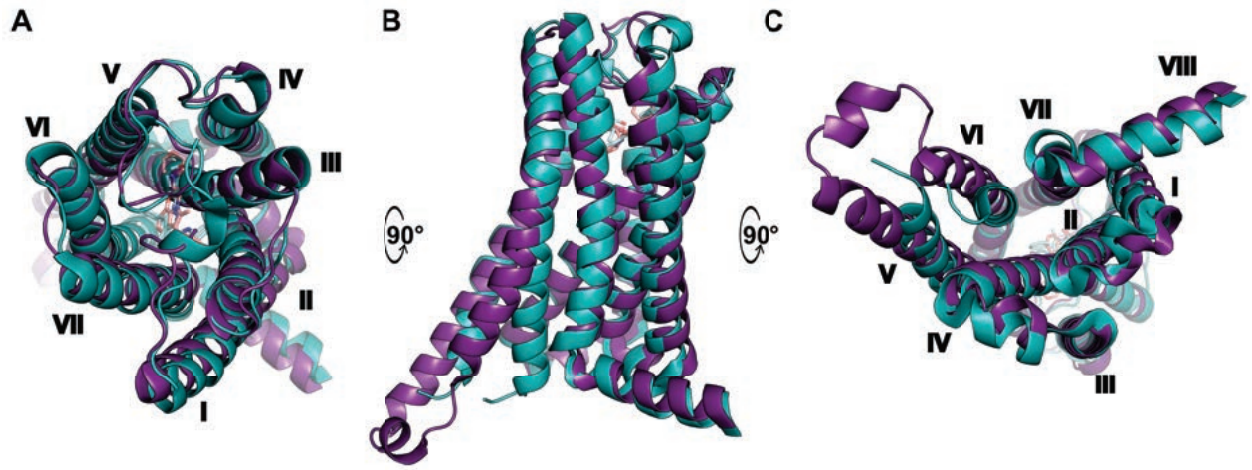


Figure 2.7: Overlay of the CB2 in an inactive conformation (teal; PDB ID 5ZTY[112]) and in an active conformation (violet; PDB ID 6PT0[114]). A) Extracellular view, turned by 90° to the back to B) side view, turned by 90° to the back to C) intracellular view. TMs are labelled in A) and C). The N-terminus and the ECL2 form a 'lid' on top of the orthosteric binding site, as can be seen in A). The noticeable outward shift of TM VI as well as the prolongation of TM V upon activation are visible on the intracellular side in C). On the extracellular side only slight TM shifts can be observed. ICL3 is only resolved in the active conformation and quite short compared to e.g. the β_2AR.

The orthosteric binding pocket is located within the seven TM core on the extracellular side of the receptor, with ECL2 and N-terminus forming a 'lid' on top of the receptor. This binding pocket is very apolar with only few polar residues such as T114$^{3.33}$ or S285$^{7.39}$. Differences in the binding pocket between active and inactive conformations are rather small and mainly visible as a slight decrease of the volume of the binding pocket in the active conformation and different orientations of the side chains. Mutational studies showed that certain residues located in the binding pocket seem to play an important role in ligand binding and receptor activation, such as F87$^{2.57}$, F91$^{2.61}$, F94$^{2.64}$, W194$^{5.43}$ and H95$^{2.65}$.[112,117] Another interesting residue in the binding pocket is W258$^{6.48}$ which is proposed to play an important role in receptor activation. This residue is highly conserved among class A GPCRs and is proposed to act as a toggle switch by taking different rotamer conformations and thereby stabilising certain receptor conformations in different activation states.[118–121] In fact, when comparing the available CB2 structures in active and inactive conformations the rotation of W258$^{6.48}$ can be observed easily, taking more space in the binding pocket in the active conformation (see Figure 2.8).[112,114]

The available structures of the CB2 were crystallised with structurally diverse ligands, which allows a comparison of ligand binding modes.[112–114] As can be seen, the binding mode for cannabinol-

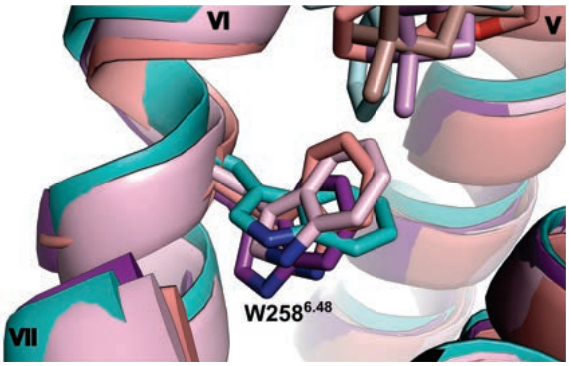


Figure 2.8: Rotamers of W258$^{6.48}$ in different conformations of the CB2. While this residue is not pointing into the binding pocket in the inactive conformation (teal; PDB ID 5ZTY[112]), it adopts a different conformation in the active conformation of the CB2. Although there are slight differences visible for the different structures in active conformations (violet: 6PT0;[114] salmon: 6KPF;[113] rose: 6KPC[113]) it is pointing into the binding pocket in each of them. These rotamer changes of W258$^{6.48}$ are proposed to play an important role in receptor activation.

like ligands such as AM12033 differs from those of WIN 55,212-2 or AM10257 (see Figure 2.9). The binding poses of all of these ligands are mainly stabilised by hydrophobic and aromatic interactions. AM12033 does also form polar interactions with S285$^{7.39}$ and the backbone of L182^{ECL2}. Interestingly, this seems not to be the case for WIN 55,212-2 and AM10257. Especially the carbonyl group of WIN 55,212-2 and the amide of AM10257 seem to be stranded without interaction partners. The residue closest to these 'stranded' functional groups is in both cases S285$^{7.39}$, which also interacts with AM12033, and it is possible that an unresolved water molecule could mediate interactions.

2.2 Regulator of G protein signalling (RGS)

Regulator for G protein signalling (RGS) proteins are a protein family that plays an important role in the regulation of G protein mediated signalling. Members of this protein family act as GTPase activating proteins (GAPs) that increase the GTPase activity of the G_α subunit.[122–126] The activated form of the G_α subunit is deactivated upon hydrolysis of the bound GTP to GDP, thereby terminating the associated signalling. Although the G_α protein inherits GTPase activity it has been shown that the hydrolysis of GTP is increased drastically upon interaction with RGS, explaining the discrepancy between hydrolysis rates of G_α *in vitro* and processes *in vivo*.[127–130] In the recent years, it became clear that RGS proteins can not only interact with and influence the G_α protein but also various other proteins in cells. This indicates that their regulatory function is not only limited to G protein mediated signalling pathways but they are also involved in a plethora

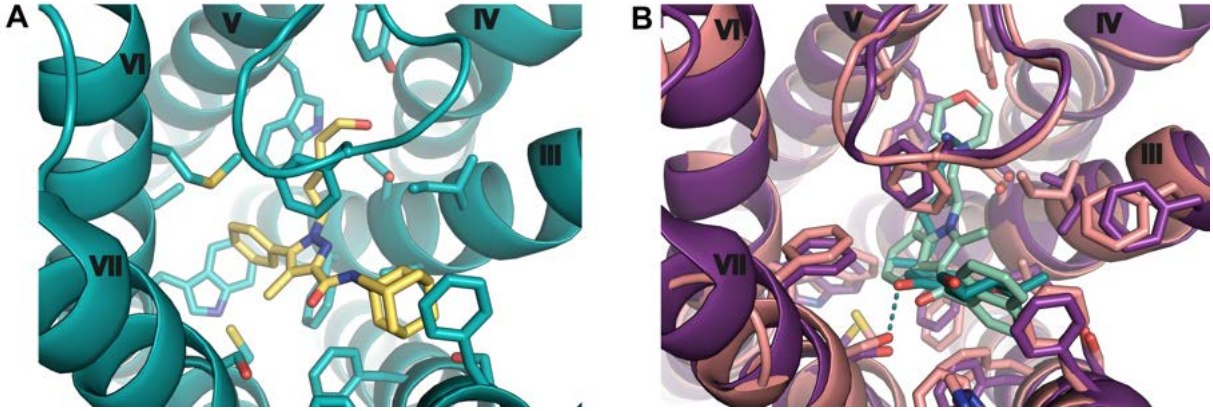


Figure 2.9: The orthosteric binding pocket of the CB2 in an A) inactive conformation and B) active conformations. A) The CB2 structure with PDB ID 5ZTY[112] in complex with antagonist AM10257. The ligand is mainly stabilised by apolar/aromatic interactions and does not seem to form polar interactions to the receptor, leaving the amide group 'stranded'. B) Overlay of the CB2 structures with PDB ID 6PT0[114] (purple) and 6KPF[113] (salmon) in complex with their respective ligands, the agonists WIN 55,212-2 (greencyan) and AM12033 (teal). Both ligands have quite dissimilar structural scaffolds and adopt different binding poses in the binding pocket. WIN 55,212-2 does not seem to form any polar interactions to the receptor and is probably mainly stabilised by apolar/aromatic interactions. AM12033 forms a polar interaction to S285$^{7.39}$ and probably also the backbone-carbonyl of L182$^{\mathrm{ECL2}}$.

of other processes. Many of these functions of the RGS proteins remain still unclear, however, progress towards elucidating them is made (see e.g. Abramow-Newerly et al.[131] on RGS and its interactions with other proteins). All members of the RGS family share the characteristic RGS domain which consists of approximately 130 amino acids and interacts with the G_α protein. Based on sequence similarity of this RGS domain, the more than 30 members of this protein family can be grouped into five[126] or six[132] subfamilies, depending on the classification method. The only member of the RGS protein family targeted in this thesis was RGS7 which belongs to the R7 or C subfamily (the members of these subfamilies are consistent between both classification methods). Hence, only this RGS subfamily will be discussed in more detail in the following.

2.2.1 The R7 RGS subfamily

Members of the R7 RGS subfamily are evolutionary conserved between a wide variety of species, from worm to vertebrates, and were shown to be essential in a wide variety of physiological processes.[133] One of the main characteristics of this subfamily are some distinct structural features that discriminate it from other RGS subfamilies. Connected to the RGS domain is the G_γ protein like (GGL) domain. Linked to this domain and forming the N-Terminus of R7 proteins are the DEP (Disheveled, Egl-10, Pleckstrin) and DHEX (DEP helical extension) domains. The GGL domain is homologous to G_γ subunits and acts very similar to it by interacting with a G_β protein. However, it is very selective towards its interaction partner, forming only complexes with the type 5 G_β protein ($G_{\beta5}$).[134] Interestingly, $G_{\beta5}$ does not only form very stable complexes with the R7 RGS proteins but also seems to be crucial for their stability and for activation of their regulatory functions.[135] The overall structure of the complex formed between RGS7 and $G_{\beta5}$ will be described in more detail below.

Additional to the $G_{\beta5}$, two other important interaction partners of the R7 RGS proteins are known, the RGS9 anchoring protein (R9AP) and the R7 binding protein (R7BP). These proteins are anchored to the membrane and can interact with the R7 RGS proteins via the DEP domain.[136,137] It has been shown that this interaction serves different functions, mainly by activating the catalytic activity of the RGS proteins and by localising them to specific cellular compartments.[135]

In their main –or at least first discovered– function the R7 RGS proteins serve as GAPs for the activated G_α proteins. As for other RGS proteins, the catalytic domain is located at the C-terminal RGS domain which is interacting with the G_α protein. Interestingly, this interaction is very specific and members of the R7 RGS subfamily were shown to only stimulate catalytic activity of $G_{i/o}$ proteins but not other G_α protein subtypes.[138,139] Aside from that, the R7 RGS proteins have been shown to interact and influence several other proteins such as GPCRs, GIRKs and many more,

highlighting their importance not only as deactivators of the G protein signalling pathway but also as regulators in a multitude of cellular processes and physiological functions (see also Anderson et al.[135]). The exact interaction sites between the R7 RGS proteins and their interaction partners are often unidentified, however, GPCR interaction might be taking place via the DEP domain while GIRK interactions are most likely mediated via the $G_{\beta 5}$ protein.[133]

R7 RGS subfamily in therapy and disease. In the recent years, RGS proteins emerged as potential drug targets of interest, although they form complexes that are often difficult to target and some of their functions are still to be fully understood.[140] However, RGS proteins are expressed in very specific localisations (even splicing variants of the same RGS are sometimes expressed in different loci), which makes it possible to e.g. influence the signalling of a specific GPCR in a specific cell type only. RGS proteins have been shown to be involved in various diseases and can serve as a downstream target for e.g. certain GPCRs.[130,140,141]

Members of the R7 RGS subfamily are, in general, expressed in the CNS and are involved in a plethora of neuronal processes but also e.g. drug addiction and tolerance.[135] RGS7 in specific has been shown to be involved in synaptic plasticity and memory processes due to its regulatory function of $GABA_BR$-GIRK signalling in the hippocampus which plays a role in several neurological diseases.[142] In principal, the R7 RGS proteins could be targeted in various positions. Drugs could either block interactions to other interaction partners or they could work as allosteric modulators influencing overall protein conformation. Allosteric modulation is known to have an effect on RGS function, e.g. regarding GIRK signalling.[142,143] However, either drug type is difficult to find especially when searching for small molecule drugs. Hindering protein-protein interactions with a small molecule is, in general, a difficult task while allosteric modulators do not only have to interact with the protein but also have to influence conformational changes. Additionally, potential drugs need to be able to pass the cell membrane to reach the intracellularly localised RGS proteins. Another difficulty for *in silico* approaches is knowing which area to target to disrupt interactions or influence RGS protein behaviour, even if the crystal structure of the protein is available. However, each molecule that acts in any way on RGS proteins is a step forward and could help understand these proteins and their involvement in physiological processes better.

Structural features of the RGS7 complex. In 2008, the first crystal structure of the full complex of a member of the R7 RGS subfamily was released, the RGS9-$G_{\beta 5}$ complex (PDB ID 2PBI).[144] Ten years later, in 2018, another R7 RGS protein structure was solved, this time of the RGS7-$G_{\beta 5}$ complex (PDB ID 6N9G; Figure 2.10).[133] Both structures show an overall similar complex organisation but also some structural differences. In both complexes the $G_{\beta 5}$ protein is sandwiched by the R7 RGS protein. The GGL domain interacts with the $G_{\beta 5}$ protein in a similar way to

the interaction between G_γ and G_β subunit in the heterotrimeric G protein complex.[133,144] In the RGS7 complex one part of the GGL domain is placed differently than in the RGS9 complex or the G protein complex, which might explain the distinct behaviour of RGS7 compared to RGS9.[133] The DEP domain and the RGS domain are interacting with the $G_{\beta5}$ from top and bottom, respectively, forming a tight and very stable complex. The DHEX-GGL linker interacts with the $G_{\beta5}$ protein in a region that is known to interact with other effector proteins in other G_β proteins.[133] The binding site of the G_α to the $G_{\beta5}$ protein is blocked by the interaction with the DEP domain and especially the DHEX-GGL linker. Binding of the G_α subunit to the RGS domain and the catalytically active C-terminus is only possible after structural rearrangement of the complex.[133]
An analysis of flexibility and structural changes within this complex by accessibility studies using HDX-MS (hydrogen deuterium exchange mass spectrometry) showed that this complex is overall very stable with the highest flexibility within the DEP/DHEX domain, the DHEX/GGL linker and the C-terminus located at the RGS domain. A differential HDX-MS analysis demonstrated that the DEP domain is also the protein part most tightly involved in binding to the R7BP protein.[133]

2.3 Molecular docking

Molecular docking is a tool to predict binding poses and interactions between a small molecule and a drug target. It allows rapid evaluation of large molecular libraries against proteins, DNA or RNA. There are also some docking algorithms that can predict protein-protein binding modes.[145] In SBDD and the Drug Discovery process, molecular docking of small molecules to protein targets plays an important role, even more with the growing rate of three dimensional structures of proteins being solved and becoming available.[146] Interactions between protein and ligand can be analysed or a ligand can be grown with additional functional groups to increase its affinity and potency based on the predicted binding pose of the ligand.[145] Another enhancement of the Drug Discovery process is the screening of large molecular libraries for novel ligands in a considerably more time efficient way than in an *in vitro* screen.[147] A schematic representation of molecular docking is shown in Figure 2.11.

2.3.1 Theoretical background of molecular docking

The idea behind molecular docking is in principle based on the "lock & key" model developed by Emil Fischer in 1894, stating that protein and ligand conformation are complementary to each other and fit each other like a key and a lock.[148] Although it was later shown that protein and ligand interactions rather rely on an induced fit mechanism[149] or even an ensemble of conformational states of a protein[150–153] the underlying idea remained the same. In principle, the docking process

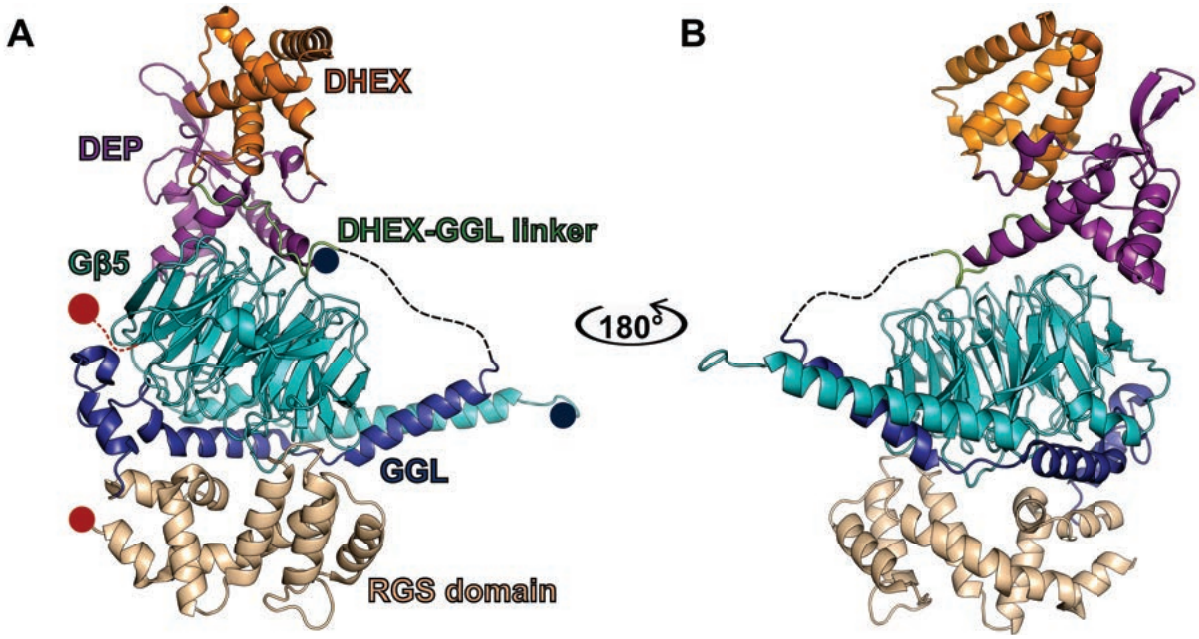


Figure 2.10: Crystal structure of the RGS7-$G_{\beta 5}$ complex (PDB ID 6N9G[133]). The different parts are coloured different and labelled in A). The $G_{\beta 5}$ (teal) is sandwiched by the RGS domain (wheat) and the DEP domain (violet). Additionally, the GGL domain (blue) is binding tightly to the $G_{\beta 5}$ protein, in a similar way as observed for the G_{β} and G_{γ} subunits in the heterotrimeric G protein complex. The DHEX-GGL linker (green) is only partially resolved and the connection of GGL domain and DHEX domain (orange) is represented by a dashed line. The N-termini are marked by blue circles and the C-termini by red circles.

consists of two parts. First, a likely binding pose of a molecule in the binding pocket of the target protein has to be found by sampling possible molecule poses and even molecule conformations. In the second step, this pose has to be scored by a scoring function to evaluate its likelihood and to compare it to other poses of the same molecule as well as the poses of other molecules in the same ligand set during the virtual screening.[147] Both steps proved to be quite challenging. A first approach for a docking algorithm has already been proposed in the early 1980s[154] and different methods to solve the sampling and the scoring problem have been suggested since then. These will be introduced in brief below and more detailed for the docking algorithms used in this thesis, DOCK3.x and SEED (solvation energy for exhaustive docking).

Molecule pose sampling. To find an optimal pose for the molecule in the protein binding pocket, a search algorithm is applied that samples the free energy landscape.[153] For this sampling process, two factors have to be considered, ligand flexibility as well as receptor flexibility. This leads to three different approaches to sampling.[153] (1) Keeping both protein and ligand conformation rigid. This

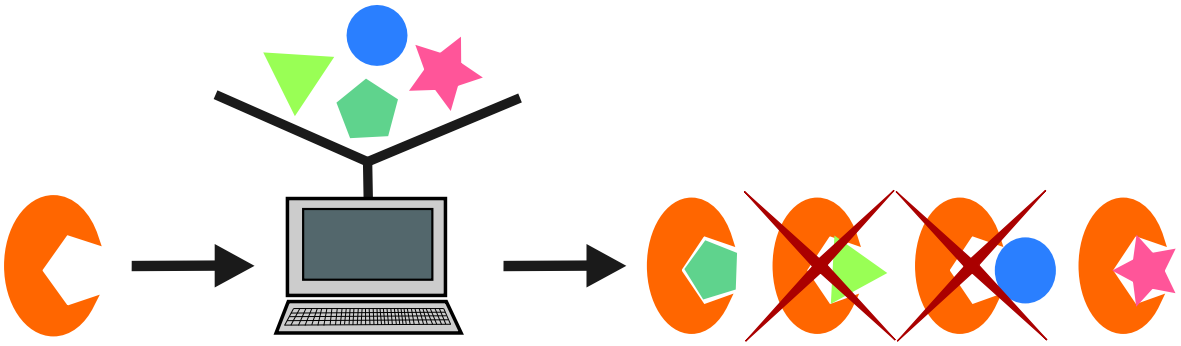


Figure 2.11: Schematic representation of molecular docking. A three-dimensional protein structure (orange) and a molecule library (pastel polygons) are fed into a docking algorithm. The algorithm then predicts the energetically most favourable binding pose of each of the molecules in the binding site of the protein. Visual inspection of the obtained binding poses can help to eliminate the ones that might seem energetically favourable but are in fact e.g. too solvent exposed or do not actually penetrate the binding pocket (red crosses).

corresponds to the "lock & key" model proposed by Fischer[148] but is rarely used nowadays. (2) Keeping the protein rigid but sampling the conformational space of the ligands. (3) The protein is partially or completely flexible and the conformational space of the ligand is also sampled. Allowing for complete flexibility of both protein and ligand is very computationally demanding. Therefore, most algorithms keep the protein rigid and bond angles and bond length of the ligands constant.

Ligand sampling algorithms can be divided into three groups: systematic, deterministic and stochastic searches.[145] In systematic search, all degrees of freedom of the ligand are explored either by an exhaustive search (testing combinatorially all combinations of different degrees of freedom; this can be very computationally demanding especially for more flexible ligands), by an incremental construction (where the ligand is built up by combining its fragments) or by using a conformational ensemble of pre-generated ligand conformations. In a deterministic search the state of the system defines the next modifications. These searches are based on energy minimisations or molecular dynamics methods. Stochastic searches are algorithms where all degrees of freedom are randomly changed at each step, thereby generating diverse conformations. Examples for stochastic searches can be Monte Carlo methods, Evolutionary algorithms, Swarm optimisation algorithms and Tabu searches.[145,153] In all cases, the sampling ideally leads to a molecule pose that is located in the global or at least a local minimum of the free energy surface.

Considering *protein flexibility* during molecular docking is more challenging than considering ligand flexibility due to the high number of degrees of freedom such a big molecule contains. However, dif-

ferent approaches to tackle this problem have been developed. The soft docking approach addresses the problem implicitly by allowing small overlaps between ligand and protein atoms by reducing the van der Waals repulsion term of the Lennard-Jones potential. This approach thereby considers small variations within the binding site that might occur due to interactions with different ligands but might also lead to unrealistic molecule poses.[152] Other strategies allow movement of side chains or even relaxation of both side chains and backbone. These allowed changes in the conformation are usually restricted to the binding site of the protein but even then are still computationally demanding. If multiple structures of a protein in different conformations are available, the flexibility of the protein can be treated implicitly by an ensemble docking approach. Flexibility of the entire protein is usually not feasible during a virtual screening since it is too time demanding or inaccurate due to simplifications.[153]

Scoring functions. Scoring functions are used to estimate the quality of a molecule pose based on an estimation of its free energy. In this context, a lower energy value corresponds to a more favourable binding mode. Of note, there are also scoring functions that are not based on free energy estimation, e.g. statistics based or machine learning based methods. However, these will not be discussed below.

Scoring functions serve two main purposes.[153] (1) Identifying the most favourable ligand conformation in the protein binding site during the molecule pose sampling process. (2) Ranking different ligands in relation to each other to be able to discriminate between potential active and inactive molecules e.g. in a virtual screening. During docking calculations the same scoring function can be used during sampling and ranking or different scoring functions can be used during both steps.

Scoring functions need to predict binding energies as exact as possible to find the most realistic binding modes of molecules while also being fast, since otherwise, the screening of a large molecule library would not be feasible in a decent amount of time. Both of these characteristics are quite controversial though and compromises have to be made. Various scoring functions based on different principles have been developed and they can be divided into three main classes: force field-based, empirical and knowledge-based scoring functions.[145,152,153]

Force field-based scoring functions use a free energy estimation which is based on classical force fields and the non-covalent interaction terms from molecular mechanics calculations, i.e. energy calculations are based on the Lennard-Jones potential (van der Waals interactions) and Coulomb interactions (electrostatic interactions). Since this classical calculation does not consider solvent or ligand entropies, force field-based scoring functions usually include correction terms to account at least partially for these energy contributions.

Empirical scoring functions are designed to reproduce binding affinity data of a training set. The

total free energy is calculated by including several different energy terms such as hydrogen bond, desolvation, electrostatic etc. and weighing them by an optimised coefficient.
Knowledge-based scoring functions are based on the analysis of interactions between atoms in known protein-ligand complexes. The frequency of interactions between atom pairs is converted to free energy values which can then be summed up during the docking process.

Limitations of molecular docking. One issue of molecular docking is the accuracy versus speed problem.[145,152] A more accurate prediction by the docking algorithm will take longer to be calculated for each molecule. However, when screening a molecule library with millions of molecules, the calculations for each molecule have to be pursued within few seconds, otherwise a docking screen could not be completed within a reasonable amount of time. On the other hand, accuracy cannot be completely sacrificed for the sake of speed because otherwise a discrimination between active and inactive molecules based on their energy scores will not be possible anymore. The focus between accuracy and speed can vary between different docking algorithms and scoring functions and can be chosen based on the needs of the user, i.e. whether the purpose of a docking calculation is to screen a large molecule library where predictions need to be made quickly or to discriminate a small set of molecules and be able to rank them according to binding affinities which would rather require accuracy.
Another issue is the prediction of absolute binding affinities which should in principle be possible based on energy calculations. However, during calculation of free energy values the entropic contribution to the total energy is often neglected or not entirely included. The entropy change upon binding of a small molecule to a protein comprises contributions from solvent and solute, such as desolvation, the hydrophobic effect or changes in translational, rotational and vibrational entropy, and can be difficult to estimate.[145] Although some of the more extensive or more specialised methods are able to predict binding affinities correctly, they are too time consuming to use in a virtual screening approach (accuracy vs. speed).[146,153]

2.3.2 Benchmarking and evaluation of docking setups

For virtual screening approaches using molecular docking, the main goal is to be able to discriminate active from inactive molecules and to have an enrichment of actives in the top ranks of the ranking list resulting from such a docking calculation. To evaluate whether this goal is achieved, the docking calculation can be benchmarked using a set of known actives of the target protein and inactive molecules. When compiling such a benchmarking set it should be considered, that in nature the number of inactive molecules is always several folds higher than the number of active molecules and this ratio should also be taken into account for the benchmarking set. However, knowledge about

inactives of a protein is rarely available which makes the compilation of a benchmarking set based on known actives and inactives rather difficult. To circumvent this problem, various methods have been developed to predict potential inactive molecules or decoys. An example for decoy prediction is the Directory of Useful Decoys-Enhanced (DUD-E),[155] which defines decoys to be similar to the corresponding active molecules in their physicochemical properties while being chemically distinct. Other methods for decoy prediction as well as target specific decoy sets are also available.[146] It should be kept in mind that these decoys do not necessarily have to be true inactives but are only predictions based on descriptors.[146]

To evaluate how well a docking calculation performs when docking such a benchmarking set to a protein, certain methods such as enrichment factors or receiver operating characteristics (ROC) plots can be used. Enrichment factors calculate the enrichment of actives in the top ranks up to a certain cutoff. In ROC plots the number of found actives is plotted against the number of found decoys going from the highest ranked to the lowest ranked. The area under the curve (AUC) can then be used to evaluate overall enrichment of actives over decoys. An AUC of 0.5 would correspond to a random enrichment of actives while a value above 0.5 corresponds to an enrichment of actives over decoys in higher ranks. Additionally, early enrichment, i.e. enrichment of actives over inactives in the top ranks, can be evaluated by inspecting the leftmost part of the curve, especially when using a logarithmic x-axis.[156] The early enrichment is specifically interesting during virtual screening campaigns of large molecule libraries in which usually only the 500 to maybe 5000 top ranked molecule poses are evaluated closer, i.e. a high enrichment of actives in the top few percent of the ranking list is desired. In the studies described in chapter 3 and chapter 4 ROC plots have been used to evaluate docking performance.

2.3.3 Virtual molecular libraries

A crucial part of successful virtual screening studies using docking calculations are the molecular libraries. These libraries can be designed in different ways, depending on the desired outcome of the screening. In general, it can be assumed that the most diverse hits can be found if the library covers a large portion of the chemical space. However, this chemical space is estimated to contain 10^{20} to 10^{180} molecules, depending on the criteria chosen for the estimations.[157–159] Obviously, these numbers are too big to screen or even to grasp for the human mind and, hence, researchers are developing strategies to cover as large portions of the chemical space as possible with virtual screening libraries while also keeping things more realistic. It should, for example, be considered that not all compounds that could theoretically exist are chemically accessible, a feature that is desirable to test the compounds against the target proteins in *in vitro* assays later on.[160] Chemical

accessibility of the molecules in a virtual library can be achieved by e.g. using vendor catalogues to constitute a virtual library from already available molecules, by pre-generating molecule libraries by combining available fragments *in silico* using robust chemical reactions or by dynamically generating specialised libraries based on prior knowledge about molecules that bind to the target protein.[159] Examples for these approaches are the ZINC Is Not Commercial (ZINC) database,[161] the Screenable Chemical Universe Based on Intuitive Data OrganizatiOn (SCUBIDOO)[162] or the Python in silico de novo growing utilities (PINGUI)[163] toolbox, respectively. Of note, the hunt for novel chemical structural scaffolds and for the accessibility of a larger portion of the chemical space can also be observed from the developments of novel drugs and synthesis routes over the past decades.[164]

In this thesis, different subsets of the ZINC database were used for the *in silico* docking screens. This database was first developed and released in 2005, aiming to provide an easy-to-use and free of charge molecule database for *in silico* docking screens.[161] Main features included the ready-to-use preparation of those molecules for docking, i.e. the molecules are protonated in the most likely protonation states, potential tautomers and enantiomers are generated and the molecules can be downloaded in the most commonly used formats for different docking programs including three dimensional conformers, as well as the availability of these compounds from commercial vendors to allow for quick *in vitro* testing of the docking predictions.[161] Over the years, this database was further developed to include molecules from the catalogues of even more vendors and to allow the user to decide which features the downloaded library should have. These features include lipophilicity (managed by calculated logP values), molecular weight (MW) ranges, potential reactivity of the molecules (based on certain functional groups), availabilty and protonation states at the desired pH ranges.[165,166] The library size is constantly growing owing to the fact that vendor catalogues are also growing with more than 700 million protomers available as of December 2020. Furthermore, different pre-generated subsets are available that contain molecules with the most commonly used and found physicochemical properties. The most important ones are the fragment sized library ($MW \leq 325$ and $-1 \leq logP \leq 3.5$), the lead-like library ($300 \leq MW \leq 350$ and $-1 \leq logP \leq 3.5$) and the drug-like library ($250 \leq MW \leq 500$ and $-1 \leq logP \leq 5$).[165]

2.3.4 The docking program DOCK

DOCK is a docking algorithm that is based on a geometric docking approach which was introduced in 1982.[154] Two main program types are in development, the DOCK6.x[167,168] and the DOCK3.x series, which function substantially different. Here, only the DOCK3.x series will be introduced further since it was used for docking calculations conducted in different projects of this thesis.

In the geometric docking approach from 1982[154] the binding site was defined by a number of receptor spheres onto which the ligand spheres were fit. Later, this approach was further refined, including shape and chemical descriptors into this matching process.[169,170] This method to define the binding pocket and facilitate placement of the ligand in the desired protein region is still applied in today's algorithm, using the so called *matching spheres* (see Figure 2.12.A). The positions of these *matching spheres* are derived from the locations of the heavy atoms of the crystallised ligand and additional spheres created by SPHGEN.[154] During ligand placement, a defined number of ligand heavy atoms needs to be placed on *matching spheres* within a certain tolerance for the pose to be accepted.[171]

An energy evaluation of the generated molecule poses was introduced in 1992, based on grids that are generated for the receptor prior to docking.[172] The scoring function that finally scores the generated molecule poses is a force field-based scoring function, calculating the non-bonding interactions between ligand and protein atoms (eq. 2.1) using the pre-generated electrostatic grids for the protein.

$$E = \sum_i \sum_j \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{D r_{ij}} \right] \tag{2.1}$$

where the sum over ligand atoms i and protein atoms j is calculated. With van der Waals attraction and repulsion parameters A_{ij} and B_{ij}, the distance between ligand atom i and protein atom j r_{ij} and the charge of ligand atom i q_i and protein atom j q_j. D is the dielectric function, i.e. a distance dependent dielectric constant, assuming a lower dielectric constant within the binding pocket compared to outside the binding pocket (defined by placement of *lowdielectric spheres* within the binding site; Figure 2.12.B). The intermolecular parameters are taken from the *AMBER* united-atom parameters.[173]

This scoring function, however, includes only the enthalpic contributions of ligand binding while completely ignoring entropic contributions, e.g. from protein and ligand desolvation. In 1999 and 2010, correction terms were introduced to the scoring function to account for these entropic contributions.[156,174]

Additional to introducing and improving scoring functions and energy calculations, molecule pose sampling methods were developed. In the first docking approaches, both protein and ligand conformation were kept rigid for docking calculations, thereby limiting possibilities for resulting molecule poses. Later, a library of pre-generated ligand conformations was used to allow for a certain flexibility of the ligand when docked to the protein.[175,176]

The most recent version of the docking algorithm, DOCK3.7, was introduced in 2013, including improved ligand sampling and code optimisations to increase the speed of docking and reduce the recquired amount of storage space.[171]

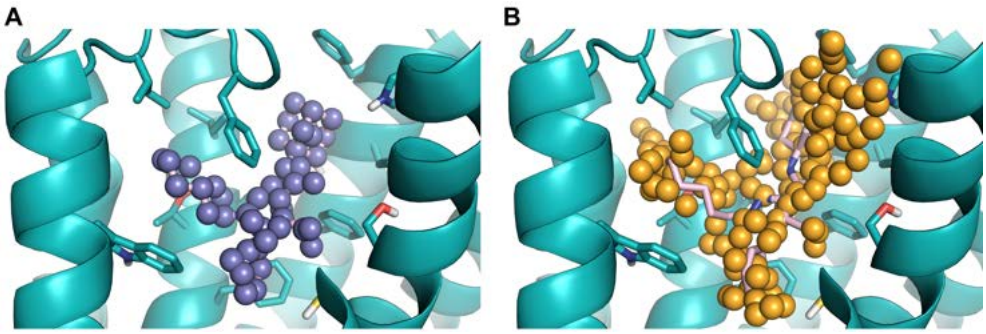


Figure 2.12: Exemplary DOCK spheres used by the DOCK algorithm to translate and rotate the docked molecule into the binding pocket. A) The *matching spheres* are based on the heavy atom positions of the template ligand and additional spheres created by SPHGEN. A defined number of heavy atoms of the docked molecules has to be placed onto these spheres within a certain threshold for a docking pose to be accepted. B) The *lowdielectric spheres* define the area of a lower dielectric constant for the distance dependent dielectric function (see also eq. 2.1).

2.3.5 Solvation Energy for Exhaustive Docking (SEED)

SEED is a docking algorithm that is specialised on docking small- to medium-sized fragments to proteins or protein binding sites. First steps towards this program were published in 1997, when *Scarsi et al.* proposed a way to calculate electrostatic energies of macromolecules in aqueous solutions based on a continuum approach and the generalised Born approximation.[177] In 1999 this method was applied in the exhaustive docking of molecular fragments to proteins or protein binding sites to score the resulting binding poses.[178] For placement of the fragments within the user specified protein region (which might be a specific binding site but also the entire protein) different methods are used for polar and apolar fragments. To allocate *polar* fragments, vectors are placed on the polar functional groups of the fragment and of polar residues within the defined protein region (see Figure 2.13.A), indicating the directionality of hydrogen bond donors and hydrogen bond acceptors. These vectors can then be matched to find the optimal geometry for hydrogen bonds between fragment and protein. Protein atoms that are already interacting with other protein atoms are excluded to increase the sampling speed. Additionally, van der Waals energies between fragment and protein are calculated to exclude fragments with energies outside a certain threshold and additional sampling is achieved by rotating the fragment around the hydrogen bond vector.
To dock *apolar* fragments to the protein binding site, points are equally distributed on the solvent accessible surface (SAS) of the fragment and the protein site of interest (see Figure 2.13.B). Additionally, low dielectric spheres are placed on these points for the protein, to calculate the receptor

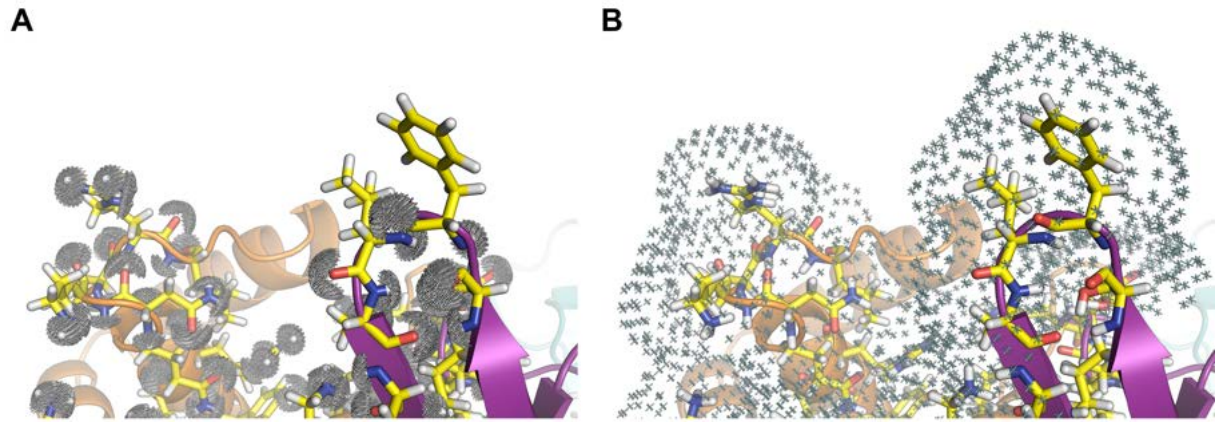


Figure 2.13: Definition of the protein structural features for SEED docking. A) For placement of polar fragments, vectors are placed on the polar functional groups which can then be used to estimate polar interactions between fragment and protein. B) Points on the SAS which are used for the placement of apoplar fragments.

desolvation energy and to evaluate van der Waals interactions. With this procedure, the hydrophobicity of the protein site of interest is evaluated and ranked before docking the apolar fragments. For both fragment and protein, vectors are then placed connecting each point on the SAS to the closest nearby atom. These atoms are then matched during docking of the fragments to find the optimal van der Waals distance to the protein. Additional sampling is achieved by rotation of the fragment around the matched receptor axis.

All fragment poses are then filtered for clashes prior to calculation of energy scores. Energy scores of the fragment are calculated based on van der Waals and Coulomb interactions with an optional additional calculation of the desolvation based on the continuum electrostatic energy approach introduced in 1997.[177]

The calculation of the desolvation energies as described above was still quite time consuming. Therefore, the scoring approach was developed further and a two step scoring was introduced in 2001.[179] To reduce the amount of calculation time, the energy score of the fragment binding poses is first calculated using a quick approach based on van der Waals and electrostatic interactions and an approximated treatment of solvation. These first estimations are then used to sort the fragment poses by their score, followed by a clustering step. Only a defined number of the poses with the highest energy scores within each cluster are then subjected to a calculation of the energy score using the more accurate solvation model. This two step scoring approach was shown to yield good results and allows the docking of fragment libraries containing more than 100 fragments within only a few hours on a desktop PC.

2.4 Molecular Dynamics simulations

As mentioned above, crystallographic approaches and the *in silico* technique of molecular docking can be useful tools to understand proteins and their activation and interaction mechanisms on a microscopic level. These approaches are mostly static and observe only snapshots of certain states of a protein or protein complex. However, in reality, biological processes are dynamic and change over time. An approach to incorporate these dynamics into *in silico* methods are Molecular Dynamics (MD) simulations which were first developed in the late 1970s.[180] This technique uses algorithms to compute how a system consisting of a protein and other molecules such as water, ligands, lipids or other proteins changes over time by calculating the movements of each atom based on the forces acting on it.[181]
In the recent years and with advancing methods and computational power, MD simulations became a useful tool in drug design and to describe systems on a microscopic level, for example discovering novel allosteric ligand binding sites,[182,183] describing opening processes of a binding pocket[184] or suggesting an activation mechanism for GPCRs.[185]

Setting up an MD simulation. To set up an MD simulation and to simulate a system, several steps have to be followed. First, the system has to be set up. The initial coordinates of the protein are usually taken from experimentally derived structures or, if there are none available, from homology models. Poses of small molecules that are simulated in complex with the protein can be derived from experimental structural data or by placement using computational methods such as molecular docking. Additionally, the natural environment surrounding the protein has to be at least approximately mimicked to obtain more realistic results. For that, proteins such as enzymes are surrounded by a water layer while membrane proteins have to be placed in a membrane-like lipid bilayer with a water layer on top and bottom (see Figure 2.14). Usually, a defined amount of salt ions is added to these water layers.[181,186,187] The water molecules can either be described implicitly, which reduces the computational cost of such a simulation drastically, or explicitly, which allows to also analyse water-protein contacts. To avoid the necessity to take surface effects caused by the system surface into consideration, periodic conditions are usually used by placing an infinite number of copies of the system next to each other.[186] Any modifications to the protein structure such as protonation states or post-translational modifications have to be defined during system setup, since they cannot change during the simulation.[181]

Before starting the production run, the system is first equilibrated. During equilibration initial velocities are randomly assigned to each atom based on the Maxwell-Boltzmann or Gaussian distribution. The system is then set to the final temperature and restraints are applied to the atoms of

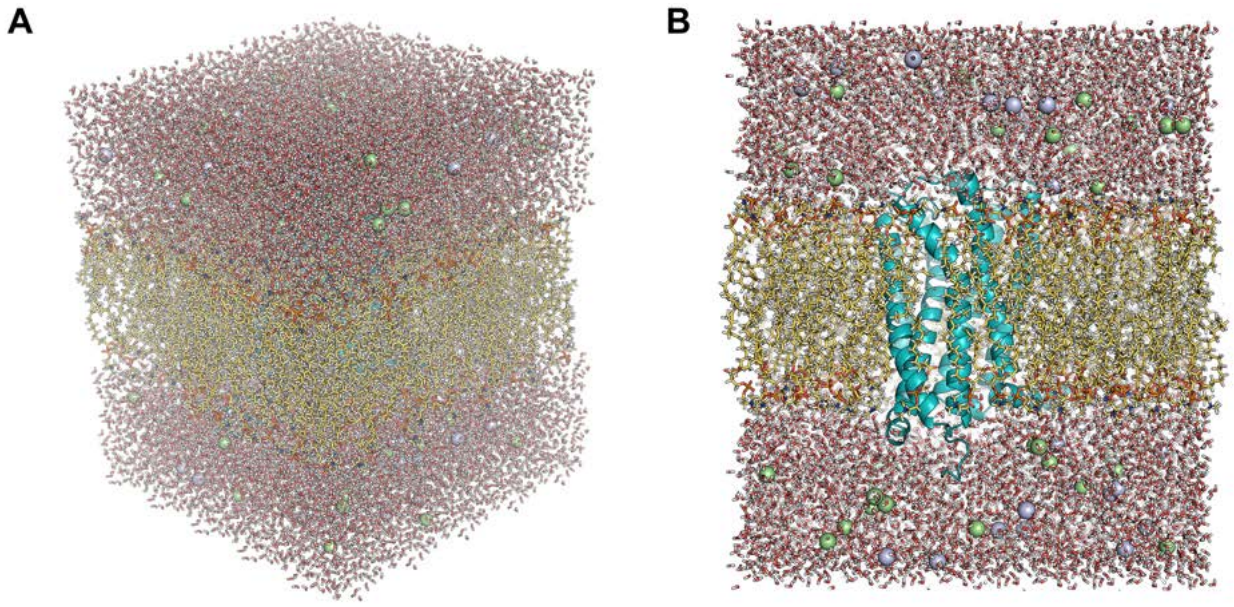


Figure 2.14: Exemplary system for MD simulations of a GPCR. The MD system forms a rectangular box (A), which can be repeated infinitely in two dimensions. In this box the receptor (teal) is placed in a lipid bilayer (yellow) and water is added on top and bottom of this lipid bilayer (B). Additionally, sodium (light blue) and chlorine (light green) ions are added to the water to obtain biologically more realistic systems.

the solute which are then step-wise loosened. This way instabilities caused by large forces can be avoided. During equilibration clashes between atoms are usually resolved and the overall system parameters can stabilise. After equilibration of the system, the production phase starts during which no restraints are applied to the atoms and which can take as long as desired by the user.[187]

Simulating the motion of a system. During the simulation the movement of the atoms within a system is calculated based on Newton's second Law of motion:[188]

$$F = m \cdot a = m \cdot \frac{dv}{dt} = m \cdot \frac{d^2x}{dt^2} \tag{2.2}$$

with the force F, mass m, acceleration a, velocity v, time t and atom position x.

A force is applied to each atom changing its acceleration and thereby affecting its velocity and defining its new position after each time step. The new positions after each time step can be derived using different algorithms, such as Verlet or leapfrog algorithm.[187] In brief, these algorithms calculate new positions and velocities of each atom only at certain points in time while leaving them constant during the time step.[188] Each of these time steps has to be shorter than the fastest movement in the system (the vibration of heavy atom bonds) and is usually around 2 fs, unless additional constraints are used.[152,181,189] The force applied to each atom is calculated from the

potential energy derived by a force field calculation which is usually based on eq. 2.3. [183]

$$E_{pot} = \sum_{bonds} k_r \cdot (r - r_{eq})^2 + \sum_{angles} k_\theta \cdot (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + cos(n\phi - \gamma] + \sum_{i<j}[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}] \tag{2.3}$$

with bond length r, bond angle θ, dihedral angle ϕ, and for atoms i and j the distance R_{ij} and partial charges q. ϵ is the dielectric constant. Equilibrium bond length r_{eq} and equilibrium angle θ_{eq} as well as constants k_r, k_θ, V_n, and A_{ij} and B_{ij} can be derived by e.g. empirical methods.

This force field function uses rather simple descriptors to describe atom-atom interactions. Bonded interactions are either described by a harmonic potential (bond length and angles) or by a sinusoidal function (dihedral angles) while non-bonded interactions are described by the Lennard-Jones potential (van der Waals interactions) and the Coulomb potential (electrostatic interactions). These force fields are then parametrised based on quantum mechanics or experimental data which is why force fields are usually suitable for specific systems or purposes only. [183]

Non-bonded interactions can be rather long-ranging which can potentially lead to extensive calculations due to the number of possible contacts. This problem is solved for the Lennard-Jones potential by using a simple distance cutoff since this potential approaches zero at a fairly short distance. Electrostatic interactions, however, are longer ranging and a simple distance cutoff would introduce errors into the calculation. Different methods to solve this issue have been introduced, for example the particle-mesh-Ewald approach. [186]

To study molecular properties of a system during MD simulation different ensembles of constant system characteristics are used. The most common ensembles are the *canonical* (constant particle number, volume and temperature; NVT) and the *isothermal-isobaric* (constant particle number, pressure and temperature; NPT) ensembles. A constant temperature is obtained by either scaling the velocities or by coupling the system to an external heat bath. In a similar approach, a constant pressure is achieved by either adapting the volume or coupling to an external 'pressure bath'. [188]

Challenges and difficulties. There are multiple challenges that complicate the comparability of MD simulations to real biological systems. One of the major challenges is the time span in which biological processes occur. These are usually in the range of several microseconds to milliseconds. Such long time spans are difficult to simulate since each time step during an MD simulation is only few femtoseconds and, hence, billions and more time steps would be necessary. A second challenge is the size of biological systems and the number of atoms that have to be simulated. Advances towards approaching both of these challenges have been made in the recent years, mostly through improved algorithms and the increasing computational power. [181,183]

Another challenge is the sampling problem. During MD simulations the potential energy surface is

sampled. However, the structures that serve as a starting point for such a simulation are usually located in a (local) minimum and it might not be possible to overcome energy barriers to sample different minima than the current one, missing out on other possible outcomes of the simulation. To overcome this problem several approaches have been developed, for example umbrella sampling, meta-dynamics simulations or accelerated MD simulations.[186]

Another issue of MD simulations is the force field itself. Since it is based on approximations, it usually does not consider the polarisability of atoms depending on their local environment but fixed partial charges are used instead.[183] Partial charges, protonation states etc. usually have to be assigned prior to simulating and do not change during the MD simulation which can also lead to differences compared to realistic systems.[181]

3 | Prediction of novel antagonists of the β_2AR by docking to structures in inactive conformations

This study has been published in M. M. Scharf, M. Zimmermann, F. Wilhelm, R. Stroe, M. Waldhoer, P. Kolb, *ChemMedChem* **2020**, *15*, 882-890.[83]

Contributions. In this study, the author of this thesis contributed to the design of the study, conducted the generation and analysis of the computational data and drew structural conclusions based on the experimental results. Furthermore, the author of this study was involved in writing the manuscript of the main publication.

All pharmacological assays were performed and assay results analysed by our collaboration partners at *InterAx Biotech AG*.

3.1 Introduction and goal of the study

As described in section 2.1.4 the β_2-adrenergic receptor (β_2AR) is one of the best-investigated GPCRs and is, therefore, very well suited as a model receptor for investigations on general concepts of class A GPCRs. In this study we wanted to explore the possibilities of docking screens against GPCRs further. In chapter 2.1 the interplay between conformational changes induced by ligand interactions and the higher affinity of ligands for certain receptor conformations was mentioned. Our aim was to support these observations by showing that docking calculations to certain receptor conformations would mainly sample ligands with the functional properties connected to this conformation. In the case described in this chapter, we wanted to show that when targeting an inactive conformation of the receptor –i.e. a conformation of the receptor that had been crystallised with an antagonist or inverse agonist– in docking calculations, this would only yield novel ligands that act as antagonists or inverse agonists towards the G_s signalling pathway. This concept has also been further explored in the reverse direction, i.e. to find activating ligands, in the study described in the following chapter 4.

Another goal of this study was to show that even for well-known receptors like the β_2AR it is still possible to find novel ligands (in this case antagonists) featuring novel structural scaffolds. To achieve this goal, we specifically searched for molecules entertaining an interaction with extracellular loop 2 (ECL2) and favoured molecules with molecular scaffolds that have not yet been described for ligands of the β_2AR. The ECL2 is both in proximity to the orthosteric binding pocket and has so far not been targeted in docking studies to the β_2AR. Finding ligands in such calculations would not only proof that it can be worth targeting even well-known targets in screening studies, but might

also result in the discovery of novel, potentially better-suited medications for diseases connected to the β_2AR.

Furthermore, the results of this study were compared to the results of a previous docking study to the β_2AR conducted in 2009.[190] Although this comparison was not a primary goal when starting the study, it allowed insights into the growth of molecule libraries throughout the years and showed the influence and opportunities that the growing number of available crystal structures might bring for future docking studies.

3.2 Methods

3.2.1 Computational methods

Structure preparation and optimisation for the docking calculations. The structures of the β_2AR with PDB IDs 2RH1[6] and 3NY9[69] were prepared for docking according to the structure preparation protocol described in section 8.1.1. To evaluate and optimise the docking performance, a set of known ligands of the β_2AR was gathered and matching decoys were generated using DUD-E[155] (see section 8.1.4 for a description of the ligand set compilation). Both ligands and decoys were then docked to the structures using DOCK3.6.[154,156,170,172,174] Docking optimisation was afterwards conducted by adjusting the positions of the spheres used by DOCK for placement of molecules in the orthosteric binding pocket until both docking poses of the known ligands in the binding pocket and enrichment of ligands over decoys were satisfying. The docking setup was considered to be sufficiently optimised if the majority of ligands entertained interactions with the protein while not showing buried polar functional groups ('stranded donors') and enrichment of ligands over decoys was above random, preferentially with a considerable early enrichment. These prepared crystal structures and their optimised spheres were then used for the docking screens.

Primary docking screen. The ZINC12[165] lead-like library containing 3,687,621 molecules was docked to the β_2AR structures with PDB IDs 2RH1 and 3NY9 which were prepared as described above using DOCK3.6.[154,156,170,172,174] Additionally, the ranking lists of the dockings to both structures were re-ranked to favour molecules ranked highly in both dockings (see section 8.1.2). The docking poses of the top 500 scored molecules of both docking calculations as well as the top 500 ranked molecules from the re-ranking were then evaluated by visual inspection to identify molecules entertaining interactions to the protein while punishing molecules with polar functional groups that were not involved in any interactions ('stranded donors'). Further, a distance filter was applied to the 10,000 top scored molecules of the docking calculation to PDB ID 3NY9 to identify molecules in proximity to certain residues in the ECL2 (2.6 Å to F193$^{45.52}$ (backbone), D192$^{45.51}$

(sidechain), C191$^{45.50}$ (backbone) and T195^{ECL2} (sidechain); numbers in superscript according to the Ballesteros-Weinstein enumeration scheme[19]). The top 500 ranked molecules from each of the resulting filtered lists were then also inspected visually. An additional focus was put on a variety of structural scaffolds to allow for the discovery of novel ligands of the β_2AR that contained scaffolds different than the well-described β-hydroxyl-amine scaffold. All selected molecules show an interaction to D113$^{3.32}$ via an amine or an amide in their docking poses. Overall, a total of 27 molecules was selected to be purchased and tested.

Secondary docking screen. To find structurally similar molecules to a hit molecule from the primary screen, compound **1 (MS008)**, the ZINC15[166] lead-like subset (5,626,190 molecules) was searched by using an Extended-Connectivity Fingerprint 4 (ECFP4)[191] and a Tanimoto cutoff of 0.5. The conformer files of the resulting set of molecules were then downloaded from ZINC12 to avoid compatibility problems between ZINC15 and DOCK3.6. The obtained molecule set was docked to the prepared structure of 3NY9 and the molecule poses were inspected visually. In total, eleven compounds were chosen from this evaluation and purchased to be subjected to *in vitro* assays.

Novelty of the compounds To confirm the novelty of the tested compounds, all molecules showing activity against the β_2AR were downloaded from ChEMBL.[192] Additionally, a second set of ligands was downloaded from ChEMBL containing all molecules showing activity against *any* of the adrenergic receptor subtypes. For each of **1 (MS008)** and its analogues the Tanimoto similarity to each of the molecules in these ligand sets was then calculated based on ECFP4.
A similar approach was used during the comparison of the previous docking study described by *Kolb et al.*[190] with this study. All compounds selected during the docking screen in 2009 as well as during the screens described here were compared to the ChEMBL set containing active molecules against *any* of the adrenergic receptor subtypes by calculating the ECFP4 Tanimoto similarity between the molecules. Additionally, the equivalent ChEMBL dataset containing all active molecules against *any* of the adrenergic receptor subtypes was downloaded from the ChEMBL database version of October 2009 and again the ECFP4 Tanimoto similarity between each molecule selected from the docking studies and each molecule from the ChEMBL dataset was calculated.

Calculation of root-mean-square deviations between different crystal structures. The structural similarity between the two crystal structures of the β_2AR used in this study, PDB IDs 2RH1 and 3NY9, was evaluated based on the root-mean-square deviations (RMSDs) between both structures. For the calculation of the RMSDs both structures were used as prepared for the docking calculations but without hydrogen atoms and adjusted to have the same number of residues. Using witnotp (Novartis Pharma AG, unpublished) the structures were aligned and RMSDs calculated for all

backbone heavy atoms (1112 atoms) and for all heavy atoms including side chains (2239 atoms). The binding pocket area was compared by calculating the RMSD value for all heavy atoms of the residues within 5 Å proximity to the ligand Carazolol crystallised in 2RH1.

3.2.2 Pharmacological characterisation

All assays were performed by our collaboration partners at *InterAx Biotech AG*. In brief, affinity was determined using ligand displacement assays with either a fluorescence-tagged or radio-labelled ligand. Antagonism of the compounds was determined using a real-time cAMP inhibition assay with Isoprenaline as stimulating agonist. A more detailed description of the experimental procedures can be found in the publication mentioned at the beginning of this chapter[83] and will not be explained further here.

3.3 Results

3.3.1 ZINC docking screen

In this study we aimed to find structurally novel ligands for the β_2AR with antagonistic behaviour towards the G_s signalling pathway. Therefore, the docking screens were performed using β_2AR structures in inactive conformations. Two different structures of the β_2AR were prepared for the docking screens: PDB ID 2RH1, crystallised with the antagonist Carazolol,[6] and PDB ID 3NY9, crystallised with an inverse agonist (compound 1).[69]

Primary docking screen. The ZINC12 lead-like subset was docked to both prepared crystal structures using DOCK3.6 and the top 500 molecules were evaluated visually. Since potential antagonists should in theory have good ranks and docking poses in the docking calculations to both structures, the ranking lists were also re-ranked to put molecules ranked highly in both lists on better re-ranks than molecules that were only ranked highly in one or none of the lists. Additionally, a special focus was put on finding molecules that entertained favourable interactions to the ECL2. To enrich ECL2-interacting molecules, the molecule poses resulting from the docking calculation to 3NY9 were filtered by applying a distance cutoff between the molecule and certain residues in the ECL2 (2.6 Å to F193$^{45.52}$ (backbone), D192$^{45.51}$ (sidechain), C191$^{45.50}$ (backbone) or T195^{ECL2} (sidechain); see Figure 3.1). The docking poses of the top ranked molecules within the distance cutoffs were then also evaluated visually. Molecules that entertained favourable interactions, i.e. interacted with the protein but did not contain non-interacting polar functional groups, were chosen from all ranking, re-ranking and filtered ranking lists. A special focus during the evaluation process was directed towards novel structural scaffolds. An interaction of the molecule with

D113$^{3.32}$ was essential for the molecules to be chosen. Finally, 27 molecules were purchased and pharmacologically characterised with respect to binding to the β_2AR (see Table A3.5). Of these molecules, five molecules interacted with T195^{ECL2}, four molecules with one of C191$^{45.50}$-F193$^{45.52}$ and 18 molecules did not interact with the ECL2 in their docking pose.

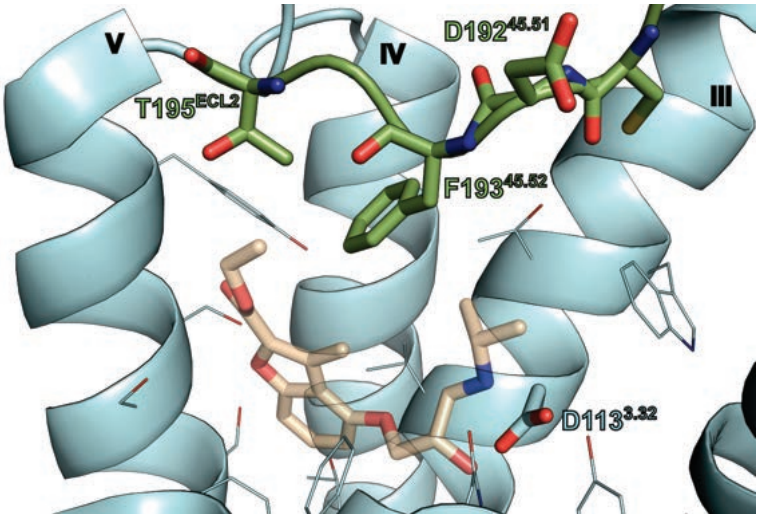


Figure 3.1: The orthosteric binding pocket of the β_2AR in an inactive conformation (PDB ID 3NY9[69]). The ECL2 is highlighted in green and the residues that were considered for the distance filtering of the docking poses (F193$^{45.52}$, D192$^{45.51}$, C191$^{45.50}$ and T195^{ECL2}) are shown as sticks. Residue D113$^{3.32}$, which was a compulsory interaction partner during selection of molecules, is also shown in stick representation. The crystallised inverse agonist compound 1 is shown in transparent sticks and all surrounding residues within 4 Å distance from it as lines to allow for a better orientation in the pocket. TMs are numbered at the top of each helix.

Secondary docking screen. The 27 molecules chosen during the primary docking screen were tested for their affinity towards the β_2AR using a fluorescence-labelled ligand displacement assay (results not shown). Although this preliminary assay was not conducted with a number of replicates to be statistically significant, it showed that only two compounds had an affinity towards the β_2AR. Of these two compounds, the structurally more novel compound –**1 (MS008)**– was chosen to be further characterised and exploited for a structure-activity-relationship (SAR) study. This molecule contains a coumaran-based scaffold (2,3-dihydro-1-benzofuran-2-ylmethanamine) that was previously undescribed for ligands of the β_2AR. Molecules with a similar structure to **1 (MS008)** were searched in the ZINC15 lead-like subset and docked to PDB ID 3NY9, since the parent molecule was selected from the docking screen to that structure during the primary screen. After visual inspection of the molecule poses, eleven molecules were selected, purchased and characterised phar-

macologically towards the β_2AR. The 2D-depictions of these molecules can be found in Table A3.1.

3.3.2 Pharmacological characterisation of 1 (MS008) and its analogues

All assays described here were performed and the data analysed by our collaboration partners at *InterAx Biotech AG*. However, the results are discussed in more detail here to evaluate the performance and significance of the docking screens. A binding assay as well as a functional assay measuring response along the G_s signalling pathway were performed for **1 (MS008)** and its analogues regarding the β_2AR. Additionally, the same functional assay was conducted for the β_1AR in order to evaluate selectivity. The results for the β_2AR are discussed in more detail below, as all predictions were made for this receptor. The assay results for the β_1AR which was not included in the predictions can be found in the additional material. All compounds show similar behaviour at the β_1AR and the β_2AR.

Ligand binding assay. To determine the affinities of **1 (MS008)** and its derivatives for the β_2AR, two different ligand displacement assays were performed. The first assay was fluorescence based, using Propranolol-green as the fluorescence-labelled ligand to be displaced from the orthosteric binding pocket. The resulting dose-response curves for selected compounds can be seen in Figure 3.2.A and the pK_i values calculated from these curves can be found in Table A3.1. In this assay, the parent compound **1 (MS008)** showed the highest affinity for the β_2AR ($pK_i = 7.0 \pm 0.1$), closely followed by compounds **2** ($pK_i = 6.9 \pm 0.1$) and **8** ($pK_i = 6.8 \pm 0.1$). These pK_i values are comparable to the affinity measured for the reference agonist Isoprenaline ($pK_i = 6.6 \pm 0.1$). All the other tested analogues of **1 (MS008)** showed lower, yet measurable, affinities for the β_2AR. Since some of the ligands showed autofluorescence in the examined wavelength region (in particular compounds **5**, **6**, **7**, **8** and **11**), an additional radioligand binding assay using $[^3H]$-DHA as displaced ligand was used to confirm the pK_i values of these compounds (see Figure 3.2.B, Table A3.1). The parent compound **1 (MS008)** as well as Isoprenaline were used as reference and to prove comparability of the fluorescence- and radioactivity-based ligand displacement assay results. As can be seen, the pK_i values calculated from the dose-response curves are indeed comparable between both assays (see Table A3.1).

Functional assay. The docking calculations were performed with structures in an inactive conformation. Hence, it was expected that **1 (MS008)** and its derivatives show antagonistic properties towards the G_s signalling pathway. This was confirmed with a real-time cAMP inhibition assay using Isoprenaline as stimulating compound. For all compounds, the IC_{50} values were calculated using the concentration-response curves at a fixed Isoprenaline concentration (500 nM for the β_2AR) and compared to the results for the reference compound ICI 118551, an antagonist of the β_2AR (see

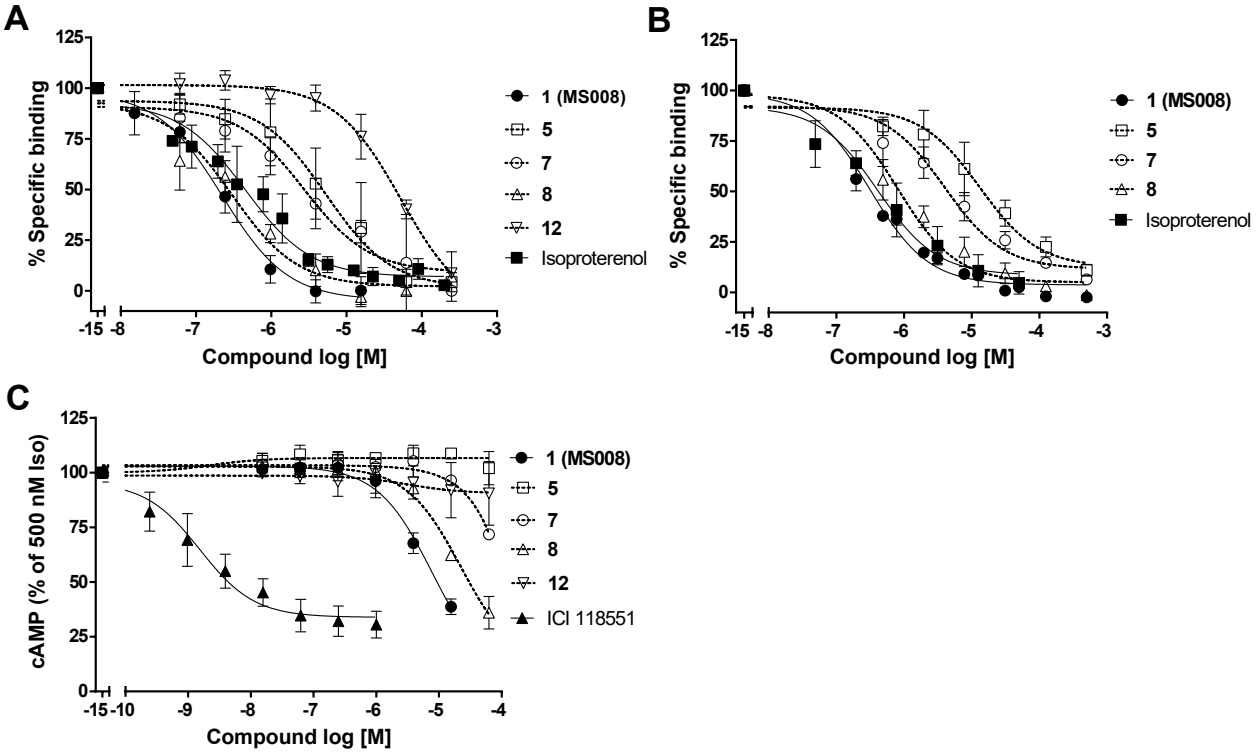


Figure 3.2: Dose-response curves in A) heterologous HTRF competition binding assay, B) heterologous radioligand competition binding assay and C) whole cell cAMP inhibition assays for selected compounds and the β_2AR. Data represents A) mean ± SD of three to five independent experiments carried out in duplicates, B) mean ± SD of three independent experiments carried out in duplicates and C) mean ± SD of three to five independent experiments carried out in duplicates. The pharmacological characterisation was conducted by our collaborators at *InterAx Biotech AG*. More details on assays can also be found in the main publication.[83]

Figure 3.2.C). The results show that although the majority of the compounds inhibits the cAMP production induced by Isoprenaline and they, therefore, act as antagonists their potency is much lower than that of reference compound ICI 118551 (see Table A3.2). The same assay was also performed with the β_1AR and similar results were obtained for this receptor (see Table A3.2).

3.4 Discussion

As described above, the ZINC12 lead-like library was screened against two structures of the β_2AR in inactive conformations, 2RH1 and 3NY9, during the primary screen (see Table A3.5). A special focus during evaluation of these docking calculations was directed towards structural novelty of the compounds and an interaction with the ECL2. As expected for a screen with a special focus on structural novelty, the hit rate of this primary screen was low (1 out of 27 molecules, i.e. 3.7%).

Nevertheless, it resulted in a molecule with a measurable affinity for the β_2AR that fulfils both criteria, i.e. structural novelty and an interaction with the ECL2, compound **1 (MS008)**. Since that primary screen did not yield any additional molecules with measurable affinity and a novel structural scaffold, it will not further be discussed here.
An analogue search was conducted with **1 (MS008)** as query and the resulting molecule set docked to the β_2AR structure with PDB ID 3NY9. Eleven additional molecules were chosen from this secondary screen after evaluation of their docking poses. All of these eleven molecules showed affinity for the β_2AR, resulting in an excellent hit rate of 100%. The overall hit rate for both screens is, therefore, 12 out of 38 molecules or 32% which is similar to what has been found in other screens against GPCR targets.[190,193–198]

Novelty of compound 1 (MS008). The structural novelty of **1 (MS008)** and its analogues was confirmed by calculating the Tanimoto similarity based on ECFP4 between each of the compounds and all molecules in ChEMBL showing activity at the β_2AR.[192] The closest match was found between **7** and CHEMBL1383731 (ECFP4 Tanimoto value of 0.35; Table A3.4). Additionally, an even broader search was conducted by comparing the twelve molecules against all molecules from ChEMBL that showed activity at *any* of the adrenergic receptor subtypes in *Homo sapiens*. The closest match found here was between **4** and CHEMBL222798 (ECFP4 Tanimoto value of 0.4; Table A3.4). None of the most similar ChEMBL ligands contains the coumaran-based scaffold (2,3-dihydro-1-benzofuran-2-ylmethanamine) which makes **1 (MS008)** and its analogues novel compared to previously described ligands of the β_2AR. To further investigate the novelty of the coumaran-based scaffold, a substructure search on the set of active molecules against any adrenergic receptor was conducted. Three different scaffolds of different sizes were used for this search, (a) coumaran, (b) the coumaran-based scaffold identified in this work (2,3-dihydro-1-benzofuran-2-ylmethanamine) and (c) the previous scaffold with an additional benzyl attached to the coumaran moiety. While search (a) resulted in nine molecules, two of which are actives of the β_2AR (Tanimoto similarity range of 0.14-0.24 to compounds **1-12**), search (c) did not yield any molecules. Search (b) resulted in the known α_2-adrenergic receptor ligand Efaroxan. However, this molecule shows a rather low Tanimoto similarity range of only 0.16-0.19 to compounds **1-12**. These results confirm the novelty of compound **1 (MS008)** and its analogues as ligands of the β_2AR, which is the result of the coumaran-based scaffold common to all these molecules.

Binding pose of 1 (MS008). According to the molecule pose predicted by the docking calculations, the coumaran-based scaffold is placed in the orthosteric binding pocket. The pose is stabilised by hydrophobic interactions between the coumaran ring system and V114$^{3.33}$, while the positively charged primary amine entertains ionic interactions with residues D113$^{3.32}$ and N312$^{7.39}$ which

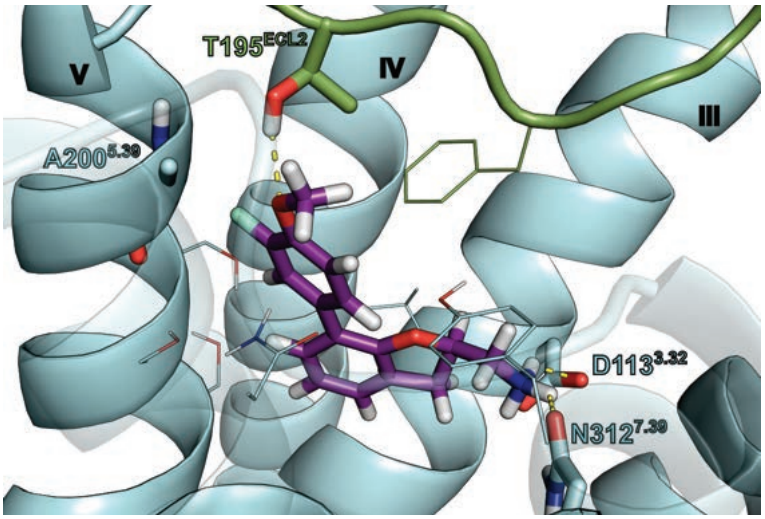


Figure 3.3: Binding pose of compound **1 (MS008)** as predicted by the docking calculation to structure PDB ID 3NY9. The compound forms hydrogen bonds with several residues in the orthosteric binding pocket which are highlighted as stick representation, i.e. D113$^{3.32}$, N312$^{7.39}$ and T195$^{\mathrm{ECL2}}$. The fluorine atom possibly interacts with A200$^{5.39}$-C$_\alpha$ ^{199} which is also highlighted as stick representation. The ECL2 is highlighted in green and several binding pocket residues that are often involved in ligand binding are shown as lines. TMs are numbered at the top of each helix and TM VI and VII which are located between the observer and the ligand are shown in transparent.

were shown to be important for ligand binding in mutational studies (see section 2.1.4). The aromatic ring attached to the coumaran is pointing towards the extracellular side of the receptor, allowing a hydrogen bond interaction between the methoxy group attached to this aromatic ring and T195$^{\mathrm{ECL2}}$ (see Figure 3.3), thereby showing the interaction with ECL2 which was specifically favoured during selection of the molecules.

Structure-activity relationship. The shared coumaran-based substructure of **1 (MS008)** and its analogues allows a closer evaluation of the influence of the different substituents on the ligand properties with respect to the β_2AR, especially their affinity (see Table A3.1). Based on the hit rate of 100% among the analogues of **1 (MS008)**, it can be concluded that the core structure of a coumaran with an attached primary amine as well as an aromatic moiety in general results in affinity for the β_2AR. Furthermore, different substituents and variations of this core substructure lead, as expected, to different affinities towards the β_2AR which allows suggestions for modifications to improve the ligand design further. It should be noted that the deductions are made based on molecule poses resulting from docking calculations which may not necessarily represent the actual binding modes of the molecules. In general, an interaction with T195$^{\mathrm{ECL2}}$ seems to stabilise the

binding mode and results in a higher affinity towards the β_2AR.

For the majority of the molecules the differences can be found in the substituents and substituation patterns of the aromatic moiety attached to the coumaran. These substituents can be located in *para*, *meta* or *ortho* position relative to the coumaran scaffold. A methoxy or hydroxyl group in *para* position led to a higher binding affinity to the β_2AR in general (**1 (MS008)**, **2**, **4**, **6**, **8**; $pK_i = 6.0$-7.0), which might be due to an interaction of this group with $T195^{ECL2}$. Surprisingly, the affinity is almost unchanged when exchanging the hydroxyl group in *para* position for an amide group ($pK_i(\mathbf{4}) = 6.2$ vs. $pK_i(\mathbf{11}) = 6.1$ and $pK_i(\mathbf{7}) = 5.9$). According to the docking pose, this might be due to a hydrogen bond between the carbonyl oxygen of the amide and $T195^{ECL2}$, forming a similar interaction as hydroxyl and methoxy group. Since a slight elongation of the amide decreases the affinity only slightly ($pK_i(\mathbf{11}) = 6.1$ vs. $pK_i(\mathbf{7}) = 5.9$), it might be interesting to investigate which influence a further elongation of the amide has, considering the possibility of maybe even targeting the exosite of the receptor (at the top of the binding pocket where salbutamol reaches[82]). Addressing this region, which is dissimilar between the β_1AR and the β_2AR might introduce selectivity against the β_1AR.

In general, compounds that show an interaction with $T195^{ECL2}$ (e.g. **1 (MS008)**, **6**, **11**) have a higher affinity to the β_2AR than compounds that lack this interaction (e.g. **5**, **3**). However, it seems like the lack of an interaction to $T195^{ECL2}$ can at least partially be compensated by an interaction to other residues, e.g. to $N293^{6.55}$, resulting in lower binding affinities ($pK_i = 4.6$-5.7) for e.g. **5** or **10** than for the compounds interacting with $T195^{ECL2}$ (e.g. **1 (MS008)** with $pK_i = 7.0$).

Furthermore it was observed that a fluorine substituent in *meta* position seems to increase binding affinity ($pK_i(\mathbf{4}) = 6.2$ vs. $pK_i(\mathbf{8}) = 6.8$). This might be due to an interaction of the slightly negatively polarised fluorine with the slightly positively polarised α-carbon of $A200^{5.39}$.[199] Upon addition of a second fluorine atom in *meta* position, no change in binding affinity can be observed. This suggests that the second fluorine atom in this position cannot interact with another matching residue in the binding pocket ($pK_i(\mathbf{1\ (MS008)}) = 7.0$ vs. $pK_i(\mathbf{2}) = 6.9$).

Although adding a fluorine substituent to the coumaran moiety itself decreases the binding affinity of the ligand ($pK_i(\mathbf{6}) = 6.0$ vs. $pK_i(\mathbf{1\ (MS008)}) = 7.0$) it does not impede ligand binding completely. Yet, this substituent is pointing towards residues $S203^{5.42}$, $S204^{5.43}$ and $S207^{5.46}$ in TM V which are involved in receptor activation (cf. chapter 2.1.4). Hence, exchanging the fluorine substituent against e.g. a hydroxyl group might lead to an interaction of the ligand with one of these residues which might not only result in an increased binding affinity but also agonistic effects.

The lowest affinity is observed for compound **12** ($pK_i(\mathbf{12}) = 4.6$), which is probably caused by the bulkier substituents attached to the aromatic moiety. Furthermore, interactions of the hydroxyl

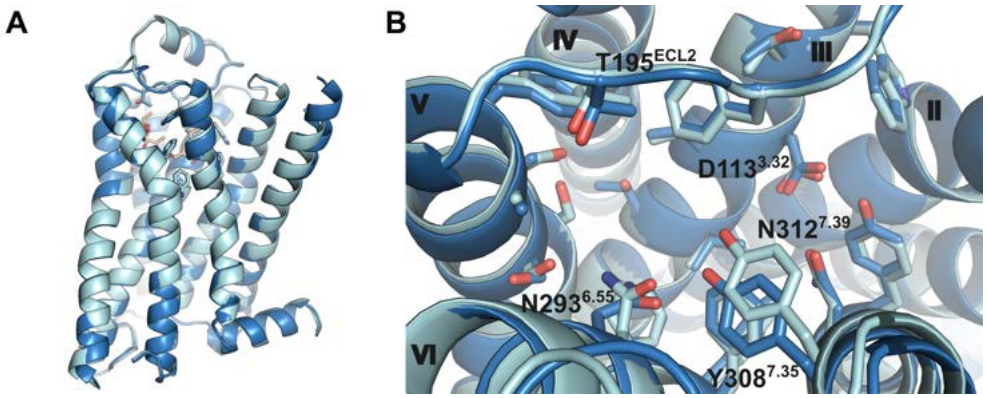


Figure 3.4: Overlay of the β_2AR structures in an inactive conformation with PDB ID 2RH1[6] (blue) and 3NY9[69] (aquamarine). Structure 2RH1 was targeted by a docking calculation in the study described in Kolb et al.[190] and both structures were targeted in the study described here. A) Side view of the overlayed structures. Only slight differences in helix placement are visible. B) Orthosteric binding pocket of the β_2AR with residues in proximity to the ligand binding site highlighted as sticks. The side chain orientation differs only slightly for most residues while bigger differences are visible for N293$^{6.55}$ and Y308$^{7.35}$.

groups with the receptor are probably less favourable due to unfavourable interaction angles caused by the longer substituent chains.

3.5 Comparison of docking studies

In 2009, *Kolb et al.*[190] conducted a docking study on the β_2AR, targeting the structure with PDB ID 2RH1 which has also been targeted in this study. *Kolb et al.* found six hit molecules, one of which was compound 1, a potent inverse agonist. Later on, the β_2AR was crystallised with this molecule resulting in the structure with PDB ID 3NY9, the second structure used for the docking calculations in this study. Since the study in 2009 was conducted by Peter Kolb, we had full access to all information and resources about ligand sets and ranking lists. This allowed us to directly compare the study from 2009 with the current study to evaluate which effects subtle changes in structures can have and how they influence the outcome of such docking studies. The differences between the two β_2AR structures 2RH1 and 3NY9 were evaluated by calculating RMSD values. The differences are fairly small (see also Figure 3.4), with a backbone RMSD of 0.388, an all-heavy-atom RMSD of 0.650 and a binding pocket RMSD of 0.357 (binding pocket defined as residues within 5 Å of Carazolol in 2RH1).

Comparison of docking results. To evaluate whether the molecules selected during the present

study could have been identified in 2009, they were searched in the molecule library used back then which was the ZINC7 database. As it turns out, the ZINC7 database contains only two out of the 27 molecules selected during the primary screen in the present study, at ranks 4611 and 40882. Conversely, it can be evaluated whether the hit molecules found in 2009 could have been discovered in the present study as well. The molecule library used in this study –the ZINC12 lead-like library– did not contain all six hit molecules from 2009 either. These molecules were then docked to both 2RH1 and 3NY9, using the same docking setups as for the library screen. The six molecules came in at ranks 345, 19519, 20332, 16928, 426387 and 3048263 in the docking calculation with structure 3NY9 (see Table A3.3), which means that only the most potent compound 1 would have been discovered in the present screen.

Novelty of the screens. The novelty of the results of both screening studies was compared by calculating an ECFP4 Tanimoto between the compounds and different ChEMBL molecule sets. Two molecule sets from ChEMBL containing all actives against *any* of the adrenergic receptor subtypes were retrieved, the first at the state of the database in 2009 (which best represents the state of knowledge during the 2009 study), the second at a more recent state using the version that has also been used to evaluate the novelty of **1 (MS008)** and its analogues. The 32 molecules selected in the study in 2009 and the molecules chosen during the primary screen in the present study were then compared to both sets by calculating the ECFP4 Tanimoto similarity. The Tanimoto ranges calculated in comparison to the ChEMBL2009 molecule set are 0.27-0.76 for the compounds from the 2009 screen and 0.28-0.60 for the current screen. With the more recent ChEMBL version, ranges are 0.3-0.79 for the first and 0.29-0.60 for the latter screen. This demonstrates that the novelty of the selected compounds during both screening studies is comparable. Furthermore, when comparing compound **1 (MS008)** and its analogues discovered in the present study to both ChEMBL sets, the Tanimoto ranges are 0.25-0.4 for ChEMBL2009 and 0.27-0.4 for the current ChEMBL version which again highlights the novelty of these compounds, as the ranges are much lower than for the molecules chosen in the 2009 screen or during the primary screen.

Impact of these results. Two main observations can be made from the comparison of the docking screen from 2009 to the current screen. First, and not surprising, molecule libraries are still growing and might therefore still feature novel ligands for well-investigated targets. The described study emphasised this with the discovery of structurally novel ligands for the β_2AR, a receptor that has been extensively studied in drug research and drug design. Second, the results show that molecules discovered in one screen are not necessarily found in a screen to another structure. This suggests that even subtle structural differences between two structures can strongly impact the outcome of a docking screen and, therefore, new structures of a receptor contain valuable information even

when differences to previous structures might be small. Hence, exhausting all available structural informations and conducting new docking studies upon retrieval of new information, even if structural differences are slight, is important and could be a powerful tool to maximise the yield of novel ligands. These different conformations could even be deliberately created with slightly differing structures from homology modelling.[200]

3.6 Conclusions

The goal of this study was to find novel ligands for the β_2AR by docking a molecule library to two crystal structures of the β_2AR in inactive conformations (PDB IDs 2RH1 and 3NY9). A special focus during selection of molecules from these docking results was directed towards interactions of the molecules with the ECL2 and novel structural scaffolds of the molecules compared to known ligands of the β_2AR. Given these additional restrictions, a low hit rate was anticipated and indeed the hit rate of the primary screen using a fluorescence binding assay for confirmation was only 3.7%. However, the goal to find an unprecedented structural scaffold for ligands of the β_2AR was reached with compound **1 (MS008)** and its coumaran-based scaffold (2,3-dihydro-1-benzofuran-2-ylmethanamine), proving that it is still possible to find novel ligands with unprecedented structures even for well-described targets such as the β_2AR. In a subsequent derivative search, more molecules containing this structural scaffold could be identified and after docking them to the structure with PDB ID 3NY9, eleven derivatives of **1 (MS008)** were selected. All molecules chosen during the secondary screen showed affinity for the β_2AR, resulting in a 100% hit rate for the secondary screen (see Table A3.1). This also demonstrates in general the affinity of the coumaran-based scaffold for the β_2AR. As was expected due to the fact that the docking calculations were done with structures in inactive conformations, all novel ligands showed antagonistic properties towards the G_s signalling pathway which was confirmed with a functional assay (see Table A3.2).
Furthermore, the docking screen was compared with a previous docking study targeting the same receptor.[190] The results from this comparison showed that new crystal structures contain new information even when structural differences to previous structures seem rather small. Hence, new structural information adds value and might be useful in the approach to find novel ligands and potentially safer drugs targeting even well-described GPCRs such as the β_2AR. In addition, it showed that molecule libraries for screening studies are growing and, thereby, cover a larger portion of chemical space, which might open up perspectives for structurally novel ligands.

3.7 Perspectives

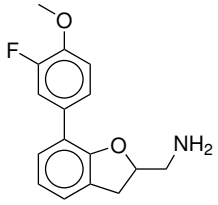
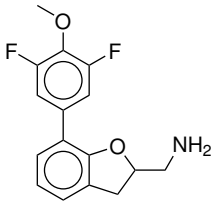
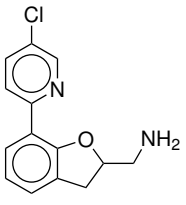
Based on this successful study, future projects could be developed. In general, the study showed that it is still possible to find novel ligands for well-described receptors such as the β_2AR and even novel ligands with unprecedented scaffolds. Therefore, new screening studies using the growing molecule libraries and the new information available due to the rising number of crystal structures could still yield novel ligands targeting well-known receptors. This could, in a longer perspective, yield new tool compounds allowing to learn more about receptor function and signalling and new drugs with potentially fewer side effects for the treatment of diseases.

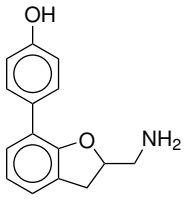
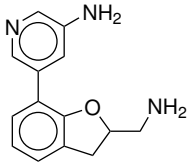
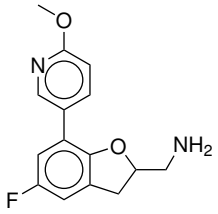
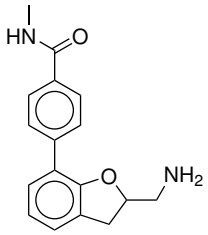
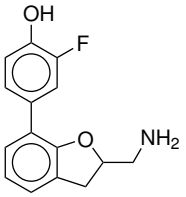
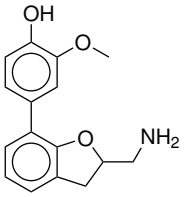
Furthermore, the novel coumaran-based scaffold (2,3-dihydro-1-benzofuran-2-ylmethanamine) discovered in this study could be modified by adding different substituents and thereby modifying e.g. the affinity for the β_2AR or efficacy towards the G_s or other signalling pathways. Several ideas for possible modifications can be based on **1 (MS008)** and its derivatives that were tested in this study. Using for example compound **7** as a starting point, the amide moiety could be elongated to then potentially reach past ECL2 and to the extracellular site to target the exosite on top of the binding pocket (similar to salmeterol[82]). This could possibly lead to a ligand with a strong affinity to the β_2AR, a selectivity for the β_2AR over the β_1AR or it might even induce unexpected signalling effects. A modulation of signalling behaviour by introducing different substituents could also be tested with compound **6**. When considering observations from mutational studies that showed that S203$^{5.42}$, S204$^{5.43}$ and S207$^{5.46}$ are possibly involved in receptor activation (see also chapter 2.1.4 and[79,80]), replacing the fluorine substituent attached to the coumaran moiety by something more polar such as a hydroxyl group might shift the ligand's behaviour from antagonistic properties to agonistic properties regarding the G_s signalling pathway. It would be interesting to explore whether such a shift of properties could be induced merely by introducing an interaction to one of these residues.

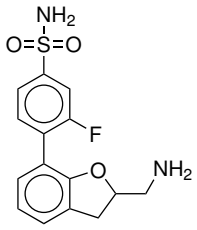
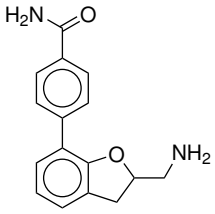
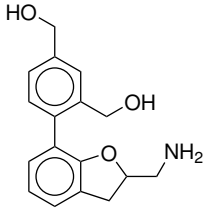
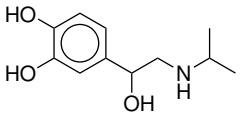
Additionally, it would be interesting to determine the properties of all ligands as well as all possible follow-up compounds not only for the β_2AR but also for the β_1AR to gain better understanding of the selectivity of these compounds towards one of the receptors. Substituent modifications could be used to modulate possible selectivity and get a better understanding of underlying interaction patterns.

3.8 Additional information

Table A3.1: pK_i values (affinity) of compound **1 (MS008)** and its analogues as well as the reference compound Isoprenaline as determined from two different assays. Fluorescence assay: pK_i values determined with a heterologous competition HTRF binding assay with 50 nM Propranolol-green. Radioligand assay: pK_i values determined with a heterologous competition radioligand binding assay with 1 nM ^{3}H-DHA. Only those compounds that showed autofluorescence were tested in the radioligand binding assay. pK_i is defined as -logK_i and values are shown as mean ± SD of n independent experiments carried out in duplicates. Assays and data analysis were performed by our collaboration partners at *InterAx Biotech AG*.

		Fluorescence assay		Radioligand assay	
Compound	Structure	pK_i	n	pK_i	n
1 (MS008)[a]		7.0±0.1	5	7.1±0.1	3
2[b]		6.9±0.1	3	n.d.	
3[c]		4.9±0.2	3	n.d.	

4		6.2±0.2	3	n.d.	
5		5.6±0.1	3	5.5±0.1	4
6		6.0±0.1	3	6.1±0.1	3
7		5.9±0.2	3	6.0±0.1	3
8		6.8±0.1	3	6.7±0.1	3
9[c]		5.8±0.2	3	n.d.	

10[c]		5.0±0.3	3	n.d.	
11		6.1±0.1	3	6.3±0.1	3
12		4.6±0.1	3	n.d.	
Isoprenaline		6.6±0.1	4	7.0±0.1	3

n.d. not determined

a Compound from primary screen that was used for analogue search.

b Compound showed weak autofluorescence, but was not tested in the radioligand binding assay due to its similarity to **1 (MS008)** and **8**.

c Compounds showed weak autofluorescence, but were not tested in the radioligand binding assay due to their low affinities.

Table A3.2: pIC_{50} values of **1 (MS008)** and its analogues as well as reference compound ICI 118551 calculated from the dose-response curves of the cAMP β_2AR inhibition assay and the cAMP β_1AR inhibition assay. Cells were stimulated with 500 nM Isoprenaline (β_2AR) or 50 nM Isoprenaline (β_1AR). pIC_{50} is defined as -logIC_{50} and values are shown as mean ± SD of n independent experiments carried out in duplicates. Assays and data analysis were performed by our collaboration partners at *InterAx Biotech AG*. n.d. not determined

	β_2AR		β_1AR	
Compound	pIC_{50}	n	pIC_{50}	n
1 (MS008)	5.1±0.1	5	4.6±0.2	5
2	5.1±0.1	3	4.9±0.1	3
3	n.d.	2	n.d.	2
4	4.7±0.1	3	4.6±0.1	3
5	n.d.	2	n.d.	2
6	4.4±0.1	4	4.4±0.3	5
7	n.d.	2	4.6±0.5	4
8	4.6±0.1	3	4.5±0.1	3
9	n.d.	2	n.d.	2
10	n.d.	2	n.d.	2
11	4.5±0.2	5	4.5±0.3	4
12	n.d.	2	n.d.	2
ICI 118551	8.8±0.1	3		

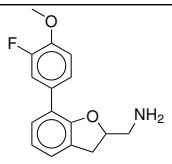
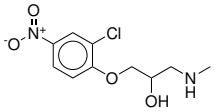
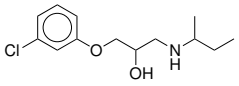
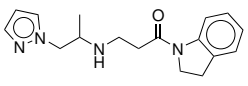
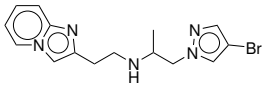
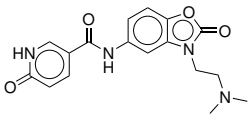
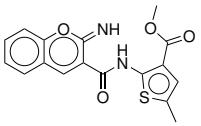
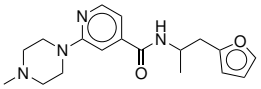
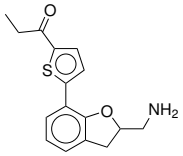
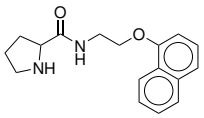
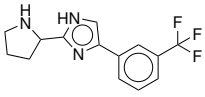
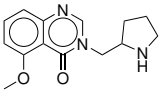
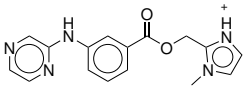
Table A3.3: Comparison of docking results with the study from 2009.[190] for this cmparison, the ranks of the six hits and their enantiomers from the docking screen against structure 2RH1[190] were compared with the ranks when using the docking setup from this study. "Rank 2009" designates the ranks of the respective molecules in Kolb et al.[190] while "Rank 2RH1" and "Rank 3NY9" are the ranks that these molecules would have obtained within the database docked in the present study.

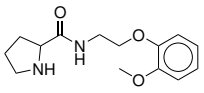
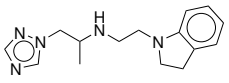
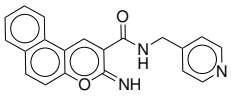
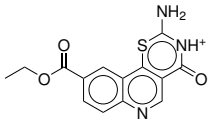
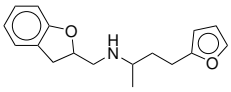
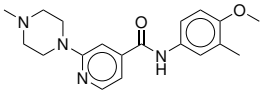
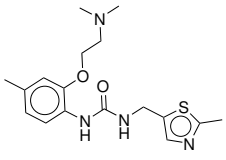
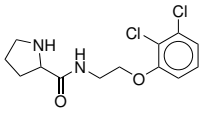
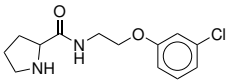
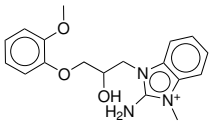
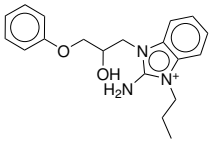
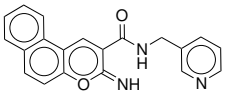
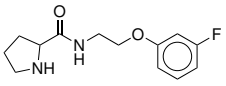
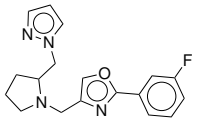
No.	ZINC ID	Rank 2009	Rank 2RH1	Rank 3NY9
1	C04008295	15	182	345
	C04008294	19	1948	1804
2	C03003177	150	9560	19519
3	C02880812	163	9881	88007
	C02880813	273	1456	20332
4	C06703239	409	16166	16928
5	C04123268	182	12790	426387
6	C20589273	-	1443595	3048263

Table A3.4: Similarity of **1 (MS008)** and its analogues in comparison to molecules from two different ChEMBL datasets. In both cases the ChEMBL ID of the most similar molecule for each of the compounds as well as the calculated ECFP4 Tanimoto coefficient for the respective molecule pair is listed. β_2AR: ChEMBL dataset containing all molecules that showed activity against the β_2AR. Any adrenergic: ChEMBL dataset containing all molecules that showed activity against *any* of the adrenergic receptor subtypes.

	β_2AR		*Any* adrenergic	
Compound	ChEMBL ID	Tanimoto	ChEMBL ID	Tanimoto
1 (MS008)	CHEMBL1704973	0.28	CHEMBL2153557	0.36
2	CHEMBL1203541	0.27	CHEMBL2153557	0.30
3	CHEMBL1541834	0.24	CHEMBL315772	0.30
4	CHEMBL1347279	0.33	CHEMBL222798	0.40
5	CHEMBL3909417	0.22	CHEMBL467259	0.30
6	CHEMBL1203541	0.24	CHEMBL481321	0.27
7	CHEMBL1383731	0.35	CHEMBL466012	0.37
8	CHEMBL1094323	0.27	CHEMBL3590203	0.30
9	CHEMBL3799593	0.30	CHEMBL2153551	0.36
10	CHEMBL433454	0.25	CHEMBL26717	0.30
11	CHEMBL1383731	0.30	CHEMBL458271	0.31
12	CHEMBL3290971	0.28	CHEMBL3590203	0.28

Table A3.5: List of ZINC IDs and 2D depiction of the compounds purchased and tested during the primary screen. No IC_{50} values from the initial binding assay are listed since the compounds were not measured in a statistically significant number of replicates. A more detailed pharmacological characterisation has only been conducted for compound **1 (MS008)**.

Name	ZINC ID	Structure	Name	ZINC ID	Structure
1 (MS008)	C67966542				
MS001	C19864514		MS002	C32110448	
MS003	C69413696		MS004	C95419952	
MS005	C58193556		MS006	C06825573	
MS007	C78998841		MS009	C67563017	
MS010	C41459488		MS011	C36874154	
MS012	C43000128		MS013	C97034351	

MS014	C41377287		MS015	C65488031	
MS016	C06644455		MS017	C34925912	
MS018	C38009606		MS019	C72281038	
MS020	C72292095		MS021	C41516858	
MS022	C41388708		MS023	C00527092	
MS024	C00301969		MS025	C06644454	
MS026	C41322021		MS027	C71875648	

4 | Comparative dockings to find novel agonists for the β_2AR

This study has been published in M. M. Scharf, M. Bünemann, J. G. Baker, P. Kolb, *Molecular Pharmacology* **2019**, *96*, 851-861.[81]

Contributions. The author of this thesis contributed to the design of the study, generated and analysed the computational data, performed and analysed *in vitro* assays, drew conclusions from the obtained results and was involved in writing the manuscript of the main publication.
The *in vitro* assays were conducted in the laboratories of Prof. Moritz Bünemann (Philipps-Universität Marburg) and Prof. Jillian Baker (University of Nottingham).

4.1 Introduction and goal of the study

G protein-coupled receptors (GPCRs) are one of the most important drug targets in humans (see also chapter 2.1) and understanding how to modulate their signalling and how to target them efficiently is therefore absolutely crucial. As described in chapter 2.1 there are certain similarities between different GPCRs within one class of this family. Therefore, it might be possible to transfer knowledge gained for one member of the GPCR family to other members within the same class which is incredibly useful to tackle the enormous task of getting a better understanding of this interesting and important family of proteins. Although this transferability of knowledge might of course not always be given it still shows up starting points and possibilities for future research. Hence, studying a well-described receptor like the β_2-adrenergic receptor (β_2AR) can lead to insights that could have not been gained by studying less well-described GPCRs but are useful for the understanding of the entire GPCR family. In this study, the β_2AR was chosen as the model receptor due to the availability of a number of crystal structures in different activation states.
The aim of this study was the prediction of ligands with agonistic properties based on conformational information contained in the receptor structure. Ligands can stabilise certain receptor conformations which might or might not lead to a certain signalling within a cell and, in reverse, it has been shown that the affinity of agonists is higher for activated receptors than for the inactive conformation.[28,29,84] Additionally, the results of previous studies showed that small changes in receptor structure can lead to entirely different docking results and that docking calculations to a receptor structure in an active conformation can increase the chance of finding molecules with agonist properties in large compound libraries.[83,194,200] With several additional structures of the β_2AR in active conformations released during the past decade we were wondering whether these

structures contained new information leading to novel agonists for the β_2AR. Furthermore, our goal was to see how to best aggregate the results of docking calculations to several structures to find more agonists.

A number of molecules was chosen and predicted to show agonistic properties using different docking strategies and these molecules were then characterised pharmacologically. This resulted not only in a remarkable ligand hit rate of 37% but all of these ligands also showed the predicted agonistic activity. Since it seems in general more difficult to predict activating ligands for the β_2AR (judging from previous docking studies[190]) it can be assumed that agonists resulting from this study are not 'accidental' hits but were predicted correct by applying the different docking strategies. Furthermore, retrospective analyses on the chosen evaluation methods gave further insights in how to find ligands with certain functions for class A GPCRs.

4.2 Methods

4.2.1 Structure preparation and docking calculations

Preparation and optimisation. Two structures of the β_2AR in an active conformation (PDB IDs 4LDL[78] and 3SN6;[8] β_2AR^{active}) as well as two structures of the β_2AR in an inactive conformation (PDB IDs 3NY9[69] and 2RH1;[6] $\beta_2AR^{inactive}$) were chosen to be prepared for docking. The preparation of the structures was done according to the general DOCK preparation protocol described in section 8.1.1. Additionally, the binding pocket of structure 3SN6 was relaxed by minimising residues W109$^{3.28}$, T110$^{3.29}$, D113$^{3.32}$, V117$^{3.36}$, F193$^{45.52}$, N293$^{6.55}$ and N312$^{7.39}$ (numbers in superscript according to the Ballesteros-Weinstein enumeration scheme[19]) in presence of the ligand crystallised within the binding pocket of this structure, BI167107, which was done to ameliorate clashes between ligand and receptor. It was also ensured that the polar hydrogen atoms of residues S207$^{5.46}$ and N293$^{6.55}$ were pointing towards the ligand to enable these residues to act as hydrogen bond donors. The optimisation of docking results and molecule poses was done by moving the spheres inside the binding pocket that are used by DOCK to translate and rotate the molecules into it. To verify the optimisation, the enrichment of a set of 125 enantiomers of known β_2AR ligands over decoys generated with DUD-E[155] was evaluated using ROC plots until a sufficiently high enrichment and early enrichment was reached (see section 8.1.4 for ligand set compilation).

Docking screen and molecule selection. The ZINC12[165] leads now subset containing 3,687,621 molecules was docked to the prepared crystal structures using the optimised docking setups and DOCK3.6.[154,156,170,172,174] Additionally, the resulting ranking lists were re-ranked in a 'dual re-ranking' comparing the ranking lists of the docking calculations to the two β_2AR^{active} to identify

molecules ranked well in both (cf. section 8.1.2). A 'selective re-ranking' comparing the ranking lists of the β_2AR^{active} 3SN6 and the $\beta_2AR^{inactive}$ 3NY9 was conducted to find molecules with more favourable scores in 3SN6 and less favourable scores in 3NY9. The top 500 molecules of the two ranking lists resulting from the docking calculations to the β_2AR^{active} 3SN6 and 4LDL and the two re-rankings were then inspected visually to account for the artificially inflated scores caused by the known deficiencies of scoring functions.[200] As an additional criterion the molecule poses of the selected molecules were also inspected in the docking calculation to the $\beta_2AR^{inactive}$ 3NY9 assuming they should adopt a similar binding pose in the β_2AR^{active} and the $\beta_2AR^{inactive}$. A few molecules showed favourable binding poses but contained bulkier molecule parts which did not fit the binding pocket ideally. Smaller derivatives of these molecules were searched in ZINC12 using a substructure search which ensured to only vary the bulky parts of the molecules. The resulting molecule set was docked to the β_2AR^{active} with PDB ID 3SN6 and the molecule poses evaluated visually. One additional molecule was selected from a docking calculation to the $\beta_2AR^{inactive}$ 2RH1. In total, 22 molecules were selected to be purchased and tested in *in vitro* assays.

Analogue search for initial hits. Two of the selected compounds showed agonistic behaviour in an initial cAMP-accumulation based assay and it was decided to search for more analogues of these two molecules. A similarity search in a library of 5,626,190 molecules from ZINC15[166] using an ECFP4[191] Tanimoto cutoff of $\geq$0.5 was conducted and a molecule set of 62 molecules was obtained. These molecules were then docked to the β_2AR^{active} 3SN6 and the resulting molecule poses were evaluated visually to select potential agonists. Finally, five molecules were selected to be purchased and tested in *in vitro* assays.

Evaluation of compound novelty. To evaluate the novelty of the selected and tested compounds, all molecules tested against the β_2AR were downloaded from ChEMBL[192] and filtered for active molecules. Each of the selected compounds was then compared to each molecule in the filtered ChEMBL dataset by calculating the Tanimoto similarity based on ECFP4. Additionally, the bioactivity dataset containing all molecules tested against *any* of the adrenergic receptor subtypes was also downloaded from ChEMBL and filtered for all active molecules. Again, all selected compounds were then compared to all molecules in this filtered ChEMBL dataset by calculating the ECFP4 Tanimoto similarity.

4.2.2 Pharmacological characterisation

To characterise the pharmacological properties of the molecules selected from the docking calculations, all compounds were subjected to three different assays. In a first approach, compounds were tested in a *Cisbio* cAMP accumulation assay in HEK293T cells for activation of the β_2AR

(laboratory of Prof. Bünemann, Philipps-Universität Marburg). A more detailed analysis of the pharmacological properties was then done using a radioligand displacement assay and a functional cAMP response element (CRE)-secreted placental alkaline phosphatase (SPAP) assay in the lab of Prof. Jillian Baker at the University of Nottingham, UK. The experimental procedures can be found in section 8.2. In the following, the three conducted assays are described in more detail.

***Cisbio* cAMP accumulation assay.** The initial characterisation of the compounds was done using the HTRF® cAMP G_s dynamic assay (Cisbio Bioassays, France). The experimental procedure is described in section 8.2. Upon stimulation of G_s coupled receptors with an agonist, cAMP is produced. A fluorescence labelled, cAMP binding antibody and a fluorescence labelled cAMP derivative are then added which emit a FRET signal if they bind to each other. The decrease of this FRET signal correlates directly with the amount of cAMP produced in the cell upon stimulation with the agonist which allows an assumption about the efficacy of the tested agonist.
Before testing the selected compounds in this assay it had to be adapted for optimal cell numbers, incubation times and a certain robustness concerning low DMSO concentrations. The measurements were pursued in Human Embryonic Kidney (HEK) 293T cells which were found to endogenously express the β_2AR at a level that allowed to measure the response upon stimulation with known agonists. The overall performance and results of the assay were confirmed by measuring known agonists of the β_2AR, in particular Isoprenaline, Epinephrine, Norepinephrine, BI167107 and Pindolol. Additional to the agonist setup, the assay can be run in an antagonist setup measuring the inhibition of agonist response by an antagonist. For the antagonist assay setup, the agonist and antagonist incubation times were optimised and the assay was validated using literature known antagonists and inverse agonists of the β_2AR (Propranolol, Carvedilol, Bucindolol, ICI 118551, Labetalol, compound 1[190]). All measurements were done in technical duplicates.

^{3}H-CGP 12177 whole cell binding assay. To confirm ligand binding, a radioligand displacement assay with ^{3}H-CGP 12177 as radioligand was conducted in Chinese Hamster Ovary (CHO) cells stably expressing either the β_2AR or the β_1AR. The experimental procedure is described in more detail in section 8.2. This assay as well as the used cell lines have been established and well described before.[201] Non-specific binding was determined with Propranolol at a final concentration of 10 µM. K_D values were then calculated using the Cheng-Prusoff equation and the K_D of ^{3}H-CGP 12177 as determined in previous experiments for these cell lines (cf. section 8.2).[201] To allow a comparison of the K_D values of the novel compounds with previously described ligands of the β_2AR and the β_1AR, seven known ligands were measured alongside the novel compounds, i.e. Cimaterol, Salbutamol, Salmeterol, Denopamine, CGP 12177, CGP20712A and ICI 118551.

CRE-SPAP production assay. To determine whether the novel compounds act as agonists to-

wards the G_s signalling pathway at the β_2AR and/or the β_1AR a CRE-SPAP production assay was used (see section 8.2 for a description of the experimental procedure). Briefly, upon production of cAMP as a result of recruitment of the G_s protein to the receptor after stimulation by an agonist, the expression of the CRE-SPAP reporter gene will be induced. The amount of expressed SPAP is directly related to the amount of produced cAMP and can be quantified by addition of *p*-Nitrophenylphosphate which is transformed to yellow, by UV-Vis spectroscopy quantifiable *p*-Nitrophenol by SPAP. Since SPAP is heat resistant to a certain degree, any interfering proteins can be degraded by heating the system prior to development with the substrate. EC_{50} and E_{max} values could be obtained from the sigmoidal dose-response curves and the maximum possible response was detected by stimulation with 10 µM Isoprenaline. Again, results obtained for the novel compounds were compared with literature known reference agonists of the β_2AR and the β_1AR, i.e. Cimaterol, Salbutamol, Salmeterol, Denopamine and CGP 12177.

To confirm that detected agonist responses were occurring via the orthosteric binding site of the βAR, the inhibition of the agonist response was investigated by an incubation with antagonists ICI 118551 (β_2AR) or CGP20712A (β_1AR) prior to stimulation with the agonists. Since the orthosteric binding site is then blocked by the antagonist, a higher concentration of an orthosterically acting agonist will be needed to induce the same response and the dose-response curve will be shifted to the right. K_D values for the antagonists can then be calculated from this right shift using the Gaddum equation and can be compared with the K_D values of the respective antagonists obtained from the ^{3}H-CGP 12177 whole cell binding assay.

These experimental assay procedures as well as cell line validation and data analysis have been described before.[202]

Efficacy ratio. As a result of determining affinity (K_D) and potency of the agonist response (EC_{50}) of the ligands in the same cell lines and in parallel experiments, the efficacy ratio ($\frac{K_D}{EC_{50}}$) at each receptor can be used as an indicator of intrinsic efficacy. The intrinsic efficacy describes the ability of a compound to stimulate a response at the observed receptor and the efficacy ratio can be used to rank ligands in order of this intrinsic efficacy at each receptor (although it cannot be compared between different cell lines).

A compound with a high value for the efficacy ratio, i.e. a lower EC_{50} than K_D value, would only need to occupy few receptors to stimulate the maximum response. This is for example the case for Cimaterol with a K_D of 81 nM and an EC_{50} of 0.21 nM at the β_2AR and hence a high efficacy ratio of 386, while a compound with a higher affinity like salmeterol can still be less efficacious (K_D 0.81 nM, EC_{50} 0.012 nM, efficacy ratio 67 at the β_2AR). Quite the opposite is the case for a partial agonist: while occupying all the receptors it would still not induce the maximum response,

i.e. the EC_{50} value would be only slightly lower than the K_D value and the efficacy ratio is therefore low. This is for example the case for CGP 12177 which can stimulate only 37% of the maximum response at a low efficacy ratio of 1.47 at the β_2AR (K_D 0.28 nM, EC_{50} 0.19 nM; see Table A4.2).

4.3 Results

4.3.1 Docking screen and molecule selection

In this study we were aiming to predict novel agonists of the β_2AR by docking calculations to structures in active conformations. For that we chose two different structures of the β_2AR in active conformations, one crystallised with an agonist and a nanobody to stabilise the active state (PDB ID 4LDL[78]) and one crystallised with an agonist and a G_s protein (PDB ID 3SN6[8]). In principle, these two structures should be in different states of activation since it can be assumed that only the receptor crystallised with the G_s protein is truly in a fully activated state. To exploit the spectrum of different active conformations better both structures were used. Additionally, two structures in inactive conformations were prepared for comparison (PDB ID 3NY9[69] and 2RH1[6]). Since the binding pocket in active conformations of the β_2AR is more contracted than in inactive conformations (see also section 2.1.4; Figure 4.1) it can be assumed that agonists docked to an inactive structure would take a very similar pose as in the active structure while the reverse should not hold true for antagonists. These docking calculations to the $\beta_2AR^{inactive}$ were also part of a related study described in chapter 3.[81]

Screening and molecule selection process. The ZINC12[165] leads now subset was docked to the prepared crystal structures using the optimised docking setups and DOCK3.6.[154,156,170,172,174] For the final selection of molecules to be characterised in the pharmacological assays, results were evaluated and compared in several ways, as will be described in the following and is also summarised in Figure 4.2. The molecule poses of the top 500 ranked molecules from the docking calculations to the β_2AR^{active}(3SN6 and 4LDL; Figure 4.2, red and yellow boxes) were evaluated visually to remove molecules ranking highly due to the known deficiencies of today's scoring functions.[200] Additionally, the ranking lists resulting from these two docking calculations were compared to each other to find molecules that are ranked highly in *both* dockings ("dual re-ranking", see section 8.1.2; Figure 4.2, orange box) and the top 500 molecules in the re-ranking list resulting from this comparison were also evaluated visually. The complementary approach comparing the ranking list from the docking calculation to the β_2AR^{active} 3SN6 to the one from the $\beta_2AR^{inactive}$ 3NY9 was pursued as a "selective re-ranking" (see section 8.1.2; Figure 4.2, green box) to find molecules that ranked high in the docking calculation to 3SN6 but low in the one to 3NY9. Again, the top 500 molecules of the

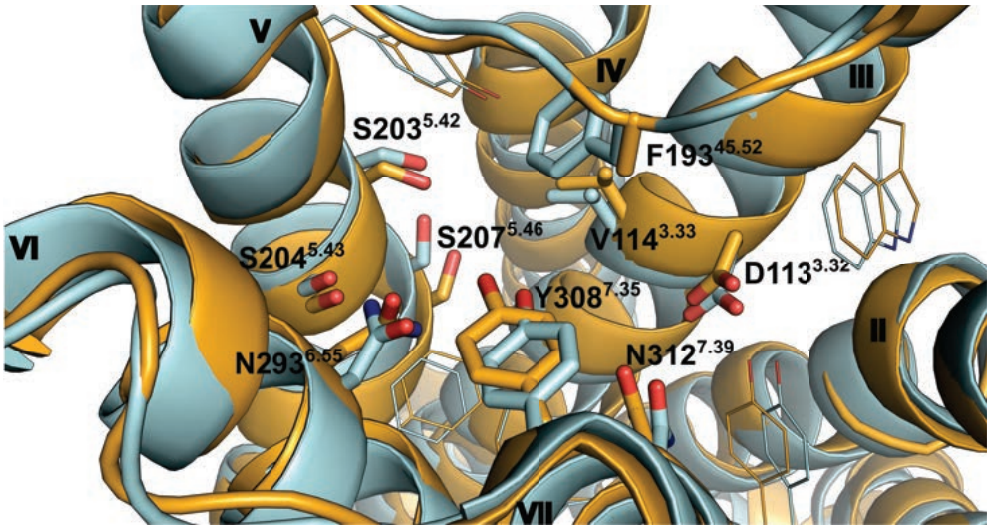


Figure 4.1: Orthosteric binding pocket of the β_2AR. Overlay of the β_2AR^{active} with PDB ID 3SN6[8] (orange) and the $\beta_2AR^{inactive}$ with PDB ID 3NY9[69] (aquamarine). Important residues are highlighted in stick representation and labelled with residue names, additional residues are shown in line representation and TMs are numbered. As can be seen, the side chain orientations between the β_2AR^{active} and the $\beta_2AR^{inactive}$ differ slightly, resulting in a smaller binding pocket volume for the β_2AR^{active} compared to the $\beta_2AR^{inactive}$. The side chain of residue F193$^{45.52}$ was not resolved in the β_2AR^{active} with PDB ID 3SN6.

resulting re-ranking list were evaluated visually. Thus, in total, the top 500 molecules of four different ranking lists were evaluated (3SN6, 4LDL, dual re-ranking, selective re-ranking; Figure 4.2, red, yellow, orange and green boxes). As an additional evaluation and selection refinement step the molecules from the first selection were also docked to the $\beta_2AR^{inactive}$ 3NY9 and their poses were evaluated visually, assuming that the likelihood of a true ligand to adopt a similar pose in the bigger binding pocket of 3NY9 compared to the poses in the smaller active binding pockets should be very high. In the end, 18 molecules were selected in this process. For five molecules that showed favourable binding poses but were too big to fit the binding pocket, smaller derivatives were searched from ZINC12. During the selection of the derivatives for redocking it was made sure that only the bulkier parts of these molecules would vary but not the parts of the molecules interacting with the receptor. The nine resulting molecules were docked to 3SN6, their poses evaluated visually and three molecules were selected for pharmacological characterisation (Figure 4.2, pink box). One additional molecule was selected from a docking calculation to the $\beta_2AR^{inactive}$ 2RH1 which was originally pursued for the study described in chapter 3. This molecule **8** was chosen because of its pose in 2RH1 which featured interactions with D113$^{3.32}$. This molecule did not adopt the same pose

in the docking calculations to one of the active structures but a pose that was flipped in the binding pocket forced by the rather bulky substituents. Anyway, it was decided to include this molecule due to its significantly smaller size which is in general considered to be important for agonists of the β_2AR. In total, 22 molecules were selected during this evaluation process to be characterised pharmacologically (see Table A4.1 and Table A4.3).

Analogue search for initial hits. Two of the selected compounds showed agonistic behaviour in an initial cAMP-accumulation based assay done in HEK293T cells, compounds **1** and **2**. These two molecules are close analogues to each other and it was decided due to their agonistic behaviour to search for more analogues of these molecules. A similarity search with a molecule library from ZINC15[166] resulted in 62 molecules. These molecules were then docked to the β_2AR^{active} 3SN6 since the parent molecules had the highest ranks in the docking calculation to that structure. The resulting molecule poses were evaluated visually and 5 additional molecules were selected to be subjected to the pharmacological characterisation (Figure 4.2, salmon box).

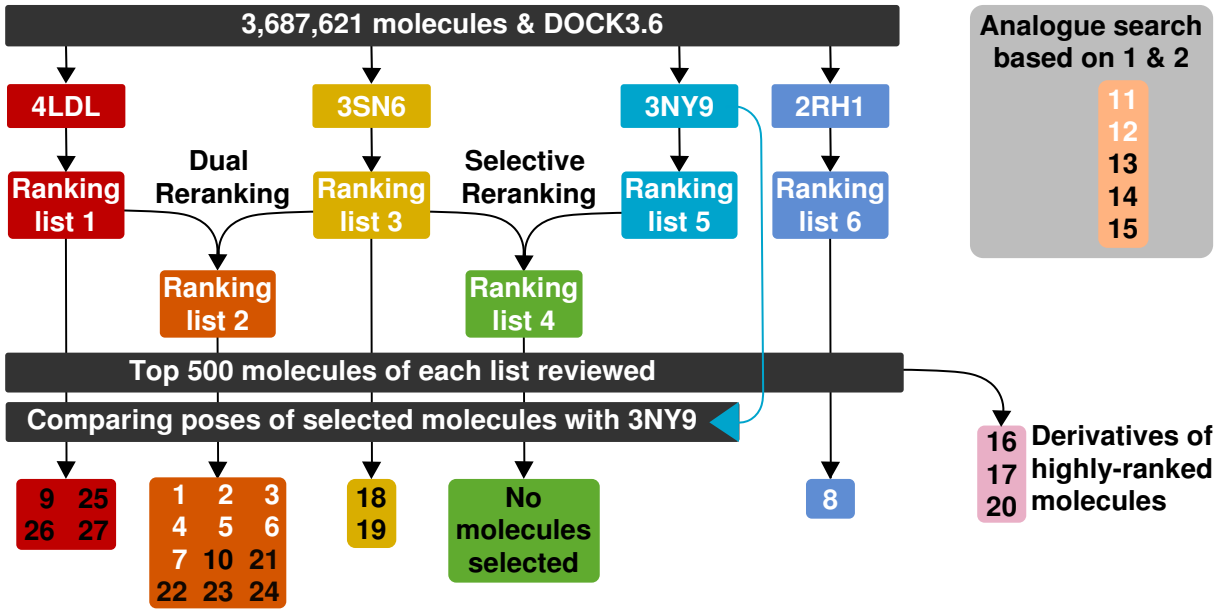


Figure 4.2: Schematic representation of the docking calculations to the β_2AR using two structures in active conformations (4LDL, 3SN6) and two structures in inactive conformations (3NY9, 2RH1). Additionally, the ranking lists from these docking calculations were re-ranked as indicated in the scheme. The molecules selected from each of these docking schemes are indicated. Molecules with names in white text where found to be agonists whereas those with names in black text did not interact with either the β_2AR or the β_1AR. In this screen, no antagonists (i.e. molecules that bind to the receptor without stimulating a response) were identified. Figure was taken from the primary publication mentioned in the beginning of this chapter.[81]

4.3.2 *In vitro* confirmation of the docking results

The 27 molecules selected from the different docking calculations were subjected to different *in vitro* assays to characterise their pharmacological properties. In a first assay, the 22 molecules from the primary screening campaign were tested for activation of the β_2AR signal transduction in HEK293T cells using a *Cisbio* assay. Based on these results the subsequent analogue search was conducted, five additional molecules were selected and also subjected to the same initial characterisation. A more detailed analysis of affinity and efficacy of all 27 molecules towards the β_2AR (and also the β_1AR) was then run in the lab of Prof. Jillian Baker at the University of Nottingham (UK) by the author of this thesis. The results of all experiments including reference experiments are described below.

***Cisbio* cAMP accumulation assay.** Since this assay was not previously optimised and validated, several respective experiments had to be run first. The experiments were conducted in HEK293T cells using the endogenously expressed β_2AR, i.e. without transfection of the cells with the β_2AR to achieve higher expression rates of the receptor. Isoprenaline was used as the reference agonist for all optimisation experiments. To maximise the assay window while still remaining within the dynamic range of the assay, the number of cells per well was optimised and 7000 cells/well was chosen as the most optimal of the tested cell concentrations. Degradation of cAMP during the experiments was inhibited by addition of the phosphodiesterase (PDE) inhibitor 3-isobutyl-1-methylxanthine (IBMX) in excess. Different IBMX concentrations were tested to ensure proper inhibition of PDE and an IBMX concentration of 0.5 mM was sufficient. To ensure equilibration of the agonist binding process prior to readout, different agonist incubation times were tested and 10 min was the minimum possible and maximum necessary incubation time. To account for the fact that the tested compounds are dissolved in DMSO, an additional experiment evaluating the influence of DMSO on assay results was conducted. The reference compound Isoprenaline is dissolved in water, which allowed to test four different DMSO concentrations of 0, 2, 5 and 10% DMSO in the wells. The results remained uninfluenced up to a final DMSO concentration of 5% which means that the DMSO from the compound stocks does not influence assay results and that it is not necessary to adjust the DMSO concentration throughout the different measurements to make them comparable.
To validate the assay results, dose-response curves of three literature known agonists of the β_2AR (Epinephrine, Norepinephrine and BI167107) as well as Pindolol, which was assumed to be a partial agonist, were measured additionally to Isoprenaline. Although the obtained EC_{50} values can not be compared to the literature due to the dependency of this value on assay system and experiment, it was possible to compare the order of the obtained EC_{50} values with the expectations

from the literature. As expected from the literature values, the order of the logEC_{50} values was BI167107 < Isoprenaline < Epinephrine < Norepinephrine. Pindolol did not induce any response, however, it depends highly on the assay whether it is reported as a partial agonist or an antagonist. During measurements using the antagonist assay setup the rather antagonistic behaviour of Pindolol in this assay system could be shown.

The antagonist assay setup was optimised using the reference compound Propranolol. Cell density and an initial guess for incubation times were taken from the agonist assay setup as well as the concentration of stimulating agonist Isoprenaline at EC_{90} (80 nM). Incubation times for both agonist and antagonist incubation were optimised and chosen to be 10 min for each of these steps. To validate the performance of the antagonist assay setup, five literature known antagonists and inverse agonists of the β_2AR were tested additionally to Propranolol (Carvedilol, Bucindolol, ICI 118551, Labetalol, Compound 1[190]). All compounds showed antagonistic behaviour as was expected, however, a distinction between antagonists and inverse agonists was not possible.

All 22 molecules selected during the primary screen were tested in the agonist setup of the assay. Two of these molecules, compounds **1** and **2**, showed partial agonism in this assay (see Figure 4.3.A). Based on these results an analogue search of these two compounds was conducted and the five compounds selected from this were also subjected to this assay. Aside from compounds **1** and **2**, none of the other compounds showed any agonistic behaviour and they were subjected to the antagonist assay setup to evaluate whether they might have antagonistic properties. Only 10 of the 25 tested compounds showed weak antagonistic behaviour (see Figure 4.3.B).

It was then decided to do the further pharmacological characterisation of all 27 compounds in the lab of Prof. Jillian Baker in Nottingham with well-established binding and functional assays to allow more detailed conclusions about their behaviour and efficacy towards the β_2AR and also the β_1AR.

As was discovered during the measurements for a later project (which is not described in this thesis) the endogenous expression level of the β_2AR in HEK293T cells is quite low resulting in a lower assay sensitivity. This assay sensitivity can be increased by transfecting the cells with the β_2AR prior to the measurement and thereby enhancing the expression level.

^{3}H-CGP 12177 whole cell binding assay. The binding affinity of all 27 selected molecules to the β_2AR and the β_1AR was determined in CHO cells stably expressing either the human β_2AR (CHO-β_2) or the human β_1AR (CHO-β_1) using a whole cell radioligand displacement assay with the radioligand ^{3}H-CGP 12177 (see Table A4.1). Additionally, the binding affinity of seven literature known ligands of the β_2AR and the β_1AR was determined as a reference. As was expected, three of these reference compounds show a clear selectivity for one of the two receptors. CGP20712A

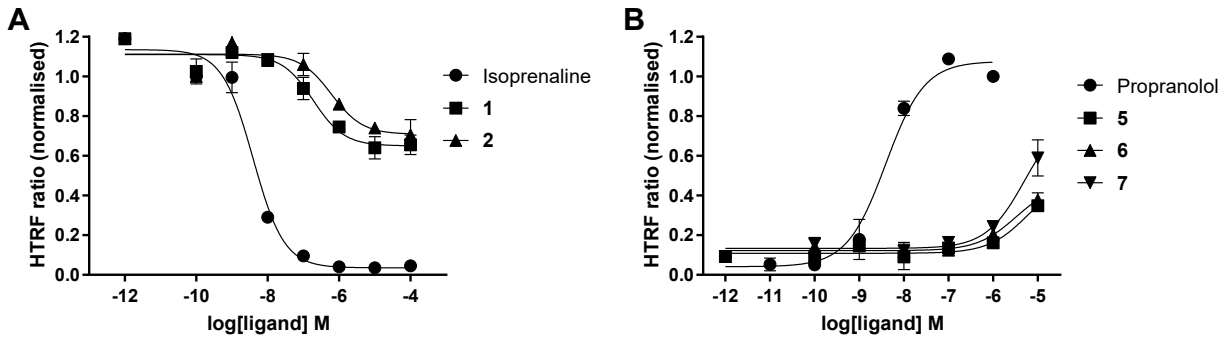


Figure 4.3: Representative dose-response curves resulting from the *Cisbio* cAMP accumulation assay. A) Agonist assay with Isoprenaline as reference agonist and compounds **1** and **2**. None of the other compounds selected from the docking calculations showed agonism in this assay. All responses were normalised to maximum Isoprenaline response. B) Antagonist assay with propranolol as reference antagonist and selected compounds. Cells were stimulated with 80 nM Isoprenaline. All responses were normalised to maximum Propranolol response. Curves are mean±SD of duplicates and represent A) two and B) one measurement. No further replicates were measured due to the subsequent and more detailed characterisation of all compounds at the University of Nottingham.

clearly shows a higher affinity for the β_1AR ($\log K_D$ = -8.6 at the β_1AR and $\log K_D$ = -5.8 at the β_2AR) while ICI 118551 has a higher affinity for the β_2AR ($\log K_D$ = -9.3 at the β_2AR and $\log K_D$ = -6.8 at the β_1AR) as well as Salmeterol ($\log K_D$ = -9.1 at the β_2AR and $\log K_D$ = -5.7 at the β_1AR). These results show that the cell lines contain either the β_2AR or the β_1AR but not both receptors and it proofs that a possible selectivity of the selected molecules towards either the β_2AR or the β_1AR could be detected. Of the 27 selected molecules, 10 show a measurable affinity towards the β_2AR and another six molecules show a weaker affinity which did not allow to calculate a K_D value. Affinities for the β_1AR are very similar to those for the β_2AR, only in a few cases and especially for the low affinity ligands slight differences in the K_D values at the β_2AR and the β_1AR are visible. Of the novel ligands compound **1** shows the highest affinity for the β_2AR and the β_1AR with a $\log K_D$ value of -6.3 and a K_D of 520 nM (see Figure 4.4.A and B), closely followed by compound **12** ($\log K_D$ = -6.2 at the β_1AR and $\log K_D$ = -6.1 at the β_2AR). The obtained results and K_D values can be found in Table A4.1.

CRE-SPAP production assay. A CRE-SPAP production assay was used to evaluate the ability of the compounds to stimulate a functional response. The assay was conducted using the same cell lines as for the binding assay. Both cell lines express the CRE-SPAP reporter gene stably. As briefly described in the methods section of this chapter, the readout of the functional response takes

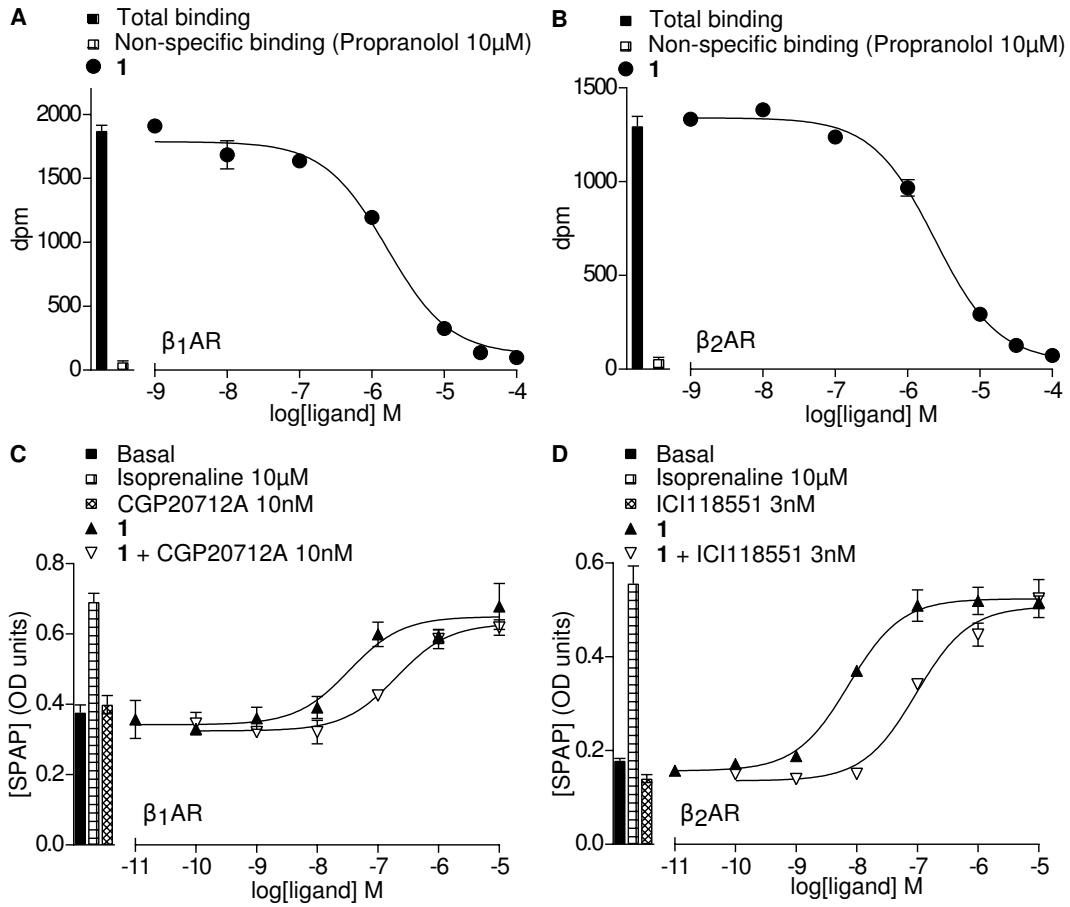


Figure 4.4: Representative dose-response curves for compound **1** in competition binding and CRE-SPAP assay. A) and B) Whole cell competition binding assay with ^{3}H-CGP 12177 in CHO-β_1 cells (A) and CHO-β_2 cells (B). Total and non-specific binding is represented by bars. In both experiments the radioligand concentration was 0.77 nM. C) and D) CRE-SPAP production assay in CHO-β_1 cells (C) and CHO-β_2 cells (D) in response to compound **1** and in absence and presence of CGP 20712A (C) and ICI 118551 (D). Basal CRE-SPAP production or production in response to 10 µM Isoprenaline or 10 nM CGP 20712A (C) or 3 nM ICI 118551 (D) are represented by bars. All data points are mean±S.E.M of triplicates. Shown dose-response curves are representatives of five (A), five (B), eight (C) and seven (D) separate experiments. OD, optical density. Figure was taken from the primary publication mentioned in the beginning of this chapter.[81]

place downstream of the receptor activation and after multiple amplification steps which allows a sensitive readout and detection of even weak partial agonists.[203] On the downside, even partial agonists may appear to be full agonists due to these amplification steps. Literature known partial agonists were included in this study to account for this known issue. The maximum possible response was obtained by measuring the response to a sufficiently high concentration of the full βAR agonist Isoprenaline (10 µM) on each plate.

Cimaterol, which is a relatively nonselective full agonist for both the β_2AR and the β_1AR shows a similar $\log EC_{50}$ value at both receptors ($\log EC_{50}$ = -8.6 at the β_1AR and $\log EC_{50}$ = -9.7 at the β_2AR) and induces a full response compared to the maximum possible response induced by Isoprenaline (see Table A4.2). As expected, the β_2AR selective agonist Salbutamol was more potent at the β_2AR than the β_1AR ($\log EC_{50}$ = -8.6 at the β_2AR and $\log EC_{50}$ = -6.5 at the β_1AR) and the known β_1AR selective agonist Denopamine induced a full agonist response at the β_1AR whereas it acted as a partial agonist at the β_2AR ($\%_{Isop} = 101$ at the β_1AR and $\%_{Isop} = 65$ at the β_2AR). CGP 12177 is a partial agonist at both receptors, however, the maximal response induced by this compound is much lower at the β_2AR than the β_1AR ($\%_{Isop} = 82$ at the β_1AR and $\%_{Isop} = 37$ at the β_2AR).

All 27 molecules were subjected to the CRE-SPAP production assay in both the CHO-β_1 and CHO-β_2 cell line. As expected, none of the molecules that did not show any affinity for either of the βARs did induce any response as did most of the ligands showing very low affinities. Only two molecules of the latter group did induce weak responses but it was not possible to calculate any EC_{50} values for them. All 10 compounds that showed a measurable affinity to the β_2AR induced agonist responses. The most potent ligand at both receptors was again compound **1** (see Figure 4.4.C and D; Table A4.2). To ensure that all agonist responses indeed occurred via the expressed βAR, all compounds were also measured using the CHO cell line expressing the CRE-SPAP reporter gene but none of the βARs in the same assay setup. As was expected, neither the novel ligands nor the reference agonists showed any response in these control assays.

To investigate whether the novel ligands indeed act via the orthosteric binding pocket, the agonist response was inhibited by the selective antagonists CGP20712A and ICI 118551 in the CHO-β_1 and CHO-β_2, respectively. From the right shift of the dose-response curves the K_D values of the antagonists were calculated via the Gaddum equation and then compared with the K_D values from the radioligand displacement assay. These values should be approximately the same if the agonist acts from the orthosteric binding pocket. In the CHO-β_2 cells, the agonist response of the reference agonists Cimaterol, Salbutamol, Salmeterol, Denopmaine and CGP 12177 were inhibited by ICI 118551, resulting in a similar K_D value for this antagonist as obtained from the binding assay

($\log K_D$ of ICI 118551 is -9.1 to -9.8 from the curve shifts of the agonists and $\log K_D$ = -9.3 from the binding experiments; see Table A4.1 and Table A4.2). Similar high K_D values were obtained for nine of the compounds with measurable affinity (see Table A4.2). These results show, that the novel agonists induce their responses via the orthosteric binding pocket of the β_2AR and via the same receptor conformation. Compound **8** induces only a partial response at the β_2AR. Although ICI 118551 also induces a right shift of the curve for **8**, the shifted curve flattens and does not reach the same maximum in the presence of ICI 118551 as it does without. Therefore, a K_D value using the Gaddum equation could not be calculated.

Similar results were obtained in the CHO-β_1AR cells in which the agonist response was inhibited by the antagonist CGP 20712A. The K_D values of CGP 20712A calculated from the curve shift of reference agonists Cimaterol, Salbutamol, Salmeterol and Denopmaine as well as for seven of the novel agonists were similar to the values obtained from the binding assay ($\log K_D$ of CGP 20712A is -8.9 to -9.2 from the curve shifts of the reference agonists and -8.6 to -9.4 for the novel compounds compared to $\log K_D$ = -8.6 from the binding assay; see Table A4.1 and Table A4.2). These results show that the novel compounds are also acting via the orthosteric binding site of the β_1AR. Compounds **3, 4** and **8** are (weak) partial agonists at the β_1AR and, therefore, no K_D values of CGP 20712A could be calculated from the shifted curves. Only reference agonist CGP 12177 shows a significantly different behaviour resulting in a lower K_D value for CGP 20712A calculated from the shifted curve than from the binding assay ($\log K_D$ = -7.0 from the shifted curve compared to $\log K_D$ = -8.6 from the binding assay; see Table A4.1 and Table A4.2)). These results were expected as it has been shown in previous studies that the CGP 12177 agonist response is occurring via a secondary conformation of the β_1AR.[204]

4.3.3 Hit rates and further results

Hit rates. In total, 27 molecules were subjected to the pharmacological characterisation. Of these molecules, 22 molecules were chosen from the primary screen and another 5 molecules from the similarity search to compounds **1** and **2**. Of the 27 tested molecules, 10 molecules (37%) showed measurable affinity to the β_2AR (see Table A4.1). Another 6 molecules (22%) might be weak ligands of the β_2AR, however, it was not possible to calculate K_D values for these molecules. All of the 10 confirmed ligands of the β_2AR showed agonistic behaviour in the functional assay, leading to a hit rate of 100% agonists among the found ligands and an overall hit rate of 37% among all tested 27 molecules (see Table A4.2). Of the found agonists, 2 were selected from the similarity search, resulting in a hit rate of 2 of 5 molecules or 40% of this similarity search. The other 8

molecules originated from the primary screen with 22 selected molecules (8 of 22 molecules or 36%). While the majority of the 8 hit molecules was selected from the dual re-ranking (7 of 22 molecules or 32%), 1 molecule was selected from the screen against $\beta_2AR^{inactive}$ 2RH1 (1/22 or 5%). The overall hit rate of 37% is at the upper end compared to earlier docking screens against aminergic receptors.[190,193–198]

Novelty of ligands. To evaluate the novelty of the discovered hit molecules as ligands of the β_2AR, the novel ligands were compared to all ligands showing affinity to or acting at the β_2AR or any adrenergic receptor as listed in ChEMBL.[192] The most potent compound **1** shows some similarity to known β_2AR agonists since it contains the same β-hydroxyl-amine, however, **1** also contains a pyrazole moiety, which it shares with six of the other novel agonists discovered during the primary docking screen. So far, molecules featuring such a pyrazole moiety have not been described as ligands of the β_2AR. Hence, the Tanimoto similarity values between the novel agonists and the set of known actives of the β_2AR from ChEMBL are low, with the highest calculated similarity value of 0.47 (**8** to CHEMBL599896) and only four molecules with a similarity > 0.4 (see Table A4.4). None of the found most similar compounds contained the same structural motif as the query molecule of the novel ligands. To obtain a better impression of the novelty of the ligands, they were also compared to the actives against *any* of adrenergic receptor subtypes. Again, a similar result was obtained with the highest similarity values of 0.49 (low-affinity **14** to CHEMBL15303) or 0.47 (**8** to CHEMBL599896) and only seven molecules with a similarity > 0.4 to a molecule in the ChEMBL dataset (see Table A4.5).

Additional to these similarity calculations, the ChEMBL database was searched for molecules containg the basic substructure of phenyl, β-hydroxyl-amine and pyrazole. None of the found molecules was linked to any bioactivity data regarding the β_2AR. Compound **7** has previously been described as a $\varkappa$-opioid receptor antagonist ($pK_i = 5.19$), however, never as a ligand for the β_2AR.[205,206]

Predicted binding poses. The β-hydroxyl-amine of compound **1** interacts with residues $D113^{3.32}$ and $N312^{7.39}$ in the predicted binding pose of the molecule (Figure 4.5.A). Other interactions between ligand and protein might be contributed by the fluorine substituents attached in *ortho* position at the aromatic ring and potentially residues $N293^{6.55}$ or $Y308^{7.35}$, which might explain its affinity and agonist activity.

Compound **8** was the only of the tested molecules that was discovered from a docking calculation to a $\beta_2AR^{inactive}$ and it also has an overall bulkier structure compared to the other discovered ligands, with an additional aromatic moiety linked via an aliphatic chain to the more standard benzene ring following the β-hydroxy-amine. This compound **8** adopts a binding pose in the larger binding pocket of the $\beta_2AR^{inactive}$ where it extends the alkyl-aromatic moiety towards the extracellular loop

2 (ECL2) while interacting with residues N293$^{6.55}$, D113$^{3.32}$ and N312$^{7.39}$ (Figure 4.5.B). However, this binding pose could not be reproduced in a docking calculation to the much narrower binding pocket of the β_2AR^{active}. Interestingly, the affinity of this ligand for the β_2AR is only slightly lower than for the highest affinity ligand **1** (logK_D(**8**) = -6.1 and logK_D(**1**) = -6.3 at the β_2AR). However, compound **8** acts only as a partial agonist at the β_2AR (35% of the maximum response induced by Isoprenaline; this is similar to the CGP 12177 response at 37%), which might explain the counterintuitive binding preferences observed from the docking calculations. Similar to CGP 12177, compound **8** can be considered a weak partial agonist due to amplification of the response in the CRE-SPAP system. Of note, compound **8** has no other moieties to interact with any of the serine residues in TM V than the weak hydrogen bond acceptor ether oxygen.

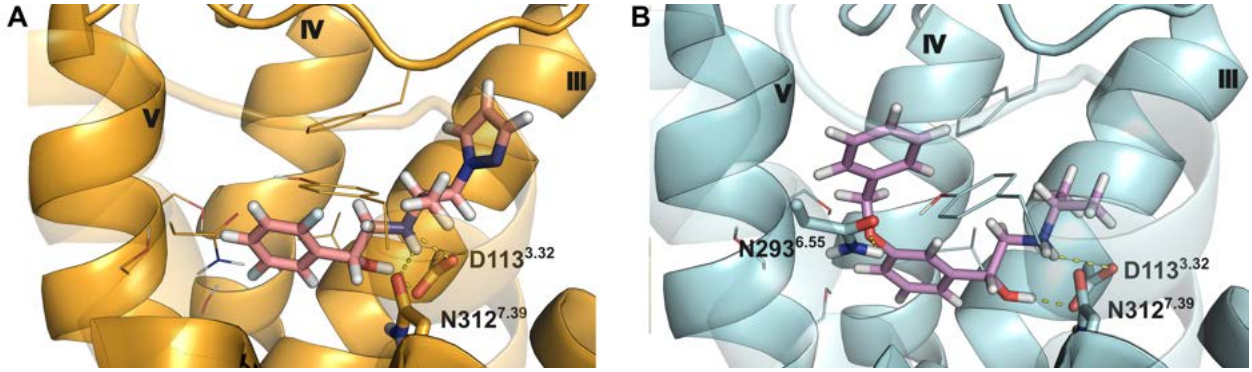


Figure 4.5: Binding poses of A) compound **1** from the docking calculation to β_2AR^{active} 3SN6 and B) compound **8** from the docking calculation to β_2AR^{inactive} 3NY9. While compound **1** interacts with D113$^{3.32}$ and N312$^{7.39}$, compound **8** forms an additional interaction with N293$^{6.55}$ (interacting residues highlighted in stick representation). TMs III, IV and V are labelled on the top of each helix. TM VI and VII would be located between the observer and the ligand and are shown in transparent. Hydrogen bonds are shown as yellow dashed linaes.

4.3.4 Evaluation of docking schemes based on retrospective ligand enrichment

As described in the previous sections, the ligands were chosen from four different docking and re-ranking schemes. However, there is a large variation in the number of molecules chosen from each scheme. To evaluate whether these variations could have been predicted, retrospective ligand enrichments using the different schemes were investigated using the same set of literature known agonists and antagonists that was used for the optimisation of the docking setup (see also section 8.1.4 for compilation of the ligand set). This set of 125 literature known ligand enantiomers as well as their decoys was docked to the two β_2AR^{active} structures and to the β_2AR^{inactive} structure

3NY9 using the same docking setups as for the screens. The resulting ranking lists were then also re-ranked the same way as done for the screens, i.e. by doing a dual re-ranking of the ranking lists from the docking calculations to 3SN6 and 4LDL and a selective re-ranking towards 3SN6 over 3NY9 (cf. section 8.1.2). ROC plots of each ranking list resulting from the different docking calculations and the re-ranking schemes were then created to evaluate enrichment of known agonists and/or antagonists over decoys (see Figure 4.6). For each scheme three different cases were evaluated: (1) treating all agonists and antagonists as actives, (2) treating only agonists as actives while antagonist were treated as decoys and (3) treating only antagonists as actives while agonists were treated as decoys. Approaches (2) and (3) were thought as a measure of agonist over anatgonist enrichment and vice versa to correlate to the activity prediction pursued in this study. The number of agonists and antagonists/ inverse agonists within the set of literature known ligands is approximately the same.

The results show, that the early and overall enrichment of agonists is approximately the same for both docking calculations to the β_2AR^{active} and also for the dual and selective re-ranking lists, and better than for antagonists. For the selective re-ranking, the overall enrichment of antagonists is below random, however, with a similar early enrichment as observed for the docking calculations to the β_2AR^{active}. For the dual re-ranking a similar overall enrichment for all three cases, i.e. either all ligands or only agonists or only antagonists as actives, to the enrichment in the three other schemes is observed. Interestingly though, no early enrichment of antagonists occurres and the first antagonist was only enriched after approximately 6% of the decoys were found. Transferring this to a screening of a library with 3.6 million molecules like the one used in this study, the first antagonists would be found after the top 200,000 ranked molecules, meaning that each ligand found in the top 200,000 ranked molecules should be an agonist. Although these calculations surely overestimate the true numbers, this observation is in line with the results and hit rates achieved during the docking screen in this study.

4.4 Discussion

In this study we investigated whether docking calculations to a certain conformation of a receptor would preferentially result in novel ligands with a certain efficacy and which docking strategy would yield the best hit rates. In specific, a molecule library was used in a comparative docking approach to active (and inactive) conformations of the β_2AR to see whether this would more likely result in novel ligands with agonistic properties, rather than just antagonists. Furthermore, the impact of docking calculations to multiple structures of the β_2AR was evaluated and how to best combine the rankings stemming from these individual docking calculations.

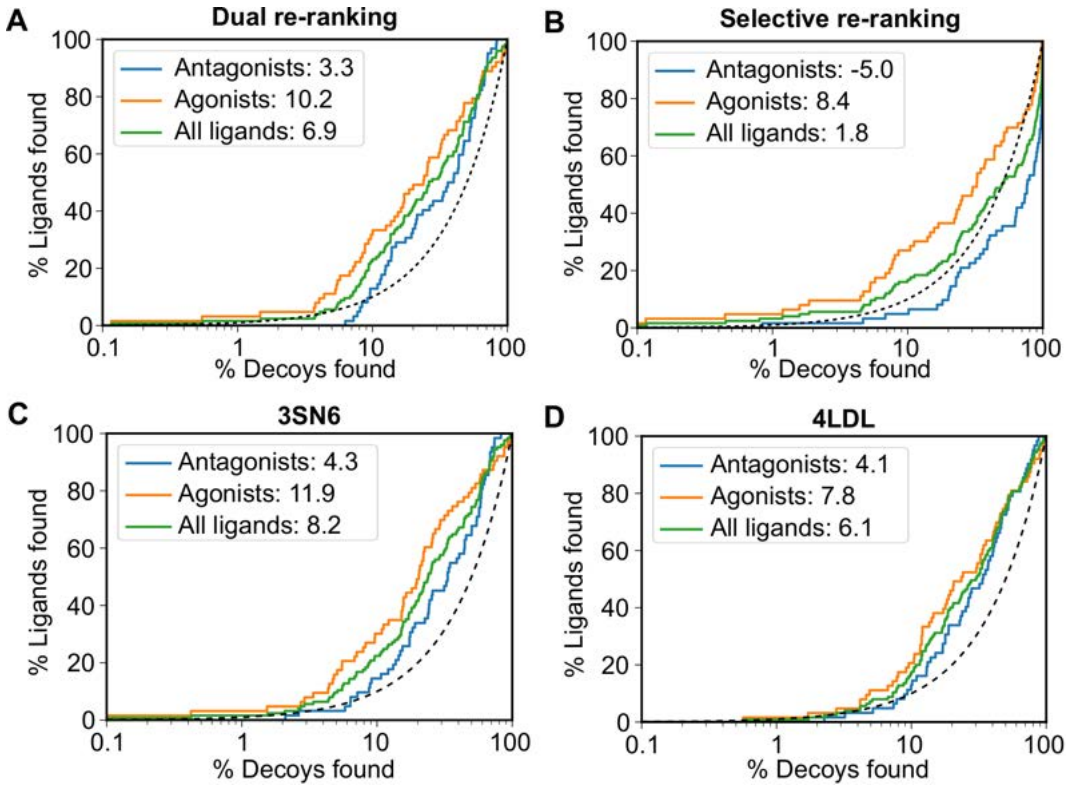


Figure 4.6: ROC plots of different (re-)ranking lists after docking calculations of a set of known β_2AR ligands and their decoys to the β_2AR^{active} and a $\beta_2AR^{inactive}$. A) Dual re-ranking of the ranking lists resulting from the docking calculations to β_2AR^{active} 3SN6 and 4LDL, B)Selective re-ranking of the ranking lists of the docking calculations to β_2AR^{active} 3SN6 over $\beta_2AR^{inactive}$ 3NY9, and ranking lists from the docking calculations to β_2AR^{active} 3SN6 (C) and 4LDL (D). The enrichment of actives over decoys was calculated, either by treating only agonists (orange) or antagonists (blue) or all ligands (green) as actives. As can be seen in A) no antagonists were enriched in the first 6% of the ranking list after dual re-ranking of two docking calculation to the β_2AR^{active}. Dotted lines represent random enrichment. Numbers represent the area under the curve.

Pharmacological evaluation. In total, 27 compounds were chosen from the docking screens and evaluated pharmacologically. Of these compounds, 10 molecules (37%) showed a measurable affinity to the β_2AR. Excitingly, all of these 10 novel ligands acted as agonists at the β_2AR, which also explains their relatively low affinity, a factor that is observed for many agonists.[207] Most of these novel ligands also act as agonists at the β_1AR which was targeted in parallel using the same assay setup as for the β_2AR to evaluate the selectivity of the compounds. It was demonstrated that all compounds act via the orthosteric binding pocket of the β_2AR and the β_1AR, since the induced responses could be inhibited by literature known selective antagonists and no responses were observed after stimulation of the parent CHO-CRE-SPAP cell line without the transfected receptor.

Since there is no direct pharmacological measure of the intrinsic efficacy of ligands, i.e. the ability of a ligand to activate the receptor, it is difficult to rank them according to this ability (in contrast to the direct measure of affinity of the ligand for a receptor which allows such a ranking). A simple way to circumvent this problem and compare intrinsic efficacies of compounds is to use their efficacy ratio. This ratio compares the EC_{50} and K_D values of the same compound and can be used if all compounds were examined in parallel under identical conditions.[207] After sorting all measured compounds according to this efficacy ratio, it can be seen that Cimaterol has the highest intrinsic efficacy at both β_2AR and β_1AR (see Table A4.2), i.e. it can activate the receptor to a degree which induces a full response even without activating all available receptors. As expected, selectivity was observed for Denopamine, which had a higher intrinsic efficacy at the β_1AR than the β_2AR, and Salbutamol, which had a higher intrinsic activity at the β_2AR compared to the β_1AR.[207] Comparing the novel agonists, compound **1** had the highest intrinsic efficacy at both β_2AR and β_1AR followed by compound **2**, while none of the compounds derived from the similarity search displayed a higher intrinsic efficacy than the parent compounds. The most potent novel agonist **1** has a comparable intrinsic efficacy to Salmeterol, a long-acting β_2AR agonist used for treatment of chronic obstructive pulmonary disease (COPD) and asthma. This shows, that the comparative docking approach used in this study was able to predict novel agonists with sufficient agonism to be potentially useful for clinical purposes. Interestingly, however, this approach did not yield compounds with high efficacies (such as catecholamines, Fenoterol, or even Cimaterol or Salbutamol[207]), but only compounds with medium efficacies.

Molecule selection and comparative docking. To obtain potential ligands with agonist properties from the ranking lists of the different docking calculations, various approaches were used. Each of these strategies can be evaluated with regard to the number of retrieved agonists. Initially, it was assumed that agonists would rank higher in the β_2AR^{active} than in the $\beta_2AR^{inactive}$, which was

confirmed by retrospective enrichment studies. Additionally to using the ranking lists as obtained from the docking, the lists were re-ranked either by dual re-ranking of the lists from the docking calculations to the two β_2AR^{active} or by selective re-ranking of the lists from one β_2AR^{active} compared to one $\beta_2AR^{inactive}$. Surprisingly, the latter did not yield any molecules that were considered worth a pharmacological characterisation, but only two smaller derivatives of compounds emerging from this ranking. Almost all of the other tested compounds were selected from the rankings of the individual dockings to the two β_2AR^{active} with PDB ID 3SN6 and 4LDL or from the dual re-ranking of the ranking lists of these two dockings. The dual re-ranking did not only yield the highest number of molecules with favourable interactions but also the highest hit rate. These results suggest that the most successful approach to retrieve agonists is by docking to more than one structure in an active conformation and combining the results by re-ranking to obtain the molecules that are ranked most highly in both individual rankings. However, the opposite approach to compare docking calculations to β_2AR^{active} and $\beta_2AR^{inactive}$ does not yield a high agonist hit rate or even number of molecules considered worth testing, which is somewhat in contrast to the initial hypothesis and earlier studies.[84,194]

This observation was further evaluated by comparing it to the results of retrospective docking calculations of known ligands of the β_2AR to the different receptor conformations. The relative enrichment of these known β_2AR agonists and antagonists over computer generated decoy molecules reflects the observations from the docking screen. Furthermore, retrospective enrichment calculations might be a useful tool to indicate which ranking or re-ranking scheme might be the most successful.

Comparing the molecule poses resulting from the docking calculations to the β_2AR^{active} and the $\beta_2AR^{inactive}$ leads to the observation that the orientation of the molecules in both is rather similar, however, with an on average higher number of favourable interactions between molecule and receptor for the β_2AR^{active}. We suggest that this could be used as an additional criterion for the evaluation of docking poses. The pose of an agonist obtained from a docking calculation to the β_2AR^{active} should have a similar orientation in a docking calculation to the $\beta_2AR^{inactive}$, since the bigger pocket of the $\beta_2AR^{inactive}$ should leave enough space for the agonist to take this orientation (see also Figure 4.1). The interaction patterns, however, should allow a differentiation.

One of the agonists from the primary screen was chosen from the ranking list of a docking calculation to the $\beta_2AR^{inactive}$ with PDB ID 2RH1 (**8**). This compound showed a significantly lower agonist efficacy than the other novel agonists. Compound **8** does not find a favourable pose in the β_2AR^{active} while it also is a partial agonist, which suggests that, in certain cases, the rigid conformation used during docking calculations can prevent the discovery of bulky ligands from

docking calculations to the β_2AR^{active}. In that case, it might be helpful to do a complementary docking to the $\beta_2AR^{inactive}$, which, on the other hand, might introduce the risk of incorrect predictions, because, statistically, a favourable rank in the $\beta_2AR^{inactive}$ but not the β_2AR^{active} would be an indicator for an antagonist or an inverse agonist. Nonetheless, this discovery of a ligand with agonistic properties using the $\beta_2AR^{inactive}$ can be treated as an exception, as is also supported by the fact that earlier docking studies to the $\beta_2AR^{inactive}$ did not yield any ligands with agonistic properties.[190,195]

Structure-activity relationship (SAR). Since several of the novel agonists (**1**-**7**, **11**, **12**) have similar chemical structures and some even only differ by one substituent, it is possible to draw some conclusions about the SAR. Most of the agonists feature different fluorine substitution patterns at the aromatic moiety closest to the hydroxyl-group of the β-hydroxyl-amine. Adding a fluorine substituent in *para* position did not change affinity or efficacy (**6** compared to **5**). However, adding an additional fluorine substituent in *meta* position leads to an increased binding affinity ($\log K_D(\mathbf{1}) = -6.3$ compared to $\log K_D(\mathbf{2}) = -5.9$ at the β_2AR) and, as expected, also a higher potency, while the intrinsic efficacy remains unchanged. Moving the fluorine substituent in *ortho* position (**1**) to *para* position (**12**) did not effect the affinity but did lead to a reduction of the efficacy ratio (especially at the β_2AR with a shift from 1.76 to 1.05; see Table A4.2). While exchanging the fluorine substituent in *ortho* position to a chlorine (**3** compared with **4**) does barely influence affinity or pEC_{50} values, it results in significantly higher percentage of maximum response for the β_2AR (79% for **4** to 95% for **3**), whereas the percentage maximum response is lowered at the β_1AR (60% for **4** to 33% for **3**). This increase can likely be explained by the ability of chlorine to form stronger interactions compared to fluorine, possibly resulting in a stabilisation of a more active conformation of the receptor. When considering the difference in response at the β_2AR and the β_1AR, an interaction of this *ortho* substituent with $Y308^{7.35}$ in TM VII, which is only present in the β_2AR but not the β_1AR, seems most likely (distances of 2.5-4 Å between the *ortho* substituents of the various ligands and the hydroxyl group of $Y308^{7.35}$ in the respective energy minimised docking poses).
Finally, the results for compounds **2** and **14** indicate, that binding is abolished when removing the hydroxyl group and introducing an additional ether oxygen between the aromatic moiety and the secondary amine.

4.5 Conclusions

The goal of this study was to investigate whether the conformational state of a receptor structure used for a docking screen would allow the prediction whether a proposed ligand would act as an agonist or antagonist at that receptor. Indeed, the results indicate, that a multiconformation

docking screen might be a productive strategy to identify novel ligands with agonism comparable to that of clinically used drugs. Although the results from docking calculations against the $\beta_2AR^{inactive}$ were not essential in terms of the ranks of the molecules, inspection of the obtained molecule poses of particular molecules could often help deciding whether this molecule was likely to be an agonist or not. The conformational information contained in a β_2AR^{active} seems to be sufficient to enrich agonists, i.e. "function follows form". Furthermore, a dual re-ranking considering the ranking lists of docking calculations to two β_2AR^{active} led to the largest number of agonists, showing that this might be a helpful strategy to enrich agonists even better in the top ranks and that new structures add new information that might be worthwhile considering and combining.

We were able to identify novel agonists for the β_2AR which show intrinsic efficacies comparable to those of clincally used drugs and also feature a for ligands of the β_2AR previously undescribed pyrazole moiety. Several derivatives of these molecules were tested throughout the study, allowing insights into their structure-activity relationship. With this multiconformation docking approach it was once more possible to obtain ligands with a novel chemotype even for such a well explored receptor as the β_2AR.[190,193–195]

The insights gained during this study should be transferable to other class A GPCRs due to their general similarities. However, it might be more difficult to apply this approach to other members of the class A subfamily due to the lack of available crystal structures in the different activation states. At the rate in which novel receptor structures are currently published, this problem might disappear soon though, allowing to test this strategy on other GPCRs. Additionally, homology models might also contain the necessary conformational information, since there are also more suitable templates available with the increasing number of receptor structures.

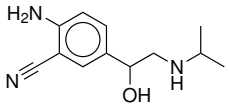
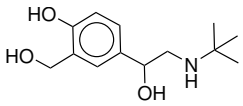
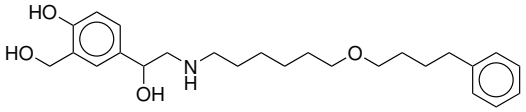
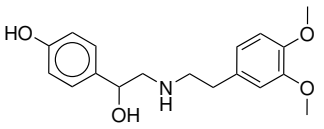
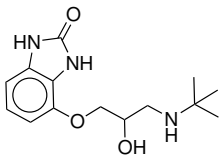
4.6 Perspectives

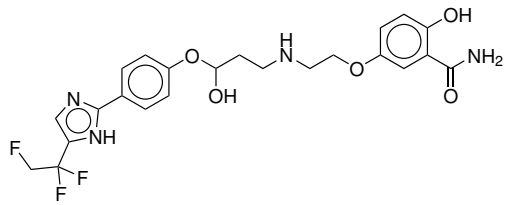
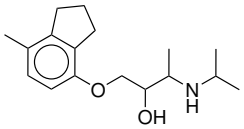
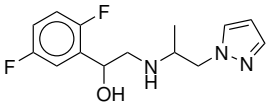
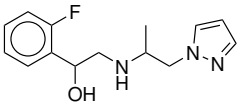
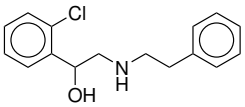
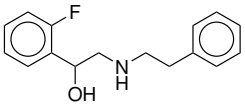
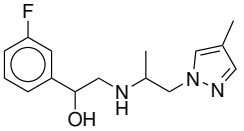
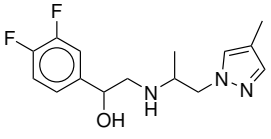
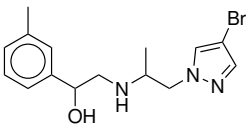
The results from this study could serve as a basis and starting point for future studies. We found out that "form defines function", i.e. that agonists can be predicted using a receptor structure in an active conformation. Furthermore, using a dual re-ranking comparing the ranking lists from docking calculations to two structures in active conformations seems to have a beneficial effect in the search for agonists. Although these results should be transferable to other class A GPCRs, it could be worthwhile to test whether this holds true for other aminergic receptors and also non-aminergic receptors. To overcome the problem of missing crystal structures, receptor structures in certain activation states could be created by homology modelling or by applying molecular dynamics (MD) simulations of the receptor in complex with an effector protein. Different clusters from these MD simulations could then be used as different activation states of the receptor and ranking lists from

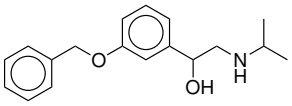
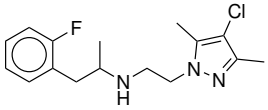
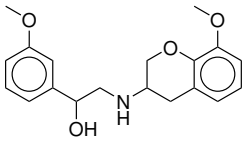
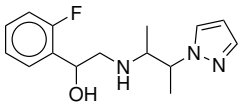
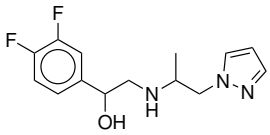
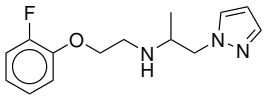
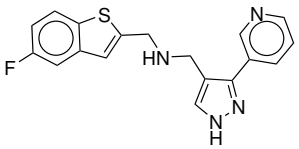
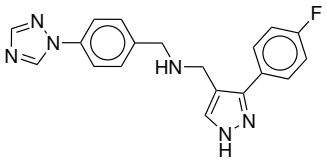
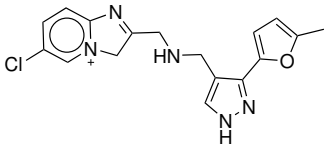
docking calculations to these different cluster structures could be used to enrich more agonists in the top ranks. Advancing this even more, the comparison of ranking lists from docking calculations to receptor conformations from MD simulations with different effector proteins might even be helpful for the prediction of ligands with a bias towards a certain signalling pathway. However, predicting biased molecules is even more challenging and might not be solvable with this approach.
Additionally, the novel ligands discovered during this project could be further developed. The structure-activity relationship data collected during this study can be used to optimise the ligands towards e.g. a higher intrinsic activity.

4.7 Additional information

Table A4.1: Affinities of the novel compounds discovered during this study and seven reference compounds for the β_1AR and the β_2AR. $\log K_D$ values were determined from ^{3}H-CGP 12177 whole cell binding (mean ±SEM of n separate experiments). The reference compounds are listed in the upper part and the novel ligands in the lower part in order of efficacy ratio at the β_2AR (a measure of intrinsic efficacy, see text and Table A4.2).

		β_1AR		β_2AR	
Compound	Structure	$\log K_D$	n	$\log K_D$	n
Literature known reference compounds.					
Cimaterol		-6.4±0.1	3	-7.1±0.1	3
Salbutamol		-5.0±0.1	5	-6.3±0.1	5
Salmeterol		-5.7±0.0	7	-9.1±0.0	7
Denopamine		-6.0±0.1	5	-5.3±0.1	5
CGP 12177		-9.4±0.0	3	-9.6±0.1	3

CGP 20712A		-8.6±0.2	6	-5.8±0.1	8
ICI 118551		-6.8±0.1	7	-9.3±0.1	7
Compounds chosen from docking calculations.					
1		-6.3±0.0	5	-6.3±0.1	5
2		-6.0±0.0	5	-5.9±0.1	5
3		-5.0±0.1	5	-5.6±0.0	5
4		-5.5±0.1	5	-5.9±0.1	5
5		-5.6±0.0	5	-5.4±0.1	5
6		-5.8±0.0	6	-5.7±0.0	5
7		-5.8±0.0	5	-5.9±0.0	5

8		-5.0±0.0	6	-6.1±0.1	6
9		-5.0±0.1	5	$IC_{50} > -4$	5
10		no binding	5	$IC_{50} > -4$	5
11[a]		-5.3±0.0	5	-5.3±0.0	5
12[a]		-6.2±0.0	6	-6.1±0.1	6
14[a]		$IC_{50} > -4$	6	$IC_{50} > -4$	6
22		$IC_{50} > -4$	5	$IC_{50} > -4$	6
25		no binding	5	$IC_{50} > -4$	5
26		no binding	5	$IC_{50} > -4$	5

a Molecules were selected from the secondary screen.

Table A4.2: Agonist responses of the reference compounds and the novel agonists. log EC_{50} values and % 10 µM Isoprenaline responses in n separate experiments from a CRE-SPAP gene transcription assay in CHO-β_1 and CHO-β_2 cells are given. Additionally, agonist responses were determined in the presence of CGP 20712A (β_1AR) or ICI 118551 (β_2AR) and log K_D values for the antagonist calculated where a parallel shift was achieved.

	β_1AR					β_2AR					log(efficacy ratio)	
Compound	log EC_{50}	% isop	n	log K_DCGP 20712A	n	log EC_{50}	% isop	n	log K_DICI 118551	n	β_1AR	β_2AR
Literature known reference compounds.												
Cimaterol	-8.6±0.1	100±9	12	-9.0±0.1	20	-9.7±0.1	98±6	12	-9.7±0.1	23	2.21	2.59
Salbutamol	-6.5±0.2	109±9	7	-9.0±0.2	6	-8.6±0.1	95±6	6	-9.8±0.1	7	1.55	2.39
Salmeterol	-6.8±0.1	110±9	7	-8.9±0.3	4	-10.9±0.2	100±7	8	-9.1±0.0	6	1.02	1.81
Denopamine	-8.1±0.1	101±4	7	-9.2±0.3	4	-6.4±0.1	65±6	7	-9.6±0.1	4	2.13	1.02
CGP 12177	-8.5±0.1	82±7	13	-7.0±0.1	7	-9.7±0.1	37±4	10	-9.6±0.1	14	-0.92	0.16

Table is continued on the following page.

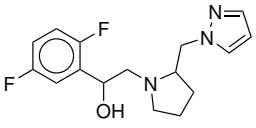
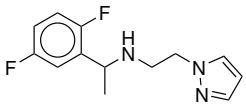
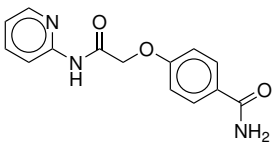
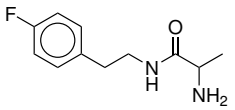
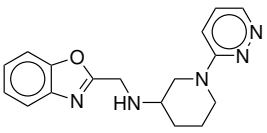
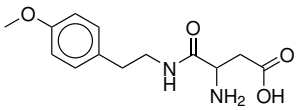
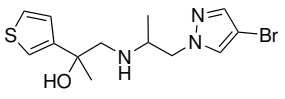
Continuation of Table A4.2

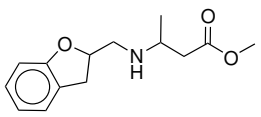
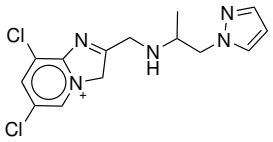
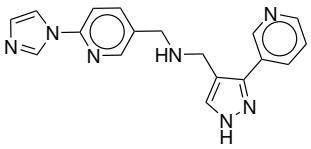
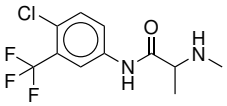
	β_1AR					β_2AR					log(efficacy ratio)	
Compound	log EC_{50}	% isop	n	log K_DCGP 20712A	n	log EC_{50}	% isop	n	log K_DICI 118551	n	β_1AR	β_2AR
Compounds chosen from docking calculations.												
1	-7.5±0.1	80±4	8	-8.6±0.1	7	-8.0±0.1	90±6	7	-9.6±0.1	6	1.2	1.76
2	-7.2±0.1	82±6	7	-8.6±0.2	5	-7.5±0.1	97±4	7	-9.5±0.1	6	1.22	1.64
3	-6.3±0.1	33±6	6			-7.0±0.1	95±8	6	-9.6±0.1	5	1.27	1.39
4	-6.7±0.1	60±6	7			-7.1±0.1	79±4	7	-9.5±0.1	6	1.25	1.24
11[a]	-6.6±0.1	113±7	6	-9.4±0.3	3	-6.5±0.1	92±9	6	-9.7±0.2	5	1.35	1.23
5	-7.0±0.2	85±7	6	-9.1±0.3	4	-6.5±0.1	86±11	6	-9.6±0.1	5	1.42	1.19
12[a]	-7.3±0.1	78±6	7	-9.3±0.1	5	-7.2±0.1	96±9	6	-9.7±0.1	6	1.10	1.05
6	-6.8±0.1	82±9	6	-9.0±0.2	3	-6.6±0.1	75±8	6	-9.8±0.1	5	0.92	0.89
7	-6.9±0.0	90±5	6	-9.1±0.2	3	-6.6±0.1	86±6	7	-9.6±0.1	6	1.09	0.73
8	$10\,\mu M$ = 21±7		8			-6.2±0.1	35±7	6				
9	$10\,\mu M$ = 30±12		7			$10\,\mu M$ = 29±4		7				
10	no response		6			$10\,\mu M$ = 50±8		7				

a Molecules were selected from the secondary screen.

No response was seen for CGP 20712A, ICI 118551 or the docking screen compounds **13**[a], **14**[a], **15**[a], **16**, **17**, **18**, **19**, **20**, **21**, **22**, **23**, **24**, **25**, **26**, **27** (n=5 for each ligand at each receptor).

Table A4.3: Compounds that were selected from the docking calculations and did not show affinity to either the β_1AR or the β_2AR. None of these molecules did induce agonist activity, either.

		β_1AR		β_2AR	
Compound	Structure	$\log K_D$	n	$\log K_D$	n
13[a]		no binding	5	no binding	5
15[a]		no binding	5	no binding	5
16		no binding	5	no binding	5
17		no binding	5	no binding	5
18		no binding	5	no binding	5
19		no binding	5	no binding	5
20		no binding	5	no binding	5

21		no binding	5	no binding	5
23		no binding	5	no binding	5
24		no binding	5	no binding	5
27		no binding	5	no binding	5

a Molecules were selected from the secondary screen.

Table A4.4: Novelty of the tested molecules. The ECFP4 Tanimoto similarity between the molecules selected from the docking screens and all molecules listed with affinity to or activity at the β_2AR in ChEMBL were calculated. ChEMBL ID and SMILES of the most similar molecule as well as the calculated ECFP4 Tanimoto coefficient are listed. If several molecules with the same smilarity were found the first entry is shown.

Mol	ChEMBL ID	Smiles	Tanimoto
1	CHEMBL16476	CC(C)NCC(O)c1ccc2ccccc2c1	0.25
2	CHEMBL1902627	CC(C)NCC(O)c1ccccc1Cl	0.31
3	CHEMBL599896	Cl.CC(CCc1ccccc1)NCC(O)c2cccc(c2)[N+](=O)[O-]	0.41
4	CHEMBL599896	Cl.CC(CCc1ccccc1)NCC(O)c2cccc(c2)[N+](=O)[O-]	0.41
5	CHEMBL327122	CC(C)NCC(O)c1cccc(O)c1	0.32
6	CHEMBL321468	CC(C)NCC(O)c1ccc(O)c(CO)c1	0.28
7	CHEMBL327122	CC(C)NCC(O)c1cccc(O)c1	0.32
8	CHEMBL599896	Cl.CC(CCc1ccccc1)NCC(O)c2cccc(c2)[N+](=O)[O-]	0.47
9	CHEMBL1159723	CCCc1ccccc1OCC(O)CNC(C)C	0.26
10	CHEMBL26183	COc1cccc2c(C[C@@H](C)NC[C@H](O)c3cccc(Cl)c3)c[nH]c12	0.36
11	CHEMBL1902627	CC(C)NCC(O)c1ccccc1Cl	0.45
12	CHEMBL16476	CC(C)NCC(O)c1ccc2ccccc2c1	0.39
13	CHEMBL1723653	Cc1ncc(CN2CCCC(C2)C(=O)Nc3ccc(cc3)c4cccc(F)c4)s1	0.24
14	CHEMBL1626224	CC(O)CNCCOc1ccccc1c2ccccc2	0.38
15	CHEMBL16476	CC(C)NCC(O)c1ccc2ccccc2c1	0.24
16	CHEMBL1577023	Cc1ccc(OCC(=O)Nc2ccc(cc2)c3oc4cccnc4n3)cc1	0.39
17	CHEMBL24	CC(C)NCC(O)COc1ccc(CC(=O)N)cc1	0.31
18	CHEMBL2441621	C1CN(CCN1)c2ccnc3ccccc23	0.30
19	CHEMBL3099658	COc1ccc(CCNC[C@@H](O)c2cc(O)cc(O)c2)cc1	0.37
20	CHEMBL1159717	CC(C)(C)NCC(O)c1ccccc1Cl	0.18
21	CHEMBL1200586	Cl.CCCNC(C)C(=O)Nc1ccccc1C	0.28
22	CHEMBL1573444	CN(Cc1ccccc1F)c2nc(nc3ccccc23)c4cccnc4	0.31
23	CHEMBL531401	CC(Oc1ccc(Cl)cc1Cl)c2onc(n2)c3ccc(NC(=O)c4occc4)cc3	0.21
24	CHEMBL3099899	COc1ccc(CNC(=O)c2cc(cnc2c3cccnc3)c4cc(Cl)cc(Cl)c4)cc1OC	0.30
25	CHEMBL305558	O[C@@H](CNCCc1ccc(NS(=O)(=O)c2ccc(cc2)n3ncc(n3)c4ccc(F)cc4)cc1)c5cccnc5	0.33
26	CHEMBL252766	CSc1ccc2nc(cn2c1)c3ccc(cc3)N(C)C	0.27
27	CHEMBL1439691	CCN(CC)S(=O)(=O)c1cc(NC(=O)C2=COCCO2)ccc1Cl	0.26

Table A4.5: Novelty of the tested molecules. The ECFP4 Tanimoto similarity between the molecules selected from the docking screens and all molecules listed with affinity to or activity at any of the adrenergic receptors in ChEMBL were calculated. ChEMBL ID and SMILES of the most similar molecule as well as the calculated ECFP4 Tanimoto coefficient are listed. If several molecules with the same smilarity were found the first entry is shown.

Mol	ChEMBL ID	Smiles	Tanimoto
1	CHEMBL15303	CC(C)NCC(O)COc1ccccc1Cn2cccn2	0.33
2	CHEMBL15303	CC(C)NCC(O)COc1ccccc1Cn2cccn2	0.42
3	CHEMBL599896	Cl.CC(CCc1ccccc1)NCC(O)c2cccc(c2)[N+](=O)[O-]	0.41
4	CHEMBL7156	OC(CNCCNC(=O)Cc1ccccc1)COC(=O)c2ccccc2F	0.42
5	CHEMBL327122	CC(C)NCC(O)c1cccc(O)c1	0.32
6	CHEMBL62072	CNCC(O)c1ccc(F)c(O)c1	0.31
7	CHEMBL25724	CC(Cc1c[nH]c2cc(C)ccc12)NCC(O)c3cccc(Cl)c3	0.32
8	CHEMBL599896	Cl.CC(CCc1ccccc1)NCC(O)c2cccc(c2)[N+](=O)[O-]	0.47
9	CHEMBL3588911	CC(C)NCc1nccn1Cc2ccccc2C	0.30
10	CHEMBL2153551	Cl.COc1cccc(OC)c1OCCNC[C@H]2COc3cccc(O)c3O2	0.36
11	CHEMBL1902627	CC(C)NCC(O)c1ccccc1Cl	0.45
12	CHEMBL16476	CC(C)NCC(O)c1ccc2ccccc2c1	0.39
13	CHEMBL1767149	CN1CCC[C@@H]1CN2N=C(Cc3ccc(F)c(F)c3)c4ccccc4C2=O	0.33
14	CHEMBL15303	CC(C)NCC(O)COc1ccccc1Cn2cccn2	0.49
15	CHEMBL15303	CC(C)NCC(O)COc1ccccc1Cn2cccn2	0.34
16	CHEMBL1577023	Cc1ccc(OCC(=O)Nc2ccc(cc2)c3oc4cccnc4n3)cc1	0.39
17	CHEMBL6863	CC(C)NCC(O)COC(=O)c1ccc(F)cc1	0.38
18	CHEMBL18772	C1CN(CCN1)c2ccc3ccccc3n2	0.31
19	CHEMBL3099658	COc1ccc(CCNC[C@@H](O)c2cc(O)cc(O)c2)cc1	0.37
20	CHEMBL371300	C[C@H](N)Cn1ncc2ccc(O)cc12	0.23
21	CHEMBL1203102	Cl.COC(=O)c1cc(ccc1O)C(O)CNCC2COc3ccccc3O2	0.31
22	CHEMBL1573444	CN(Cc1ccccc1F)c2nc(nc3ccccc23)c4cccnc4	0.31
23	CHEMBL1327	OC(Cn1ccnc1)c2ccc(Cl)cc2Cl	0.30
24	CHEMBL3588905	C(NCc1c[nH]c(n1)c2ccccc2)c3cccnc3	0.42
25	CHEMBL305558	O[C@@H](CNCCc1ccc(NS(=O)(=O)c2ccc(cc2)n3ncc(n3)c4ccc(F)cc4)cc1)c5cccnc5	0.33
26	CHEMBL351483	CC1=C(CCN2CCc3oc4ccccc4c3C2)C(=O)N5C=C(Cl)C=CC5=N1	0.30
27	CHEMBL101340	FC(F)(F)c1ccc(Cl)c(NC2=NCCN2)c1	0.31

5 | *In silico* prediction of novel ligands for the Cannabinoid receptor 2

Contributions. The author of this study conducted the major part of the work described in this chapter, i.e. optimising docking setups as well as performing and analysing docking screens.
The *in vitro* assays of the reference compounds used as a basis for the computational efforts were performed and analysed by Tamara Miljuš from the lab of Prof. Dmitry Veprintsev at the University of Nottingham.[208]
The pharmacological characterisation of the compounds selected from the docking screen was conducted by Leire Borrega Román and David Sykes in the lab of Prof. Dmitry Veprintsev at the University of Nottingham.

5.1 Introduction and goal of the study

As described in chapter 2.1.5 the Cannabinoid receptor 2 (CB2) is involved in important physiological processes such as inflammations or immunoregulation and might be an interesting drug target for various neurodegenerative diseases.[102–106] It has been shown that the CB2 recruits different effector proteins and is involved in various signalling pathways which might also be of interest for the action of drugs targeting this receptor.[208] It is, therefore, interesting to understand the underlying mechanisms that lead to initiation of certain signalling pathways. Ideally, this comprehension of the underlying processes could even result in their rational manipulation to induce only the desired signalling pathways. With the only recently released crystal structures of the CB2 this investigation of the underlying mechanisms can be taken to a structural level and interactions between ligands and receptor can be observed more closely.[112–114]
The general aim of the studies described in this as well as the next chapter 6 is to investigate how and which interactions between ligand and receptor lead to a bias towards a certain signalling pathway, in this case of $G_{i/o}$ proteins and β-arrestins. To approach this, several different techniques were employed. Assay data regarding the recruitment of different effector proteins for a set of known ligands of the CB2 was provided by our collaborators from the Veprintsev lab at the University of Nottingham. This data can serve as a basis to connect structural observations to *in vitro* behaviour of the ligands regarding the different signalling pathways. These tested molecules were docked to the different available crystal structures and all docking setups were optimised to obtain reasonable docking poses for each molecule, as described in more detail below. These docking poses will then be used for an interaction analysis to see whether certain interaction patterns can be connected

to a certain effector protein recruitment bias. This analysis will be done by our collaborators at the University of Nottingham. In the further process, these observations can then be employed to make predictions for novel ligands.
Additionally, the optimised docking setups were used to screen a molecule library using docking calculations. However, this docking screen was not aiming to find ligands with a bias towards a certain signalling pathway but rather to find CB2 ligands with novel structural scaffolds. Furthermore, the library was screened against CB2 structures in active and inactive conformations. This will allow a further evaluation to see whether agonists rather emerge from docking calculations to an active conformation while antagonists rather emerge from docking calculations to an inactive conformation of the CB2, like it was observed for the β_2AR in chapter 4. In general, novel agonists for the CB2 are more desirable than novel antagonists, since their novel structural scaffolds might inherit interesting functional properties.

Previous work. Prior to starting this structural approach of the project, Tamara Miljuš from the Veprintsev lab at the University of Nottingham measured recruitment of different effector proteins to the CB2 after stimulation with ligands from a set of 35 previously described ligands of the CB2 (see Table A5.1 for molecule structures).[208] To quantify the recruitment of the effector proteins to the receptor a BRET assay was used. The recruitment was measured for five different members of the $G_{i/o}$ subfamily (G_{i1}, G_{i2}, G_{i3}, G_{oA} and G_{oB}) as well as β-arrestin 1 and β-arrestin 2. The results of these assays can be used to identify ligands inducing a certain recruitment bias which can then be correlated to interaction patterns derived from the docking poses created as described in this chapter. These patterns will be analysed by the Veprintsev lab.

5.2 Methods

5.2.1 Docking setup and optimisation

Structure preparation. In total, one CB2 structure in an inactive conformation (PDB ID 5ZTY[112]) as well as two CB2 structures in active conformations (PDB IDs 6PT0[114] and 6KPF[113]) were prepared for docking according to the general structure preparation described in chapter 8.1.1. Each of these structures was prepared without water molecules as well as with specific water molecules predicted as described below. Additionally, 5ZTY was also prepared with a water molecule resolved in the crystal structure and located in proximity to the binding pocket. To determine likely orientations for the water molecules, the hydrogen atoms were first oriented into seemingly favourable directions to allow for interactions with protein and/or ligand before minimisation. Each structure with placed water molecules was minimised using different restraints by adapting the harmonic

force constants (see Table A5.2). The protonation of the histidine located inside the binding pocket ($H95^{2.65}$; numbers in superscript according to the Ballesteros-Weinstein enumeration scheme[19]) was specified as ε protonated for all structures and only 5ZTY was additionally prepared with δ protonated $H95^{2.65}$. In total, 9 structures were prepared using 5ZTY, 3 structures using 6PT0 and 4 structures using 6KPF.

Structure preparation with ligand AM841. The CB2 structure with PDB ID 5ZTY was also prepared with ligand AM841, which was crystallised with the Cannabinoid receptor 1 (CB1) in PDB ID 5XR8, as template ligand. This ligand was selected due to its higher similarity with Cannabinol and cannabinol-like molecules compared to the ligand originally crystallised in 5ZTY. For that, the ligand from the CB1 structure was copied to the receptor structure of 5ZTY after aligning both receptor structures. Additionally, ligand AM841 was truncated by exchanging the long alkyl chain for a methyl substituent ($AM841_{truncated}$) to decrease the flexibility of this ligand part and to account for its low resolution in the crystal structure. The structure was then prepared according to the protocol described in chapter 8.1.1 but with an additional minimisation of $AM841_{truncated}$ in the orthosteric binding pocket of the CB2. The structure was prepared once without a water molecule and once with a water molecule placed between $S285^{7.39}$ and the ligand. Minimisation was then conducted using CHARMM and the CHARMm22 force field[209] in witnotp (Novartis Pharma AG, unpublished) with different harmonic restraints set for ligand, water molecule and the closest surrounding residues (see Table A5.2).

Docking optimisation. The docking setups were optimised to obtain molecule poses with reasonable interactions in at least one active and one inactive docking setup for the majority of the molecules from the reference molecule set (see Table A5.1) and using DOCK3.7.[171] Of note, these optimised molecule poses were also a part of a different project which will not be further described here. Since the orthosteric binding pocket of the CB2 is quite apolar with only few polar residues, it is comparably difficult to score binding poses, and polar molecule parts ended up stranded, i.e. without interaction partners, in several docking poses. To increase the polarity of the binding pocket, possible water positions within the binding pocket were predicted using various tools and water molecules were placed in selected positions in proximity to the molecule binding sites. As an additional help to judge the molecule poses, the knowledge of the binding poses of the crystallised ligands was employed. The docking setups themselves were manipulated by moving the spheres in the binding pocket that are used by DOCK to translate and place the molecules within the binding pocket. The most promising docking setups with the highest number of proposed optimal molecule poses were then selected after visual inspection based on a number of selection criteria: punishing stranded donors (i.e. functional groups with hydrogen bond donor properties that do not interact

with the protein or any other molecule/molecule part in the binding pocket), considering geometries between aromatic/hydrophobic molecule parts that might allow aromatic and hydrophobic interactions and favouring a certain similarity to binding poses of the crystallised ligands. In total, two docking setups based on 5ZTY, two based on 6PT0 and one based on 6KPF were selected to screen a molecule library for novel ligands. Aside from that, molecule poses from four docking setups based on 5ZTY, three based on 6PT0 and one based on 6KPF were selected for an interaction analysis in collaboration with the Veprintsev lab at the University of Nottingham which will not be described further in this thesis.

SEED analysis. To explore the orthosteric binding pocket of the CB2 for hydrophobic and hydrophilic hotspots, SEED was employed. For that, the general procedure described in chapter 8.1.3 was followed and the CB2 structure with PDB ID 5ZTY used ($H95^{2.65}$ protonated as HIE). All water molecules resolved in the crystal structure (including the one located in the orthosteric binding pocket) were removed and the residues within 8 Å distance of the crystallised ligand defined as the protein region of interest. The complete set of 160 fragments was then docked to this protein region using SEED (v.3.3.4).[178] However, only the results of a subset of these fragments were evaluated more closely by visual inspection of their distribution in the orthosteric binding pocket. A special focus during evaluation was directed towards small polar fragments such as methanol or ethanol, their binding locations and their orientation.

Water position prediction with OpenEye SZMAP. The OpenEye software SZMAP[210] was used to predict possible water binding sites within the orthosteric binding pocket. In brief, SZMAP uses a semi-continuum solvent approach to sample possible water binding sites and water orientations. Single water molecules are treated explicitly during energy and probability calculations while the rest of the solvent is treated as a continuum solvent.
All three structures used for the different docking calculations (PDB IDs 5ZTY, 6PT0 and 6KPF) were analysed with SZMAP. For that, protein structures were basically prepared as described in chapter 8.1.1 until after minimisation of the hydrogens. Additionally, the ligand was prepared in a similar way by protonation and subsequent minimisation of the hydrogen atoms using the CHARMm22 force field.[209] Parametrisation of protein and ligand was then done using OpenEye's PCH.[210] SZMAP was then run using the default parameters and results were analysed using OpenEye VIDA.[210] The most interesting water positions were chosen based on negative ΔG values, placement of the water position in proximity to ligand and receptor and comparability to the results derived by other methods (SEED, MOE Solvent Analysis) or for the other analysed structures of the CB2.

Water position prediction with the MOE Solvent Analysis. The MOE Solvent Analysis[211] tool

was used as an additional tool to predict possible water binding sites. This analysis tool uses the three-dimensional reference interaction site model (3D-RISM) method to analyse potential solvent distributions in macromolecular systems. In brief, the 3D-RISM method uses a system of self-consistent equations to calculate solvent particle densities.[212] Binding energies are calculated based on a solvent binding free energy map. The distinct water sites are determined by fitting three-dimensional gaussian functions to the solvent oxygen density and using their center as the site center.

The prediction was pursued for all three structures used in the docking calculations (PDB IDs 5ZTY, 6PT0 and 6KPF). The protein was prepared as for docking (see chapter 8.1.1) and the crystallised ligand was protonated with a subsequent minimisation of the added hydrogens in the same way. The default parameters were used for the MOE Solvet Analysis prediction. The predicted water binding sites were then chosen based on their interaction energy values and hydration free energy (negative values for all contributed energies), proximity to ligand and receptor and comparability to results obtained using SEED and OpenEye SZMAP.

5.2.2 Docking screen

The ZINC15[166] drug-like library consisting of 10,844,842 molecules was screened against the CB2 using DOCK3.7.[171] In total, the library was docked against five of the optimised docking setups, two based on inactive 5ZTY (one with a water molecule close to $S285^{7.39}$ ($5ZTY^{S285}$), one using $AM841_{truncated}$ as template ligand ($5ZTY^{AM841}$)), two based on active 6PT0 (one without a water molecule ($6PT0^{no\text{-}water}$), one with a water molecule close to $S285^{7.39}$ ($6PT0^{S285}$)) and one based on active 6KPF (with two water molecules: one close to $S285^{7.39}$ and one close to $L182^{ECL2}$ ($6KPF^{S285\text{-}L182}$)). The top 500 ranked molecule poses resulting from each of these docking calculations were evaluated visually. Additionally, the ranking lists of some of these dockings were re-ranked to potentially increase the probability of discovering agonists over antagonists. In total, three re-rankings were conducted: Two dual re-rankings and one selective re-ranking (see section 8.1.2). For the dual re-rankings the ranking lists of the docking calculations to $6PT0^{no\text{-}water}$ and $6PT0^{S285}$ were compared to each other and the ranking list of the docking calculation to $6PT0^{S285}$ was compared to the one resulting from the docking calculation to $6KPF^{S285\text{-}L182}$. The selective re-ranking was conducted using the ranking list of the docking calculation to $6PT0^{S285}$ compared to the ranking list from the docking calculation to $5ZTY^{S285}$, favouring towards the 6PT0 based structure. The top 500 ranked molecule poses of each of the three resulting re-ranking lists were then evaluated visually for each of the docking calculations included in the specific re-ranking and based on an active conformation of the receptor.

To aim for structurally novel CB2 ligands, all ranking and re-ranking lists were additionally filtered using a Tanimoto filter. For that, bioactivity data against the human CB2 was downloaded from ChEMBL[192] and filtered to keep only those molecules that showed activity against the CB2. The Tanimoto similarity based on ECFP4 between each of the molecules from the ChEMBL dataset and the top 2000 ranked molecules of each (re-)ranking list was then calculated. Only those molecules from the (re-)ranking list were kept that had a Tanimoto similarity $\leq$0.45 to the most similar ChEMBL molecule. The top 500 ranked molecule poses of each filtered (re-)ranking list were then evaluated visually.

A total of 37 molecules was selected from all the different (re-)ranking lists to be tested experimentally in *in vitro* assays. Finally, only 29 of these molecules were purchased due to inavailability or high costs.

5.2.3 Pharmacological characterisation

The pharmacological characterisation of the compounds selected from the ZINC15 docking screen was conducted by Leire Borrega Román and David Sykes in the lab of Prof. Dmitry Veprintsev at the University of Nottingham.

Binding affinity of the compounds was measured using a Time Resolved Fluorescence Resonance Energy Transfer (TR-FRET) assay. In brief, displacement of a fluorescence tracer (NBD-D77) from the orthosteric binding site by the measured compounds was detected based on the decrease of the FRET signal between the tracer and a fluorescence labelled SNAP-CB2 in a time resolved manner. For that, the compound dilution (in DMSO) was was mixed with the tracer (900 nM) in the assay buffer (HBSS, 5 mM HEPES, 0.5% BSA, 0.02% Pluronic Acid; pH 7.5) in 384-well Optiplate (PerkinEmler). To each well 1 µg HEK TR cell membranes containing the Terbium-labelled SNAP-CB2 was added and simultaneous measurement of the fluorescence signals at 520 nm and 620 nm was immediately started at a PHERAstar FSX microplate reader. After 15 min the equilibrium was reached and the measurement stopped.

The data was then analysed using Graphpad Prism 8[213] and fitting the Motulsky and Mahan equation[214] ('kinetics of competitive binding') to the time resolved FRET data to obtain k_{on} and k_{off} values for the measured compound. With these values the K_D value could be calulated using $K_d = \frac{k_{off}}{k_{on}}$. Additionally, K_i values were calculated from the IC_{50} values of the dose-response curve after equilibrium binding was reached. The K_D of the tracer was determined to be 438 nM and SR144528 was used as positive control and to determine non-specific binding.

For each compound the influence on the determined fluorescence signals was measured to account for false positives due to assay interference.

5.3 Results

5.3.1 Optimisation of docking setups

As described in the introduction to this chapter, the ultimate aim is to correlate interactions between ligands and the CB2 in the orthosteric binding site to recruitment and recruitment bias of different effector proteins to the receptor after stimulation with the ligand. For that, it was necessary to predict reliable binding poses of the tested ligands in the orthosteric binding pocket of the CB2, which was pursued using docking calculations. The setup of these docking calculations had to be optimised for each of the three used CB2 structures (PDB IDs 5ZTY, 6PT0 and 6KPF) to obtain believable molecule poses for at least the majority of the ligands. Since the orthosteric binding pocket of the CB2 is quite apolar, the evaluation of the molecule docking poses and choice of reliable ones was rather challenging. To get basic ideas of ligand orientation within the orthosteric pocket, the binding poses of the ligands in the available crystal structures were used. Molecule poses with 'stranded donors', i.e. polar functional groups that did not entertain any interaction, were punished unless it was a stranded amide or carbonyl group in a similar position as observed for ligand AM10257 in the structure with PDB ID 5ZTY. Any kind of possible aromatic interaction was rated as positive.

The different approaches to optimisation of docking setups and the decisions made during this process are described for each of the used CB2 structures in more detail below. The main approach was the prediction of water positions in the orthosteric binding pocket. Water molecules were placed in likely positions to mediate polar interactions between ligand and receptor, thereby ameliorating docking poses and reducing the number of stranded donors. Of note, this analysis was based on a visual evaluation of the docking results and molecule poses.

Molecule docking poses were optimised for the set of 35 ligands that has previously been tested by our collaborators at the Veprintsev lab at the University of Nottingham (see Table A5.1). This molecule set had to be adapted to the restrictions inherited by the docking approach. Ligands with very long, flexible molecule parts cannot be docked reliably due to their high number of degrees of freedom and the resulting high amount of possible docking poses it could adopt in a protein binding pocket. Therefore, a number of molecules had to be excluded prior to docking, such as the endogenous CB2 ligands Anandamide and 2-AG. A set of molecules with a core binding part connected to a fluorescence tag by a long, flexible alkyl chain had to be truncated to eliminate the flexible part and only the core binding anchor was kept. Additionally, the ligands crystallised with the CB2 in the used crystal structures (AM10257, AM12033) or used for docking setup (AM841) were included unless they were already in the set of tested ligands (WIN 55,212-2). Finally, a set

of 30 + 3 dockable ligands was obtained.

Docking optimisation with an inactive conformation of the CB2 (PDB ID 5ZTY).

First approaches. The structure of the CB2 in an inactive conformation was the first one to be prepared and optimised for docking calculations. The optimisation procedure of this docking setup was, therefore, more extended compared to the procedure for the other prepared CB2 structures. First optimisation steps were conducted without including any water molecules in the structure but, unfortunately, the resulting molecule poses did not meet the above described criteria for a majority of the docked molecules. Apolar interactions are more challenging to predict with DOCK3 than polar interactions which makes placement of rather lipophilic compounds in a quite apolar binding pocket like the one of the CB2 more difficult. This is because DOCK3 estimates the apolar interactions based on van der Waals interactions and apolar desolvation terms, however, other types of apolar interactions such as aromatic interactions are not included in the interaction energies.[145] To add polarity into the binding pocket and to allow for more polar interactions, the binding site can be explored for potential binding positions of water molecules which might not be resolved in the crystal structure but could mediate interactions between ligand and receptor. In the binding pocket of the CB2 structure with PDB ID 5ZTY only one water molecule was resolved which is positioned at a quite far edge of the binding pocket close to the hydroxyl-group of ligand AM10257 and in proximity to T114$^{3.33}$ and Y190$^{5.39}$ (see Figure 5.1.A, red sphere). Including this crystallised water molecule in the docking setup did not lead to an improvement of molecule poses though, probably because it is out of reach for most of the ligands. Since no other water molecules are resolved in the binding pocket of the CB2 in structure 5ZTY, an approach to predict possible water molecule binding sites had to be found and pursued.

Water molecules from crystal structures of other GPCRs. A convenient approach to predict possible water molecule binding sites is to have a look at the structures of other GPCRs with a high similarity to the target receptor and search for water molecules crystallised in their orthosteric binding pockets. The GPCR closest related to the CB2 is, obviously, the CB1, however, unfortunately no water molecules can be found in the orthosteric binding pocket of any of the released CB1 structures. Other related GPCRs with available crystal structures are the sphingosine receptors (up to 29.5% identity to the CB2) and the lysophosphatidic acid receptors (up to 27.5% identity to the CB2) but neither of the structures of these receptors contains water molecules in useful positions of the binding pocket (i.e. in proximity to the ligand in the CB2 structure). Additionally, water molecules resolved in the orthosteric binding site of six opioid receptor structures were inspected (identity of 24-25% between the CB2 and the opioid receptors). After aligning CB2 and opioid

structures, some of the water molecules from the opioid structures were located in positions in the CB2 binding pocket where they could interact with the crystallised ligand AM10257 (water molecules from the structures with PDB IDs 4N6H[215] and 5C1M;[216] see Figure 5.1.A, light blue spheres). Finally, it was decided to not use these water molecules for docking after all, due to the rather low similarity of the binding pocket region between the CB2 and the opioid receptors.

SEED predictions. To gain, in general, a better idea of the binding pocket of the CB2, i.e. which parts are more hydrophobic and which more hydrophilic, and to get a rough idea where water molecules might be located, a SEED analysis was pursued. Although the entire set of 160 fragments that was available internally was docked to the orthosteric binding pocket of the CB2, the main focus during evaluation of the results was then put on small polar fragments, especially methanol and ethanol. Most interesting was a cluster of both of these fragments pointing towards the hydroxyl-group of S285$^{7.39}$ and located between S285$^{7.39}$ and the spot where the carbonyl-group of the crystallised ligand-amide would be located (see Figure 5.1.A, white lines). This location of a methanol/ethanol cluster might indicate that a water molecule could be placed in the same position mediating a polar interaction between S285$^{7.39}$ and the amide-carbonyl of the ligand.

Water prediction using OpenEye SZMAP. To supplement the previous approaches, two different softwares predicting possible water binding sites within protein binding pockets were applied to the CB2 as well. One of these softwares is SZMAP by OpenEye. The water binding positions predicted by this program were visualised in VIDA and analysed. Within the binding pocket and in proximity to the ligand three water positions with negative ΔG values were found. One water position was predicted in proximity to T114$^{3.33}$ and Y190$^{5.39}$, one close to S285$^{7.39}$ and one roughly between H95$^{2.65}$ and the ligand carbonyl-group (see Figure 5.1.A, yellow spheres).

Water position prediction using MOE Solvent Analysis. Another tool to predict water positions in protein binding sites is the MOE Solvent Analysis tool. During evaluation of the predicted water positions only those molecules were considered further that showed negative values for all energy contributions and were in proximity to ligand and receptor. Three water positions were found in the binding pocket and in proximity to the ligand binding site: one close to T114$^{3.33}$ and Y190$^{5.39}$, one between S285$^{7.39}$ and the ligand carbonyl-group and one located roughly between H95$^{2.65}$ and the ligand carbonyl (see Figure 5.1.A, green spheres).

Comparison of the predicted water positions. The possible water binding sites predicted by the different methods were compared to see whether the results overlap (see Figure 5.1.A). Indeed, both SZMAP and the MOE Solvent Analysis predicted a water position in roughly the same position in which the water molecule in the crystal structure is located (close to T114$^{3.33}$ and Y190$^{5.39}$). Another water position common between both methods and also in the same location

where the methanol/ethanol clusters from SEED were found, was predicted between S285$^{7.39}$ and the ligand carbonyl-group, where it could mediate a hydrogen bond between ligand and receptor. The predicted water binding sites in both of these locations were placed in approximately the same positions for all methods. A third predicted water binding site located roughly between H95$^{2.65}$ and the ligand carbonyl was found for both SZMAP and the MOE Solvent Analysis but the position was not as well conserved as observed for the two other water position locations. Overall, these similar results for the different methods indicate, that water molecules in these positions are probably quite likely. Since the docking calculation containing the water molecule from the crystal structure has already been tested before, the predicted water position in that same location was not followed up on. For each of the other predicted water positions, close to S285$^{7.39}$ and close to H95$^{2.65}$, a structure containing a water molecule in that approximate position was prepared and the molecule set docked to these structures.

Protonation state of H95$^{2.65}$. The protonation state of histidines in proteins is an important issue for docking calculations because it can influence the energy grids used in DOCK3.7 which might have an influence on the molecule poses resulting from docking. The protonation state of histidines is usually chosen based on the local environment. Particularly tricky is the decision upon the protonation state of a histidine if this residue is located within the binding pocket because it can influence interactions between molecules and this residue and, hence, overall molecule poses. In the CB2, H95$^{2.65}$ is located in the orthosteric binding pocket and might directly interact with ligands. Therefore, choosing the correct protonation state can be critical to obtain correct molecule poses. However, protonation can also vary depending on the environment, i.e. protonation of histidines located in the binding pocket might also depend on ligands bound to the receptor. In all docking approaches described above, H95$^{2.65}$ was protonated at the nitrogen atom in ε-position (HIE), i.e. with a hydrogen atom pointing into the binding pocket. To account for the uncertainty of this protonation state, the CB2 structure was also prepared with H95$^{2.65}$ protonated in δ-position (HID). The docking poses of only few molecules changed compared to the results from the docking calculation to the structure with HIE95$^{2.65}$ and docking poses were overall not improved. Based on these observations and the already satisfying results after placing water molecules in the binding pocket, it was decided to not prepare any other structure with HID95$^{2.65}$ instead of HIE95$^{2.65}$.

Using AM841$_{\text{truncated}}$ from CB1 structure 5XR8 as template ligand. During optimisation of the docking setups it became obvious that placement of ligands with a cannabinol-like structure did not work very well, resulting in stranded donors and unlikely molecule poses (see Figure 5.2.B). The ligand crystallised in structure 5ZTY is structurally very different from the cannabinol-like ligands which was thought to be one of the reasons for the problems in positioning these ligands.

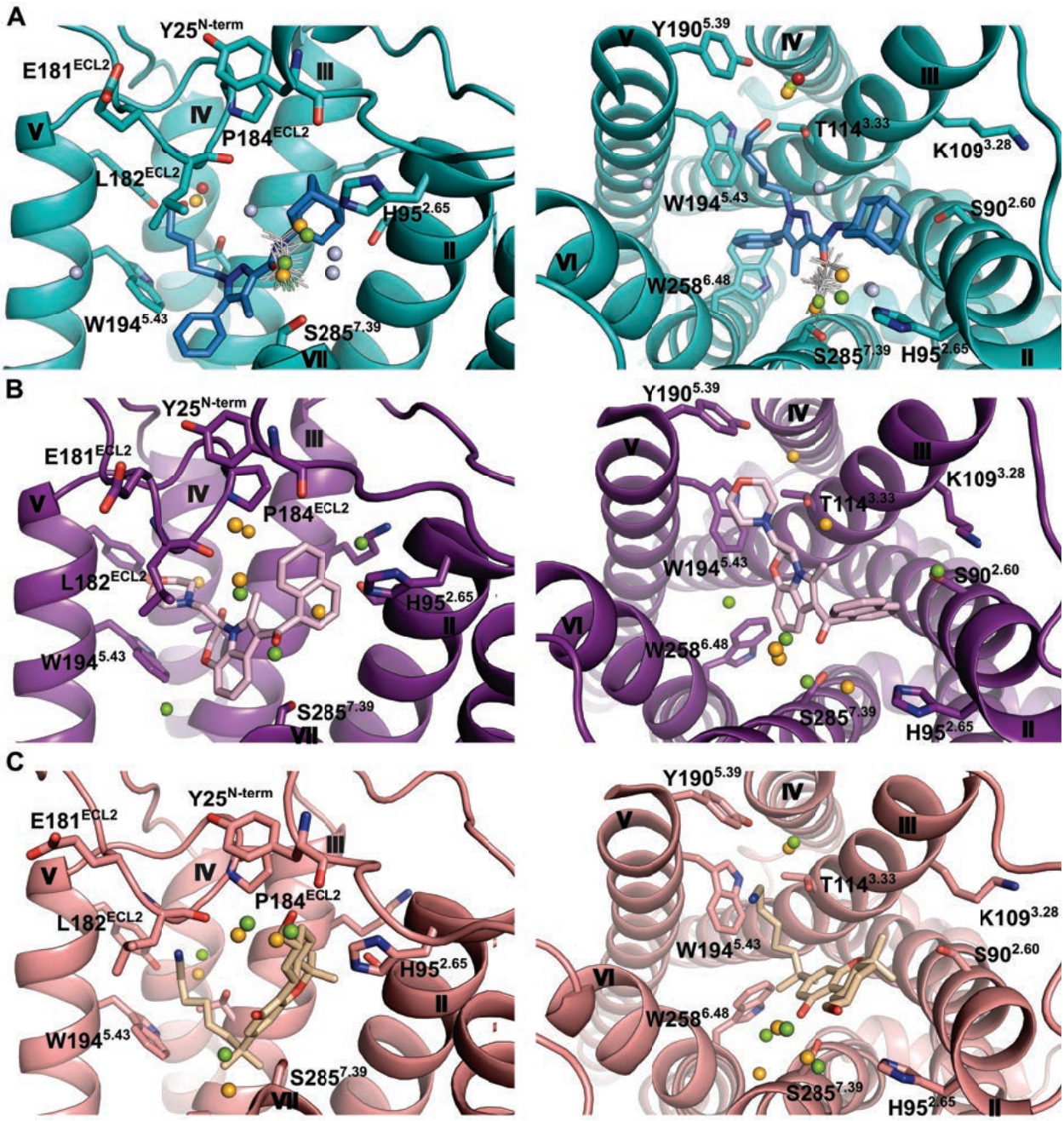


Figure 5.1: Predicted water positions in the orthosteric binding pocket of the CB2 that were selected for a further evaluation in the structures with A)PDB ID 5ZTY, B) PDB ID 6PT0 and C) PDB ID 6KPF. Various methods were used for the prediction of water positions: original crystal structure (red), opioid receptor structures (PDB IDs 4N6H[215] and 5C1M;[216] light blue), SEED analysis (white lines), OpenEye SZMAP (yellow) and MOE Solvent Analysis (green). All residues with a water position predicted in their proximity in any of the analyses and structures are shown as sticks and labelled. TMs are labelled at the top of the helices. Left: side view of the binding pocket. TM VI and VII were omitted for clarity. Right: View from the extracellular side onto the receptor. ECL2 and N-terminus were omitted for clarity. The pictures were taken from the same point of view for all structures for better comparability.

At that time, no CB2 structures containing a cannabinol-like ligand had been released, however, for the CB1 several structures with similar ligands were available. Therefore, it was decided to use the cannabinol-like ligand AM841 crystallised with the CB1 in the structure with PDB ID 5XR8[217] as a template ligand for docking calculations to the CB2. The flexible alkyl chain of AM841 was truncated to a methyl group ($AM841_{truncated}$; see Figure 5.2.A) since the electron density of that molecule part was not resolved in the template structure 5XR8 and the ligand was copied to the CB2 structure after aligning the CB1 and the CB2. $AM841_{truncated}$ was then minimised in the binding pocket of the CB2 and used as template ligand for the DOCK preparation, once without a water molecule and once with a water molecule placed close to $S285^{7.39}$ in a position predicted as described above. Especially the docking setup containing the water molecule yielded improved docking poses for the cannabinol-like ligands which were placed similar to $AM841_{truncated}$ in the binding pocket and could form polar contacts to the placed water molecule (see Figure 5.2.C).

Docking optimisation with an active conformation of the CB2 (PDB ID 6PT0).

Since the majority of molecules in the ligand test set are agonists they should fit better into the binding pocket of the receptor in an active conformation compared to an inactive conformation. Therefore, the structures of the CB2 in active conformations were also prepared as soon as they were released. Similar to the docking calculations to the inactive CB2 structure (PDB ID 5ZTY) the docking setup had to be optimised and, again, it was decided to place predicted water molecules in the binding pocket. This time, only OpenEye SZMAP and the MOE Solvent Analysis were used to find possible water binding sites, since comparison to crystal structures of other GPCRs and SEED did not result in any additional information when optimising the CB2 structure 5ZTY. For all docking calculations and tested docking setups with structure 6PT0, $H95^{2.65}$ was protonated in ε-position and it was not tested how results could change if it was protonated in δ-position.

Water prediction using OpenEye SZMAP. As described for structure 5ZTY, OpenEye SZMAP was used to predict possible water binding sites and the results were analysed in VIDA. During this analysis, five potentially interesting water positions were found in the binding pocket: One position close to $T114^{3.33}$/$Y190^{5.39}$, one in proximity to $HIE95^{2.65}$ and three in proximity to different parts of the ECL2, in particular the backbone carbonyl-groups of $E181^{ECL2}$/ $L182^{ECL2}$, $L182^{ECL2}$ or $P184^{ECL2}$ (see Figure 5.1.B, yellow spheres).

Water position prediction using MOE Solvent Analysis. Additionally, a second water position prediction was conducted using the MOE Solvent Analysis and water positions with all negative energy contributions and located within the binding pocket were chosen. In total, four water positions were selected for further evaluation and comparison: One close to the carbonyl-group of

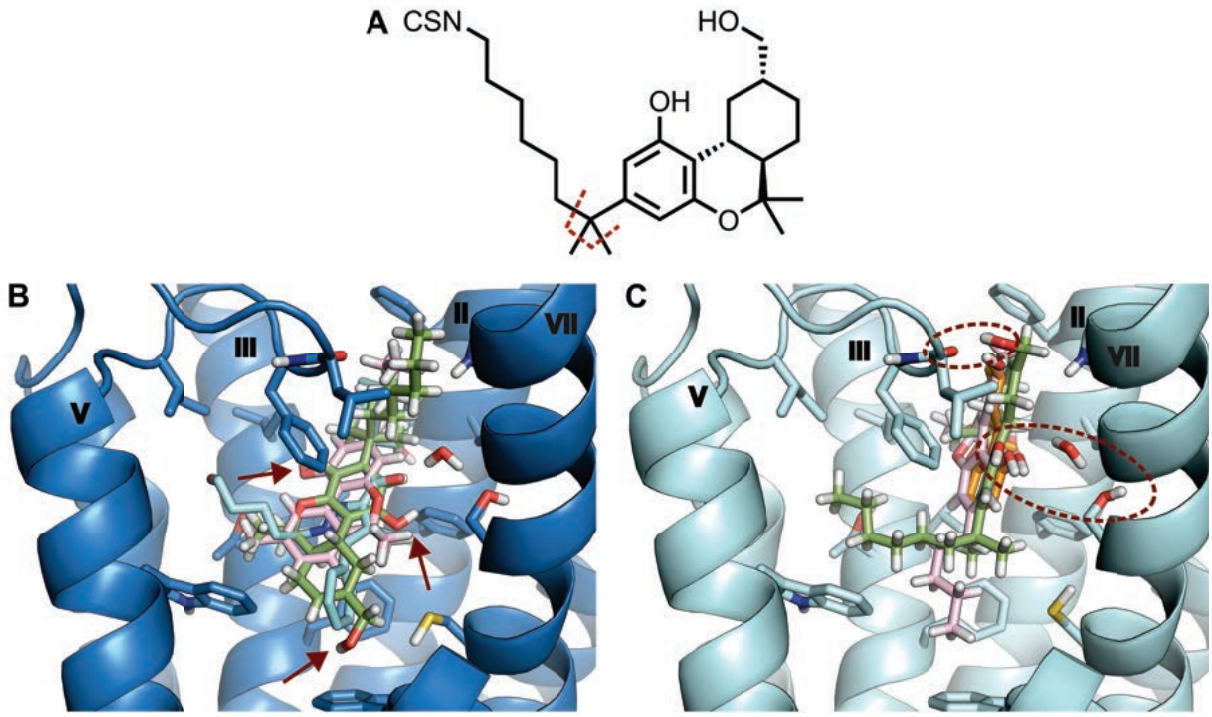


Figure 5.2: Changes in the docking poses of cannabinol-like molecules upon using AM841 as a template ligand. A) 2D-depiction of ligand AM841 which was crystallised in the CB1 structure with PDB ID 5XR8.[217] The ligand was truncated as indicated by the red dashed lines ($AM841_{truncated}$) and used as template ligand during preparation of the CB2 structure 5ZTY for DOCK. B) Docking poses of Cannabinol (pink) and the cannabinol-like ligand HU210 (green) when using the originally crystallised ligand AM10257 (aquamarine) as template. The red arrows indicate stranded hydroxyl groups of Cannabinol and HU210. C) Docking poses of Cannabinol (pink) and the cannabinol-like ligand HU210 (green) when using $AM841_{truncated}$ (orange) as template ligand. Both docked molecules adopt a similar pose to $AM841_{truncated}$ and can form polar interactions to the receptor/water molecule as indicated by the red circles.

the crystallised ligand and in proximity to S285$^{7.39}$, one close to the backbone carbonyl-group of L182^{ECL2}, one in proximity to the toggle switch residue W258$^{6.48}$ and to W194$^{5.43}$ and one on the far edge of the binding pocket close to K109$^{3.28}$ and S90$^{2.60}$ (see Figure 5.1.B, green spheres).

Comparison of the predicted water positions. Unlike observed for the predicted water positions in the CB2 structure 5ZTY, the water positions predicted by both methods did not overlap for structure 6PT0 in most cases (see Figure 5.1.B). Only one water position close to L182^{ECL2} was predicted by both methods. The water position close to T114$^{3.33}$/Y190$^{5.39}$ which was predicted by SZMAP is close to where the crystal water is located in structure 5ZTY while the spot close to S285$^{7.39}$ predicted by the MOE Solvent Analysis overlaps with a water position predicted for structure 5ZTY. All other predicted water positions do not overlap between methods or with the results from the analysis of structure 5ZTY. Due to time limitations it was not possible to prepare structures for all predicted water molecule positions that were taken into closer consideration and the selection was limited to the positions located in the most likely positions to interact with the docked molecules, in particular the positions close to S285$^{7.39}$ and to L182^{ECL2}.

Docking optimisation with an active conformation of the CB2 (PDB ID 6KPF).

Additional to the CB2 structure in an active conformation with PDB ID 6PT0, another structure in an active conformation with PDB ID 6KPF was prepared. The ligands crystallised in both structures, WIN 55,212-2 in 6PT0 and AM12033 in 6KPF, are structurally quite different. AM12033 has a cannabinol-like structure and, as observed for 5ZTY, the structural differences of the used template ligands can play a role for molecule poses resulting from docking calculations. Aside from that, the binding pockets of 6PT0 and 6KPF differ slightly with slight changes of side chain orientations which were likely caused by the adaptation of the receptor binding pocket to the bound ligand. Therefore, it is interesting to dock to both structures in active conformations and use the different structural information inherited in each of them. Again, water positions were predicted using OpenEye SZMAP and the MOE Solvent Analysis and H95$^{2.65}$ was protonated in ε-position in all prepared structures.

Water prediction using OpenEye SZMAP. Four water binding sites with reasonable energy values were predicted in proximity to the crystallised ligand by OpenEye SZMAP: one close to T114$^{3.33}$/ Y190$^{5.39}$, one close to the backbone carbonyl-group of L182^{ECL2}, one close to S285$^{7.39}$ but not to the ligand and one in between the ligand and Y25$^{N\text{-term}}$ (see Figure 5.1.C, yellow spheres). Especially the predicted water molecules in the positions close to L182^{ECL2} and Y25$^{N\text{-term}}$ could be mediating interactions between ligand and receptor, although the distance for the latter to the receptor might be too big for efficient interactions.

Water position prediction using MOE Solvent Analysis. After analysing the MOE Solvent Analysis results as described above, three possible water binding sites were selected for closer evaluation: One close to T114$^{3.33}$/Y190$^{5.39}$, one close to the backbone carbonyl-group of L182^{ECL2} and one located between the ligand and Y25$^{N\text{-}term}$ (see Figure 5.1.C, green spheres). Additionally, a water position close to S285$^{7.39}$ and within hydrogen-bond distance of the ligand was found, however, not with the desired energy values. This position was included into a further analysis anyway since water positions in a similar region were also found for 5ZTY and 6PT0.

Comparison of the predicted water positions. Comparing the water position predictions from both methods it can be seen that they mostly overlap (see Figure 5.1.C). Again, water positions were found in the region where the crystallised water in structure 5ZTY is located. Also, both methods predicted water positions close to the backbone carbonyl-group of L182^{ECL2}, a position that was also found in the analysis with structure 6PT0. Interestingly, the water position close to Y25$^{N\text{-}term}$ was consistently predicted by both methods and has not been found in the analysis of any of the other two structures. Anyhow, since it is located quite far from the ligand and might therefore not interact efficiently with the docked molecules, no structure containing a water molecule in that position was prepared. The water positions predicted close to S285$^{7.39}$ do not overlap between both methods and also not with the positions predicted for structures 5ZTY or 6PT0. However, it was decided to prepare a structure containing a water molecule in this location as predicted by the MOE Solvent Analysis anyway, since a water molecule in this position has also been predicted for the other two structures and since it could interact with the crystallised ligands and, hence, also the docked molecules.

Docking setups chosen for further evaluation.

In total, seven different structures were prepared based on the CB2 structure with PDB ID 5ZTY: (1a) no water molecule in the binding pocket, (1b) with the crystallised water close to T114$^{3.33}$/ Y190$^{5.39}$, (1c) with a predicted water molecule close to S285$^{7.39}$, (1d) with a predicted water molecule close to H95$^{2.65}$, (1e) as 1d but water minimised in absence of the ligand, (1f) with H95$^{2.65}$ protonated as HID (for all other it is HIE), (1g) with AM841$_{truncated}$ as template ligand and a predicted water molecule close to S285$^{7.39}$.

Based on the structure with PDB ID 6PT0 three different structures were prepared: (2a) with no water molecule in the binding pocket, (2b) with a water molecule placed at the position close to the backbone carbonyl of L182^{ECL2} as predicted by SZMAP and (2c) with a water molecule placed close to the ligand carbonyl-group as predicted by the MOE Solvent Analysis.

Using the other structure in an active conformation with PDB ID 6KPF as a basis, four different

structures were prepared: (3a) with no water molecule in the binding pocket, (3b) with a water molecule placed at the position close to the backbone carbonyl of $L182^{ECL2}$ as predicted by the MOE Solvent Analysis, (3c) with a water molecule placed in proximity of $S285^{7.39}$ as predicted by the MOE Solvent Analysis and (3d) with both water molecules from 3b and 3c.
Docking results for each of these structures were further optimised by moving the spheres used by DOCK to rotate and translate the molecules in the binding pocket. Finally, the results of all docking calculations were compared and the docking setups that yielded the highest number of molecule poses that met the initially mentioned criteria were chosen. It was made sure that a molecule pose that met the criteria was obtained from at least one of the docking setups for inactive as well as active conformation for the majority of molecules. In Table 5.1 all tested docking setups are listed and the ones that were chosen for the docking screens are highlighted.
Molecule docking poses of the reference molecules for the interaction analysis by the Veprintsev lab were finally chosen from the docking calculations to four docking setups based on PDB ID 5ZTY, three based on PDB ID 6PT0 and one based on PDB ID 6KPF. For the ZINC15 library screen the number of used docking setups was reduced to the ones with the highest number of favourable molecule poses of the reference molecules, while also taking the diversity of the setups with different template ligands into account. Finally, two docking setups based on PDB ID 5ZTY, two based on PDB ID 6PT0 and one based on PDB ID 6KPF were used for the docking screen (see Table 5.1).

5.3.2 Prediction of novel ligands for the CB2

A docking screen was conducted to predict novel ligands of the CB2 using the optimised docking setups described above. There were two main aims for this screen. One aim was to find novel ligands with previously undescribed structural scaffolds for ligands of the CB2. Ligands with previously undescribed structural scaffold can potentially adopt novel binding modes which could reveal new insights to CB2 signalling. The second aim was to rather discover agonists than antagonists since there is more to learn from their binding mode and *in vitro* behaviour. To achieve these goals, various re-ranking and filtering steps were conducted using the ranking lists from the docking calculations to the different docking setups. These additional steps are described further below and are shown schematically in Figure 5.3.

Screening and molecule selection. The ZINC15 drug-like library consisting of 10,844,842 molecules was docked to five different docking setups of the CB2: two based on the structure in an inactive conformation with PDB ID 5ZTY (both with a predicted water molecule placed in proximity to S285 but one prepared with the originally crystallised ligand ($5ZTY^{S285}$) and one with $AM841_{truncated}$ ($5ZTY^{AM841}$); see Figure 5.3, blue boxes) and three docking setups based on structures of the

Table 5.1: List of the tested docking setups based on crystal structures of the CB2 in an inactive conformation (PDB ID 5ZTY) or active conformation (PDB ID 6PT0 and 6KPF). The differences between the setups are indicated. The five docking setups that were selected for the docking screen are highlighted in bold and were assigned an abbreviation.

No.	PDB ID	Water molecule	Abbreviation
1a	5ZTY	None	
1b	5ZTY	Crystal water close to T114$^{3.33}$ and Y190$^{5.39}$	
1c	5ZTY	**Predicted water molecule close to S285$^{7.39}$**	**5ZTY^{S285}**
1d	5ZTY	Predicted water molecule close to H95$^{2.65}$	
1e	5ZTY	Predicted water molecule close to H95$^{2.65}$ (minimised in absence of the ligand)	
1f	5ZTY	With HID95 $^{2.65}$	
1g	5ZTY	**Predicted water molecule close to S285$^{7.39}$ & AM841$_{truncated}$ as template ligand**	**5ZTY^{AM841}**
2a	6PT0	**None**	**6PT0$^{no\text{-}water}$**
2b	6PT0	Predicted water molecule close to L182^{ECL2} backbone carbonyl	
2c	6PT0	**Predicted water molecule close to ligand carbonyl/S285$^{7.39}$**	**6PT0^{S285}**
3a	6KPF	None	
3b	6KPF	Predicted water close to L182^{ECL2} backbone carbonyl	
3c	6KPF	Predicted water molecule close to S285$^{7.39}$	
3d	6KPF	**Two predicted water molecules close to S285$^{7.39}$ and to L182^{ECL2} backbone carbonyl**	**6KPF$^{S285\text{-}L182}$**

CB2 in active conformations with PDB ID 6PT0 (one without water molecules ($6PT0^{no\text{-}water}$), one with a predicted water molecule in proximity to S285 ($6PT0^{S285}$); see Figure 5.3, yellow boxes) and PDB ID 6KPF (with two predicted water molecules placed in proximity to S285 and L182 ($6KPF^{S285\text{-}L182}$); see Figure 5.3, violet boxes). To increase the probibility of enriching agonists in the top ranks and thereby their discovery, several re-rankings were conducted (see also chapter 4 and Scharf et al.[81] for a comparable approach). Two dual re-rankings were conducted comparing the ranking lists resulting from the docking calculations to $6PT0^{no\text{-}water}$ and $6PT0^{S285}$ (see Figure 5.3, orange box) and comparing the ranking lists from the docking calculation to $6PT0^{S285}$ to the ranking list from the docking calculation to $6KPF^{S285\text{-}L182}$ (see Figure 5.3, pink box). Additionally, a selective re-ranking was conducted, comparing the ranking list resulting from the docking calculation to $6PT0^{S285}$ to the one from the docking calculation to $5ZTY^{S285}$, with a selectivity towards the 6PT0 based setup (see Figure 5.3, green box). All ranking and re-ranking lists were then also filtered using a Tanimoto similarity filter, requiring an ECFP4 Tanimoto ≤ 0.45 to a set of known ligands of the CB2 obtained from ChEMBL.

The top 500 molecule poses of each of the five initial ranking lists, the three re-ranking lists (poses from in total five docking setups based on active conformations) and of the filtered lists of each of these (re-)ranking lists were evaluated visually. This evaluation was based on the same criteria as described for the selection of the docking setup, i.e. no stranded donors were allowed, unless it was an amide or carbonyl group placed in the same region as observed for the crystallised ligand in the structure with PDB ID 5ZTY, and with a special focus on potential aromatic and hydrophobic interactions. Finally, 39 molecules were selected for further evaluation, of which 29 compounds were purchased to be tested in *in vitro* assays by our collaboration partners at the Veprintsev lab (see Table A5.3).

Selected compounds. Of the 29 compounds that were finally selected and purchased (see Table A5.3), eight compounds were selected from the docking calculations to the docking setups based on the CB2 structure in an inactive conformation (PDB ID 5ZTY). Six of these molecules have an ECFP4 Tanimoto below 0.45 to the known CB2 ligands obtained from ChEMBL and are, therefore, considered as molecules with novel scaffolds for potential CB2 ligands. Since these eight compounds stem from docking calculations to the CB2 in an inactive conformation, they are expected to be antagonists or inverse agonists.

The bigger part of 21 compounds was selected from the docking calculations to the CB2 in active conformations. From each of the docking calculations to $6PT0^{S285}$ as well as from the docking calculation to $6KPF^{S285\text{-}L182}$, eight molecules were selected. The remaining five molecules were selected from the docking calculation to $6PT0^{no\text{-}water}$. While only two of five molecules from the

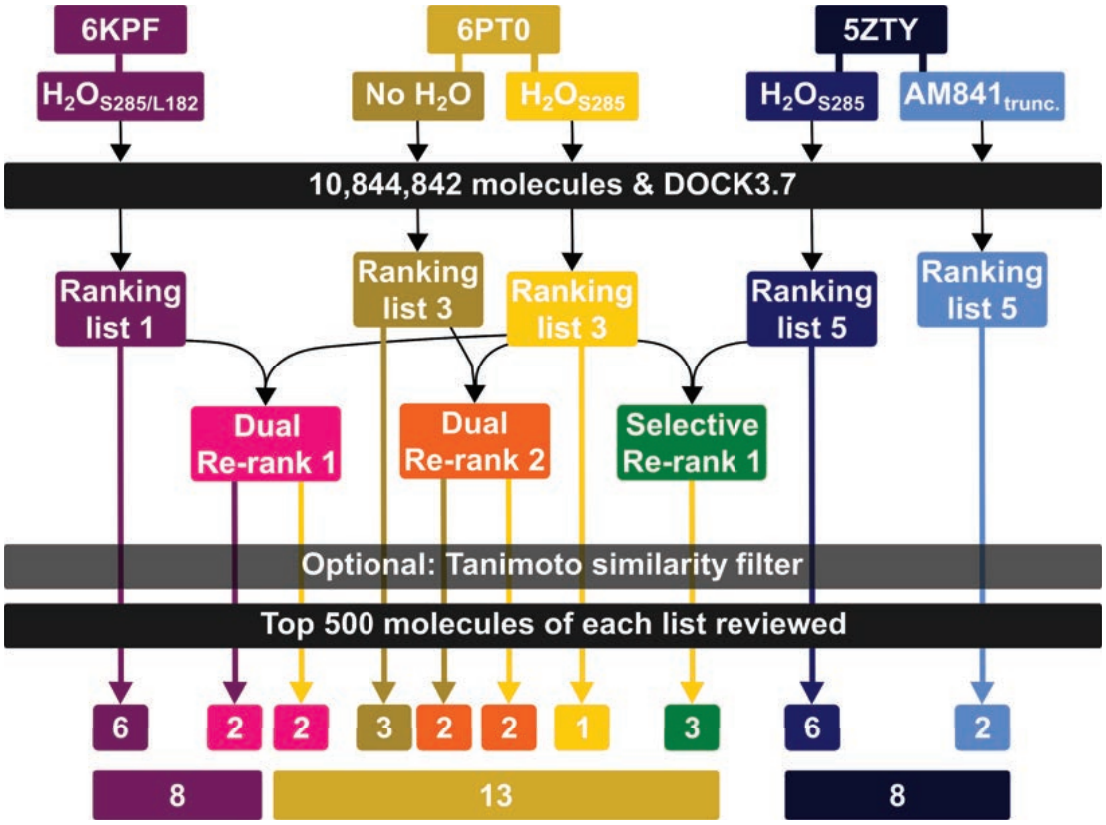


Figure 5.3: Schematic representation of the workflow of the docking screen against the CB2. A ZINC15 library was docked using five different docking setups of the CB2 based on one structure of the CB2 in an inactive conformation (PDB ID 5ZTY) and two structures of the CB2 in active conformations (PDB IDs 6PT0 and 6KPF). The ranking lists resulting from the docking calculations were compared in three re-rankings and the top 500 molecule poses from each (re-)ranking list were evaluated before and after applying an additional known ligand similarity filter. The number of molecules that was selected and purchased from each (re-)ranking list and for each of the used docking setups can be found at the bottom.

docking calculation to $6PT0^{no\text{-}water}$ and two of eight molecules from the docking calculation to $6KPF^{S285\text{-}L182}$ were chosen from re-ranking lists, seven of eight molecules chosen from the docking calculation to $6PT0^{S285}$ resulted from the re-rankings. However, all three conducted re-rankings included the ranking list from the docking calculation to $6PT0^{S285}$ which was not the case using the other docking setups. Comparing the selected molecules to the set of known ligands from ChEMBL by calculating the ECFP4 Tanimoto similarity, 16 of the 21 molecules can be considered as novel for CB2 ligands with a Tanimoto coefficient below 0.45 to any of the molecules from the ChEMBL set. All of the selected 21 molecules are expected to be agonists towards the G_i signalling pathway since they result from docking calculations to CB2 structures in active conformations in complex with G_i proteins. A certain agonism towards other signalling pathways cannot be concluded from this but is also likely.

It should be noted that although the main aim was to find ligands with novel structural scaffolds, molecules with a higher Tanimoto similarity than 0.45 to known ligands of the CB2 were also selected. This was thought as an internal control to make sure the docking setups work in general, since it can be assumed that it is more likely to discover hits among molecules that are more similar to known CB2 ligands than among dissimilar molecules.

Pharmacological characterisation. The binding affinity of the selected compounds was determined by our collaborators at the Veprintsev lab at the University of Nottingham using a TR-FRET assay. It should be noted, that all results presented in the following are preliminary and were only measured once. In this assay, a reduction of the FRET signal was observed for 14 compounds. However, a closer evaluation of these results revealed the interference of four to five of these compounds with the assay readout and, thus, four compounds were labelled as false positives and one compound was marked as questionable. For the other nine compounds either no assay interference was observed or although the compound interfered with the assay readout at high concentrations, it could clearly be identified as a ligand of the CB2. For each of these nine potential hit molecules, k_{on}, k_{off} and $K_D{}^{kinetic}$ values were calculated based on the time resolved data and K_i values were obtained from the equilibrium binding dose-response curves (see Table 5.2). The $logK_D{}^{kinetic}$ and $logK_i$ values are comparable for each compound between these two determination methods.

Of the nine potential hit molecules, six showed K_D values that were estimated to be higher than 1 µM and can, therefore, be considered as ligands with a low affinity for the CB2. The other three compounds seem to have K_D values below 1 µM and can be considered as compounds with higher affinities for the CB2. It should be noted though, that these results are based on one measurement and for some compounds on few data points only, since higher concentrations showed assay interference. Therefore, K_D values and results might still change after more measurements of

Table 5.2: Binding affinity estimations for the nine potential hits from the ZINC15 library docking screen against the CB2. A TR-FRET assay was conducted to measure affinity of the compounds for the CB2. k_{on}, k_{off} and $K_D^{kinetic}$ values were calculated from the time resolved FRET signal. Additionally, K_i values were calculated from the equilibrium binding dose-response curves. All listed values are estimations and were measured with n=1. The assay was conducted and analysed by Leire Borrega Román and David Sykes from the lab of Prof. Dmitry Veprintsev at the University of Nottingham.

Compound	k_{on} ($M^{-1}min^{-1}$)	k_{off} (min^{-1})	$K_D^{kinetic}$	$logK_D^{kinetic}$	$logK_i$
K030MS004	$1.2 \cdot 10^5$	1.4	$1.2 \cdot 10^{-5}$	-4.9	-4.9
K030MS005[b]	$4.8 \cdot 10^7$	1.6	$3.4 \cdot 10^{-8}$	-7.5	-7.5
K030MS009[b]	$2.2 \cdot 10^6$	1.2	$5.5 \cdot 10^{-7}$	-6.3	-6.3
K030MS013	$1.3 \cdot 10^5$	2.2	$1.7 \cdot 10^{-5}$	-4.8	-4.7
K030MS014	$7.8 \cdot 10^4$	1.1	$1.4 \cdot 10^{-5}$	-4.9	-4.9
K030MS019	$1.1 \cdot 10^5$	2.8	$2.5 \cdot 10^{-5}$	-4.6	-4.6
K030MS021	$5.9 \cdot 10^5$	3.6	$6.1 \cdot 10^{-6}$	-5.2	-5.2
K030MS022	$3.5 \cdot 10^5$	1.9	$5.6 \cdot 10^{-6}$	-5.3	-5.3
K030MS025[b]	$4.9 \cdot 10^7$	1.1	$2.3 \cdot 10^{-8}$	-7.6	-7.6
K030MS028[a]	$3.5 \cdot 10^4$	0.2	$6.3 \cdot 10^{-6}$	-5.2	-5.1

a Molecule might be a false positive.

b Values estimated based on data points of three or less usable concentrations.

the same compounds have been conducted.

5.4 Discussion

5.4.1 Docking optimisation and water position prediction

The binding pocket of the CB2 consists mostly of apolar residues while the polar residues are often out of reach of the ligands. In the CB2 crystal structure with PDB ID 5ZTY it even seems like the amide group of ligand AM10257 does not interact with the receptor and seems stranded –a molecule pose that would immediately be discarded during docking evaluation. This ligand moiety is in proximity to the polar residue $S285^{7.39}$ but too far from it to form a hydrogen bond (see Figure 5.4). Two possible conclusions can be drawn from this observation of the ligand binding pose. First of all, ligand binding in the orthosteric binding pocket of the CB2 is mostly caused by apolar interactions while polar interactions do not seem to play a major role in this. This can also be concluded from a look at the binding pocket itself and from the mostly apolar ligands

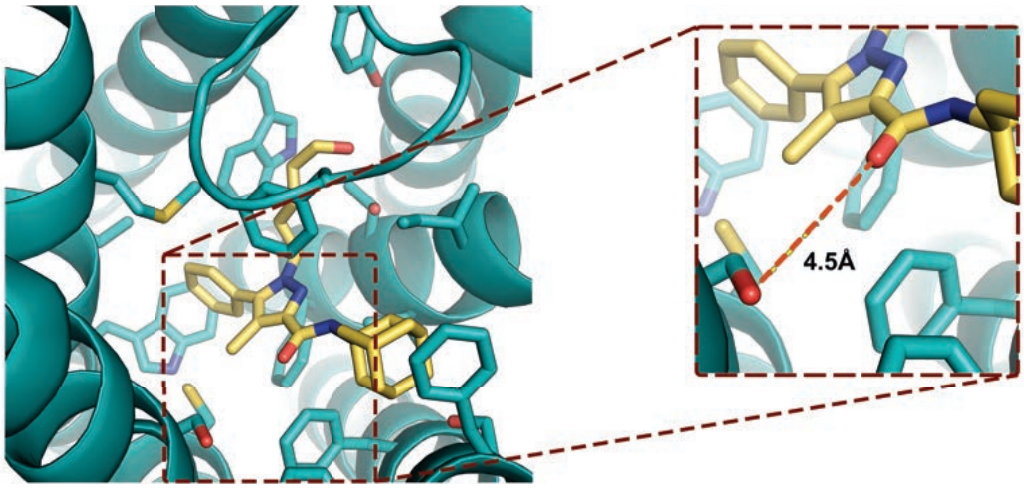


Figure 5.4: Binding pose of ligand AM10257 in the CB2 structure with PDB ID 5ZTY. The amide group does not form any polar contacts and is, thus, 'stranded'. The enlargement of this area shows that the amide-carbonyl of the ligand has a distance of approximately 4.5 Å to the receptor residue S285$^{7.39}$. This distance is too big for a direct polar interaction between ligand and receptor but a water molecule could be placed there and mediate polar interactions between both.

of the CB2 (agonist JWH133 even has no polar functional groups whatsoever, see Table A5.1). Secondly, there might be water molecules in the orthosteric binding pocket that were not resolved in crystallography and that could mediate polar contacts between the amide-carbonyl of AM10257 and residue S285$^{7.39}$. This led to the idea to predict water molecule positions in the binding pocket to allow for mediated ligand receptor interactions and thereby improve molecule docking poses.

Prediction of water molecules. Three tools were used to predict possible water positions: SEED, OpenEye SZMAP and the MOE Solvent Analysis. While SEED is not particularly good at predicting water positions, it can be used to find polar interaction sites. The interaction sites for small polar fragments (Methanol, Ethanol) were used as an estimation for water positions. Interestingly, all three methods predicted placement of small polar fragments or water molecules between S285$^{7.39}$ and the amide-carbonyl of AM10257 which is in concordance with the bridging water molecule hypothesis mentioned above. Furthermore, both OpenEye SZMAP and the MOE Solvent Analysis placed water molecules in the same spot in which a crystallographically resolved water molecule can be found in structure 5ZTY. These results indicate, that the used water position prediction methods yield reasonable results. OpenEye SZMAP and the MOE Solvent Analysis were then also applied to the other two structures used for docking (PDB IDs 6PT0 and 6KPF).
Alternatively to these prediction methods, a more extended approach using short MD simulations could have been used to predict possible water positions. Due to the reasonable results using the

other tools this was not considered further, however, with the data from the subsequent MD simulations described in chapter 6, such an analysis of water positions in the orthosteric binding pocket could be considered further.

Choice of the template ligand. The results from the docking calculations of the reference ligands to 5ZTY showed, that even when including the predicted water molecules no reasonable poses for ligands with a cannabinol-like structure could be obtained (i.e. the ligands had stranded donors). To improve the docking poses of these molecules it was decided to try using a different template ligand for the generation of the DOCK spheres. Ligand AM841 crystallised with the CB1 (PDB ID 5XR8) has a cannabinol-like structure and a truncated version of this ligand was used as template ligand for the DOCK preparation of 5ZTY. The docking poses of cannabinol-like ligands were improved in these docking setups. These results show that the choice of the template ligand can influence docking results and induce bias into docking screens. Although this could have been expected to a certain degree based on how the DOCK algorithm places ligands on spheres that were generated from the template ligand, it is still surprising how much the chosen template ligand influenced the results. However, in this project part the main goal was to optimise molecule poses of the reference ligands which was accomplished using this approach, and the potentially introduced bias does not have a major influence here.

Improvement of docking results. Overall, the placement of water molecules in the predicted positions led to an improvement of docking results with less stranded donors and more potential interactions between molecules and receptor for most of the docked molecules. Furthermore, it was also shown that the choice of the template ligand for DOCK sphere generation can influence the docking results and improve docking poses. The employed methods can, therefore, be deemed as successful, independently from the results of the subsequent ZINC15 library screen.

5.4.2 Novel ligands of the CB2 by docking screen predictions

In the second part of the project the ZINC15 drug-like library was screened against five of the optimised docking setups to find novel ligands of the CB2 with a preferably low similarity to known CB2 ligands. During the selection process, several re-ranking and filtering steps were employed (see Figure 5.3). Finally, 29 molecules were selected to be tested *in vitro* and a first measurement regarding the affinity of the selected molecules for the CB2 was conducted in the Veprintsev lab at the University of Nottingham. Although the results are preliminary and might change during further testing, they will already be discussed here.

Hit rates. The preliminary binding affinity data shows that 9-10 of the compounds are potential hit molecules. This would result in a hit rate of 9-10 of 29 tested molecules (31-34%). When

considering only the compounds that were estimated to have a K_D below 1 µM, the hit rate would be 3 of 29 molecules or 10%. Since the main aim of this screen was to find ligands with novel structural scaffolds compared to known CB2 ligands, a lower hit rate could be expected. Therefore, the screen can be considered very successful with the observed hit rate even if several of the compounds show a rather low affinity. This is also confirmed in comparison to the hit rates achieved in the study described in chapter 3[83] or other screens against GPCR targets.[190,193–198]

Docking setups. The tested compounds were selected from docking results of a docking screen of a molecular library against five different docking setups of CB2 structures in different conformations. The potential hit molecules resulted from the docking calculations against different of these setups: two hits from the $5ZTY^{S285}$, one hit from the $5ZTY^{AM841}$, two hits from the $6PT0^{no\text{-}water}$, four hits from the $6PT0^{S285}$ and one questionable hit from the $6KPF^{S285\text{-}L182}$, i.e. three molecules from docking calculations against the CB2 in an inactive conformation and seven molecules from docking calculations against the CB2 in active conformations. Therefore, it would be expected to find three antagonists and seven agonists towards the G_i signalling pathway in an assay measuring effector protein recruitment or downstream signalling. It should be considered though that low affinity agonists might not induce a measurable response and could be classified as antagonists. Furthermore, these results and the high hit rate might indicate that the placed water molecules and the docking optimisation steps indeed helped to improve the docking results and achieve higher hit rates. Additionally, the re-ranking potentially improved the enrichment of ligands in the top ranks, since four of the ligands were selected from re-ranking lists. Of the higher affinity compounds, two molecules were selected from docking calculations to the CB2 in an inactive conformation (one from the $5ZTY^{S285}$ and one from the $5ZTY^{AM841}$) and one was selected from the docking calculation to the CB2 in an active conformation ($6PT0^{S285}$), i.e. it would be expected that two are antagonists and one is an agonist.

Novelty of potential hit molecules. During the screening, a special focus was directed towards novelty of the selected compounds as ligands of the CB2. For that, the ECFP4 Tanimoto coefficients to a set of known ligands of the CB2 retrieved from ChEMBL were calculated and all compounds with a Tanimoto similarity equal or below 0.45 to any of the molecules from the ChEMBL dataset were considered as structurally novel for ligands of the CB2. Two of the hit molecules have a Tanimoto coefficient above 0.45 to molecules in the ChEMBL dataset (K030MS0013: 0.49 to CHEMBL371117; K030MS021: 0.83 to CHEMBL466429) and would not be considered as structurally novel. The other eight compounds –including the compounds with the highest estimated affinities for the CB2– show a Tanimoto similarity of 0.45 or less to the molecules from the ChEMBL dataset (lowest is 0.33 between K030MS005 and CHEMBL497325 and highest is 0.45 between K030MS025

and CHEMBL2441472) and can, therefore, be considered as ligands with novel structural scaffolds. Hence, the goal to find ligands with novel structural scaffolds for ligands of the CB2 was reached.

Ligand binding poses. When evaluating the docking poses of the molecules with the highest estimated affinity for the CB2, it can be observed that the two molecules retrieved from the docking calculations to the 5ZTY^{S285} and the 5ZTY^{AM841} (K030MS005 and K030MS009) both interact with the water molecule placed close to S285$^{7.39}$ via an amide-carbonyl group, similar to what is observed for the crystallised ligand AM10257 (see Figure 5.5.A and B). Both molecules contain a similar structural motif which is, however, connected to other molecule parts in a different way. Judging by their docking pose, both ligands probably entertain aromatic/apolar interactions with F117$^{3.36}$, F183^{ECL2}, W194$^{5.43}$ and F94$^{2.64}$. Other aromatic and apolar residues are also close enough to the molecules to stabilise its binding pose via apolar interactions, and K030MS009 might also interact with W258$^{6.48}$. The molecule selected from the docking calculation to the CB2 in an active conformation (6PT0^{S285}; K030MS025) might form a polar interaction to H95$^{2.65}$ and is otherwise stabilised by apolar interactions with mainly F117$^{3.36}$, F183^{ECL2} and W194$^{5.43}$ (see Figure 5.5.C). Its overall pose is slightly different from the poses observed for the other two molecules, however, at the current state of the pharmacological characterisation of these compounds no further conclusions can be drawn from this.

Another potential hit has also an interesting docking pose. Compound K030MS014, which was selected from a docking calculation to the 6PT0$^{no\text{-}water}$, is only a weak binder with an estimated logK_D of -4.9, however, its docking pose differs significantly from that of the other molecules. This molecule binds at the top of the orthosteric binding pocket, only separated from the solvent by the N-terminus of the CB2. It forms polar interactions with the backbone carbonyl of I27$^{N\text{-}term}$ and the side chain of Q32$^{1.31}$ via an amide functional group and is otherwise only stabilised by apolar contacts (see Figure 5.5.D). Although this is only a predicted binding pose and might not be the actual binding mode of this molecule, it is an interesting starting point for further investigations since a different binding mode might lead to interesting observations regarding the induced downstream signalling. Based on further assay results this molecule could be interesting to use for further developments and SAR studies.

5.5 Conclusions

Overall, the study described in this chapter was very successful. To optimise the docking setups, water positions in the orthosteric binding site of the CB2 were predicted and in the case of 5ZTY the structure was also prepared using a different template ligand. These optimisation steps indeed resulted in improved molecule docking poses of the reference molecules. This shows that the

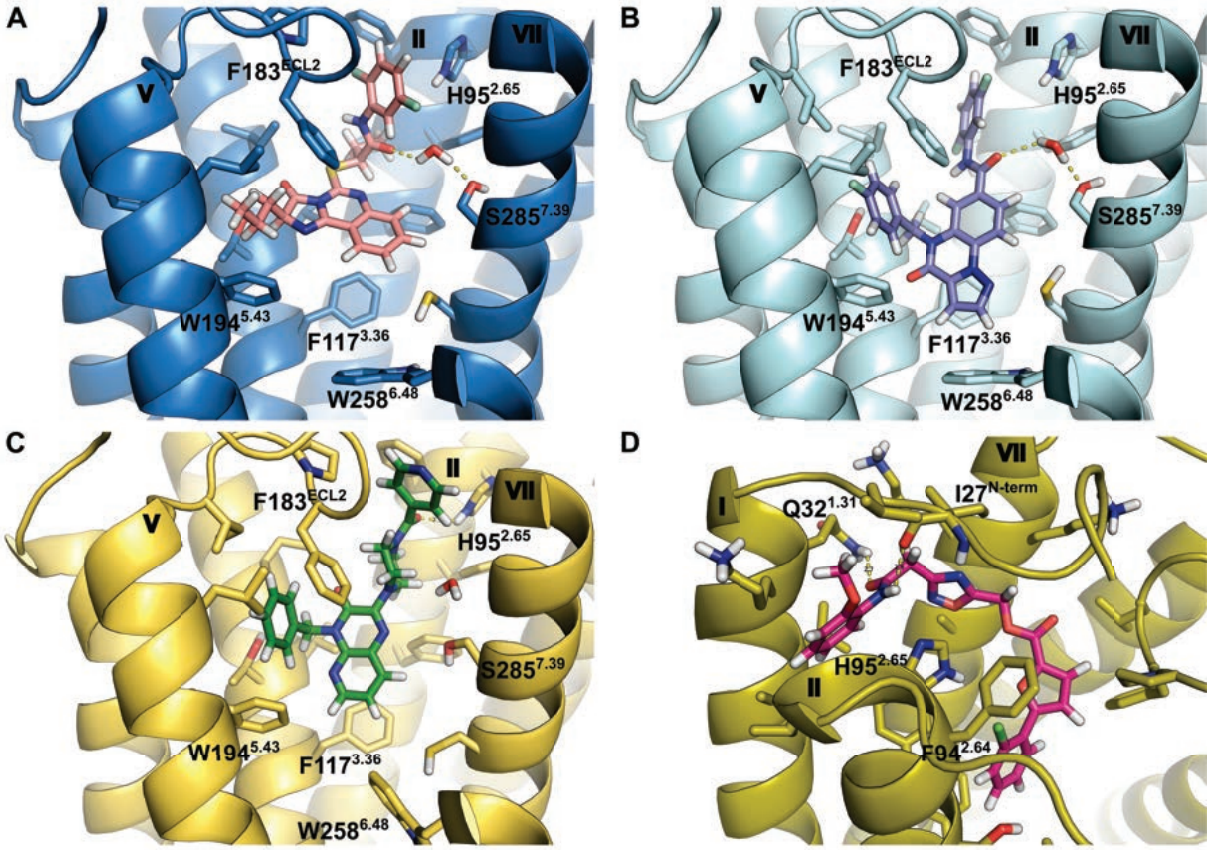


Figure 5.5: Docking poses of four potential hits from the ZINC15 drug-like library screen against the CB2. The compounds in A-C) have the highest estimated affinities for the CB2, while the compound in D) shows an interesting binding pose. TMs are numbered at the top of each helix and some residues are labelled for better orientation in the binding pocket. A) Compound K030MS005 was selected from a docking calculation against the $5ZTY^{S285}$. It forms a polar interaction with the water molecule placed close to $S285^{7.39}$. B) Compound K030MS009 was selected from a docking calculation against the $5ZTY^{AM841}$. Like the compound in A) it forms polar interactions with the water molecule placed close to $S285^{7.39}$. C) Compound K030MS025 was selected from the docking calculation to the $6PT0^{S285}$. It might form a polar interaction to $H95^{2.65}$ and is expected to act as an agonist towards the G_i signalling pathway since it was selected from a docking to the CB2 in an active conformation. D) Compound K030MS014 has a different binding pose compared to the other molecules. It is not located within the orthosteric binding pocket but rather at its edges. The molecule might form polar interactions with the backbone carbonyl of $I27^{N\text{-term}}$ and the side chain of $Q32^{1.31}$.

placement of water molecules in the binding site can improve docking results, especially if the binding site is overall very apolar. Furthermore, these improved docking setups were then used to screen a molecular library for novel ligands of the CB2. The preliminary results of affinity binding assays conducted with molecules selected from this docking screen suggest 9-10 potential hits out of 29 molecules and a hit rate of 31-34%. This hit rate is higher than expected for a docking screen aiming to find structurally novel ligands, indicating that the optimisation of the docking setups also improved the screening results. Furthermore, one of the potential hit molecules has an interesting binding pose in the docking calculations which might potentially lead to interesting observations regarding downstream signalling. In conclusion, it can be worth investing time into optimisation of docking setups to improve docking screening results, and the placement of water molecules predicted by diverse tools can be one useful methods to achieve this.

5.6 Perspectives

The study described in this chapter can be used as a basis for further developments. The molecule binding poses obtained from the docking calculations of the reference molecules to the different docking setups can be used for an interaction analysis. The interaction profile obtained for each ligand can then be connected to its *in vitro* behaviour regarding recruitment of different effector proteins. This data can then be used to search for patterns which can potentially be applied for predictions of recruitment bias induced by known or novel ligands. This analysis is currently being worked on by our collaborators at the Veprintsev lab at the University of Nottingham.

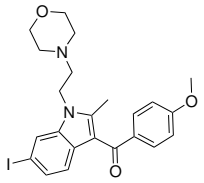
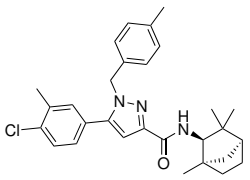
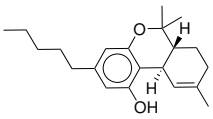
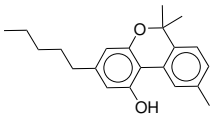
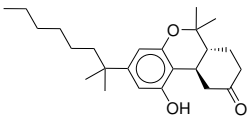
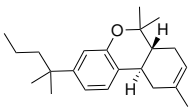
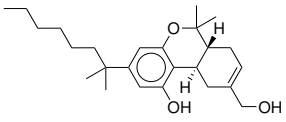
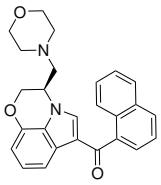
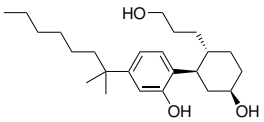
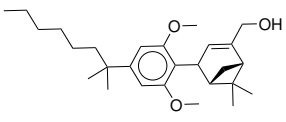
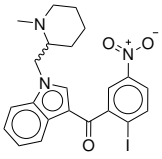
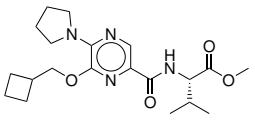
The results of the docking screen can also serve as a basis for further projects. First of all, the compounds need to be characterised more thoroughly and the affinities of the potential hits for the CB2 need to be confirmed by repeating the TR-FRET assays. Additionally, the efficacy of these compounds should be evaluated either by measuring downstream signalling or recruitment of effector proteins. This further characterisation is currently conducted by our collaborators at the Veprintsev lab at the University of Nottingham. Based on the results of these assays, the most interesting novel ligands can be selected and developed further by searching for structurally similar compounds and evaluating their SAR. This might lead to novel insights on CB2 ligands and the activation of the receptor and could potentially also result in novel potent ligands.

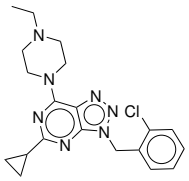
One interesting starting point for an SAR analysis might also be the potential hit molecule for which a different binding mode was found in the docking calculations (K030MS014). Molecules with a different binding mode to the usually observed molecule poses can potentially also induce a different downstream signalling. The found compound only shows a low affinity for the CB2 and it might be difficult to observe downstream effects for this compound. In spite of that, it might still

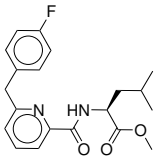
be interesting to modify this compound with functional groups aiming to increase its affinity for the CB2 and, thereby, increasing the probability to observe downstream signalling effects. However, first a more detailed pharmacological characterisation is necessary to confirm this potential hit as a ligand for the CB2.

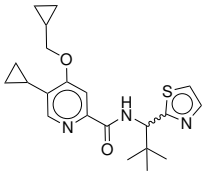
5.7 Additional information

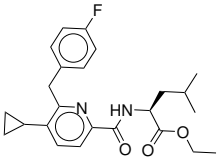
Table A5.1: List of known reference ligands of the CB2 that were tested for ligand-induced recruitment of different effector proteins by the Veprintsev lab.[208] These molecules were then used as reference molecules during the optimisation of docking setups and single molecules from this set were selected to be included in the MD simulations described in chapter 6.

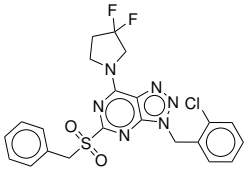
Name	Structure	Name	Structure
AM630[a]		SR144528[a,b]	
THC		Cannabinol	
Nabilone		JWH133	
HU210		WIN55212-2	
CP55940[b]		HU308[b]	
AM1242[c]		RO5135445	

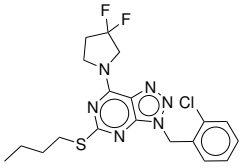
RO6435559 

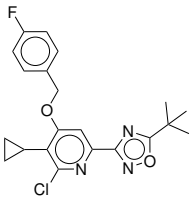
RO6844112[b] 

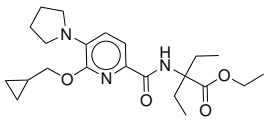
RO6850007[c] 

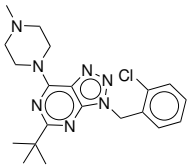
RO6853973 

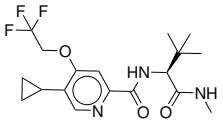
RO6871487 

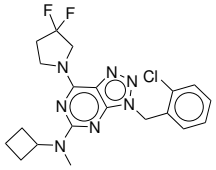
RO6883666 

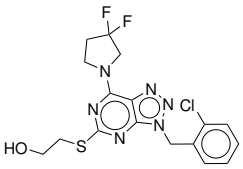
RO6926274 

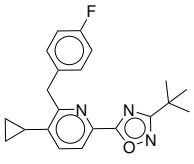
RO6843766 

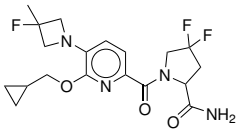
RO6844395 

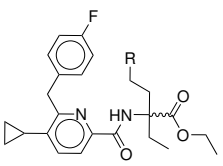
RO6853457 

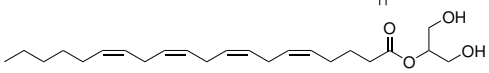
RO6869094 

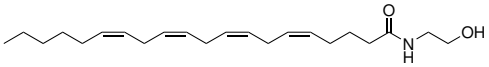
RO6878558 

RO6892033 

RO7032019 

RO723469x[c,d] 

2-AG[e] 

Anandamide[e] 

---

a Molecule is an inverse agonists.

b Molecule was selected for the MD simulations.

c Both enantiomers were included in the docking calculation.

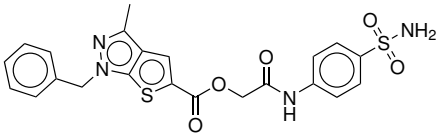
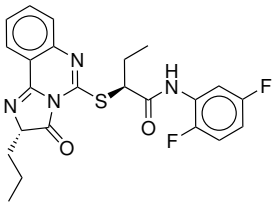
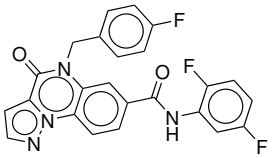
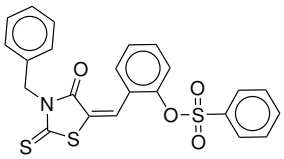
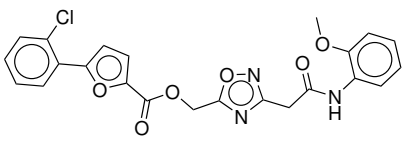
d Molecule was truncated compared to originally tested molecules. At position *R* a linker of varying length to a fluorescence tag would be attached.

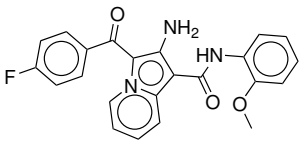
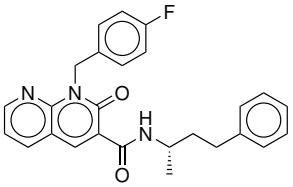
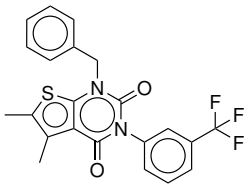
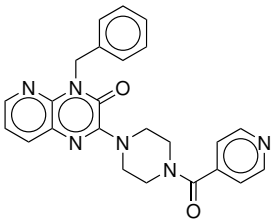
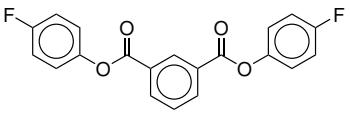
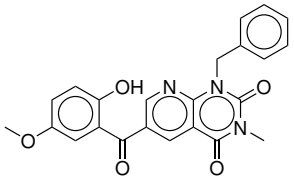
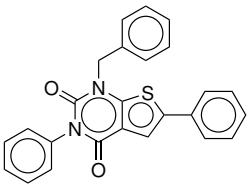
e Molecule was not included in docking calculation due to high flexibility.

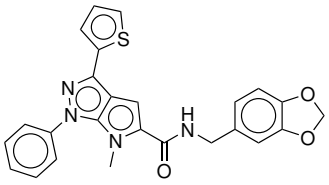
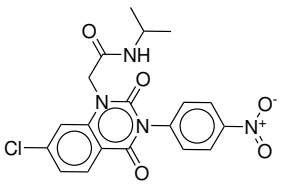
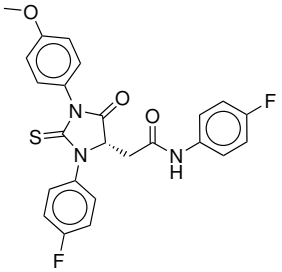
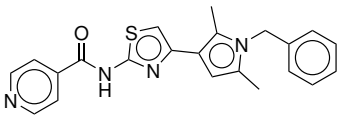
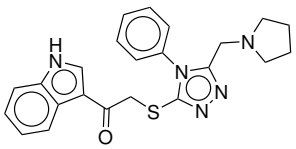
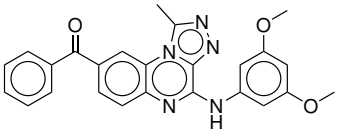
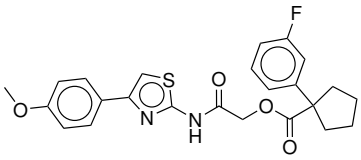
Table A5.2: Restraints applied to the CB2 and parts of the structure while minimising during the docking preparation. Hydrogen atoms were not restrained unless stated otherwise. *harm* is the harmonic force constant used for the positional restraints. The higher the value of this constant is, the more restraint is an atom to its position. Water molecules were oriented to allow for potential interactions prior to minimisation. Residues in proximity to the placed water molecule were included in most cases to allow for an adaptation of potential interactions. For more details on the minimised structures see also Table 5.1.

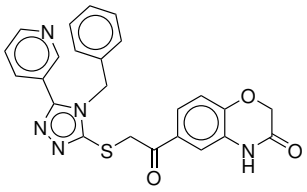
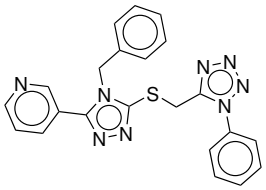
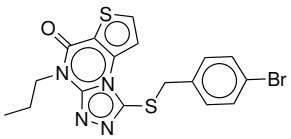
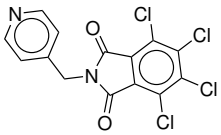
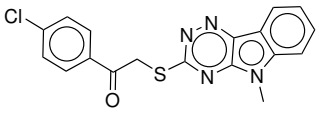
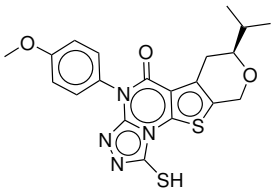
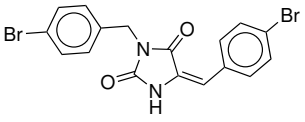
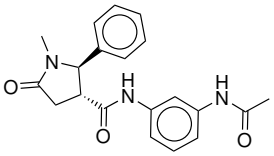
No.	PDB ID	Restraints
1a	5ZTY	*harm*=1 on repaired residues and H atoms, rest fixed
1b	5ZTY	All atoms fixed except water molecule, $Y190^{5.39}$ and ligand-OH
1c	5ZTY	All atoms fixed except water molecule
1d	5ZTY	All atoms fixed except water molecule, in presence of the ligand
1e	5ZTY	All atoms fixed except water molecule, in absence of the ligand
1f	5ZTY	*harm*=1 on repaired residues and H atoms, rest fixed
1g	5ZTY	no restraints on water molecule; *harm*=1 on ligand, $S285^{7.39}$; *harm*=2 on residues in 6Å distance of the ligand and repaired residues; *harm*=5 on all other atoms
2a	6PT0	*harm*=1 on repaired residues and H atoms, rest fixed
2b	6PT0	*harm*=2 on water molecule, $L182^{ECL2}$, H atoms and repaired; *harm*=5 on all other atoms
2c	6PT0	*harm*=2 on water molecule, $F281^{7.35}$, $S285^{7.39}$, H atoms and repaired residues; *harm*=5 on all other atoms
3a	6KPF	*harm*=1 on repaired residues and H atoms, rest fixed
3b	6KPF	*harm*=2 on water molecule, $L182^{ECL2}$, H atoms, repaired and lig-CH_2-OH; *harm*=5 on all other atoms
3c	6KPF	*harm*=2 on water molecule, $F281^{7.35}$-backbone, $S285^{7.39}$, H atoms and repaired residues; *harm*=5 for all other atoms
3d	6KPF	*harm*=2 on water molecule, $L182^{ECL2}$-backbone, $F281^{7.35}$-backbone, $S285^{7.39}$, H atoms, lig-CH_2-OH and repaired; *harm*=5 for all other atoms

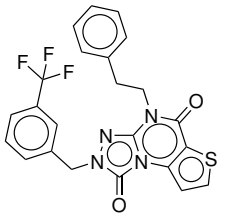
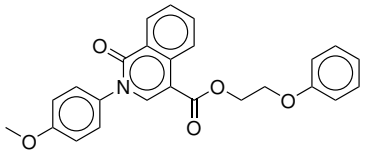
Table A5.3: List of molecules that were selected from the INC15 docking screen targeting the CB2. Compound IDs, ZINC IDs and 2D-depictions are listed as well as the docking setup from which the compound was chosen. The last column indicates whether the compound can be considered to have a novel structural scaffold for ligands of the CB2, i.e. whether or not the compound has an ECFP4 Tanimoto similarity ≤ 0.45 to a set of known ligands of the CB2 retrieved from ChEMBL. The compounds were sorted according to potential hits or non-binders as determined by the preliminary assay results provided by the Veprintsev lab.

Compound	ZINC ID	Structure	Docking setup	Tanimoto $\leq$0.45
Potential hits.				
K030MS004	ZINC 000011066749		5ZTY^{S285}	Yes
K030MS005	ZINC 000021936235		5ZTY^{S285}	Yes
K030MS009	ZINC 000169681222		5ZTY^{AM841}	Yes
K030MS013	ZINC 000009296638		6PT0$^{no\text{-}water}$	No
K030MS014	ZINC 000021446877		6PT0$^{no\text{-}water}$	Yes

K030MS019	ZINC 000004585236		6PT0^{S285}	Yes
K030MS021	ZINC 000015883342		6PT0^{S285}	No
K030MS022	ZINC 000020151725		6PT0^{S285}	Yes
K030MS025	ZINC 000064578980		6PT0^{S285}	Yes
K030MS028	ZINC 000000851644		6KPF$^{S285\text{-}L182}$	Yes
Potential non-binders.				
K030MS002	ZINC 000003324004		5ZTY^{S285}	Yes
K030MS003	ZINC 000004335477		5ZTY^{S285}	No

K030MS006	ZINC 000033277421		5ZTY^{S285}	Yes
K030MS007	ZINC 000096291888		5ZTY^{S285}	Yes
K030MS008	ZINC 000001443485		5ZTY^{AM841}	No
K030MS010	ZINC 000006944261		6PT0$^{no\text{-}water}$	Yes
K030MS011	ZINC 000008726148		6PT0$^{no\text{-}water}$	Yes
K030MS012	ZINC 000009224706		6PT0$^{no\text{-}water}$	Yes
K030MS018	ZINC 000003283505		6PT0^{S285}	Yes

K030MS020	ZINC 000008861306		$6PT0^{S285}$	Yes
K030MS023	ZINC 000032741611		$6PT0^{S285}$	Yes
K030MS024	ZINC 000035529165		$6PT0^{S285}$	Yes
K030MS029	ZINC 000000984147		$6KPF^{S285\text{-}L182}$	Yes
K030MS030	ZINC 000001069258		$6KPF^{S285\text{-}L182}$	No
K030MS031	ZINC 000002314407		$6KPF^{S285\text{-}L182}$	Yes
K030MS033	ZINC 000009260762		$6KPF^{S285\text{-}L182}$	No
K030MS034	ZINC 000020745274		$6KPF^{S285\text{-}L182}$	Yes

K030MS035	ZINC 000064946340		6KPF$^{S285\text{-}L182}$	Yes
K030MS036	ZINC 000064977159		6KPF$^{S285\text{-}L182}$	No

6 | Evaluation of structural indicators for recruitment bias using MD simulations of the CB2

Contributions. The author of this study evaluated ligand recruitment data to identify ligands that induced recruitment bias and conducted and analysed the Molecular Dynamics simulations. The *in vitro* assays of the reference compounds used as a basis for the computational efforts were performed and analysed by Tamara Miljuš from the Veprintsev lab at the University of Nottingham.[208]

6.1 Introduction and goal of the study

The study described in this chapter is closely connected to the study described in the previous chapter 5. Again, the general aim is to find structural clues on how and why some ligands induce a recruitment bias of certain effector proteins over others. While in the previous chapter a rather static docking approach was applied using the fixed receptor conformations of crystal structures, a more dynamic approach was conducted in this chapter by applying Molecular Dynamics (MD) simulations. The recruitment data for a set of reference compounds provided by the Veprintsev lab (see also Table A5.1 for molecule structures) was analysed to identify the ligands inducing a certain recruitment bias. Based on these results, four molecules were selected as biased or unbiased reference ligands for the MD simulations. The CB2 was then simulated in complex with different effector proteins (none, $G_{i,\alpha}$ or β-arrestin 2) and the selected ligands. The optimised docking setups and obtained docking poses of the selected molecules from the study described in chapter 5 were used as initial starting points for these MD simulations. The MD simulations are intended to give further insights into differences of ligand-receptor interactions for ligands inducing a different recruitment bias. Furthermore, they might yield receptor conformations of the different activation states with different effector proteins which might be helpful for the prediction of novel ligands with a certain bias based on docking calculations. Finally, changes of the receptor conformation and ligand binding poses might elucidate differences between the ligands and their binding behaviour and might give further hints why certain ligands induce certain responses upon binding to the receptor.

Previous work. As described in the introduction of chapter 5, recruitment data of five different members of the $G_{i/o}$ subfamily (G_{i1}, G_{i2}, G_{i3}, G_{oA} and G_{oB}) as well as β-arrestin 1 and β-arrestin 2 for a set of known ligands of the CB2 (see Table A5.1 for molecule structures) was provided by the Veprintsev lab at the University of Nottingham.[208] The E_{max} values resulting from these assays were used to identify ligands with a bias towards recruitment of a certain effector protein. These

ligands were then used in the MD simulations described in this chapter.

6.2 Methods

6.2.1 Analysis of the correlation of effector protein recruitment

As described above, recruitment of different effector proteins (G_{i1}, G_{i2}, G_{i3}, G_{oA}, G_{oB}, β-arrestin 1 and β-arrestin 2) to the receptor after stimulation with different agonists was measured by the Veprintsev lab using BRET recruitment assays and the data was kindly forwarded to us.[208] This data was used to analyse whether any of the 35 tested molecules induced a recruitment bias, i.e. stronger or weaker recruitment of one effector protein over another effector protein after stimulation with the compound. For this analysis, the E_{max} values from the recruitment data of each combination of the measured effector proteins were plotted for all molecules and the plots analysed for visible correlations between the datasets of the different effector proteins. Molecules deviating from the observed correlations were chosen and marked as biased towards recruitment of one or the other effector protein. The results from all correlation plots were compiled and molecules inducing recruitment of certain effector protein subfamilies identified. The molecules used for the MD simulations were selected based on these results.

6.2.2 Molecular dynamics simulations

Structure preparation. Three different CB2 structures in complex with different effector proteins were prepared for MD simulations, namely inactive (unbound; MD^{inact}), $G_{i,\alpha}$ bound (MD^{Gi}) or β-arrestin 2 bound ($MD^{\beta Arr}$). The inactive structure was simulated with two different ligands (RO6844112, SR144528) while both $G_{i,\alpha}$ bound and β-arrestin 2 bound structures were simulated with four different ligands (RO6844112, SR144528, HU308, CP55940). Residue $E50^{1.49}$ which is pointing into the membrane was protonated to be uncharged for all complex structures.

For the MD^{inact}, the CB2 structure with PDB ID 5ZTY was prepared. The missing intracellular loop 3 (ICL3), which consists of 12 residues, was modelled using MOE.[211]The protonation states of histidines from docking were used for the MD simulations and hydrogens were added to the structure but not minimised. A water molecule close to $S285^{7.39}$ from the water position prediction was placed with the same orientation as used for docking (cf. chapter 5). The allosteric sodium ion that is known to bind to the inactive state of Class A GPCRs[21] was placed in proximity to $D80^{2.50}$ by copying it from the D4 dopamine receptor structure with PDB ID 5WIV.[218]

For the MD^{Gi}, receptor and $G_{i,\alpha}$ subunit conformations from PDB ID 6PT0 were used. The proteins were protonated and histidine protonation suitable for the local environment was selected for

the $G_{i,\alpha}$ subunit while for the receptor the protonation states from docking were used. A water molecule was placed close to $S285^{7.39}$ in the same position and orientation as used for the docking calculations described in chapter 5.

For the $MD^{\beta Arr}$, the receptor conformation was taken from PDB ID 6PT0. Since there is no structure of β-arrestin 2 in complex with a GPCR available yet, the structure and orientation was modelled from β-arrestin 1 in complex with the muscarinic receptor 2 (PDB ID 6U1N[219]) using Modeller[220] which was appropriate due to the high sequence similarity between β-arrestin 1 and β-arrestin 2 (77% identity between β-arrestin 1_{rat} and β-arrestin 2_{human}). Clashes between receptor and β-arrestin 2 were resolved by minimising the residues in the interface area using CHARMM and the CHARMm22 force field[209] in witnotp (Novartis Pharma AG, unpublished). During minimisation a positional restraint with a harmonic force constant of *harm*=1 was applied to the side chains of the residues at the CB2-β-arrestin 2 interface and *harm*=3 to the backbone of these residues while all other atom positions were fixed. The proteins were then protonated and histidine protonation suitable for the local environment was selected for β-arrestin 2 while for the receptor the protonation states were used as for docking unless a different protonation was more reasonable due to a changed local environment. Again, a water molecule was placed close to $S285^{7.39}$ in the same position and orientation as used for the docking calculations described in chapter 5.

The initial ligand poses of SR144528, RO6844112 and HU308 were derived from the docking calculations to the respective CB2 structures (5ZTY or 6PT0) containing the water molecule close to $S285^{7.39}$ as described in the previous chapter 5. Since no good docking pose was obtained for CP55940, the ligand pose was taken from the CB1 crystal structure with PDB ID 6KQI[221] and minimised in the binding pocket of CB2 structure 6PT0 containing the water molecule close to $S285^{7.39}$ using CHARMM and the CHARMm22 force field[209] in witnotp (Novartis Pharma AG, unpublished). During this minimisation the receptor atoms and the water molecule were fixed to their positions while a positional restraint with a harmonic force constant of *harm*=1 w as applied to the ligand CP55940.

System setup. The system including water molecules, ions and the lipid bilayer was prepared using the CHARMM-GUI Membrane Builder.[222–225] Ligands were parametrised using the CHARMM General Force Field (CGenFF).[226,227] Chain termini were patched as ACE (N-terminus) or CT3 (C-terminus) except for the N-terminus of the $G_{i,\alpha}$ subunit which was left unpatched. Hydrogen coordinates were preserved and the disulfide bond between $C174^{ECL2}$ and $C179^{ECL2}$ were specified. In the inactive structure, $D80^{2.50}$ was deprotonated while it was protonated in the structures in complex with the effector proteins and $E50^{1.49}$ was protonated in all structures. Post-translational modifications were added to residues G2 (myristoylated) and C3 (palmitoylated) of the $G_{i,\alpha}$ sub-

unit. Residues T48 and S360 of the β-arrestin 2 were phosphorylated with a dianionic phosphate group. The protein complexes were placed in a bilayer consisting of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and the z-axis was oriented along a vector through C288$^{7.42}$ and S285$^{7.39}$. The box size was chosen to accommodate the complete protein complex and 0.15 M NaCl was added to the aqueous phase.

Structure and system setup with SR144528. The system setup with ligand SR144528 was different due to certain limitations and compatibility problems of the used ligand parametrisation and the simulation software ACEMD (v.3.2.4).[189] In CHARMM-GUI which was used for the MD system setup is the CGenFF v.4.1 for ligand parametrisation implemented. In this force field a dummy atom is added to halogen atoms to account for the σ-hole in the electron density of halogens. However, this dummy atom is not yet implemented in ACEMD causing the simulations to fail. For this reason an alternative preparation route was used to be able to simulate the chlorine-containing SR144528 within ACEMD. For the preparation of the system with CHARMM-GUI the chlorine atom in SR144528 was first replaced by a fluorine atom to avoid addition of the dummy atom to the system and the system was prepared as described above. In a subsequent step, the fluorine atom was again replaced by the chlorine atom to revert the ligand to the correct structure. Additionally, partial charges in the .psf files were replaced with partial charges from the general parameter file of CGenFF v.3.0.1 using parameters of *p*-chlorotoluene.

Equilibration and Production. The systems were equilibrated with NAMD (v.2.12)[228] using the CHARMM36 force field[229] and input parameters as generated by CHARMM-GUI.[222,230] The production simulations were run with ACEMD (v.3.2.4)[189] using the CHARMM36 force field.[229] Each system was run in three independent replicates based on the individually equilibrated systems and each replicate was simulated for a duration of 1 µs with time steps of 4 fs. The simulation temperature was set to 300 K and maintained using a Langevin thermostat. The Particle Mesh Ewald (PME) method was used for the calculation of long-range electrostatic interactions. For Van-der-Waals and short-range electrostatic interactions a cut off of 9 Å was used.

Analysis. The trajectories of each replicate were pre-processed by wrapping the frames, aligning them to the first frame of the replicate and removing all lipids using VMD (v.1.9.2).[231] For each system, the trajectories of the three replicates were then concatenated and analysed using cpptraj (v.15.00).[232] The frames were clustered into five clusters based on the binding pocket residues as well as based on the ligand pose using the *cluster* command. Overall changes of TMs, receptor and effector protein parts and the ligand were monitored using the *rms*, *atomicfluct* and *2drms* commands. Distances between atoms were monitored using the *distance* command and dihedral angles were determined along the C_α-C_β bond using the *dihedral* command. Contact frequencies

between ligand and protein were calculated based on different properties. Potential hydrogen bonds were determined using the *hbond* command. To measure potential aromatic interactions, a vector was placed orthorgonal to the plane fitted to the heavy atoms of the aromatic ring using the *vector* command. The dot-product between the vectors of an aromatic ring pair (*vectormath* command) as well as the distance of their centers of mass (*distance* command) was then calculated. The frequency of the potential interactions was subsequently determined based on the fraction of frames in which distance and angle between the observed aromatic ring pair were within a certain threshold (distance ≤ 5.5 Å and $150° \leq$ angle $\leq 30°$ for potential face-to-face interactions; distance ≤ 5.5 Å and $60° \leq$ angle $\leq 120°$ for potential face-to-edge interactions). The fraction of native contacts was determined between the ligand and all atoms of apolar residues (Phe, Tyr, Val, Ile, Ala, Leu, Trp) within 4 Å distance around the ligand using the *nativecontacts* command. For the analysis only native contacts with a frequency above 5% were considered.
RMSDs between the structures in different clusters were calculated using UCSF Chimera.[233] Structures were aligned to the C_α atoms of the TMs using the *MatchMaker*[234] and RMSDs for TMs, binding pocket residues, ligand and W258$^{6.48}$ were calculated using the *rmsd* command. The representative structures of the main cluster were compared for systems with the same ligand and for systems with the same effector protein. Other clusters were included if their RMSD was notably high compared to the main cluster. Additionally, all four available crystal structures were compared to each other in the same way.

6.3 Results

6.3.1 Identification of ligands inducing a recruitment bias

To analyse whether any of the 35 molecules measured at the Veprintsev lab induced a stronger or weaker recruitment of one effector protein or effector protein subfamily to the CB2 compared to other effector proteins/subfamilies, the correlation between the different pathways (i.e. recruitment of the different effector proteins) was evaluated.[208] For that, the E_{max} values of each of the measured effector protein recruitments (G_{i1}, G_{i2}, G_{i3}, G_{oA}, G_{oB}, β-arrestin 1 and β-arrestin 2) were plotted against each other for each of the tested molecules. The data was taken from Miljuš[208] and will not be shown in this thesis.
In general, a linear correlation between all pathways could be observed (see Figures A6.1 and A6.2). This linear correlation was more distinct when comparing the recruitment of the different G protein subtypes to each other than when plotting G protein recruitment against β-arrestin recruitment. For single molecules some data points deviated from the observed correlation and could, therefore,

be marked as inducing stronger recruitment of one of the effector proteins over the other effector protein in the same plot. This data was then compiled to single out tendencies for certain molecules towards recruitment of a certain effector protein or effector protein subfamily compared to the other effector proteins.

Strong trends throughout different correlation plots were visible for seven of the 35 tested molecules. While RO7032019 induced overall weaker recruitment of G_{i1} proteins compared to all other effector proteins, Nabilone did the same for G_{i2} proteins and RO7234694 for β-arrestin 1. Interesting behaviour was also observed for the two endogenous ligands 2-AG and Anandamide with a stronger recruitment of either β-arrestin compared to G proteins for the first while the latter only showed stronger recruitment of β-arrestin 1. A recruitment bias for one of the β-arrestins, in this case β-arrestin 2, compared to all other measured effector proteins was also observed for CP55940. The opposite case, inducing a stronger recruitment of G proteins over β-arrestins, was observed for HU308. In this analysis, only the E_{max} values were considered and not the EC_{50} values since the focus was on the maximum recruitment of an effector protein but not the concentration of ligand needed to induce the recruitment. A strong correlation was also seen for the EC_{50} values when comparing the different pathways (plots not shown).

6.3.2 Analysis of the MD simulations

To explore changes in the receptor structure and in protein-ligand interaction patterns for ligands with different recruitment bias, MD simulations were employed. Various systems of the receptor in complex with different effector proteins (simulation based on an inactive receptor conformation without effector protein, MD^{inact}; simulation based on an active receptor conformation in complex with $G_{i,\alpha}$, MD^{Gi}; simulation based on an active receptor conformation in complex with β-arrestin 2, $MD^{\beta Arr}$; see Figure 6.1.A) and different ligands (see Figure 6.1.B) were simulated and analysed. The selection of the ligands for these MD simulations was based on the results of the pathway correlation analysis described above and a diverse set of four ligands with a bias towards recruitment of different effector proteins was chosen. As an unbiased reference the agonist RO6844112 was selected, while CP55940 and HU308 were selected as biased reference compounds with a recruitment bias towards β-arrestin 2 or G proteins, respectively. The inverse agonist SR144528 was selected as a negative control showing no recruitment of any of the effector proteins.

RMSD and 2D-RMSD plots. After aligning and concatenating the frames of all replicates for each system, certain structural features were analysed using cpptraj to evaluate the overall behaviour of each system and to compare with other systems. When analysing the RMSD of certain parts of the

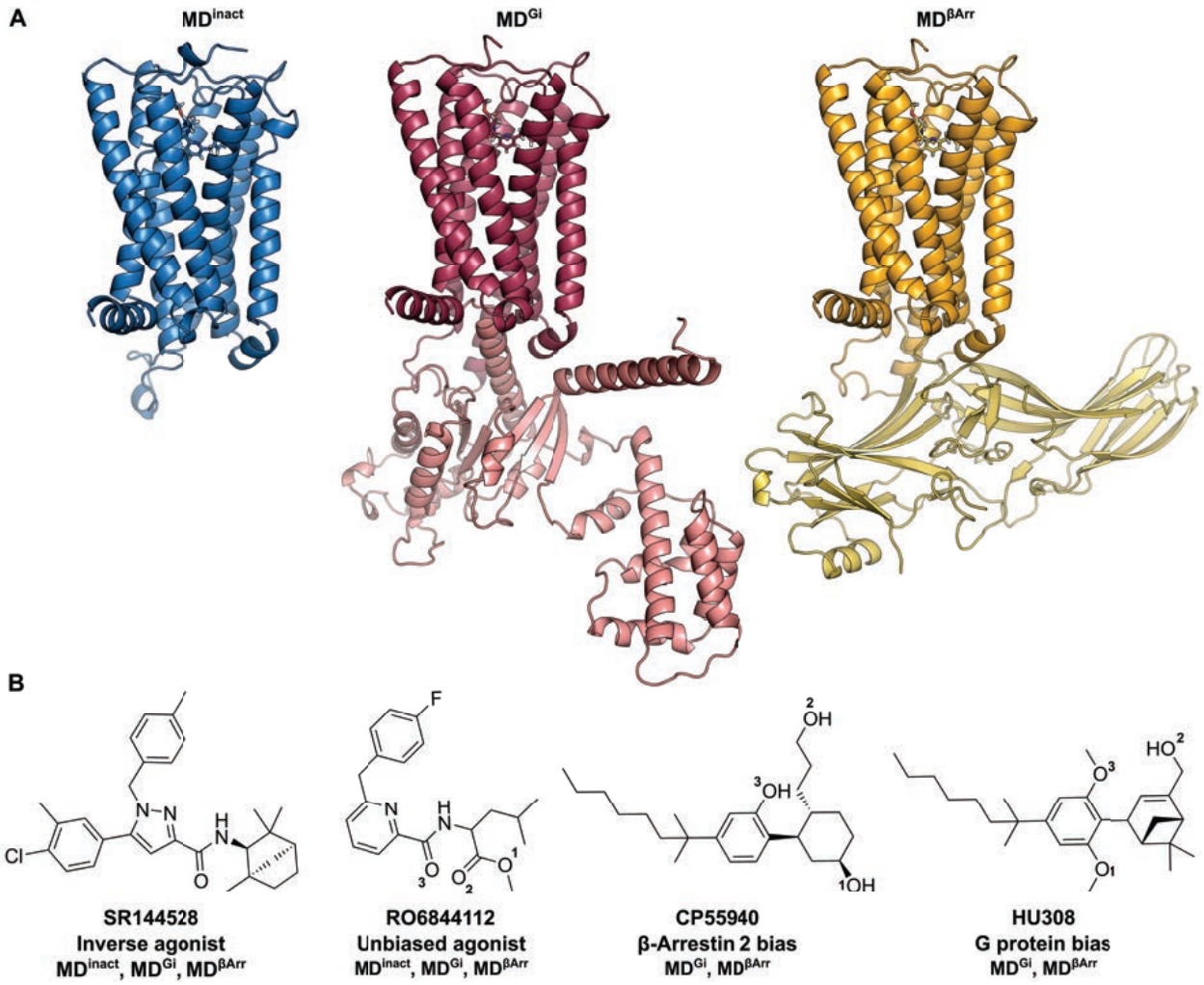


Figure 6.1: Depiction of the simulated CB2 complexes and ligands. A) Three different systems of the CB2 were simulated: an inactive conformation of the CB2 without effector protein (blue), MD^{inact}; the CB2 in an active conformation (dark red) in complex with a $G_{i,\alpha}$ protein (salmon), MD^{Gi}; and the CB2 in an active conformation (yelloworange) in complex with β-arrestin 2 (light yellow), $MD^{\beta Arr}$. B) 2D-depictions of the four CB2 ligands selected for the MD simulations. The ligands were selected based on experimental results provided by the Veprintsev lab.[208] SR144528 is an inverse agonists, RO6844112 was selected as the unbiased reference agonist and CP55940 and HU308 were selected based on their recruitment bias towards β-arrestin 2 or G proteins, respectively. The ligands were simulated in complex with the MD^{inact}, the $MD^{Gi}$and the $MD^{\beta Arr}$as indicated below the 2D depictions. Numbers next to the oxygen atoms are as used for the *hydrogen* bond analysis.

structure (excluding the effector proteins) using the first frame as a reference, all systems seemed stable without notable changes. Only the intracellular loop 3 (ICL3) showed changes in RMSD of up to 6 Å throughout the simulations (plots not shown). Variations in the RMSD of TM V were observed in the MD^{inact} for both ligands SR144528 and RO6844112.

This behaviour of TM V could also be observed in the 2D-RMSD plots (see Figure A6.3-A6.6.B). An RMSD above 2 Å of this helix was mainly observed for the MD^{inact} and with a higher RMSD for agonist RO6844112 compared to inverse agonist SR144528. Single replicates of other systems also showed higher RMSDs for TM V (SR144528 and HU308 in the MD^{Gi} and HU308 in the $MD^{\beta Arr}$). RMSDs for the overall receptor, the TM-C_α atoms, TM VI-C_α atoms and the ligand poses were comparable in all systems based on the 2D-RMSD plots (see Figure A6.3-A6.6.A and E; results for overall receptor and TM VI-C_α atoms not shown). An exception is the MD^{inact} with RO6844112 for which a higher RMSD for the TM-C_α atoms was observed compared to the results of the other simulations.

Higher RMSDs for the binding pocket residues can mainly be observed for ligands SR144528 and RO6844112 (see Figure A6.3.C and Figure A6.4.C). Interestingly, these differences are higher for agonist RO6844112 compared to inverse agonist SR144528 in the MD^{inact}. The 2D-RMSD plots also revealed differences between replicates of the same system. The biggest differences were observed for the replicates of the $MD^{\beta Arr}$ with ligand HU308 (see Figure A6.6). This ligand adopted substantially different poses in different replicates of the $MD^{\beta Arr}$ which is discussed further below.

Cluster analysis. The protein complex was clustered into 5 clusters, once based on the conformation of the binding pocket residues and once based on the ligand binding pose. RMSDs between the main clusters as well as clusters with a notable difference from the main cluster were calculated (see Tables A6.1-A6.4). The aim of this analysis was, to see whether one of the effector proteins or one of the ligands induced a significantly different conformation of the overall receptor or the binding pocket and to compare the ligand poses in the different systems. For that the structures were first aligned based on the TM-C_α atoms and the RMSDs between the TM-C_α atoms, the binding pocket residues and – if applicable – the ligand poses were then calculated. Additionally, RMSDs were calculated for the toggle switch residue $W258^{6.48}$ to evaluate its movement. As a comparison, the four available crystal structures of the CB2 were compared to each other using the same procedure (Table A6.1). While most RMSDs of the clusters were in the same range as for the crystal structures, there were a few interesting observations.

Differences of the RMSD of toggle switch residue $W258^{6.48}$ can be observed between the clusters of the simulations with inverse agonist SR144528 compared to the ones with the agonists (see Table A6.4). While the RMSD of this residue is below 1 Å when comparing the main clusters of any

of the agonist simulations to each other, it is is notably higher when comparing the main clusters of the simulations with inverse agonist SR144528 to the clusters of the simulations with the three agonists (RMSD above 1.5 Å). This is interesting since this residue is believed to play a role in receptor activation and should, thus, take a different conformation in complexes with an inverse agonist. Of note, when simulating agonist RO6844112 with the inactive structure this residue also moves to a different conformation compared to the simulations with SR144528 (comparing the main clusters only). This observation was also reflected in the results from the analysis of the dihedral angles, as will be described further below.

While for most systems a substantially more populated cluster was found, this is not the case for RO6844112 when clustered based on the binding pocket residue conformations. In all systems with this ligand the main cluster contained less than 50% of the frames and for the MD^{inact} and $MD^{\beta Arr}$ a second cluster based on at least 15% of the frames showed a higher RMSD to the main cluster than other clusters, especially in the binding pocket (see Table A6.3. This phenomenon was not observed for other ligands. Interestingly, while these differences for RO6844112 could be seen when clustering based on the binding pocket conformation, this was not observed when clustering based on the ligand pose (see Table A6.2). This could indicate that while the pose of ligand RO6844112 is quite stable, the binding pocket is not stabilised in one certain conformation. This observation also corresponds with the observations from the 2D-RMSD plots, where higher RMSD values for the binding pocket residue side chains as well as backbone could be observed throughout all simulations with this ligand (see Figure A6.4.C and D). It should be noted, that the conformation of $W258^{6.48}$ is also rather different in these clusters compared to the main clusters and this residue even moves completely out of the binding pocket in the case of the second analysed cluster for the MD^{inact}.

Aside from that, only in two other cases clusters substantially different to the main cluster were observed. In the case of SR144528 in the MD^{Gi} a second cluster of the ligand conformation was observed, leading to an only slightly changed binding pocket conformation (see Table A6.2). A second considerably different cluster was also observed for ligand HU308 in the $MD^{\beta Arr}$. Approximately one third of the frames form a second cluster (based on binding pocket residues as well as ligand pose) with a significantly different ligand binding pose (RMSD of 5.2-5.6 Å compared to the main cluster and the main cluster from the MD^{Gi}, see Tables A6.2 and A6.3; in the further text referred to as alternative binding pose) and slight differences in the binding pocket. This is also reflected in the 2D-RMSD plots which indicate substantial differences of the binding pocket and the ligand pose for one of the three replicates (see Figure A6.6.C-E). The binding pose of HU308 is notably different in this replicate compared to the other two replicates and the binding pose of this ligand in the MD^{Gi} (see Figure 6.2). In the replicate with the alternative binding pose the ligand

still adopts a quite similar pose to the one from the equilibration after the first time step, although already migrating towards the final pose (see Figure 6.2.A and B). The final poses are then reached after 50 time steps for the alternative pose and almost after the first time step for the other two replicates (see Figure 6.2.B). These poses then seem to be quite stable, as can also be seen from the similarity of the ligand poses after the last time step compared to the centroid poses from the ligand based clustering (see Figure 6.2.C).

Dihedral angles. The dihedral angles along the C_α-C_β bond were monitored for a number of binding pocket residues to evaluate differences between the different simulated systems (see Figures A6.7-A6.10). Overall, the dihedral angles in all systems were similar with a few exceptions. In all systems the dihedral angles of the apolar residues I110$^{3.29}$ and L182^{ECL2} fluctuated considerably. Interestingly, the dihedral angle of residue S90$^{2.60}$ was mainly changing in the MD^{inact} and in the simulations with ligand CP55940 which is the only of the ligands to form polar interactions with this residue. In the MD$^{\beta Arr}$ with ligand HU308 a different dihedral angle of this residue was observed in the majority of the frames of the replicate with the alternative ligand pose compared to the other two replicates (see Figure A6.10.B). Fluctuations of the dihedral angle of residue S285$^{7.39}$ which is proposed to interact with most of the ligands were mainly observed for agonist RO6844112 and for inverse agonist SR144528 (see Figures A6.7 and A6.8). However, both of these ligands can only form hydrogen bonds via a carbonyl-group which is probably not able to form strong polar interactions with the receptor.

Changes in the dihedral angle of toggle switch residue W258$^{6.48}$ can mainly be observed for the MD^{inact} with agonist RO6844112 (see Figure A6.8) which is correspondent with the results from the RMSD and cluster analysis. Interestingly, for inverse agonist SR144528 no changes of the dihedral angle of this residue are visible except for one of the replicates of the MD$^{\beta Arr}$ (see Figure A6.7) . However, the RMSD values of this residue in the cluster analysis would suggest more differences. This could imply a slight movement of the TM VI backbone in this area instead of a movement of only residue W258$^{6.48}$. A different dihedral angle of this residue can also be observed for the MD^{Gi} with ligand HU308 which is not reflected in the results of the cluster analysis.

The dihedral angles of the phenylalanine residues are quite stable throughout all simulations. Changes are mainly seen for the MD^{inact} (F87$^{2.57}$, F91$^{2.61}$, F94$^{2.64}$, F117$^{3.36}$) and MD^{Gi} (F87$^{2.57}$, F94$^{2.64}$, F281$^{7.35}$) with ligand RO6844112 (Figure A6.8). In other systems, fluctuations are mainly seen for F117$^{3.36}$, F281$^{7.35}$ and F 94$^{2.64}$, however, not to the same extent.

Interaction analysis. To evaluate whether interactions and contacts between ligand and receptor might play a role in receptor activation, potential hydrogen bond and aromatic contacts as well as native contacts were analysed. It should be noted that these hydrogen bonds and aromatic

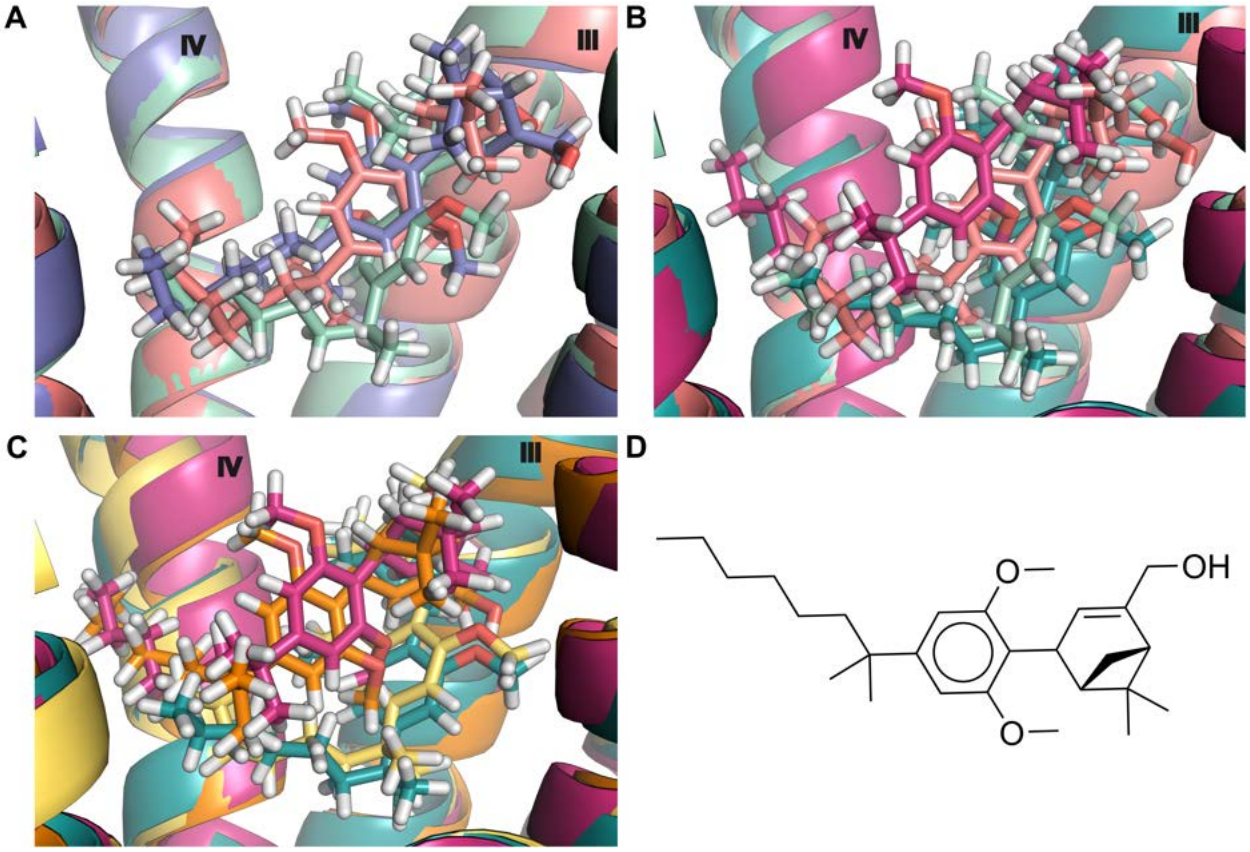


Figure 6.2: Ligand binding poses of HU308 in the different replicates of the $MD^{\beta Arr}$. A) In the alternative replicate (salmon) the ligand pose differs only slightly from the pose after equilibration (blue-violet) after the first time step. In the other two replicates the ligand adopts a different pose (mint; pose of one replicate shown as representative for both). B) The ligand moves to an entirely different binding pose in the alternative replicate (pink; after the last time step) which was already indicated after the first time step (salmon). In the other two replicates the final ligand pose (teal) is almost the same as observed after the first time step (mint). C) After clustering based on the ligand pose, approximately a third of the frames forms a cluster with a binding pose (orange) which is similar to the final ligand pose of the alternative replicate (pink). The final ligand pose of the other two replicates (teal) is similar to the centroid pose of the main cluster (yellow) which is formed by approximately two thirds of the frames. This pose is also similar to the binding pose of HU308 from the MD^{Gi}. A-C) TM III and IV are labelled on top of each helix. TM VI would be located between the observer and the ligand and was hidden for clarity. D) 2D depiction of HU308.

interactions are only calculated based on a definition of vector angle and distance ranges for such interactions and might, therefore, not reflect true interactions. However, for an ease of readability these potential interactions will be called *hydrogen* bonds and aromatic *face-to-face* or *face-to-edge* interactions in the following passages. Additional to these potential interactions, the native contacts between ligand and apolar residues within a 4 Å distance were determined. Evaluating the frequencies of *hydrogen* bond contacts for each ligand, only small differences are visible for the simulations with different activation states and effector proteins (see Figure A6.11). Interestingly, the frequency of *hydrogen* bond contacts between the receptor and ligands SR144528 and RO6844112 is quite low (below 15% except for one replicate in the $MD^{\beta Arr}$ and one in the MD^{Gi} of SR144528) especially in comparison to the other two ligands. Each of the ligands entertains *hydrogen* bonds with water molecules. The water molecule that was placed in the initial system to mediate contacts between ligand and protein does not seem to stay in this position throughout the simulation. A comparison of the *hydrogen* bonds of each of the ligands to the ones of another ligand cannot be done due to the different number of hydrogen bond donors and acceptors of each of these ligands. Ligand HU308 in $MD^{\beta Arr}$ shows huge variations of *hydrogen* bond frequencies in the different replicates which were a result of the alternative binding pose.

More differences between interaction frequencies can be observed for the *face-to-face* and *face-to-edge* interactions (see Figure A6.12). While interaction frequencies for the MD^{Gi} and $MD^{\beta Arr}$ are similar for almost all observed aromatic ring pairs and all ligands, differences can be seen for the MD^{inact}. The trends of these differences are the same for both SR144528 and RO6844112 with a few exceptions. For ligand SR144528 the *face-to-face* interaction frequency between $F183^{ECL2}$ and the ligand-pyrazole is higher in the MD^{Gi} compared to the MD^{inact} and the $MD^{\beta Arr}$, while the *face-to-edge* interaction frequency of this aromat pair is higher in the MD^{inact}. These trends are different compared to the equivalent interactions of $F183^{ECL2}$ to the RO6844112-pyridine. Additionally, for RO6844112 a frequent *face-to-edge* interaction of $F117^{3.36}$ to the ligand-pyridine can be observed (frequency up to 50%), especially in the MD^{Gi} and the $MD^{\beta Arr}$. The equivalent *face-to-edge* interaction of $F117^{3.36}$ to the SR144528-pyrazole cannot be observed. It should be noted, that the aromatic contacts with the RO6844112-pyridine and the SR144528-pyrazole were treated as equivalent based on their similar position when comparing the ligand binding poses and the fact that only contact frequencies are observed, however, due to the different electronical properties of these aromatic rings the interactions themselves are not comparable. Interaction frequencies of CP55940 and HU308 do not differ between the MD^{Gi} and $MD^{\beta Arr}$, except for the replicate of the HU308-$MD^{\beta Arr}$ in which the ligand adopted the alternative pose.

To estimate apolar interactions between the ligands and the receptor, the frequency of native con-

tacts was also evaluated. First of all, the total number of contacts (above a frequency threshold of 5%) in all three replicates were evaluated. It should be noted, that the number of contacts corresponds to the number of atom-atom contacts, i.e. ligands with more atoms likely have a higher number of contacts. Therefore, these numbers should not be compared between ligands but only for the same ligand in different systems. The inverse agonist SR144528, which is also the ligand with the highest molecular weight, has the highest number of contacts with 425 contacts in the MD^{inact}. Both ligand SR144528 and RO6844112 have a higher number of contacts in the MD^{inact} (SR:425; RO:346) compared to the MD^{Gi} (SR:373; RO:246) and the $MD^{\beta Arr}$ (SR:281; RO:273). While the number of contacts is comparable between the MD^{Gi} and the $MD^{\beta Arr}$ for agonists RO6844112 (246/273) and CP55940 (315/276), this is not the case for inverse agonist SR144528 (373/281) and agonist HU308 (325/208). However, the alternative ligand binding pose of HU308 in the $MD^{\beta Arr}$ has to be considered here. If the total number of contacts is analysed per replicate, some differences between the replicates become visible. Differences of almost 100 contacts can be observed between the replicates of the $MD^{\beta Arr}$ of SR144528 and CP55940 and the MD^{Gi} of RO6844112, while the number of contacts differs by more than 100 for the MD^{inact} of RO6844112 and both the MD^{Gi} and the $MD^{\beta Arr}$ of HU308. While these results might allow some deductions about the systems it should be kept in mind that this is only the total number of contacts and does not tell anything about the frequency of contacts and the distribution of contacts among the apolar binding pocket residues. The frequency of contacts per residue was then plotted for each replicate and each system, to evaluate differences (see Figure A6.13-A6.15). While the contact frequencies for each residue seem quite consistent for SR144528 and CP55940 when comparing the replicates to each other, they are less consistent for RO6844112 and considerably less consistent for HU308. Of note, this inconsistency for HU308 cannot only be observed for the $MD^{\beta Arr}$ and does not seem to be linked to the alternative binding pose of HU308 observed in the $MD^{\beta Arr}$.

Contacts with residues $F87^{2.57}$, $F91^{2.61}$ and $F94^{2.64}$ are considerably less frequent for ligand RO6844112 compared to the other ligands and seem more frequent in the $MD^{\beta Arr}$ compared to the MD^{inact} and MD^{Gi}. RO6844112 also seems to have more frequent contacts with $I110^{3.29}$ and less frequent contacts with $F281^{7.35}$ in the MD^{inact} compared to the MD^{Gi} and $MD^{\beta Arr}$. Contacts with $F117^{3.36}$ are frequent for all the agonists while inverse agonist SR144528 shows low contact frequencies in the MD^{Gi} and the $MD^{\beta Arr}$ and high contact frequencies in the MD^{inact}. Contacts with $F183^{ECL2}$ are more frequent for RO6844112 and SR144528 compared to CP55940 and HU308. CP55940 shows more frequent contacts with $W258^{6.48}$ and $V261^{6.51}$ in the $MD^{\beta Arr}$ compared to the MD^{Gi} but overall less frequent than observed for RO6844112 and especially SR144528. Interestingly, no contacts can be observed between $W258^{6.48}$ and HU308. Compared to all other ligands HU308

also shows less frequent contacts with residues Y190$^{5.39}$, L191$^{5.40}$ and W194$^{5.43}$. However, a detailed analysis of the contacts of HU308 is more complex due to the high inconsistency between the different replicates.

6.4 Discussion

6.4.1 Correlation of effector protein recruitment and identification of ligands inducing recruitment bias

To evaluate the correlation of the ligand-induced recruitment of different effector proteins, the E_{max} values of the different pathways were plotted against each other (see Figures A6.1 and A6.2). The results of this analysis led to three main conclusions. First of all, a correlation between the different pathways could be observed, especially when comparing the recruitment of the different G proteins while the correlation of G protein and β-arrestin recruitment is not as distinct. This suggests that most ligands induce recruitment of effector proteins in a similar –i.e. unbiased– way. The second conclusion is that there are certain ligands that induce a stronger or weaker recruitment of one effector protein or effector protein subfamily compared the other effector proteins. Several ligands showed clear trends throughout the different plots and it can be assumed that they are recruitment biased ligands. The ligands for the MD simulations were selected based on this assumption. The third and not really surprising conclusion is that most compounds do not show recruitment bias of different effector proteins. However, the number of 4-7 biased compounds (depending on the definition of bias) out of a total of 35 measured compounds (11-20%) probably still exceeds expectations.

6.4.2 Structural indicators for recruitment bias from MD simulations

MD simulations of the CB2 in complex with different ligands and effector proteins were carried out to look for structural indications for recruitment bias of different effector proteins. Ligands with a different recruitment bias (see Figure 6.1) were selected based on experimental assay data provided by the Veprintsev lab[208] and simulated in complex with the CB2 in an inactive conformation (MD^{inact}) or in complex with the CB2 in an active conformation and either the $G_{i,\alpha}$ protein (MD^{Gi}) or β-arrestin 2 ($MD^{\beta Arr}$). In specific, the focus was on an analysis of protein-ligand interactions and conformational changes within the binding pocket or overall structural differences. The aim was to identify potential clues why a ligand induces a biased recruitment and to potentially generate a 'β-arrestin biased conformation' of the CB2 that could be used for a further structural analysis or structure based approaches such as a docking screen to find biased ligands.

Comparability of the simulated ligands. The ligands that were selected for the MD simulations are depicted in Figure 6.1. Inverse agonist SR144528 and unbiased agonist RO6844112 are structurally more similar to each other than to the two other molecules. They contain an amide which links an aromatic ring to a bulkier, alkylic molecule part. Both molecules are rather apolar with the amide being the only molecule part to potentially form polar interactions in SR144528, while RO6844112 contains an additional ester moiety which could also form polar contacts. The other two molecules, HU308 and CP55940, are structurally similar to each other. Both are more polar than SR144528 or RO6844112 and can form stronger polar contacts/hydrogen bonds via hydroxyl-groups. Especially CP55940 can form strong polar interactions with the receptor since it has three hydroxyl-groups which might also be a reason for its different recruitment behaviour. Judging from this comparison of the molecule structures it might be difficult to compare interaction patterns between the molecules and draw conclusions regarding recruitment bias from this. However, since the inverse agonist SR144528 and the unbiased agonist RO6844112 are structurally comparable it might be possible to use the results obtained from these simulations to gain some insights to receptor activation and inactivation. In the same way, the comparison of interactions and structural observations from the simulations with CP55940 (biased towards β-arrestin 2 recruitment) and HU308 (biased towards G protein recruitment) might reveal some estimations regarding a certain bias, since these two ligands have comparable structures. It should be noted, that any conclusions drawn from simulations using only four ligands with different properties should be treated with caution, since certain observations might not be reproduced for other ligands.

Overall structural changes. Different structural aspects of the MD simulations were analysed, mainly focusing on the ligand and the binding pocket. Overall structural changes were only briefly analysed to ensure that the overall system behaved as expected. Only slight changes of the overall receptor structure were observed, which was to be expected. The movement of TM V can only be observed in the MD^{inact} since this helix is probably stabilised by the effector proteins in the MD^{Gi} and $MD^{\beta Arr}$. The ICL3 is known to be quite flexible and to move throughout simulations.[24] Thus, it was to be expected to see RMSD changes of this loop in all systems. No overall structural changes were observed that would hint to an artificially strange behaviour of the system.

Activation and inactivation of the receptor. Agonist RO6844112 was simulated with the inactive conformation of the CB2 and inverse agonist SR144528 was simulated with active conformations of the receptor. In both cases, overall structural changes towards the other activation state could not be expected since it the ligand can influence but not dictate activations states[24] This means that an agonist alone cannot stabilise a completely activated conformation of the receptor (i.e. similar to the conformation of the receptor in complex with a G protein) while an inverse agonist

cannot lead to an inactive receptor conformation if the receptor is in complex with an effector protein. This is what was also observed in the simulation results. Interesting is, however, the behaviour of the toggle switch residue W258$^{6.48}$. This residue has been suggested to be involved in the activation of class A GPCRs such as the CB2.[118–121] In the starting structures of the MDs (and all available crystal structures; see Figure 2.8) this residue has an entirely different conformation in the inactive and active conformations of the receptor. While it is pointing into the binding pocket in the active conformation of the receptor, it is turned away to leave more space for the ligand in the inactive conformation of the receptor (see Figure 6.3). Based on the analysis of the structural clusters of the systems, this residue behaves differently when simulated with an agonist or inverse agonist. In the MD$^{\text{inact}}$ with agonist RO6844112 the movement of W258$^{6.48}$ from the inactive conformation to an active-like conformation or an entirely different conformation –where the residue is completely moving out of the binding pocket– can be observed (see Figure 6.3.A). These changes are not observed for the MD$^{\text{inact}}$ with inverse agonist SR144528, where the residue stays in the inactive conformation. The reverse case can be observed for the MD$^{\text{Gi}}$ and MD$^{\beta\text{Arr}}$, in which residue W258$^{6.48}$ stays in the active conformation during the simulations with any of the agonists while it moves out of the binding pocket when simulating with inverse agonist SR144528 (see Figure 6.3.B and C). The residue does not turn and flip into a more inactive-like conformation though. This movement of W258$^{6.48}$ is probably caused by the bulky chlorinated aromatic ring of SR144528 pushing the residue away and out of the binding pocket.

Interesting is also the movement of the TM VI backbone in proximity of this residue. The biggest differences to the starting structures in the MD$^{\text{inact}}$ can be observed for the alternative residue conformation resulting from the simulation with RO6844112 (see Figure 6.3.A, mint structure). Although the movement of this residue completely out of the binding pocket might be an artefact (after all the clusters were built on the entire binding pocket and not just W258$^{6.48}$) it still shows the instability of W258$^{6.48}$ in an inactive conformation when simulated with an agonist. In reverse, in the MD$^{\text{Gi}}$ and the MD$^{\beta\text{Arr}}$ the biggest changes of the TM VI backbone around W258$^{6.48}$ can be seen when simulating with inverse agonist SR144528. This observation could mean that the energy barrier for the W258$^{6.48}$ rotation from an active-like to an inactive-like state is too big to be overcome in the simulation which leads to the movement of the entire helix backbone to still move the residue out of the binding pocket. In the reverse case, the energy barrier for the rotation of W258$^{6.48}$ from an inactive to an active-like conformation seems to be lower since this flip can be seen for the MD$^{\text{inact}}$ with agonist RO6844112. This is particularly interesting since these conformational changes of W258$^{6.48}$ are associated with receptor activation and from these observations it seems like once the residue is in an active conformation it is not as easy to deactivate again while the

opposite process seems much easier to accomplish. This is somewhat contrary to the expectation that the inactive conformation would be easier to stabilise than the active conformation. Possible interpretations for this might be that either the toggle switch residue W258$^{6.48}$ does not play an as important role in receptor activation as initially assumed or that the CB2 might in fact be easier to activate than to deactivate with either needing a longer time span than simulated or higher energies to overcome barriers to move W258$^{6.48}$ from an active to an inactive conformation. However, it should also be kept in mind that the active receptor conformations in the MD simulations were stabilised by effector proteins which might also reflect in the stabilisation of a more active-like binding pocket conformation in the MD$^{\text{Gi}}$ and MD$^{\beta\text{Arr}}$ with SR144528.
Additional to these observations regarding toggle switch residue W258$^{6.48}$, the analysis of the dihedral angles revealed changes for residues F87$^{2.57}$, F91$^{2.61}$ and F94$^{2.64}$ in the MD$^{\text{inact}}$ with agonist RO6844112. These residues were shown to play a role in receptor activation[112] and the observed changes of their dihedral angles might also indicate the transition to a more active-like binding pocket conformation when simulating an agonist in the MD$^{\text{inact}}$. It should be noted though, that only changes in dihedral angles were evaluated and not the final dihedral angles and residue conformations. Furthermore, changes in the dihedral angles of F87$^{2.57}$ and F94$^{2.64}$ were also observed for the MD$^{\text{Gi}}$ with the same ligand.

RMSD and cluster analysis. The 2D-RMSD and cluster analysis for the unbiased agonist RO6844112 (see Figure A6.4 and Tables A6.2-A6.4) suggest that while the ligand pose is rather stable throughout the simulations, the conformation of the binding pocket residues surrounding it change considerably. While this might have been expected for the MD$^{\text{inact}}$ where the inactive receptor conformation might need to adjust more to an agonist induced active-like conformation, it was not expected for the MD$^{\text{Gi}}$ and the MD$^{\beta\text{Arr}}$. Similar differences for the other two agonists CP55940 and HU308 could not be observed to the same extent. This variability of the binding pocket around RO6844112 is observed for side chains as well as the C_α atoms of the binding pocket residues and for all replicates with differences between replicates in all three MD$^{\text{inact}}$, MD$^{\text{Gi}}$ and MD$^{\beta\text{Arr}}$(see Figure A6.4.C and D). When clustering based on the binding pocket there was no clear main cluster based on a majority of the frames for either MD$^{\text{inact}}$, MD$^{\text{Gi}}$ or MD$^{\beta\text{Arr}}$, confirming the variability of the binding pocket residues surrounding RO6844112. An explanation for the larger conformational changes of the binding pocket around RO6844112 might be its small and rather apolar character which might leave more wiggle room for the surrounding residues. However, it might also be a characteristic of unbiased agonists. In this case, an unbiased agonist might overall stabilise activated conformations of the receptor but not fix the binding pocket in a way that it would shift the equilibrium towards a certain active conformation of the receptor, thereby inducing

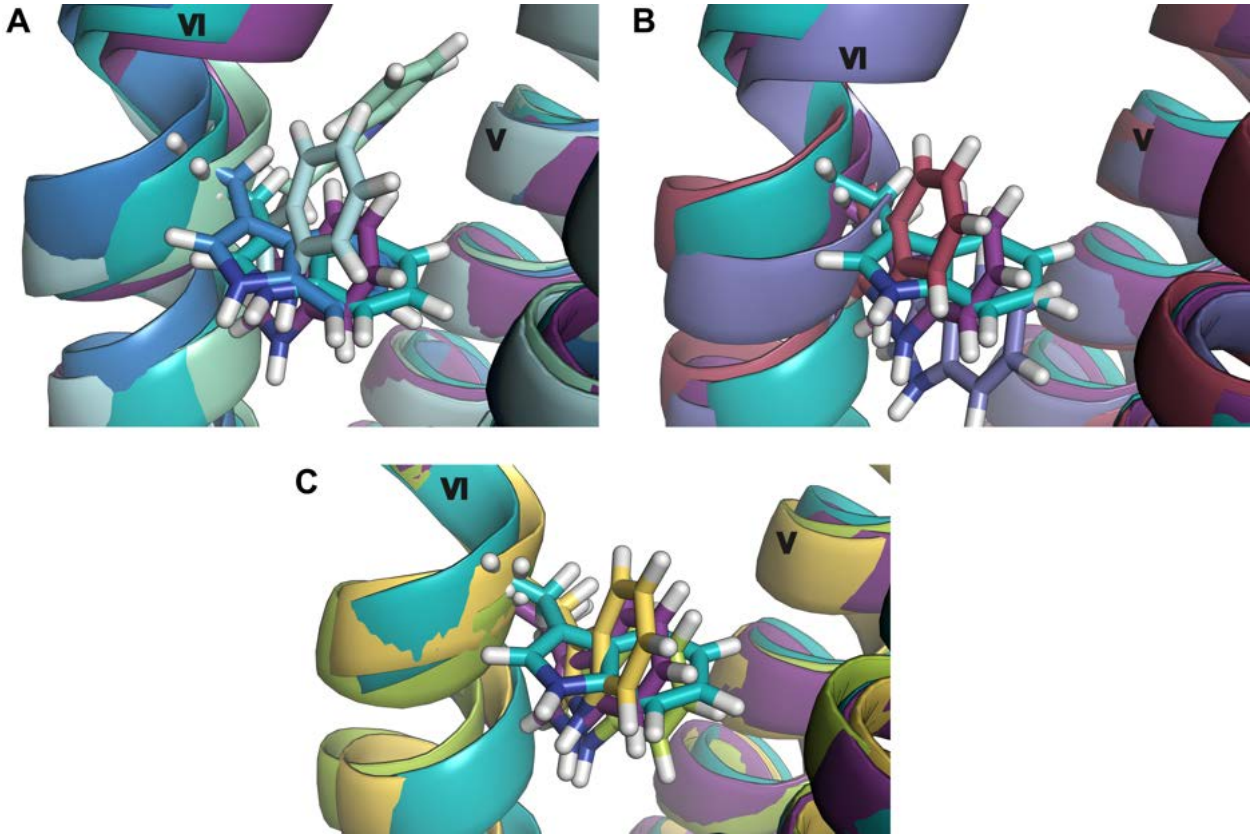


Figure 6.3: Conformations of toggle switch residue W258$^{6.48}$ compared between a CB2 structure in an inactive conformation (PDB ID 5ZTY; teal) and in an active conformation (PDB ID 6PT0; violet), which were used as starting structures for the MD simulations, to the main binding pocket based cluster centroids of the simulations with agonist RO6844112 and inverse agonist SR144528 in the A) MD$^{\text{inact}}$ (RO6844112 main cluster: light blue; RO6844112 alternative cluster: mint; SR144528: blue), B) MD$^{\text{Gi}}$ (RO6844112: dark red; SR144528: blue-violet) and C) MD$^{\beta\text{Arr}}$ (RO6844112: light yellow; SR144528: lime green). Movements cannot only be observed for residue W258$^{6.48}$ but also for the TM VI backbone in proximity of this residue.

a recruitment bias towards certain effector proteins. Agonists with a recruitment bias could then not only be identified based on interaction patterns but also by whether they stabilise the binding pocket conformation or not. To confirm or disprove this theory, more unbiased ligands with similar and dissimilar structures to RO6844112 need to be simulated in the different MD systems and the variability of the binding pocket conformation then needs to be analysed and compared between the systems.

For inverse agonist SR144528 the 2D-RMSD plots suggest bigger changes in the $MD^{\beta Arr}$ compared to the MD^{inact} (see Figure A6.3) which can probably be explained with the adaptation of the binding pocket in an active conformation to an inverse agonist induced inactive-like conformation of the binding pocket. Interestingly, in the MD^{Gi} this can only be observed in one of the replicates to the same extent as in the $MD^{\beta Arr}$. In this one replicate the ligand seems to have adopted a slightly different binding pose. To evaluate whether this alternative binding pose might be relevant and whether differences between the simulation of an inverse agonist with the MD^{Gi} and $MD^{\beta Arr}$ might be reproducible, the simulation of additional replicates of the MD^{Gi} and/or the simulation of a different inverse agonist in all systems might be of interest.

An alternative ligand pose was also found in one of the replicates of the $MD^{\beta Arr}$ with HU308. As can be seen from the 2D-RMSD plots (see Figure A6.6), this alternative ligand pose also induces notable changes in the binding pocket. The cluster analysis showed that while the pose of HU308 in the other replicates is comparable to the pose in the MD^{Gi}, this alternative pose is not. Both ligand poses seem quite stable throughout the $MD^{\beta Arr}$ (see also Figure 6.2) and the final poses are already reached within less than the first 5% of the frames in all replicates. However, it is not clear whether this observation might be connected to the recruitment bias this ligand induces and a true alternative binding pose was found or whether this is an artefact caused by, for example, a wrong initial ligand pose. To evaluate the stability of these binding poses further, more replicates need to be simulated starting from both obtained ligand poses. Furthermore, it should be evaluated whether this behaviour might be related to the bound β-arrestin 2 protein, by running additional simulations starting from both ligand poses in the MD^{Gi} and potentially also in the MD^{inact}.

Analysis of ligand receptor contacts. Since the orthosteric binding pocket of the CB2 is quite apolar it was expected to gain more insights from the analysis of apolar than polar contacts. Overall only few clear observations could be made and it is difficult to link any observations to a recruitment bias. *Hydrogen* bond contacts between ligand and receptor were mainly observed for CP55940 and HU308 while RO6844112 and SR144528 seem to rarely interact with the receptor via polar contacts (see Figure A6.11). This suggests that RO6844112 and SR144528 are mainly stabilised by forming apolar interactions with the receptor which is congruent with the results of the aromatic

interaction analysis. Both ligands display a wider variety of aromatic contacts and contact partners than CP55940 and HU308 (see Figure A6.12). The analysis of the aromatic *face-to-face* and *face-to-edge* interactions also shows that RO6844112 forms more aromatic contacts with F117$^{3.36}$ than SR144528 does. Interestingly, this residue also displays less native contacts with inverse agonist SR144528 in the MD$^{\mathrm{Gi}}$ and MD$^{\beta\mathrm{Arr}}$ than in the MD$^{\mathrm{inact}}$ and than observed for the agonists. This suggests that F117$^{3.36}$ might be involved in agonist recognition and maybe activation of the receptor. Toggle switch residue W258$^{6.48}$ shows high native contact frequencies with RO6844112 and especially SR144528 which was both expected (see Figures A6.15-A6.14). Contacts between CP55940 and W258$^{6.48}$ are almost as frequent as for these two ligands. However, although HU308 is structurally similar to CP55940, no contacts between this ligand and W258$^{6.48}$ can be observed, which might be one reason for the different recruitment behaviour of these two ligands.

Interesting observations from the native contact analysis were also made for the overall number of contacts of each ligand and the consistency of contact frequencies for each residue among the different replicates of a system. SR144528 displays the highest number of contacts of all ligands –especially in the MD$^{\mathrm{inact}}$– and the contact frequencies between the different replicates seem quite consistent. This might be caused by the size of the ligand which fills the orthosteric binding pocket quite well, resulting in fewer changes of the binding pose and contacts with more residues. Interestingly, this inverse agonist has a higher number of contacts in the MD$^{\mathrm{inact}}$ which corresponds to the expectation that an inverse agonist should be more affine to an inactive than an active conformation. A high consistency of contact frequencies between replicates was also observed for CP55940. This ligand is the most polar of the four evaluated ligands and can form three hydrogen bonds with the receptor. This likely leads to a stable binding pose and, hence, consistency in the native contacts. The contact frequencies of ligand RO6844112 are slightly less consistent compared to the other two ligands. This matches the observation from the 2D-RMSD plots that suggest more changes in the binding pocket conformation in all simulations with this ligand. The most interesting results were observed for ligand HU308. In both MD$^{\mathrm{Gi}}$ and MD$^{\beta\mathrm{Arr}}$ low consistencies of contact frequencies were seen between the different replicates. Surprisingly, this seems to be independent from the alternative binding pose found in one replicate of the MD$^{\beta\mathrm{Arr}}$. These inconsistencies might suggest instabilities in the binding pocket and ligand pose in the simulations with HU308 which might either hint to interesting behaviour of this ligand causing the observed recruitment bias or to an incorrect initial ligand pose. Therefore, results from simulations with HU308 should be treated carefully for now and more simulations with this ligand could be helpful to clear doubts.

It should be kept in mind that the initial poses for all ligands except CP55940 were obtained by docking calculations and are therefore only predictions that do not necessarily reflect the true bind-

ing pose. This should especially be considered in context to alternative binding poses and could explain pose changes during the simulation. It could also explain the inconsistencies and variability observed for ligand HU308. The initial pose of CP55940 was based on the pose of the ligand in a crystal structure of the CB1 (PDB ID 6KQI[221]) and seems to be stable throughout all replicates. The initial ligand poses and the poses of the ligands in the main cluster of the ligand-based clustering are shown in Figures 6.4 and 6.5.

Ligand-water contacts and the initially placed water molecule. In the initial system setup a predicted water molecule was placed in the orthosteric binding pocket to mediate contacts between the ligands and receptor residue S285$^{7.39}$. In all simulations, the water molecule moves out of the binding pocket during the simulation. This could indicate that a water molecule in that position is not stable or might not be bound in this position after all. However, all ligands display frequent polar contacts with water molecules in the *hydrogen* bond analysis and also with the polar molecule parts pointing into the direction in which the initial water molecule was placed. This could suggest that a) the water molecule position is stable enough for the initial molecule to not leave the binding pocket immediately and/or b) that other water molecules might move into that same region to interact with the ligand. Therefore, it might be interesting to analyse ligand-water contacts further, evaluate possible water networks in the orthosteric binding pocket and examine how water molecules enter the binding pocket from the outside to interact with the ligand. Furthermore, it might be worth investigating how the system behaves and how water contacts change if the initial water molecule is not placed.
Even if these results could suggest that there is no water molecule bound in the predicted position this should not influence the results of the previous docking study described in chapter 5 since the placement of the water molecule did lead to overall improved molecule docking poses.

A 'β-arrestin biased conformation'. Overall, the structural differences between the structures from the MD$^{\mathrm{Gi}}$ and the MD$^{\beta\mathrm{Arr}}$ are rather small which makes it difficult to find a 'β-arrestin biased conformation' to improve chances of finding β-arrestin biased ligands. There are several possible explanations for this. First of all the simulation time of 1 µs might be too short to induce bigger structural changes. This process could take considerably longer and might therefore not have been completed. Another question is whether there even is something like a 'β-arrestin biased conformation'. A debate on whether G protein/β-arrestin bias even exists is going on and even then it is not clear whether this is something structurally induced or might stem from other processes and influences. From recently published structures of the Neurotensin receptor 1 and the Muscarinic M2 receptor in complex with β-arrestin 1 it was observed that differences between the G protein

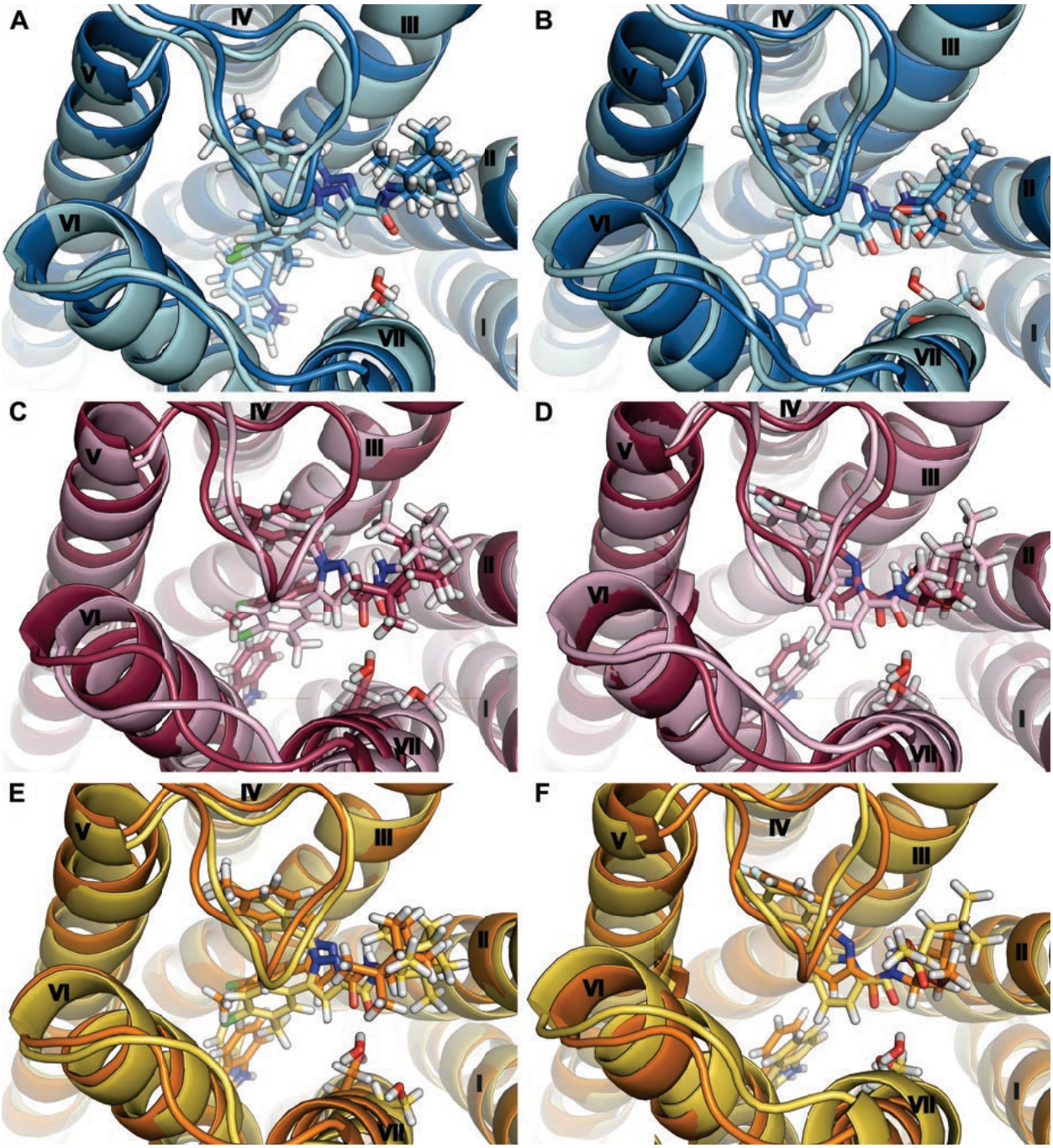


Figure 6.4: Poses of SR144528 and RO6844112 before equilibration and after simulating. The poses of SR144528 are shown on the left (A,C,E) and for RO6844112 on the right (B, D, F) for the MD^{inact} (A,B), the MD^{Gi} (C, D) and the $MD^{\beta Arr}$ (E, F). The initial ligand pose before equilibrating the systems is shown in the darker colour (A,B: dark blue; C,D: purple; E,F: orange) and the ligand pose of the centroid structure of the main ligand-based cluster is shown in a lighter colour (A,B: light blue; C,D: rose; E,F: yellow). The initially placed water molecule is shown in all structures, as well as residues $W258^{6.48}$ and $S285^{7.39}$ for better orientation of the observer in the binding site. All pictures were taken from the extracellular side of the receptor and the same point of view.

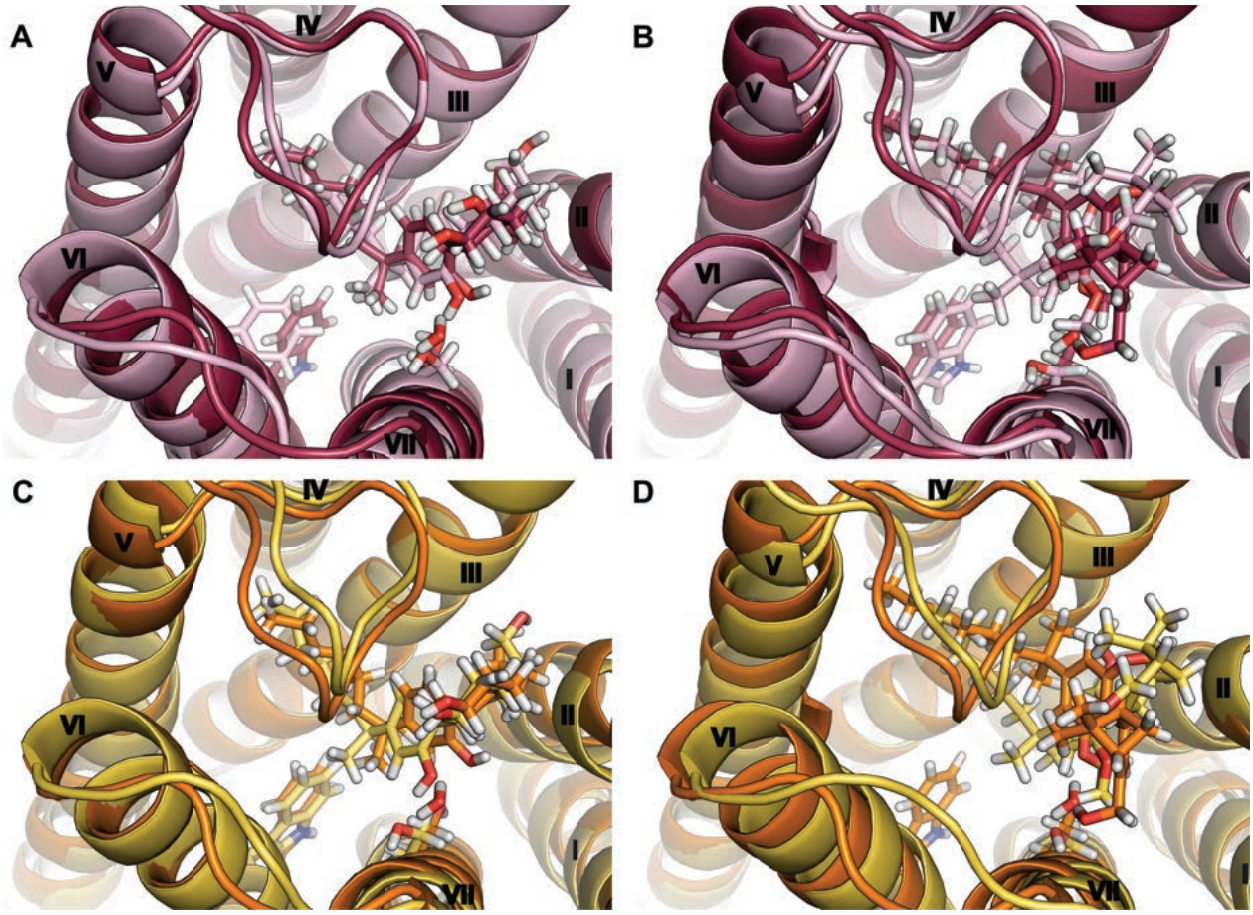


Figure 6.5: Poses of CP55940 and HU308 before equilibration and after simulating. The poses of CP55940 are shown on the left (A,C) and for HU308 on the right (B, D) for the the MD^{Gi} (A, C) and the $MD^{\beta Arr}$ (B, D). The initial ligand pose before equilibrating the systems is shown in the darker colour (A,B: purple; C,D: orange) and the ligand pose of the centroid structure of the main ligand-based cluster is shown in a lighter colour (A,B: rose; C,D: yellow). The alternative binding pose of HU308 in the $MD^{\beta Arr}$ is not shown. The initially placed water molecule is shown in all structures, as well as residues $W258^{6.48}$ and $S285^{7.39}$ for better orientation of the observer in the binding site. All pictures were taken from the extracellular side of the receptor and the same point of view.

bound active conformation and the β-arrestin 1 bound active conformation are small.[219,235] This suggests that it might be difficult to track bias on the structural basis with a simple MD simulation and other methods such as metadynamics might be more useful to track the ligand binding in more detail. It is reassuring though, that in principle activation and inactivation can be tracked using MD simulations[24] and some indicators regarding this were also found in this study.

6.5 Conclusions

The aim of the study described in this chapter was to elucidate on a structural level how and why certain ligands induce a recruitment bias towards a certain effector protein or effector protein subfamily. The first step was to identify such biased ligands within a set of known CB2 agonists. For that, the correlation between the recruitment of different effector proteins was analysed based on recruitment data provided by the Veprintsev lab.[208] The most important observation from this correlation analysis is that ligands inducing a recruitment bias do indeed exist and 4-7 biased ligands could be identified from the set of 35 analysed ligands. This is reassuring since it shows that biased ligands might not be as rare as assumed and that the search for such biased ligands is reasonable. It should be kept in mind though that only recruitment of effector proteins and recruitment bias was analysed here and no statement about effects on downstream signalling can be made.

MD simulations were then used to analyse, why some of these ligands induce a recruitment bias. A set of four representative ligands –an inverse agonist, an unbiased agonist, a β-arrestin 2 biased agonist and a G protein biased agonist– were then simulated in complex with the CB2 and either no effector protein, the $G_{i,\alpha}$ protein or β-arrestin 2. These simulations were then analysed with a main focus on the binding pocket conformation, the ligand pose and the interactions and contacts between ligand and receptor. This analysis revealed differences in the behaviour of the different ligands. The activation of the receptor when simulating an inactive conformation with an agonist could be observed within the binding pocket, based on the movement of residues that were previously shown to play a role in receptor activation. The reverse case, i.e. the deactivation when simulating an active conformation with an inverse agonist, could not be observed. This might be due to the stabilisation of the receptor conformation by the effector proteins or due to a slower deactivation of the receptor. However, in any case, it was possible to observe differences between the simulations with an agonist and an inverse agonist, showing that it is possible to simulate receptor activation and inactivation and to discriminate between activating and deactivating ligands.

The discrimination between ligands inducing a different recruitment bias proved to be more difficult. However, the behaviour of the three agonists in the simulations is quite different. While the β-arrestin 2 biased ligand (CP55940) is very stable throughout all simulations, the unbiased ligand

(RO6844112) does not seem to stabilise the binding pocket conformation and the G protein biased ligand (HU308) adopts an alternative binding pose in the $MD^{\beta Arr}$. This behaviour of the ligands in the MD simulations might be linked to the recruitment they induce or it might be a ligand-specific property. To analyse this further more simulations might be necessary. The CB2 conformations resulting from the MD^{Gi} and the $MD^{\beta Arr}$ do not differ by much from each other, however, this reproduces structural observations from other receptors.[219,235]

In conclusion, the results of this study are promising. Although it was not possible to link recruitment bias to certain interactions between ligand and receptor, the behaviour of the ligands in the MD simulations might give hints towards how and why the ligands induce biased effector protein recruitment. Further studies and simulations will be necessary to support these observations but a first step towards elucidating the different activation effects has been taken with these results.

6.6 Perspectives

With the study described in this chapter a basis is set which can be used to build up on. The analysis of activation and deactivation of the receptor was mainly built on the movement of toggle switch residue $W258^{6.48}$. This movement could be evaluated further by clustering based on this residue conformation to see whether one specific conformation is stabilised or whether the residue is moving a lot in either of the simulations. The movement of the TM VI backbone in this area could also be analysed in more detail to see whether the observations from the clusters are representative. Further, it could be interesting to simulate different agonists (e.g. CP55940 and HU308) in the MD^{inact} or a different antagonist in the MD^{Gi} and the $MD^{\beta Arr}$ to confirm or disprove the observations. However, this would only show activation or inactivation of the receptor but not ligand bias which was the main aim of this study.

For the unbiased agonist RO6844112 stronger variations within the binding pocket were observed than for any of the other ligands. To evaluate whether this is a property of unbiased agonists or a ligand-specific property, more MD simulations with other unbiased ligands would be necessary. For that, unbiased agonists with similar as well as dissimilar structures to RO6844112 should be simulated in complex with the different CB2-effector protein complexes.

The alternative binding pose of HU308 in the $MD^{\beta Arr}$ is also an interesting starting point for further investigations. The reproducibility of this ligand pose could be evaluated by simulating more replicates of the same system. Furthermore, the stability of both ligand poses could be evaluated by starting simulations from both poses in all systems. Additionally, metadynamics simulations could give more insights on the stability of either of these ligand poses. Should both binding poses of HU308 still seem likely after these investigations, it could be interesting to investigate whether

this ligand actually also adopts two different binding poses in the real biological system. For that, mutational studies could be employed and it could be investigated further whether this might be a reason for the observed recruitment bias.

Aside from that, water molecules might also play an important role in ligand binding and receptor activation. It could be interesting to evaluate water networks in the binding pocket and the movements of water molecules into and out of the binding pocket further. The influence of the initially placed water molecule on the results of the MD simulations might also be worth further investigations.

6.7 Additional information

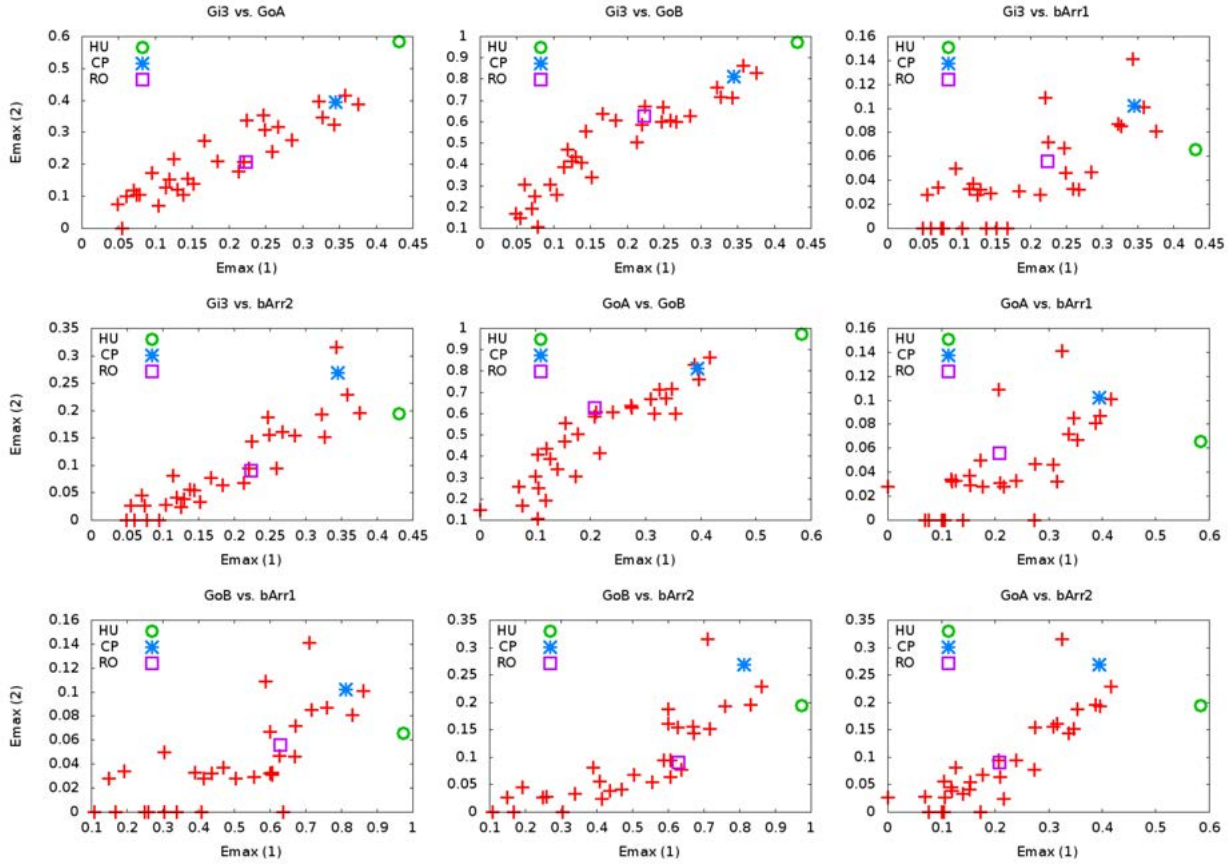


Figure A6.1: Plots to analyse the correlation between recruitment of different effector proteins upon stimulation of the CB2 with an agonist (part 1 of 2). Recruitment of five different members of the $G_{i/o}$ protein subfamily (G_{i1}, G_{i2}, G_{i3}, G_{oA} and G_{oB}), β-arrestin 1 and β-arrestin 2 was measured for a set of known ligands of the CB2 (see Table A5.1) and the data provided by the Veprintsev lab at the University of Nottingham.[208] The E_{max} values for the recruitment of different effector proteins were then plotted against each other and the correlation between the pathways was analysed. The ligands that were selected for the MD simulations (RO6844112, CP55940 and HU308) are marked with different symbols. The inverse agonists were not included in the plots.

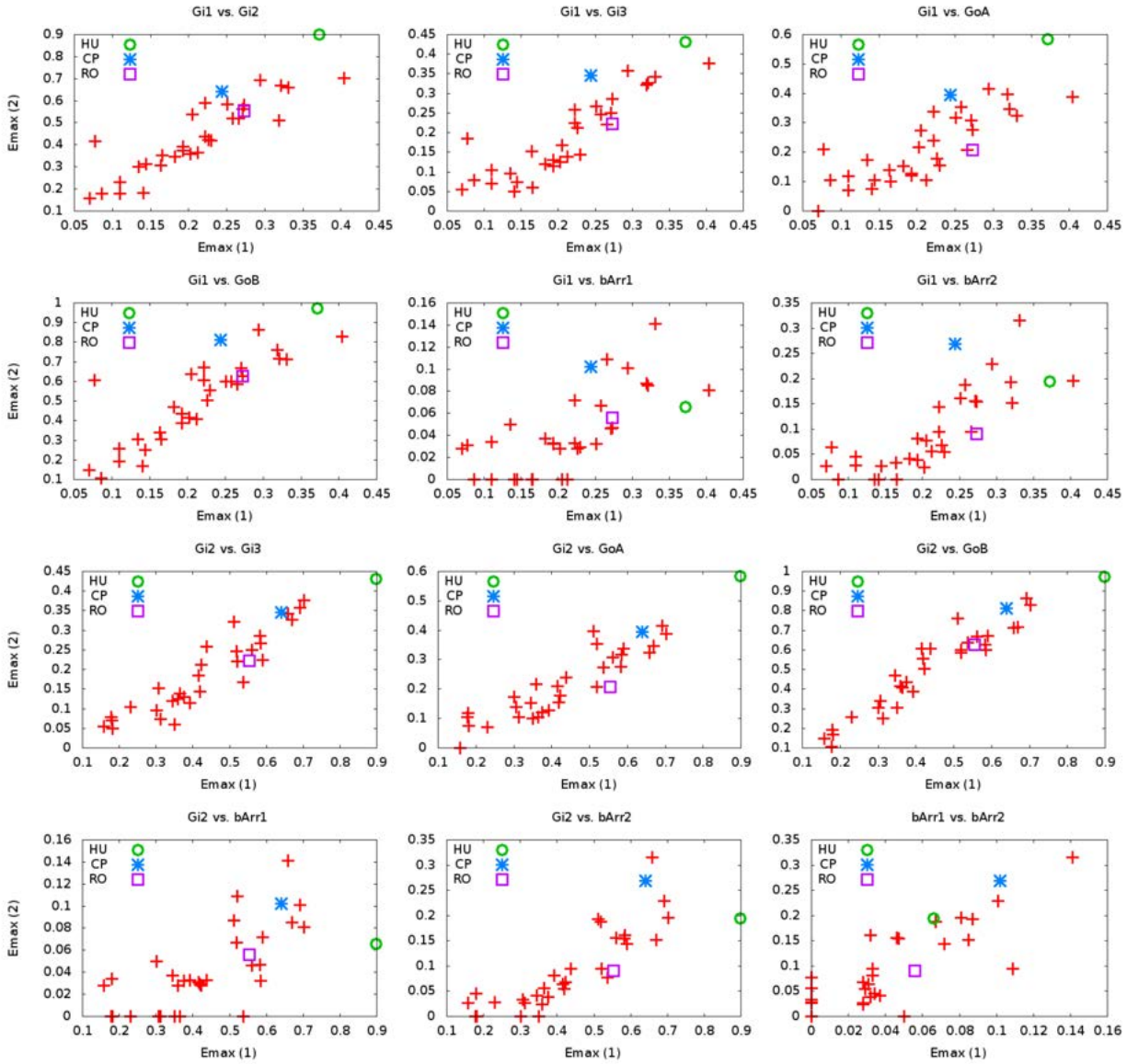


Figure A6.2: Plots to analyse the correlation between recruitment of different effector proteins upon stimulation of the CB2 with an agonist (part 2 of 2). Recruitment of five different members of the $G_{i/o}$ protein subfamily (G_{i1}, G_{i2}, G_{i3}, G_{oA} and G_{oB}), β-arrestin 1 and β-arrestin 2 was measured for a set of known ligands of the CB2 (see Table A5.1) and the data provided by the Veprintsev lab at the University of Nottingham.[208] The E_{max} values for the recruitment of different effector proteins were then plotted against each other and the correlation between the pathways was analysed. The ligands that were selected for the MD simulations (RO6844112, CP55940 and HU308) are marked with different symbols. The inverse agonists were not included in the plots.

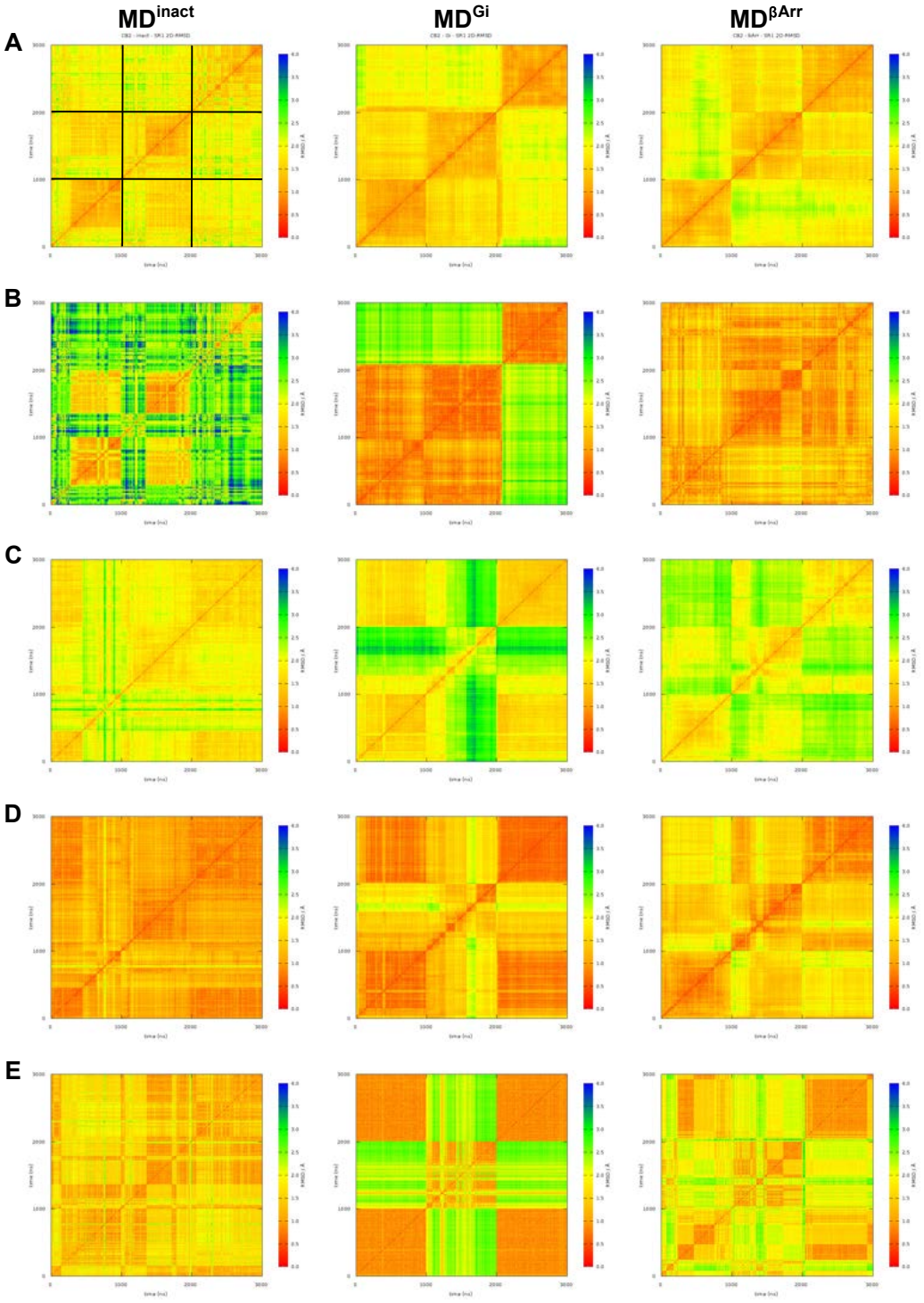


Figure A6.3: 2D-RMSD plots of inverse agonist SR144528 for A: TM-C_α atoms, B: TM V-C_α atoms, C: all atoms of the binding pocket residues, D: C_α atoms of the binding pocket residues and E: ligand SR144528. The respective simulation systems are indicated at the top. All three replicates were plotted in sequence and the time in ns is depicted on x- and y-axis with 0 ns as the lowest and 3000 ns as the highest value. New replicates started as indicated by the lines in the top left plot. The colours represent the RMSD value with red = 0 Å, yellow = 2 Å and blue = 4 Å.

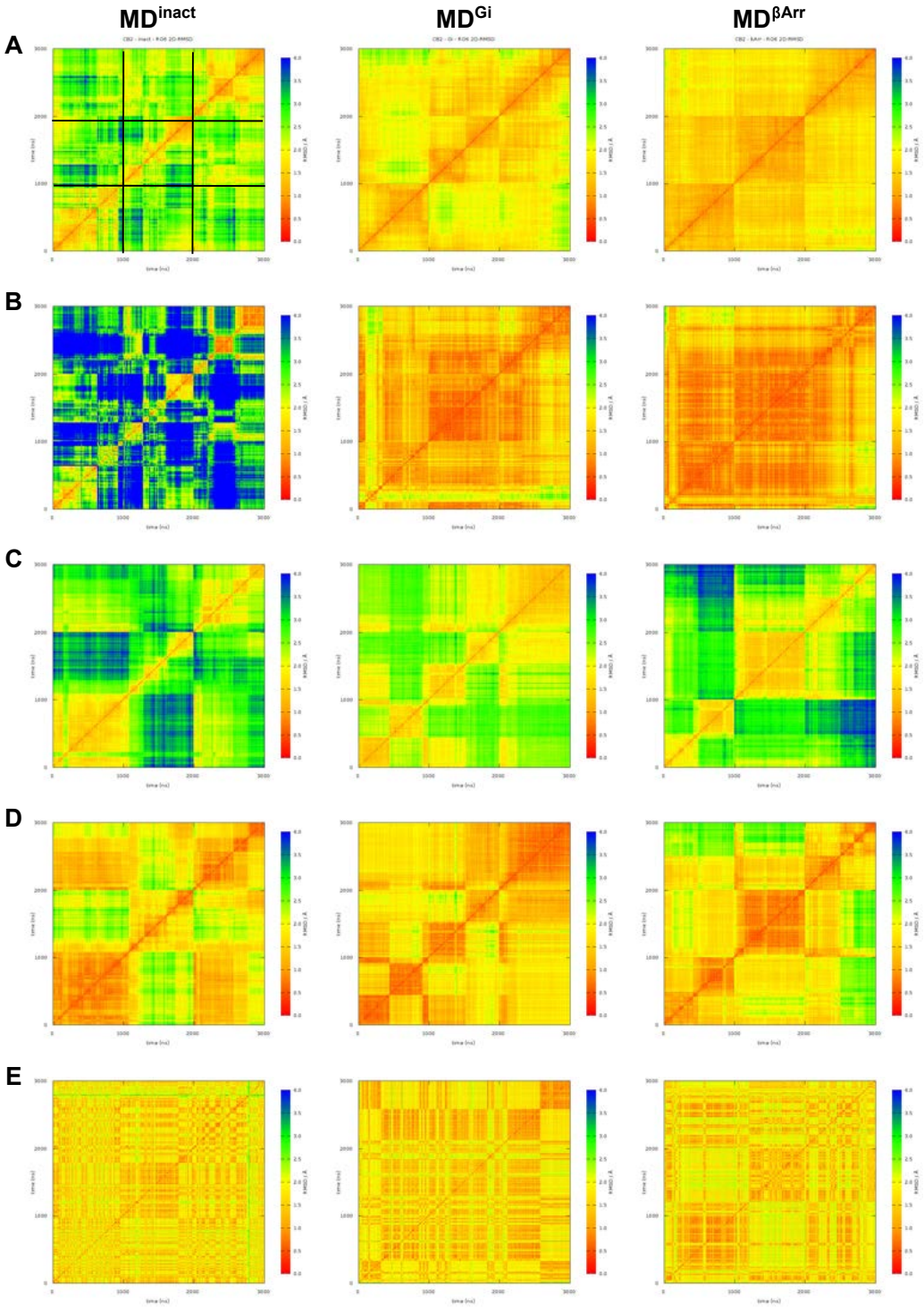


Figure A6.4: 2D-RMSD plots of the unbiased agonist RO6844112 for A: TM-C_α atoms, B: TM V-C_α atoms, C: all atoms of the binding pocket residues, D: C_α atoms of the binding pocket residues and E: ligand RO6844112. The respective simulation systems are indicated at the top. All three replicates were plotted in sequence and the time in ns is depicted on x- and y-axis with 0 ns as the lowest and 3000 ns as the highest value. New replicates started as indicated by the lines in the top left plot. The colours represent the RMSD value with red = 0 Å, yellow = 2 Å and blue = 4 Å.

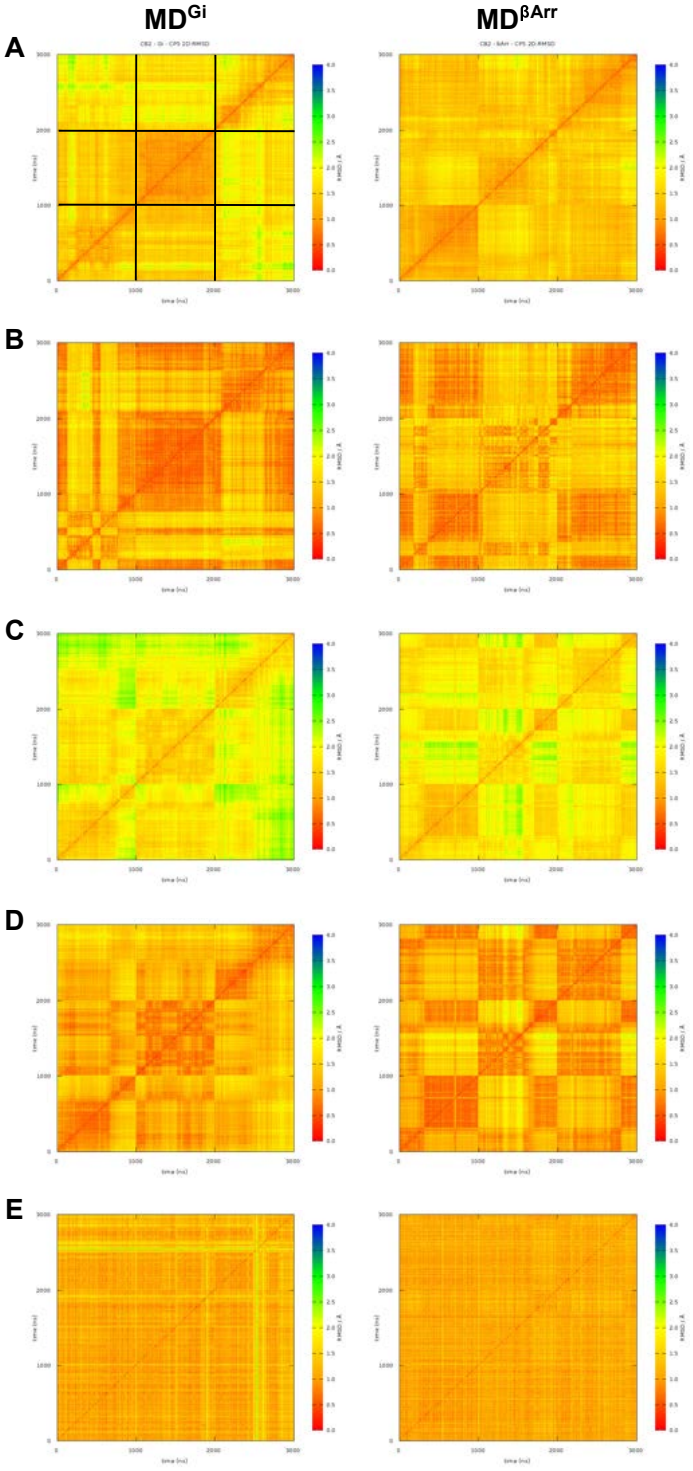


Figure A6.5: 2D-RMSD plots of agonist CP55940, which is biased towards β-arrestin 2 recruitment, for A: TM-C_α atoms, B: TM V-C_α atoms, C: all atoms of the binding pocket residues, D: C_α atoms of the binding pocket residues and E: ligand CP55940. The respective simulation systems are indicated at the top. All three replicates were plotted in sequence and the time in ns is depicted on x- and y-axis with 0 ns as the lowest and 3000 ns as the highest value. New replicates started as indicated by the lines in the top left plot. The colours represent the RMSD value with red = 0 Å, yellow = 2 Å and blue = 4 Å.

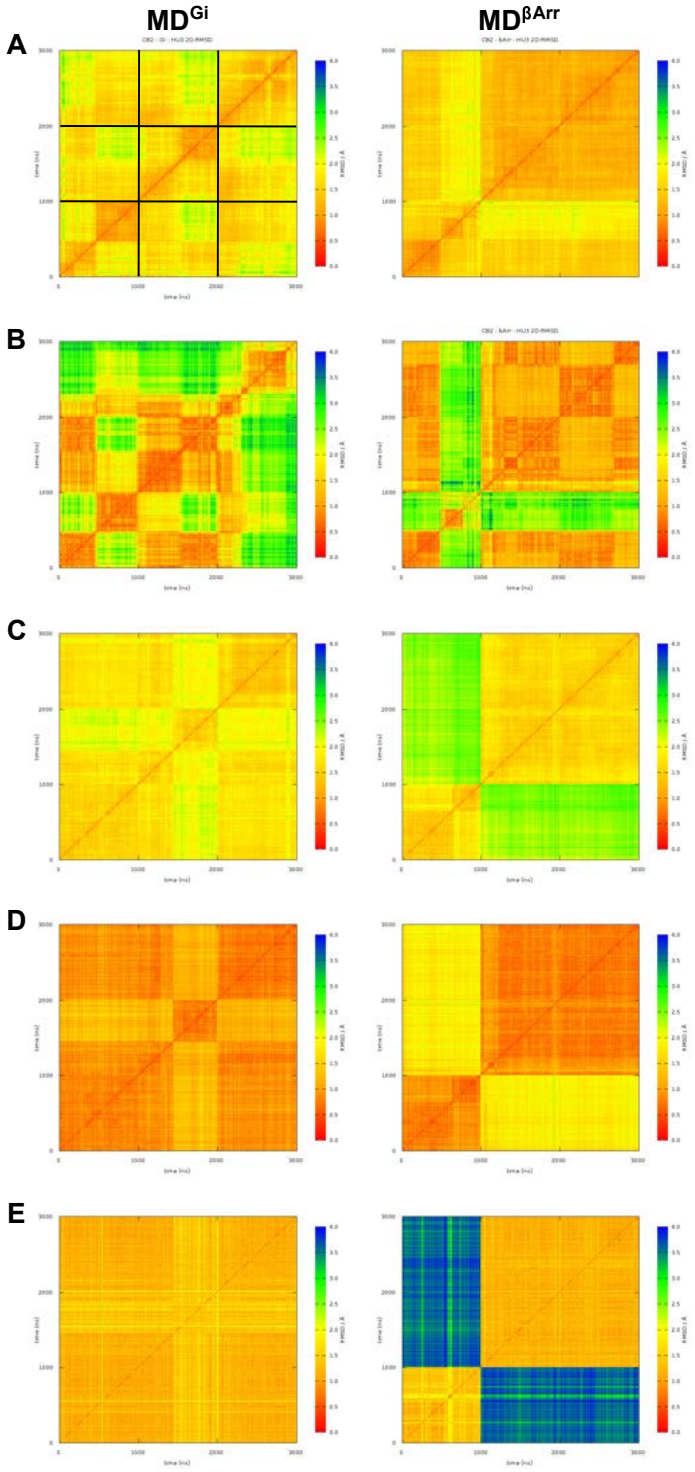


Figure A6.6: 2D-RMSD plots of agonist HU308, which is biased towards G protein recruitment, for A: TM-C_α atoms, B: TM V-C_α atoms, C: all atoms of the binding pocket residues, D: C_α atoms of the binding pocket residues and E: ligand HU308. The respective simulation systems are indicated at the top. All three replicates were plotted in sequence and the time in ns is depicted on x- and y-axis with 0 ns as the lowest and 3000 ns as the highest value. New replicates started as indicated by the lines in the top left plot. The colours represent the RMSD value with red = 0 Å, yellow = 2 Å and blue = 4 Å.

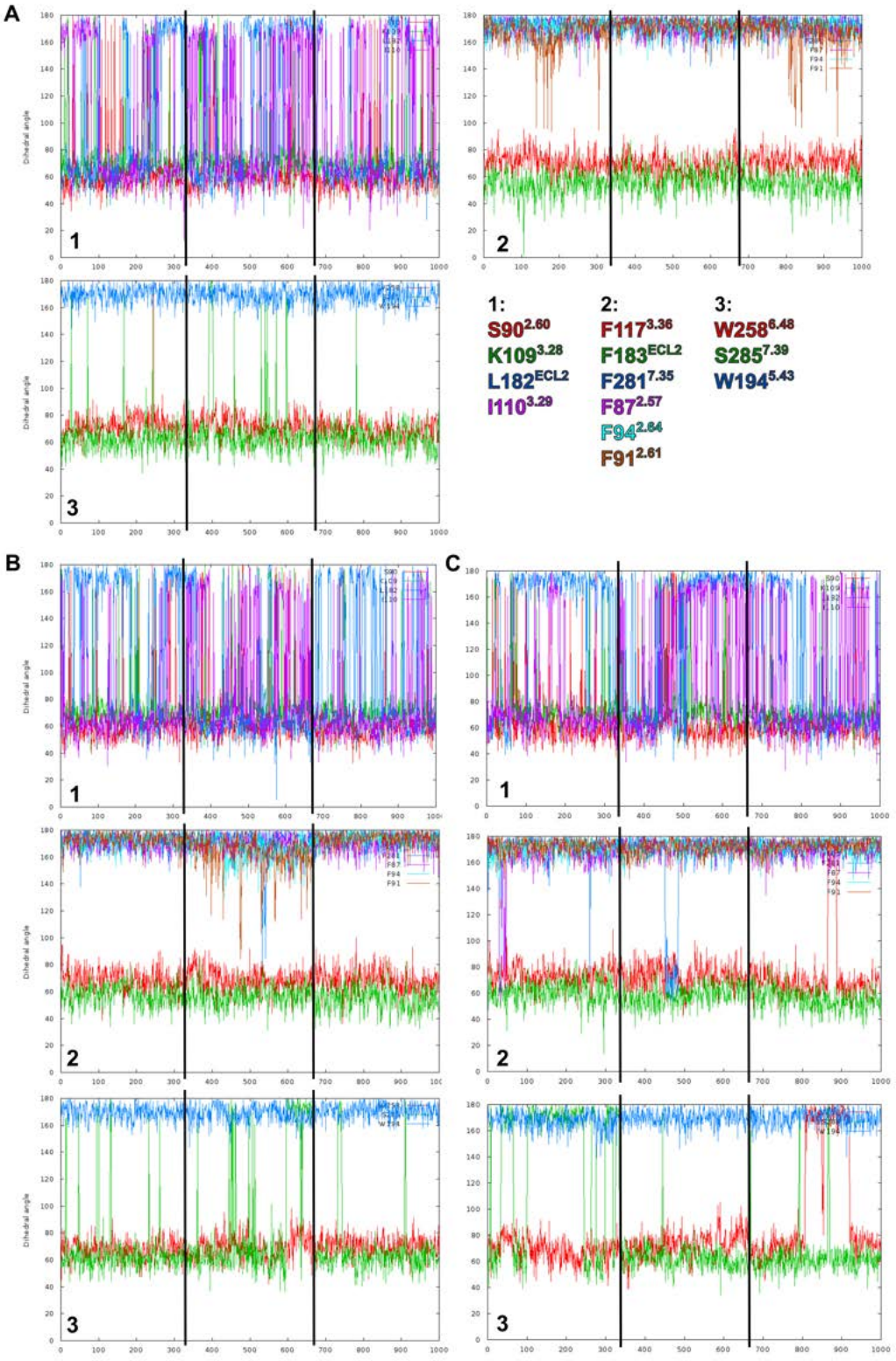


Figure A6.7: Plot of the dihedral angles of selected residues over time for inverse agonist SR144528 for the A) MD^{inact}, B) MD^{Gi} and C) $MD^{\beta Arr}$. The selected residues in each plot are as indicated in the legend. The time steps are plotted on the x-axis and the absolute values of the dihedral angle on the y-axis. Dihedral angles were determined every third time step. Vertical lines indicate cuts between the three replicates.

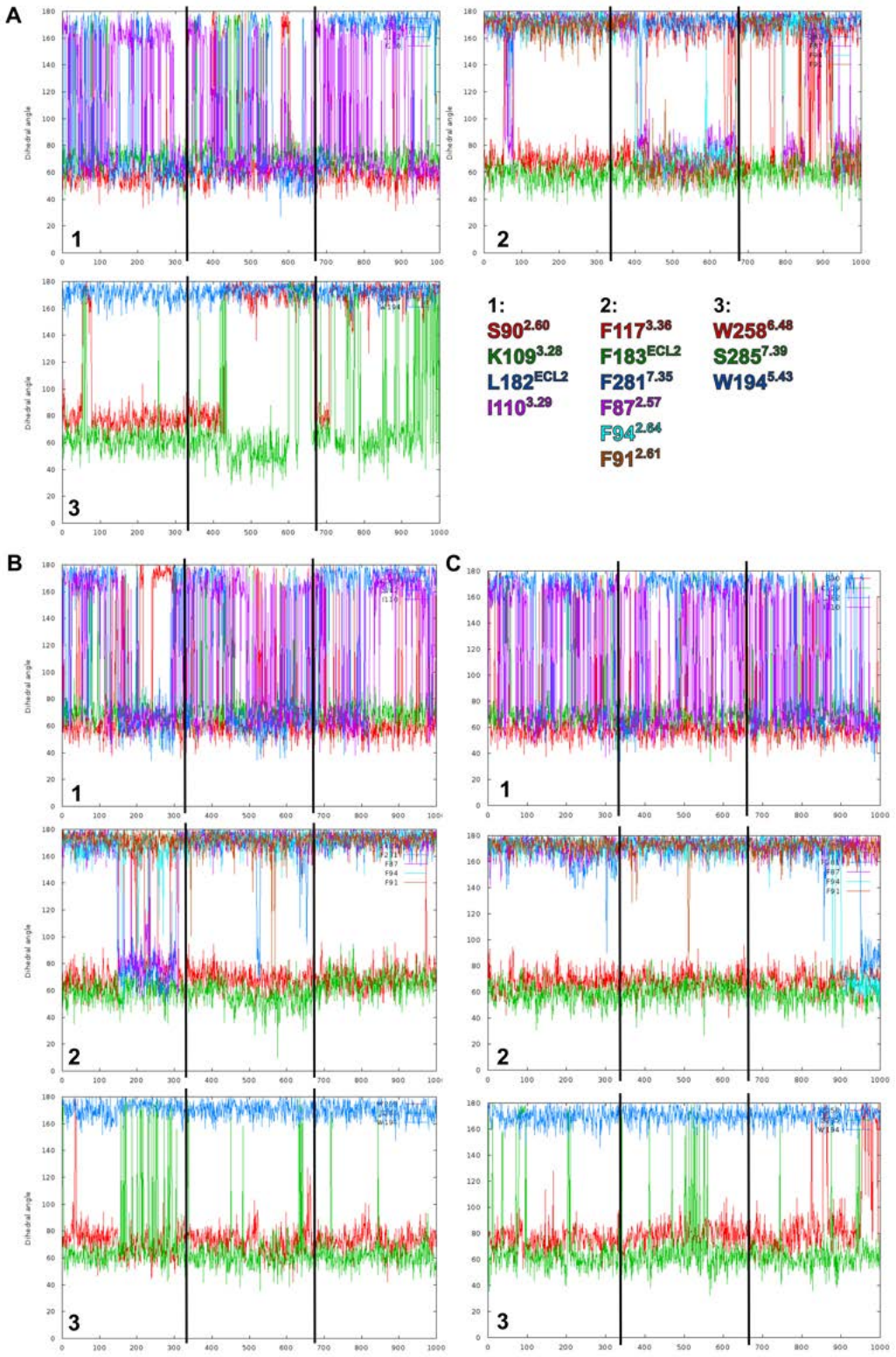


Figure A6.8: Plot of the dihedral angles of selected residues over time for unbiased agonist RO6844112 for the A) MD^{inact}, B) MD^{Gi} and C) $MD^{\beta Arr}$. The selected residues in each plot are as indicated in the legend. The time steps are plotted on the x-axis and the absolute values of the dihedral angle on the y-axis. Dihedral angles were determined every third time step. Vertical lines indicate cuts between the three replicates.

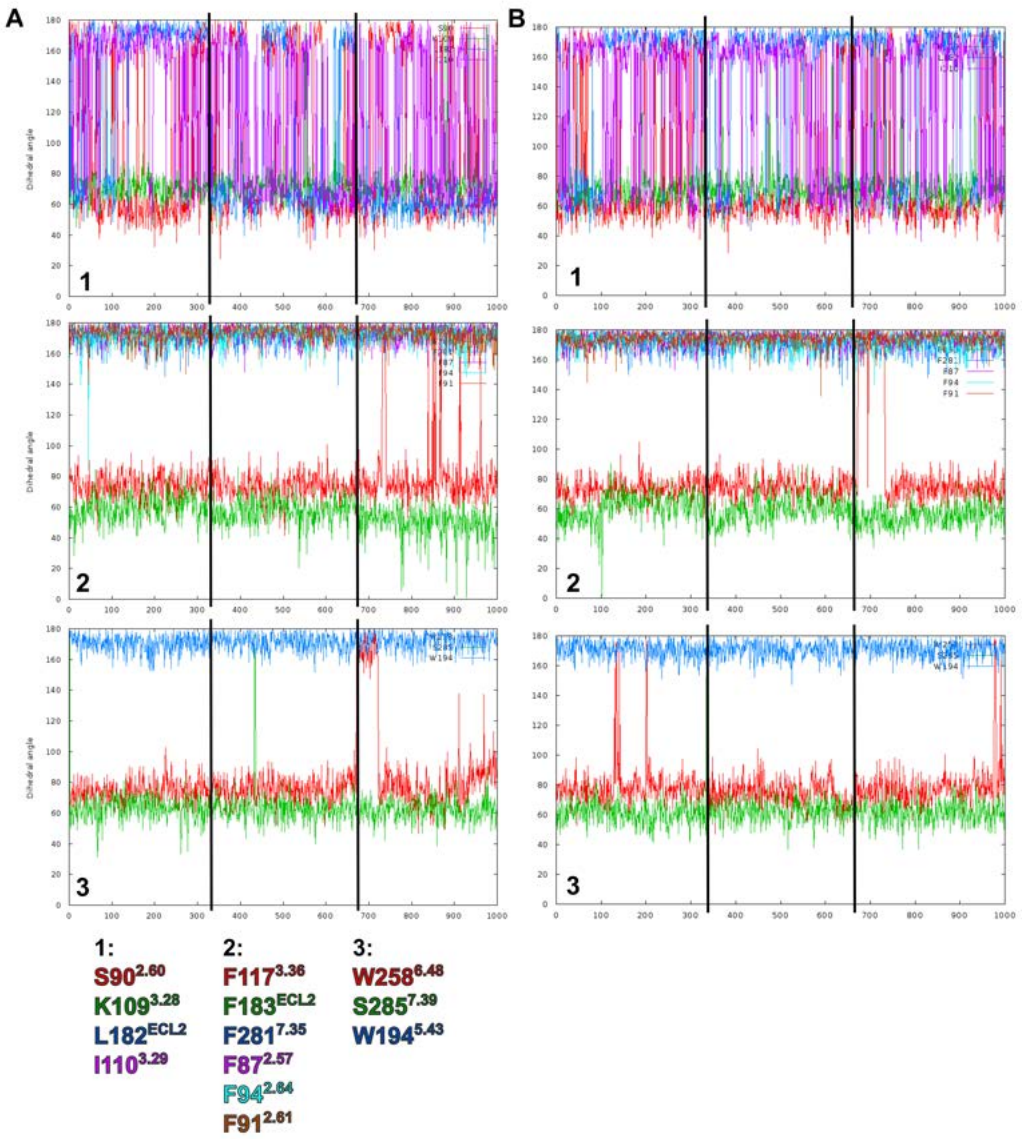


Figure A6.9: Plot of the dihedral angles of selected residues over time for agonist CP55940 for the A) MD^{Gi} and B) $MD^{\beta Arr}$. The selected residues in each plot are as indicated in the legend. The time steps are plotted on the x-axis and the absolute values of the dihedral angle on the y-axis. Dihedral angles were determined every third time step. Vertical lines indicate cuts between the three replicates.

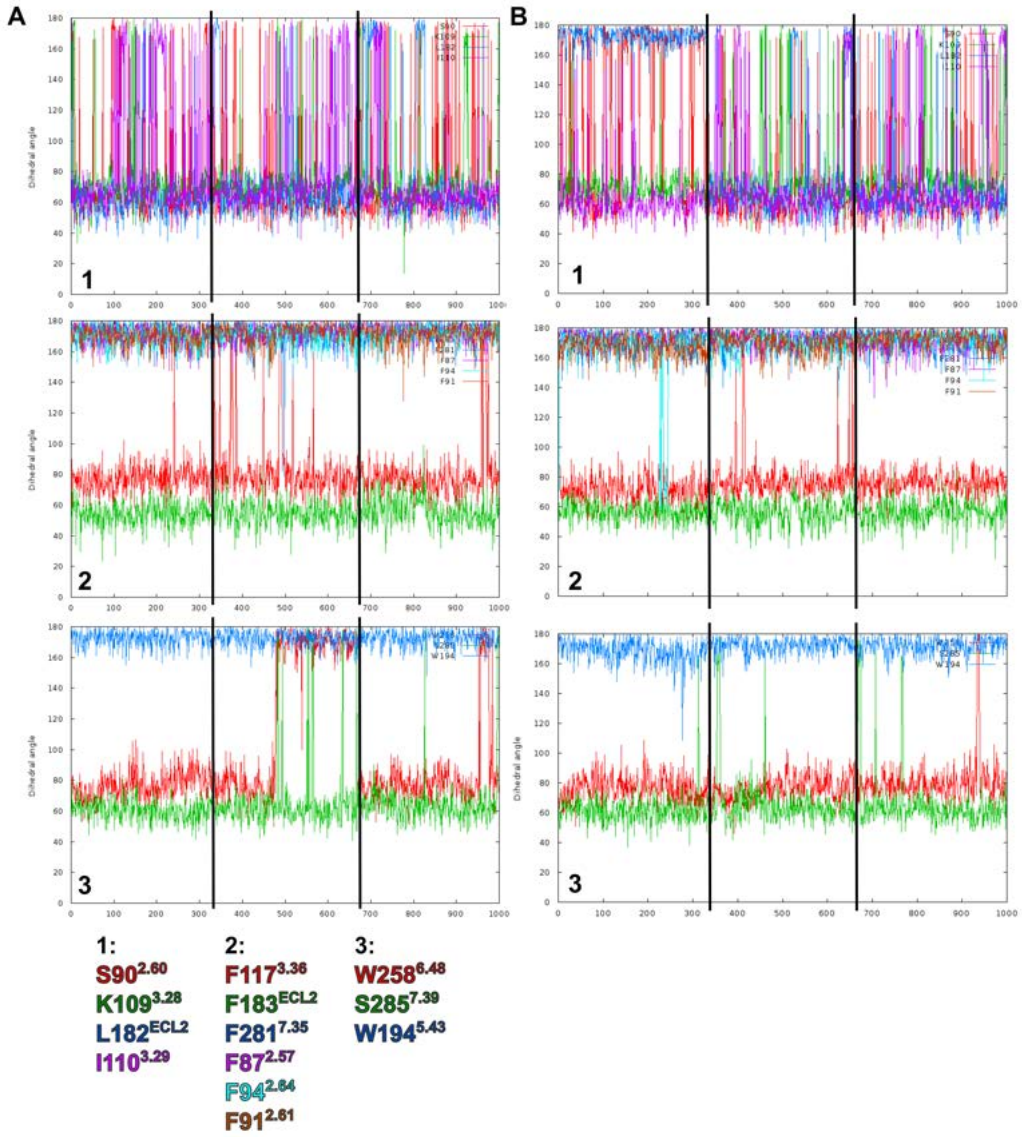


Figure A6.10: Plot of the dihedral angles of selected residues over time for agonist HU308 for the A) MD^{Gi} and B) $MD^{\beta Arr}$. The selected residues in each plot are as indicated in the legend. The time steps are plotted on the x-axis and the absolute values of the dihedral angle on the y-axis. Dihedral angles were determined every third time step. Vertical lines indicate cuts between the three replicates.

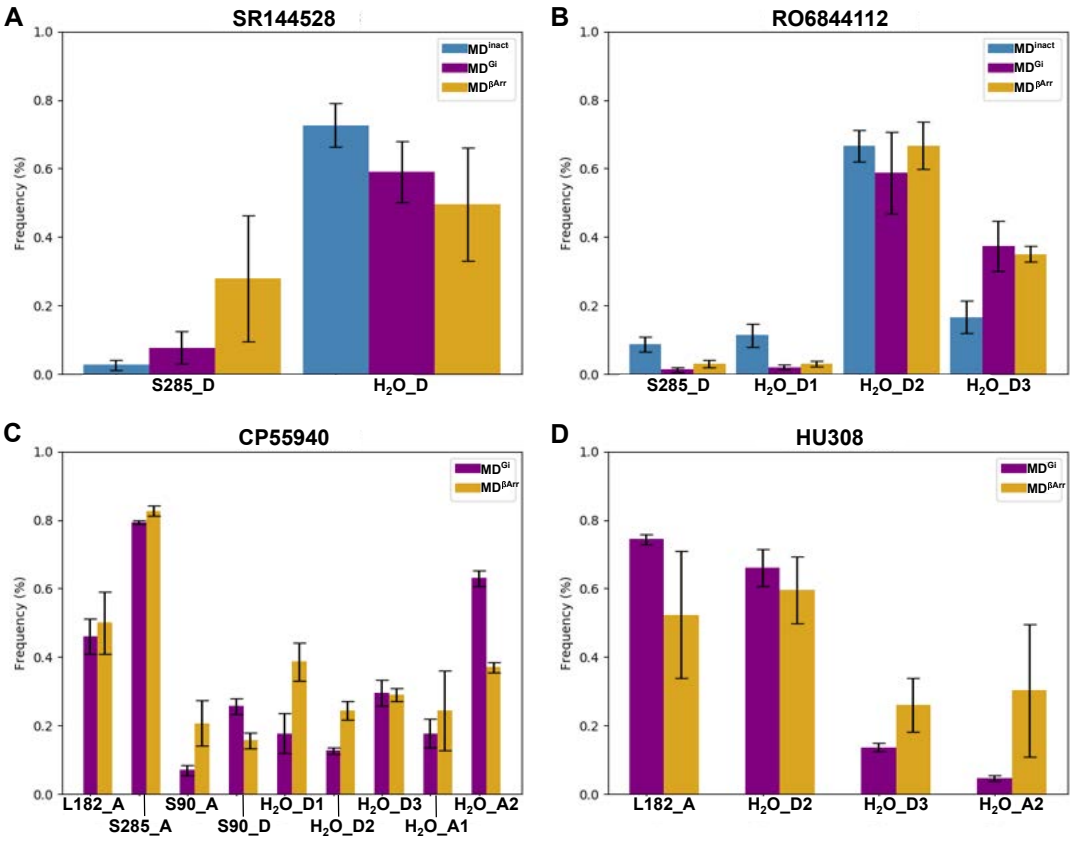


Figure A6.11: Frequency of *hydrogen* bond contacts between receptor or water and A: inverse agonist SR144528, B: agonist RO6844112, C: agonist CP55940 and D: agonist HU308. Bond frequencies were calculated for specific atom pairs of receptor and ligand. The interacting residue is labeled on the x-axis, where *D* indicates that the residue acts as *hydrogen* bond donor and *A* that it acts as *hydrogen* bond acceptor. Water molecules interacting with different polar groups of the ligand were numbered according to atom numbering of the oxygen atoms in the ligand (see Figure 6.1). Blue bars show the frequency in the MD^{inact}, violet in the MD^{Gi} and golden in the $MD^{\beta Arr}$. Values are plotted as mean of three replicates and error bars indicate S.E.M.

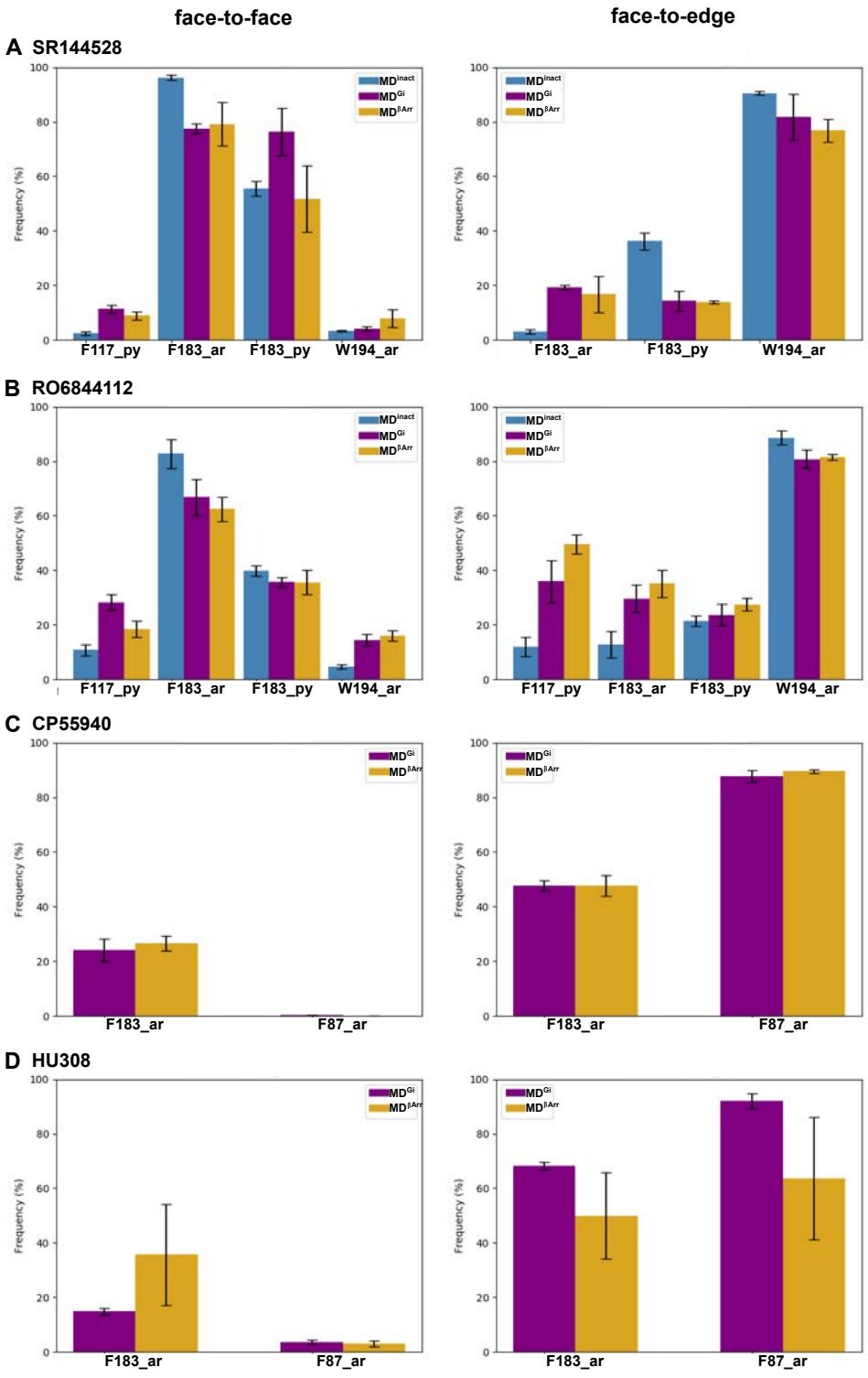


Figure A6.12: Frequency of aromatic *face-to-face* (left) and *face-to-edge* (right) contacts between the receptor and A: inverse agonist SR144528, B: agonist RO6844112, C: agonist CP55940 and D: agonist HU308. The interacting residue and ligand part are labelled on the x-axis. SR144528: py is pyrazole moiety, ar is the non-chlorinated aromatic moiety. RO6844112: py is pyridine moiety, ar is the other aromatic moiety. Blue bars show the frequency in the MD^{inact}, violet in the MD^{Gi} and golden in the $MD^{\beta Arr}$. Values are plotted as mean of three replicates and error bars indicate S.E.M.

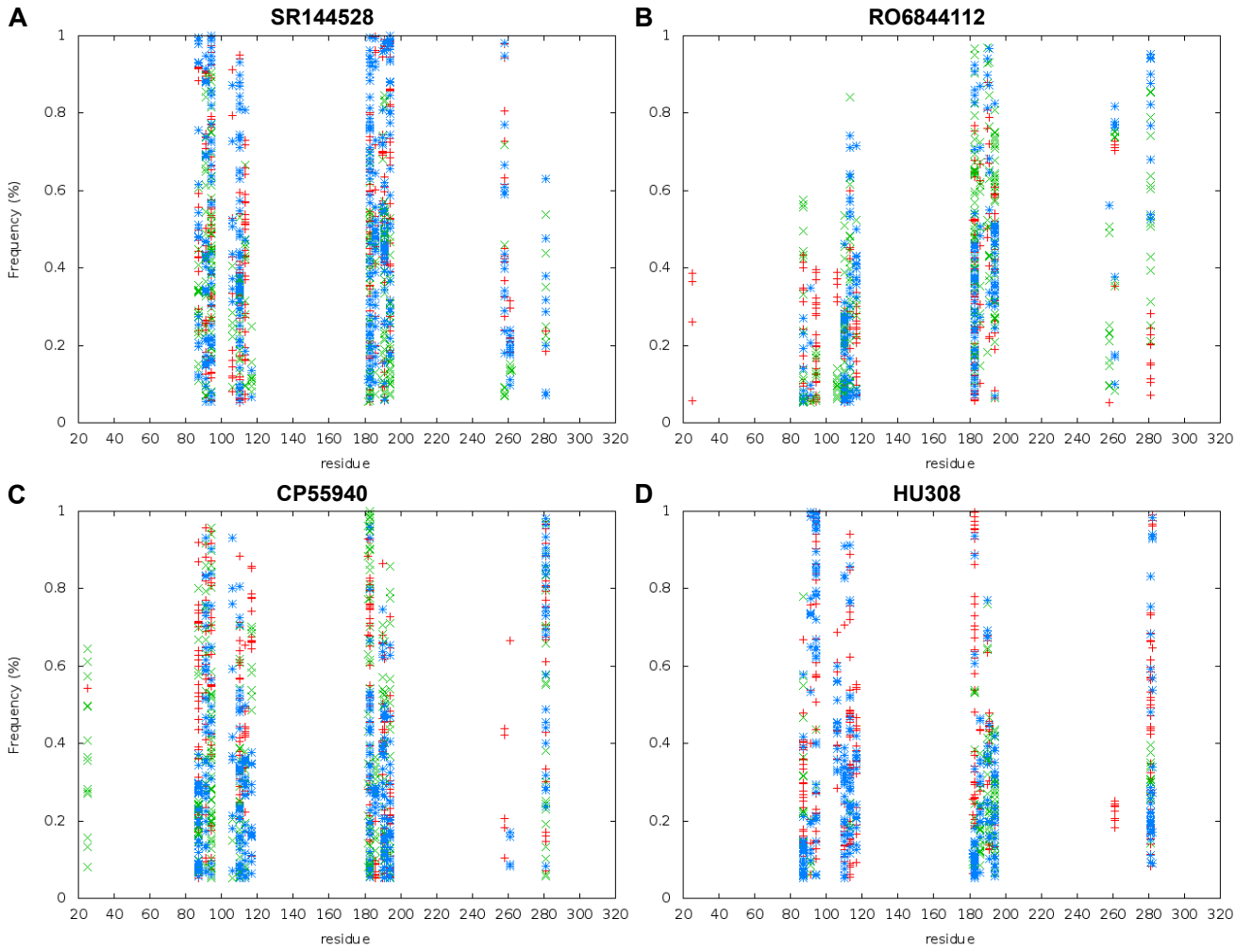


Figure A6.13: Frequency of native atom-atom contacts per residue between receptor and A: inverse agonist SR144528, B: agonist RO6844112, C: agonist CP55940 and D: agonist HU308 in the MD^{Gi}. For each replicate, native contacts between the ligand and any apolar residue within a distance of 4 Å were analysed. All native contacts above a frequency cutoff of 5% were extracted and plotted for each replicate in a different colour.

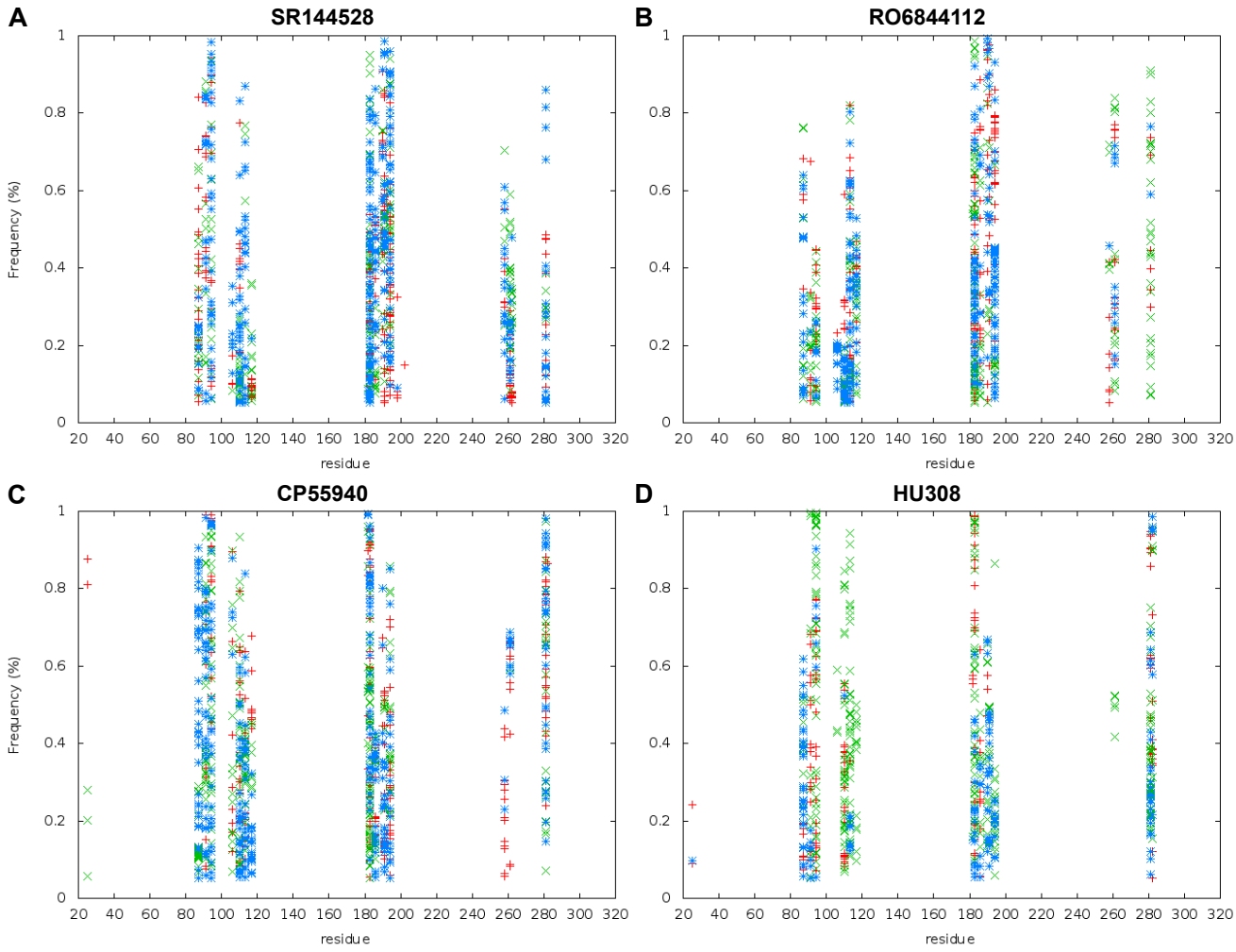


Figure A6.14: Frequency of native atom-atom contacts per residue between receptor and A: inverse agonist SR144528, B: agonist RO6844112, C: agonist CP55940 and D: agonist HU308 in the $MD^{\beta Arr}$. For each replicate, native contacts between the ligand and any apolar residue within a distance of 4 Å were analysed. All native contacts above a frequency cutoff of 5% were extracted and plotted for each replicate in a different colour.

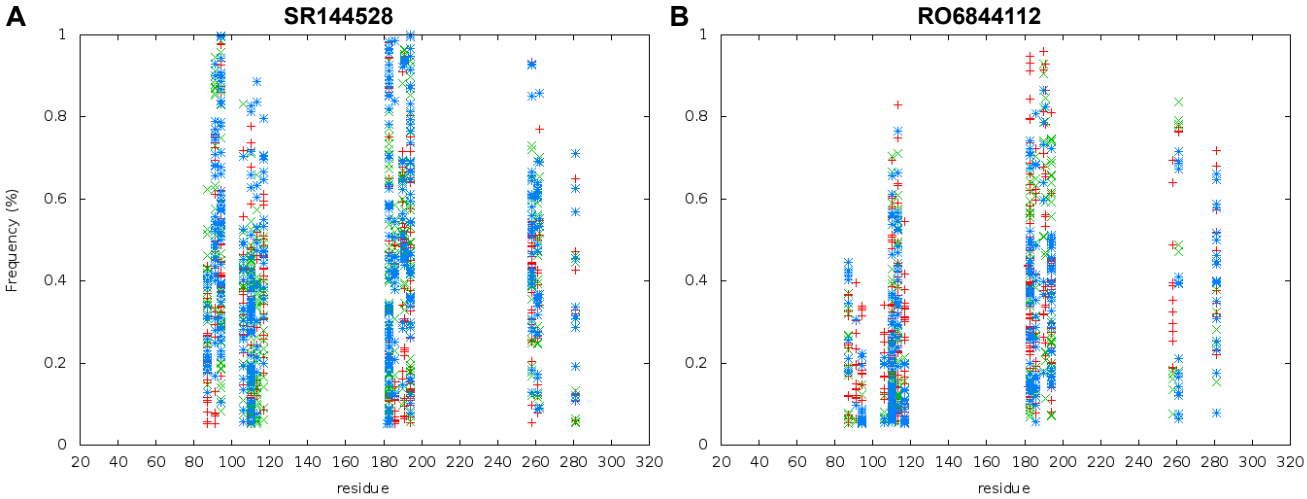


Figure A6.15: Frequency of native atom-atom contacts per residue between receptor and A: inverse agonist SR144528 and B: agonist RO6844112 in the MD^{inact}. For each replicate, native contacts between the ligand and any apolar residue within a distance of 4 Å were analysed. All native contacts above a frequency cutoff of 5% were extracted and plotted for each replicate in a different colour.

Table A6.1: RMSDs between the available crystal structures of the CB2. RMSDs were calculated for TM-C_α atoms, all binding site residue atoms and for residue $W258^{6.48}$.

	6PT0			6KPF			6KPC		
	TM	bp	W258	TM	bp	W258	TM	bp	W258
5ZTY	2.29	1.34	1.55	2.46	1.50	1.77	0.99	1.04	1.36
6PT0				0.94	1.39	1.39	2.12	1.33	1.57
6KPF							2.22	1.14	0.87
6KPC									

Table A6.2: RMSDs between cluster centroids from the ligand-based clustering. The main clusters were compared and additional clusters included if they showed substantial RMSDs to the main cluster. The fraction of frames forming each cluster are listed. RMSDs were calculated for TM-C_α atoms, all binding site residue atoms and the ligand.

SR144528

		$MD^{Gi}{}_0$			$MD^{Gi}{}_1$			$MD^{\beta Arr}{}_0$		
	Frac.	TM	bp	Lig	TM	bp	Lig	TM	bp	Lig
$MD^{inact}{}_0$	0.58	2.68	2.572	1.66	2.41	2.63	2.65	2.54	2.48	1.91
$MD^{Gi}{}_0$	0.68				1.75	2.45	2.96	1.32	1.60	1.67
$MD^{Gi}{}_1$	0.23							1.67	2.23	3.04
$MD^{\beta Arr}{}_0$	0.57									

RO6844112

		$MD^{Gi}{}_0$			$MD^{\beta Arr}{}_0$		
	Frac.	TM	bp	Lig	TM	bp	Lig
$MD^{inact}{}_0$	0.98	3.20	3.48	2.20	3.18	3.72	2.70
$MD^{Gi}{}_0$	0.62				1.41	1.93	1.51
$MD^{\beta Arr}{}_0$	0.53						

CP55940

		$MD^{\beta Arr}{}_0$		
	Frac.	TM	bp	Lig
$MD^{Gi}{}_0$	0.95	1.39	2.25	1.96
$MD^{\beta Arr}{}_0$	0.99			

HU308

		$MD^{\beta Arr}{}_0$			$MD^{\beta Arr}{}_1$		
	Frac.	TM	bp	Lig	TM	bp	Lig
$MD^{Gi}{}_0$	0.91	1.59	1.74	1.03	1.85	2.55	5.47
$MD^{\beta Arr}{}_0$	0.67				1.57	2.57	5.55
$MD^{\beta Arr}{}_1$	0.30						

Table A6.3: RMSDs between cluster centroids from the binding-site-based clustering. The main clusters were compared and additional clusters included if they showed substantial RMSDs to the main cluster. The fraction of frames forming each cluster are listed. RMSDs were calculated for TM-C_α atoms, all binding site residue atoms, the ligand and for residue W258$^{6.48}$.

SR144528

		$MD^{Gi}{}_0$				$MD^{\beta Arr}{}_0$											
	Frac.	TM	bp	Lig	W258	TM	bp	Lig	W258								
$MD^{inact}{}_0$	0.98	2.48	2.21	1.97	3.22	2.31	2.18	1.97	2.62								
$MD^{Gi}{}_0$	0.76					1.71	1.84	1.61	1.43								
$MD^{\beta Arr}{}_0$	0.41																

RO6844112

		$MD^{inact}{}_1$				$MD^{Gi}{}_0$				$MD^{\beta Arr}{}_0$				$MD^{\beta Arr}{}_3$			
	Frac.	TM	bp	Lig	W258	TM	bp	Lig	W258	TM	bp	Lig	W258	TM	bp	Lig	W258
$MD^{inact}{}_0$	0.37	2.59	3.91	1.88	5.39	3.62	2.85	1.90	1.31	2.77	2.31	2.44	1.34	2.38	2.60	2.08	1.16
$MD^{inact}{}_2$	0.20					3.22	3.49	2.62	5.99	2.96	3.70	3.34	6.07	2.89	4.80	2.76	5.65
$MD^{Gi}{}_0$	0.47									1.64	1.69	2.69	0.86	2.47	3.80	2.62	1.43
$MD^{\beta Arr}{}_0$	0.32													1.89	3.34	2.76	1.13
$MD^{\beta Arr}{}_3$	0.17																

CP55940

		$MD^{\beta Arr}{}_0$															
	Frac.	TM	bp	Lig													
$MD^{Gi}{}_0$	0.60	1.39	1.65	2.18													
$MD^{\beta Arr}{}_0$	0.61																

HU308

		$MD^{\beta Arr}{}_0$				$MD^{\beta Arr}{}_1$											
	Frac.	TM	bp	Lig		TM	bp	Lig									
$MD^{Gi}{}_0$	0.79	2.04	1.54	1.16		2.13	2.38	5.27									
$MD^{\beta Arr}{}_0$	0.65					1.67	2.28	5.18									
$MD^{\beta Arr}{}_1$	0.33																

Table A6.4: RMSDs between cluster centroids from the binding-site-based clustering, sorted and compared by effector protein. The main clusters were compared and additional clusters included if they showed substantial RMSDs to the main cluster. The fraction of frames forming each cluster are listed. RMSDs were calculated for TM-C_α atoms, all binding site residue atoms and for residue W258$^{6.48}$.

MD$^{\text{inact}}$

		$RO6_0$			$RO6_2$										
	Frac.	TM	bp	W258	TM	bp	W258								
$SR1_0$	0.98	2.621	1.885	2.329	2.649	4.001	5.815								
$RO6_0$	0.373				2.588	3.908	5.392								
$RO6_2$	0.201														

MD$^{\text{Gi}}$

		$RO6_0$			$CP5_0$			$HU3_0$							
	Frac.	TM	bp	W258	TM	bp	W258	TM	bp	W258					
$SR1_0$	0.761	1.689	2.712	2.695	1.548	2.061	2.558	2.014	2.064	2.501					
$RO6_0$	0.469				1.505	1.983	0.613	2.065	2.516	0.996					
$CP5_0$	0.6							2.083	1.979	0.541					
$HU3_0$	0.789														

MD$^{\beta\text{Arr}}$

		$RO6_0$			$RO6_3$			$CP5_0$			$HU3_0$			$HU3_1$		
	Frac.	TM	bp	W258	TM	bp	W258	TM	bp	W258	TM	bp	W258	TM	bp	W258
$SR1_0$	0.413	1.46	2.382	1.746	1.624	2.71	2.321	1.505	2.001	1.823	1.285	1.775	1.615	1.858	2.561	1.702
$RO6_0$	0.316				1.891	3.344	1.126	1.491	1.932	0.734	1.221	1.846	0.739	1.901	2.69	0.63
$RO6_3$	0.167							1.628	2.597	1.291	1.397	2.674	1.166	1.852	2.38	0.886
$CP5_0$	0.614										1.046	1.669	0.998	1.676	2.01	0.828
$HU3_0$	0.646													1.672	2.283	0.521
$HU3_1$	0.333															

7 | Prediction of small molecule binding sites to target the RGS7-$G_{\beta 5}$ complex

Contributions. The main part of the study described in this chapter was conducted by the author of this thesis. The following parts mentioned here were contributed by other people: An initial suggestion of possible binding sites was done by Julius Perschel and the collaborators at the Martemyanov lab. One of three structures used for docking calculations was prepared by Julius Perschel. The docking screen against proposed molecule binding site *region 2* was conducted and analysed by Jonas Kammertöns, supervised by the author of this thesis. All docking screens of the SDDL library against different proposed molecule binding sites were first evaluated by Theresa Jünemann to filter for molecules with more likely binding poses, followed by a more detailed analysis by the author of this thesis.

7.1 Introduction and goal of the study

Regulator of G protein signalling (RGS) proteins play an important role in an impressive number of cellular processes and diseases.[140] Although it is not entirely understood which physiological functions they are involved in and how they act exactly, they could serve as important drug targets in the treatment of a plethora of diseases (cf. section 2.2).[130,140,141] To develop a better understanding of their function and how they could be targeted in diseases, it is important to find small molecule tool compounds which interact with these proteins and modulate them. RGS7, one member of the R7 RGS protein family, has recently been crystallised in complex with the $G_{\beta 5}$ protein (PDB ID 6N9G[133]). This crystal structure can be used as a starting point for computational studies to target this protein complex.

In this study, we aimed to find such small molecules interacting with the RGS7 protein complex, thereby either inhibiting interactions with other proteins or inducing conformational changes, both leading to a changed behaviour of this protein *in vitro*. To achieve this goal, an *in silico* approach was used, screening a molecule library against the RGS7-$G_{\beta 5}$ complex structure using docking calculations. Any potential binders identified by this approach could then be tested *in vitro* by our collaboration partners from the Martemyanov lab at The Scripps Research Institute in Florida.

The search for such small molecules that could modulate the RGS7 function comprises challenges that make this screening project particularly difficult. Although it is known which parts of the RGS7-$G_{\beta 5}$ complex are most likely involved in interactions with other proteins such as the G_{α} protein, R7BP, GIRKs or GPCRs, the exact interaction areas remain unknown. Furthermore, there

is no knowledge about changes of the complex structure while interacting with other proteins or while acting as a GTPase activating protein (GAP). Most importantly, there is no knowledge where a small molecule could bind to the complex and how it could induce conformational changes by interacting with the protein complex.

The first obstacle in this project was to identify cavities in the protein complex that might not only serve as binding sites for small molecules but are also located in parts of the complex where small molecule binding could either inhibit interactions with other proteins or where it might influence the conformation of the complex and thereby modulate its function.

An additional challenge is that the discovered molecules need to be able to reach the protein complex, i.e. the molecules need to be able to pass the cell membrane to get to the intracellularly expressed and acting RGS7. This challenge bears certain difficulties by itself since screening libraries are not specifically designed to contain cell membrane permeable molecules. Furthermore, there are no reliable tools or methods available to predict computationally how well a molecule could pass a cell membrane. This prediction can be based on calculated logP values but this is not more than an initial guess.

Despite of these difficulties on top of the mere challenge of predictions made by docking calculations, this project is not only very interesting and challenging but potentially also very rewarding. Any small molecule that binds to the RGS7-$G_{\beta5}$ complex and modulates its function by changes in the complex conformation would help to understand these proteins better, especially if the actual molecule binding site is also known. A small molecule inducing such a modulation could be a strong tool compound in the further investigation of this fascinating protein family and maybe even result in a medical drug targeting this protein class.

Previous work. Before I started working on this project, some preliminary work had been done already. The solved crystal structure of the RGS7-$G_{\beta5}$ complex was forwarded to our lab by the Martemyanov lab before it was officially released and Julius Perschel started to work on the project. He prepared chains AD of the crystal structure (PDB ID 6N9G [133]) for docking and did a preliminary visual analysis with the complex structure to identify cavities that could potentially bind small molecules. Parts of the cavity analysis conducted by the author of this thesis were based on these cavities proposed by Julius Perschel as well as cavities proposed by the Martemyanov lab (namely *J1, region 1, region 2, J3/region 3, J4* and *long-cav*).

7.2 Methods

7.2.1 Structure preparation

The structure with PDB ID 6N9G was prepared according to the general protocol described in section 8.1.1. Since the complex was crystallised as a dimer and both chains showed slight differences in some of the closer evaluated regions, both chain AD and chain BC were prepared separately for docking. Additionally, chain AD was prepared twice, with histidines H88 and H97 in chain A either protonated in δ-position (HID) or in both δ- and ε-position (HIP), since these histidines are located in potentially interesting cavities and might, therefore, influence interactions and binding modes of potential ligands. The protein structure of chain AD with one protonation state of these histidines was prepared by Julius Perschel.

To prepare the protein complex for DOCK targeting the cavities which were selected for further evaluation, dummy ligands located in the desired cavity areas were used. In most cases, those dummy ligands were based on fragment poses generated during the prior analysis using SEED. The spheres used by the DOCK algorithm for placement of the docked molecules within the protein area of interest were then moved to fill the desired parts in the cavities. Varying locations of spheres within a cavity were especially important for cavity M1 which is wide enough to potentially bind molecules in different parts of the cavity.

7.2.2 Prediction of cavities

SEED analysis. To find and evaluate cavities that might potentially act as binding sites for ligands, fragments were docked to different regions of the complex structure using SEED (v.3.3.4).[178] The general procedure described in chapter 8.1.3 was used for preparation and execution. SEED was applied to chain AD of the structure with PDB ID 6N9G (H80 and H97 in chain A protonated as HIP; different protonation states of these residues were used in some of the docking screenings). Possible cavities to be explored were initially suggested by Julius Perschel and our collaborators at the Martemyanov lab or were found during the further analysis. The complete protein complex was explored using a small subset of 6 diverse fragments (acetate, benzene, ethanol, phenol, propane, water) to locate possible additional cavities. The complete set of 160 available fragments was then docked to each of the 9 suggested cavities. A representative subset of fragment poses and centers of mass was analysed visually for each of these suggested cavities to evaluate qualitatively whether the cavity might be suitable to bind ligands. As criteria for that, the placement of polar fragments as indicator for possible polar interactions and the placement of larger fragments as indicator for pocket volume (i.e. whether the cavity could accommodate small molecules) were mainly evaluated.

fpocket analysis. The fast cavity detection algorithm fpocket[236] was used as an alternative approach to find cavities and to evaluate whether cavities might be targeted with ligands. In brief, this detection algorithm uses alpha spheres (i.e. spheres contacting four atoms on their boundary) that are generated from Voronoi tesselation of the heavy atoms. These alpha spheres are then filtered by their radius based on the theory that spheres within a protein cavity should have an intermediate radius. The remaining spheres are than clustered in three steps based on their distance to other spheres and to members of other sphere clusters.[236] Additionally, a druggability score is implemented that assesses the likeliness of a drug-like molecule to bind to the cavity based on several pocket descriptors calculated by fpocket.[237]

The algorithm allows to set different parameters to influence which pockets are found. Here, different settings for minimal alpha sphere radius m and maximal alpha sphere radius M and for minimal distance between sphere clusters s to not be joined were tested (default: m=3.4 Å, M=6.2 Å and s=2.5 Å; setting 2: m=2.9 Å, M=5.5 Å and s=2.5 Å; setting 3: m=3 Å, M=5 Å and s=2 Å). These parameters influence mainly how voluminous and burried the found cavities are. As input, chain AD was used as prepared for docking. Very small cavities and cavities on the protein surface were excluded from further analysis. The druggability score calculated by fpocket was evaluated for the resulting set of cavities as well as all cavities chosen from the analysis with SEED.

7.2.3 Docking screens

ZINC15 *lead-like* library. The ZINC15[166] *lead-like* library was docked to two of the proposed cavities, *region 2* and *M1*, using DOCK3.7[171] aiming for different objectives.

The docking calculations targeting *region 2* aimed to find possible ligands binding to and modulating the RGS7 protein complex. This docking screen was conducted and analysed by Jonas Kammertöns, supervised by the author of this thesis. In brief, he docked the molecule library to two conformations of *region 2*, i.e. chain AD and chain BC, and evaluated the molecule poses of the top 500 ranked molecules from the resulting ranking lists by visual inspection. This visual inspection is necessary to account for known deficiencies of the used scoring functions.[200]Finally, 18 molecules were selected to be purchased and were sent to the Martemyanov lab and tested in *in vitro* assays.

The docking calculations of the molecule library to cavity *M1* served a different perspective. The initial aim was to identify potential ligands targeting this cavity. However, this screen was then mainly used to evaluate different docking setups. Since this potential binding site is quite spacious, it was possible to not only address the complete cavity but also parts of it. This was done by placement of the spheres used by DOCK to translate and rotate molecules into the protein binding

site in certain locations of the cavity. In total, 6 different docking setups targeting *M1* were used and at least the top 100 ranked molecule poses of each ranking list were inspected visually to evaluate each docking setup. During visual inspection a focus was put on e.g. the occurence of stranded donors and possible solvent exposure of the docked molecules. Based on these preliminary docking calculations two docking setups were chosen to be targeted by the SDDL molecule library described below.

SDDL library. Molecule data for an in-house library of The Scripps Research Institute in Florida (SDDL library) was provided to us by our collaborators at the Martemyanov lab. Conformer files for DOCK were created using the db2 preparation pipeline, resulting in a dockable molecule library of 628,596 molecules (approx. 98% of the originally provided library). This molecule library was then docked to different cavities of the protein complex using DOCK3.7.[171] In total, the library was docked to seven different cavity preparations: two of M1 (with different sphere setup), three of region 2 (chain BC, chain AD and chain AD with HID97) and two of M2 (with H88 as either HID or HIP). An initial analysis of the top 1000 molecules of each of the resulting ranking lists was done visually by Theresa Jünemann. These pre-selected molecules were then further evaluated by the author of this thesis to select the most targetable cavity.

7.3 Results

The project aimed to find small molecules that could interact with the RGS7-$G_{\beta 5}$ complex and potentially modulate its function by inducing conformational changes. The first step was to find cavities and analyse whether they could serve as binding sites for small molecules. For this analysis, small molecular fragments were docked to the entire structure as well as nine selected cavities using SEED. Additionally, the fpocket algorithm was applied to the complex structure to predict possible binding sites and evaluate their druggability. Based on these analyses, the cavities that seemed most likely to serve as binding sites for small molecules were selected and targeted with docking calculations. The visual evaluation of the molecule poses resulting from these docking screens did then serve as a second layer for the analysis of the druggability of the addressed cavities and to screen for possible ligands to be tested in *in vitro* assays against the target complex. The cavities that were evaluated during the different analysis steps are listed in Table 7.1.

7.3.1 Cavity prediction and analysis

Six possible binding sites were suggested by Julius Perschel and the Martemyanov lab prior to the cavity analysis (*J1, region 1, region 2, J3/region 3, J4* and *long-cav*; see Figure 7.1.A). These initial suggestions consisted mainly of a few residues in each of the mentioned spots. Therefore, the

Table 7.1: List of evaluated cavities, how they were identified and their approximate location in the RGS7 protein complex. All cavities except *long-cav* are located in the interface region of the mentioned parts of the protein complex. Cavities highlighted in bold were included for subsequent docking screens.

Cavity	Identifying method	Location
Region 1	initial suggestion	DEP/$G_{\beta5}$
Region 2	**initial suggestion**	**DHEX/DEP**
J3/region 3	initial suggestion	$G_{\beta5}$/RGS
J1	initial suggestion	DHEX/DEP/DHEX-GGL linker
J4	initial suggestion	$G_{\beta5}$/GGL/RGS
$J4_{small}$	initial suggestion	$G_{\beta5}$/GGL/RGS
long-cav	initial suggestion	within $G_{\beta5}$
M1	**SEED**	**$G_{\beta5}$/RGS**
M2	**SEED**	**DEP/DHEX-GGL linker/$G_{\beta5}$**

definitions of these suggested cavities were expanded to cover the complete potential cavity prior to the SEED analysis.

SEED analysis. The complete set of 160 fragments was docked to the six initially proposed cavities using SEED (see Table 7.1; Figure 7.1). The results were then analysed visually by evaluating the distribution and binding locations of a representative selection of the fragments in the cavity. It was made sure for each cavity that fragments of different molecular weight and different polarity including charged fragments were evaluated to gain a more comprehensive picture of fragment distributions and whether the cavity could accommodate a small molecule. Three of the proposed cavities (*J1, region 1, J3/region 3*) did either not fit bigger fragments or there was no space to expand the bound fragments into bigger molecules. These pockets were, therefore, excluded from any further evaluation or subsequent docking calculations. Fragment distributions in cavity *J4* were not satisfying since fragment clusters were mainly found at the edges of the cavity. Since this might have been caused by a too broad definition of this cavity, a smaller version of cavity *J4*, $J4_{small}$, was additionally explored using SEED. Fragments were binding deeper within the cavity of $J4_{small}$ compared to the original definition of *J4*, however, in some parts of the cavity no fragments bound at all. Hence, cavity $J4_{small}$ was also excluded from a further evaluation and docking calculations. Two cavities, *region 2* and *long-cav*, were selected for a further evaluation. Both cavities are spacious enough to accommodate small molecules and polar fragments were clustering in certain parts of the cavities, indicating that a potential ligand could be stabilised in the binding pocket by forming polar interactions with the protein.

Additional to these SEED runs targeting specific parts of the protein, an analysis for the complete protein complex was conducted. Six fragments with different polarities (acetate, water, ethanol, phenole, propane, benzene) were docked to the entire complex structure to discover overall binding hotspots to reveal possible additional cavities to the six previously proposed ones. Several interesting fragment clusters showed up, some of which were located within the proposed cavities. Furthermore, fragment clusters were found in two additional cavities, *M1* and *M2* (see Table 7.1; Figure 7.1.B). Each of these cavities was addressed and evaluated by docking the complete fragment library using SEED. Cavity *M1* seemed sufficiently spacious to target it with small molecules and fragment clusters indicated that polar interactions of molecules within this cavity should be possible. Although cavity *M2* was less spacious, fragment clusters within this cavity indicated possible interaction sites and enough space to extend bound fragments into bigger molecules.

fpocket analysis. As an additional method to predict and analyse possible ligand binding sites, the fpocket algorithm was applied to the complete protein complex. Predictions of this algorithm can vary depending on chosen parameter combinations. Therefore, different parameter sets were chosen, resulting mainly in differences in the size of the predicted cavities and in solvent exposure (see Figure 7.2). The results obtained with each setting were analysed visually. All cavities predicted by fpocket that seemed too small to accommodate small molecules or that were located on the protein surface and, therefore, too exposed to the solvent were discarded and not included for further analysis. The majority of the remaining pockets predicted by fpocket were similar to the cavities that were previously analysed with SEED (see Figure 7.2). Based on the fpocket results, no additional possible ligand binding sites were found.
As an additional measure to analyse the predicted binding sites, fpocket gives a druggability score. Since fpocket predictions were in similar regions of the complex as the suggested ligand binding sites *region 2, M1, M2* and *long-cav*, the druggability score for these cavities could be analysed. Cavities *M2* and *long-cav* showed good druggability scores (i.e. between 0.5 and 1) and, therefore, seem likely as small molecule binding sites. Cavity *M1* is quite spacious but showed good druggability scores especially in proximity to the RGS domain. Surprisingly, *region 2* was only partially found by the fpocket algorithm which might indicate problems in the druggability of this possible binding site. It should be kept in mind though that this druggability score mainly evaluates the likelihood of drug-like molecules to bind to the cavity and results might be different for smaller or larger molecules.

Cavity selection for docking screens. Based on the results of the SEED and fpocket analysis, the proposed cavities were evaluated and the ones that seemed most likely to serve as small molecule binding sites were taken into closer consideration. Four cavities were selected for this (*region 2,*

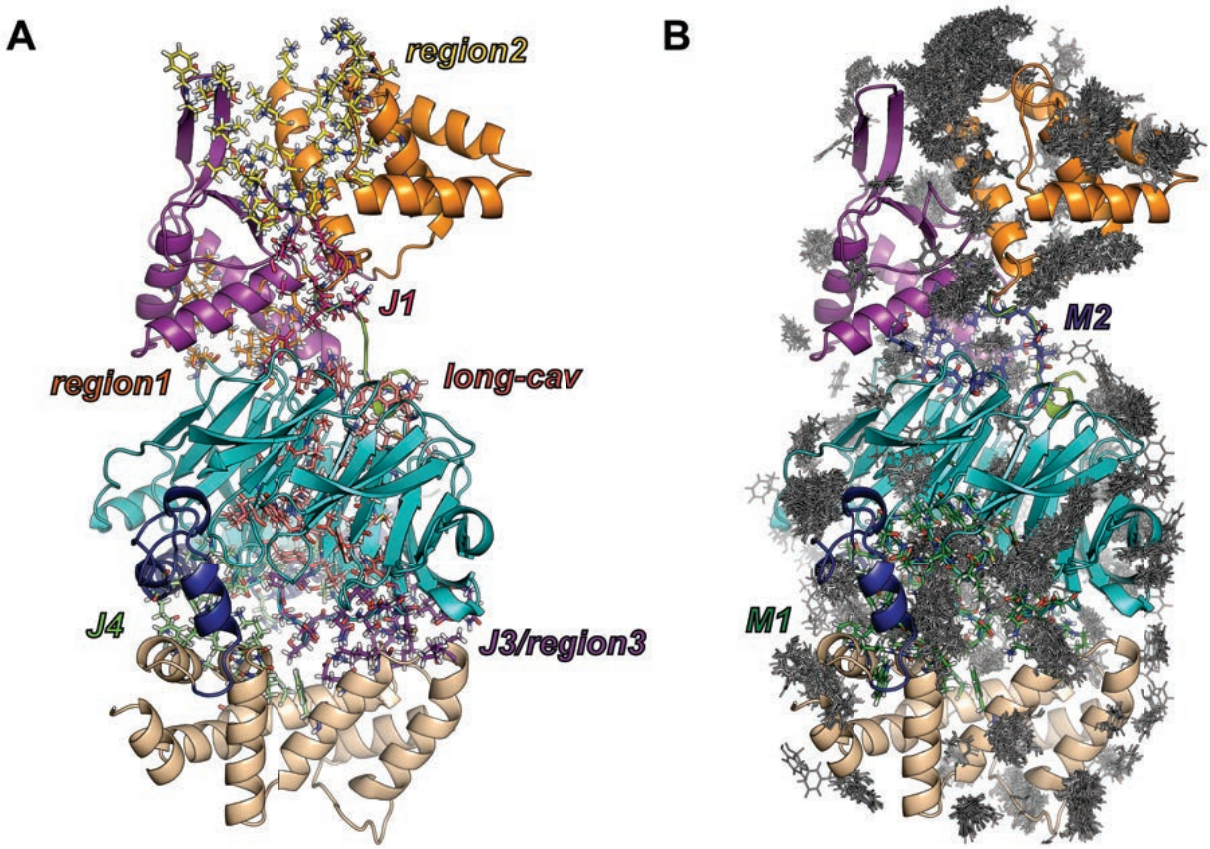


Figure 7.1: Proposed cavities that were evaluated by a SEED fragment docking. A) Cavities were initially proposed by Julius Perschel and the Martemyanov lab. These cavities are located in different parts of the RGS7-$G_{\beta 5}$ complex and were initially suggested based on visual inspection of the crystal structure (PDB ID 6N9G). B) To find additional potential binding pockets, the complete RGS7-$G_{\beta 5}$ complex was subjected to a SEED analysis with six fragments (acetate, water, ethanol, phenole, propane, benzene; all clusters are shown). Based on the fragment clusters, two additional cavities were identified.

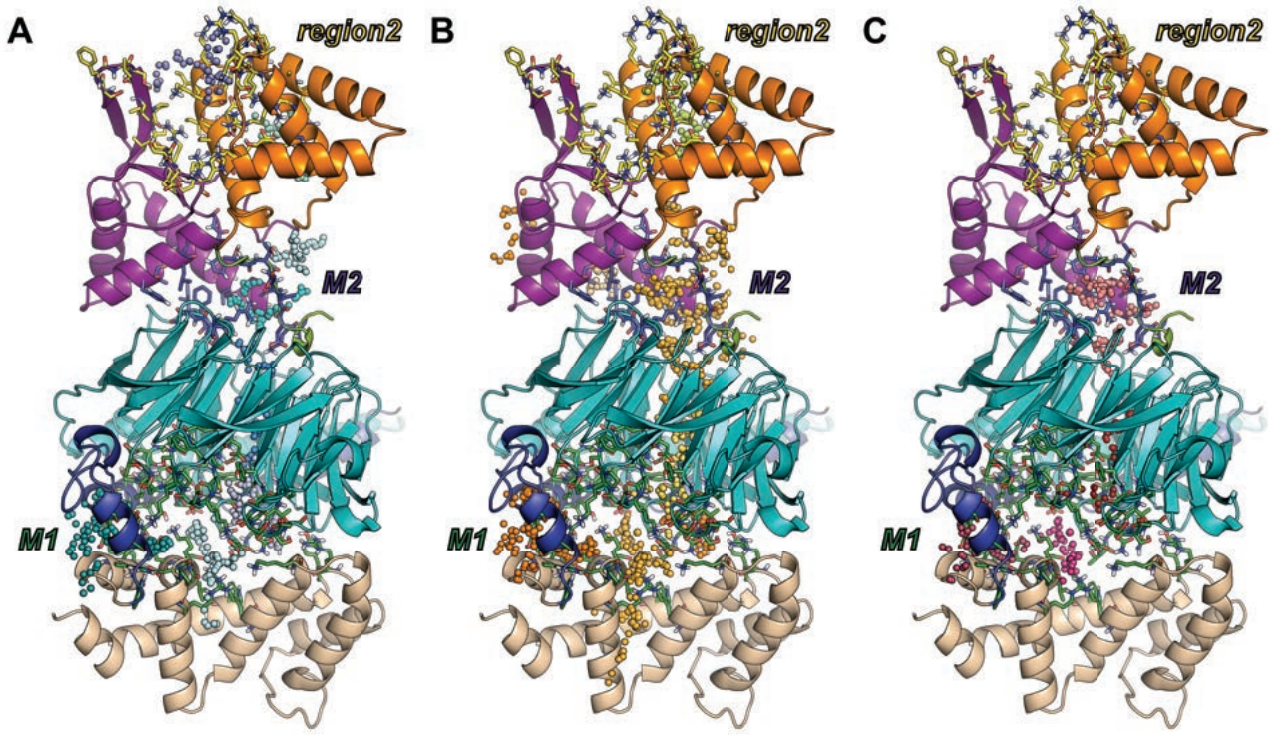


Figure 7.2: The fpocket algorithm was applied to the RGS7-$G_{\beta 5}$ complex to identify possible ligand binding site. The algorithm was either run with default parameters (a) or parameters were changed to identify smaller and more buried cavities using setting 2 (B) or setting 3 (C). While cavities *M1* and *M2* were identified with all parameter combinations, cavity *region 2* was only partially identified when using default parameters. The detected cavities are shown by coloured spheres. Only those potential cavities that were not too small or solvent exposed to be targetable by small molecules were kept.

M1, M2 and *long-cav*; see Figure 7.3) and further evaluated regarding their location in the protein complex and other features, before making the final selection of cavities to be targeted with docking calculations.

Region 2 is located at the interface of the DEP and DHEX domain (see Figure 7.3.B) which is a potentially interesting spot to target with small molecules. As described in section 2.2, the DEP domain is involved in interactions with other proteins such as R7BP and GPCRs. Hence, a small molecule binding at the DEP/DHEX interface might potentially inhibit an interaction of RGS7 with other proteins or it could induce conformational changes that might lead to a modulated function of the protein. On the other hand, some of the features of this cavity might also be problematic. *Region 2* consists of a relatively large number of positively charged residues that point into the cavity. Although this might in principle result in strong polar interactions between small molecule and protein which would lead to strong binding of the ligand, it might be problematic to

transport such polar small molecules through the cell membrane to reach the intracellularly located target. Another problem is that *region 2* is quite solvent exposed which can be a problem for ligand binding. Additionally, the cavity was only partially detected during the fpocket analysis, indicating problems with the druggability of this cavity. Nonetheless, *region 2* was selected to be targeted in docking calculations due to its interesting location in the protein complex and its potentially good properties for strong interactions with small molecules.

Cavity *M1* is located at the interface of the RGS domain and the $G_{\beta5}$ protein (see Figure 7.3.C). Interactions with the RGS domain might lead to a modulation of its catalytic function and, therefore, influence the deactivation of the G_{α} protein. Additionally, molecules binding in this region might also influence the formation of the RGS7-$G_{\beta5}$ complex which might result in a changed behaviour of the complex. The cavity is quite spacious and fpocket predicted a good druggability score in proximity to the RGS domain. It is not as solvent exposed as *region 2* but accessible from the solvent. The SEED analysis showed binding of polar fragments in several parts of the cavity which indicates the possibility of polar interactions between a small molecule and the protein. Due to the spacious nature of this cavity it might be possible to target it with drug-like molecules with a bigger molecular weight. Overall, cavity *M1* seems interesting to target in a docking screen with molecules of various molecular weights.

Cavity *M2* is located at the interface of several parts of the complex between DEP, DHEX-GGL linker and the $G_{\beta5}$ protein (see Figure 7.3.A). A small molecule interacting with the protein complex in this region could modulate the RGS7 function in various ways. It could influence the conformation of the DEP domain which plays a role for interactions with other proteins. Another possibility is that it could have an impact on the complex formation with the $G_{\beta5}$ protein. Finally, it could modulate the conformation and flexibility of the DHEX-GGL linker which might, again, impact complex formation with the $G_{\beta5}$ or other proteins or which might lead to overall conformational changes modulating the function of RGS7. The cavity is quite buried but accessible from the solvent. It is not very spacious but big enough to fit small molecules which could form polar and apolar interactions with several residues within the cavity. Since cavity *M2* seems quite addressable especially also for smaller molecules and the cavity is also druggable according to the fpocket predictions, it was decided to target this cavity in docking screens.

The last cavity that was evaluated more closely is cavity *long-cav* (see Figure 7.3.D). It is located in the center of the $G_{\beta5}$ protein within the propeller-like oriented β-sheets. The druggability score predicted by fpocket for this cavity is quite high and fragment distributions from SEED analysis seem favourable within this cavity. The cavity is spacious enough to accommodate a small molecule which could form several polar and apolar interactions with the residues within the cavity. Cavity

long-cav is buried and not solvent exposed but accessible. In principle, cavity *long-cav* seems very addressable by small molecules and could be interesting to target in a docking screen. Its location within the protein complex, however, is probably not useful for the desired purposes. The cavity is located in a structurally quite stable protein fold and a small molecule binding in this region would probably not be able to disrupt this structural scaffold or induce conformational changes of the $G_{\beta 5}$ protein. Aside from that, it is located within the $G_{\beta 5}$ protein and does not have contact with the RGS7 protein, hence, it would probably not induce conformational changes that could lead to a modulated RGS7 function. Therefore, it was decided to not target this cavity in a docking screen.

7.3.2 Docking screens and cavity evaluation

From the SEED and fpocket analysis three cavities (*region 2*, *M1* and *M2*) were chosen to target in docking screens. These screens served two purposes. First of all, the druggability of the cavities was evaluated further based on the results of these screens. This evaluation was based on the number of molecule poses forming favourable interactions with the protein, a low number of molecule poses containing stranded hydrogen bond donors and the overall consistency of molecule poses inside the cavities as an indicator of the overall quality of the docking results. Secondly, the docking calculations served as a screening for molecules that might bind to and influence the RGS7-$G_{\beta 5}$ complex.

Two molecule libraries were docked to the different proposed cavities. The ZINC15 *lead-like* subset consisting of 7,495,140 molecules was docked to cavity *region 2* in chain AD as well as chain BC and to cavity *M1* in different docking setups. Furthermore, an in-house library of The Scripps Research Institute in Florida (SDDL library; 628,596 molecules) forwarded to us by our collaboration partners at the Martemyanov lab was docked to all three selected cavities. In total, the SDDL library was docked to seven different setups of these cavities: *region 2* in chain AD with HID97, chain AD with HIP97 and chain BC; *M1* in two different docking setups; and *M2* with HID88 and with HIP88. Of note, chain BC was only included in docking calculations targeting *region 2* since the conformation of chain AD and chain BC differs only in that cavity and in none of the other cavities.

Cavity *region 2*. Cavity *region 2* was targeted by two molecule libraries. The ZINC15 *lead-like* library screen was conducted and analysed by Jonas Kammertöns, supervised by the author of this thesis. Therefore, it will only be mentioned briefly for a more complete picture. Since the conformation of *region 2* differs between the two different dimers, the library was docked to both, chain AD and chain BC. After visual evaluation of the molecule poses by Jonas Kammertöns, a total of 18 molecules was selected to be tested in *in vitro* assays by our collaborations partners in the Martemyanov lab. All selected molecules except one contained at least one negatively charged

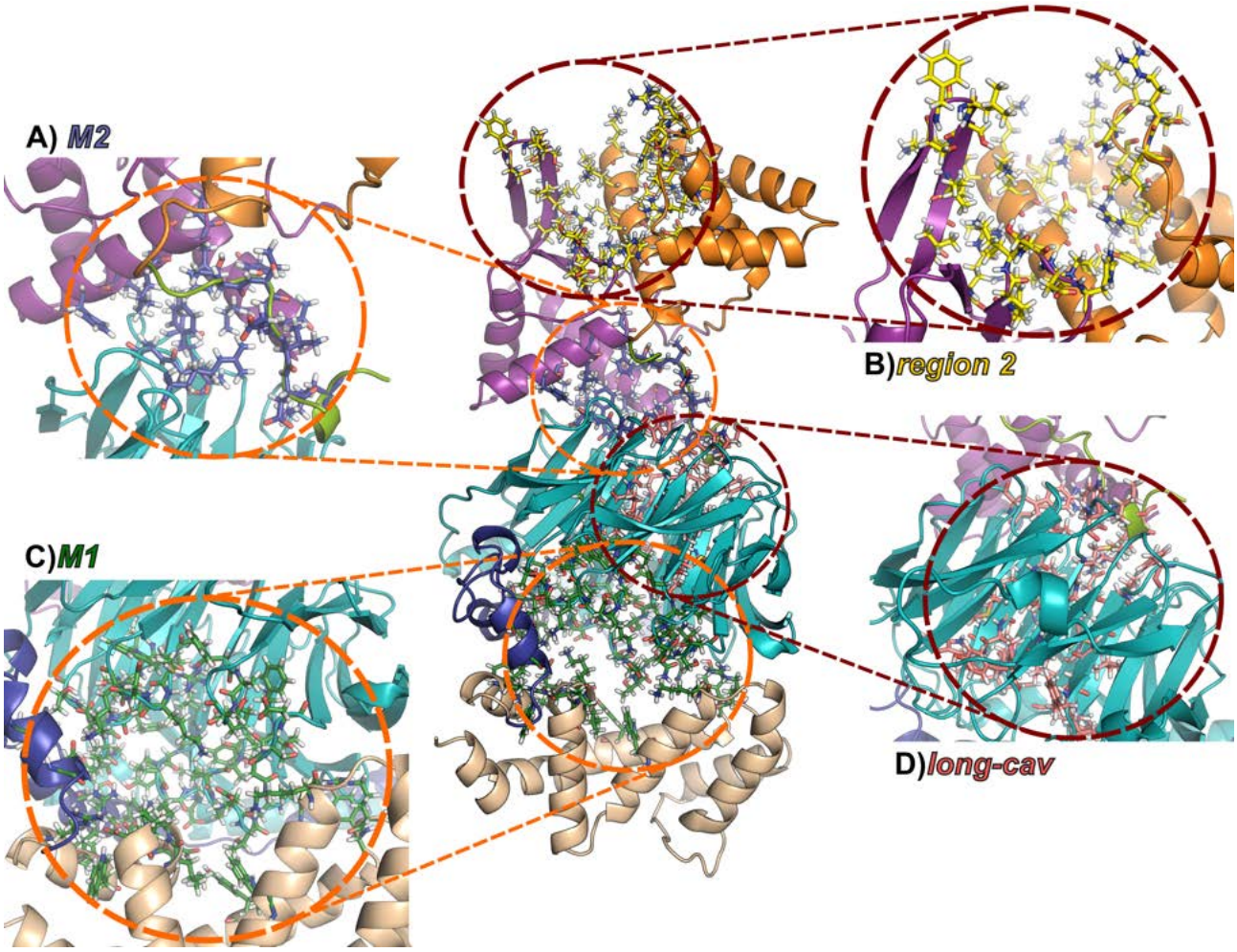


Figure 7.3: Location of cavities in the RGS7-$G_{\beta5}$ complex that were evaluated further based on their position in the complex before being targeted with a docking calculation. A) Cavity *M2* is located at the interface of DEP (violet), DHEX-GGL linker (green) and $G_{\beta5}$ (cyan). B) Cavity *region 2* is located at the complex surface of DEP (violet) and DHEX (orange). C) Cavity *M1* covers the region between $G_{\beta5}$ (cyan) and RGS (wheat). D) Cavity *long-cav* is located within the $G_{\beta5}$ protein (cyan). While *M2*, *region 2* and *M1* were selected as target sites for docking calculations, *long-cav* was excluded from this analysis due to its position in the RGS7-$G_{\beta5}$ complex.

carbonic acid group. Unfortunately, none of the selected compounds showed an effect in the assays (results not shown). Of note, this does not necessarily mean that the compounds do not show any affinity to the protein, since the conducted assays do not measure affinity but rather conformational changes of the protein complex. However, with the current limitations regarding assays to measure binding affinity of a small molecule to this protein complex, this cannot be elucidated further and the molecules are considered as non-binders.
In a second docking screen, the SDDL library was also docked to cavity *region 2*. Again, the docking calculations were conducted using both conformations, i.e. chain AD and chain BC. Additionally, two different protonation states of H97 in chain AD were tested (HID and HIP) since this residue is pointing into the cavity and might, therefore, influence molecule poses and interactions. Overall, the results were quite similar in all three docking calculations. The majority of the molecules pre-selected by Theresa Jünemann from the top 1000 ranked molecules contained carbonic acid groups which can form ionic interactions with the positively charged residues in the cavity. In general, the pre-selected compounds were polar and formed several polar interactions with the protein. Due to the shallow and solvent exposed nature of the cavity, molecule poses of smaller molecules were more favourable, since molecule parts of the bigger molecules were often extended into the solvent which is unfavourable for ligand binding. Overall, only few of the molecules seem worth considering to be tested in *in vitro* assays.

Cavity *M1*. Cavity *M1* has a quite big volume and is too big to be completely occupied by one molecule. This allows targeting only certain parts of the cavity by placement of the DOCK spheres in the desired regions. In total, six different setups with DOCK spheres in different regions of the cavity were created. In three of the tested setups the sphere locations were based on fragment clusters of the SEED analysis, while the other three were not directly based on results of that analysis. The ZINC15 *lead-like* subset was docked to the protein using each of the setups. At least the top 100 ranked molecule poses resulting from each of these docking calculations was then inspected visually to evaluate the docking setup. This evaluation was based on molecule interactions, stranded donors and overall consistency of docking poses. Overall, stranded donors were observed frequently for the molecule poses resulting from each of the docking setups, especially if the spheres were distributed within the complete cavity or located close to the $G_{\beta 5}$ protein. Molecule poses resulting from the docking setups with spheres located closer to the RGS domain had less stranded donors and poses were more consistent. Overall, the docking results of none of the docking setups was considered particularly good based on this evaluation. However, it was decided to include the two setups with the highest numbers of tolerably good poses for the SDDL library screen. In both of these setups, the DOCK spheres were located close to the RGS domain.

The overall results of the SDDL docking screen using the two selected docking setups of cavity *M1* were similar to the results of the ZINC15 *lead-like* library docking calculations. As observed for the ZINC15 *lead-like* library docking calculations, many of the top ranked molecule poses had stranded donors or molecule parts that were exposed to the solvent. The molecules found in the top ranks of both SDDL docking calculations to *M1* had more extended structures and bigger molecular weights than the molecules found in the top ranks of the screens against the other two cavities, *region 2* and *M2*, i.e. they were rather drug-like than lead-like molecules. Comparing the results of these two different docking setups based on the molecule poses pre-selected by Theresa Jünemann, one of these setups yielded better results than the other one. Results from this docking calculation contained fewer stranded donors and molecule poses were overall more consistent compared to the results using the other docking setup. However, overall none of the two docking setups did yield docking results which made any of the molecules seem to be worth selecting and testing in *in vitro* assays.

Cavity *M2*. The SDDL library was also docked to two different setups of cavity *M2*, one with HID88 and one with HIP88. In both cases, docking results were quite similar. The top ranked molecules were of medium, lead-like size and contained polar groups. Based on the molecule poses pre-selected by Theresa Jünemann, the molecules formed polar interactions to the protein with fewer stranded donors than observed in the results from the docking calculations to cavity *M1*, and molecule poses were overall quite consistent. However, cavity *M2* has a much smaller volume than *M1* and is more closed which might be a reason for more consistent molecule poses and also results in molecule poses with less solvent exposure. A recurring interaction motif of several of the molecule poses was a bridging interaction between D259 and R327 (both in chain D). Differences in the docking results for the setups with HID88 compared to HIP88 were rather small with in general similar top ranked molecules and comparable molecule poses. Overall, both docking setups yielded good results according to the evaluation criteria and the selection of molecules to be tested in *in vitro* assays is conceivable.

7.4 Discussion

Cavity prediction and analysis. Most of the suggested cavities were initially selected by visual inspection of the protein complex (six out of nine evaluated cavities; see Table 7.1) and further prediction and analysis was mainly based on the results of fragment docking using SEED, supplemented by an analysis using the fpocket algorithm. Fragment clusters within each cavity gave a more informed idea about polarity, approachability and volume of each cavity. As expected, these results showed that a cavity selection purely based on the visual inspection of a protein might not

yield cavities that are targetable by molecules in a docking screen. It was observed for several of the initially proposed cavities that fragments did either not enter the cavity at all or that there was no space to extend the binding fragments into bigger molecules. On the other hand, visual inspection of the protein complex did not reveal all possible molecule binding areas. Two additional cavities were discovered from a SEED fragment docking to the entire protein complex. Additional to the SEED fragment docking calculations, the protein complex was analysed using the fpocket algorithm to find possible molecule binding sites. The results from this analysis confirmed the selection of promising cavities based on the SEED results, except for *region 2* which is not druggable according to the fpocket analysis.

From this analysis using SEED and fpocket, four cavities were chosen as possible ligand binding sites: *long-cav*, *region 2*, *M1* and *M2*. Considering the goal of this binding site prediction and subsequent docking screens, which is not only to find molecules binding to the RGS7-$G_{\beta 5}$ complex but also to modulate its conformation, these cavities also had to be analysed concerning the probability that a molecule could induce such a conformational change when binding to the inspected cavity. This evaluation was based on the cavity location and knowledge about structural stability of certain protein folds. This way cavity *long-cav* was excluded based on its location within the structurally stable $G_{\beta 5}$ protein, while the other cavities are located in structurally more flexible regions and/or interface areas of different parts of the protein complex. This evaluation based on prior knowledge of the protein complex was therefore a valuable addition to the SEED and fpocket analysis, to find the most likely molecule binding sites resulting in the desired *in vitro* outcome.

Cavity evaluation based on docking calculations. A further analysis of the three selected cavities (*region 2*, *M1* and *M2*) was based on docking screens against these cavities and the resulting molecule poses. From the docking calculations of the ZINC15 *lead-like* library as well as the SDDL library, it became obvious that cavity *M1* might not be suitable as a ligand binding site. Molecules docked into various parts of this cavity could often not entertain favourable interactions with the protein. Due to the voluminous and accessible nature of the cavity, it is possible that solvent molecules could enter the binding pocket and compensate for missing interactions of stranded molecule donors with the protein. However, this is difficult to predict and might still result in weaker protein-molecule interactions. Therefore, cavity *M1* might not be very well suited for docking screens and selection of molecules to be tested in *in vitro* assays.

Molecule poses resulting from the docking calculations to cavities *region 2* and *M2* are much more likely based on favourable polar interactions, molecule pose consistency and general molecule placement within the cavities. For both cavities certain advantages and disadvantages should be considered. Cavity *region 2* is very polar with a comparatively high number of positively charged

residues pointing into the cavity. This could result in a strong binding affinity of polar molecules to this region, especially if the molecules contain negatively charged functional groups. However, the RGS7 protein complex is expressed intracellularly and might not be reached by such polar or charged molecules, since they might be unable to pass the cell membrane. Another disadvantage of cavity *region 2* is its shallow nature and exposure to the solvent. Especially the latter can lead to a weakening of protein-molecule interactions, although molecules binding to such areas are not completely unlikely. Even though these features of cavity *region 2* might be problematic for docking screens, its location within a quite flexible part of the protein complex and at a possible interaction interface for other proteins still make it an interesting target. A first screen of the ZINC15 *lead-like* library against this region conducted by Jonas Kammertöns did unfortunately not yield the desired results. It should be noted that the fpocket algorithm detected this cavity only partially which implies a lower druggability of this protein region and is consistent with the experimental results. In spite of this, *region 2* is still potentially interesting due to the location and possible potential of this cavity and it might be worth testing some more, structurally different molecules.
Cavity *M2* probably has the best prerequisites as a small molecule binding site compared to the other two closer evaluated cavities. It is not very voluminous but leaves enough space for small molecule binding, it is accessible from the solvent but not solvent exposed and it contains polar residues which could form polar interactions with a small molecule. Furthermore, it is located at the interface of different parts of the protein complex and might stabilise or allosterically modulate the complex conformation. These features make cavity *M2* an interesting target for docking screens. However, certain parts of the cavity might also be problematic for molecule binding. Two residues within the cavity, D259 and R327, are probably forming an ionic interaction with each other, judging from their distance. For the molecule poses resulting from the docking calculations to this cavity it was observed frequently that the molecule was placed in between these two residues, thereby possibly disturbing the ionic interaction (see Figure 7.4). Although molecule binding in that position is not completely unlikely, especially if the molecule forms polar interactions with both residues, molecule poses disturbing such an interaction are more uncertain. In spite of that, testing molecules adopting such poses might still be worth a try and the disturbance of this inter-residue interaction might even result in unexpected *in vitro* observations.

Cavity selection for docking screens. The ultimate goal of the docking screens is to select molecules to be tested in *in vitro* assays. The overall results of the docking calculations and the considerations discussed above can be used as a basis to decide which cavity to select molecules for. As discussed, cavity *M1* is probably not suitable as a target for small molecules and, hence, none will be selected for this cavity. Molecules tested from the previous ZINC15 *lead-like* library

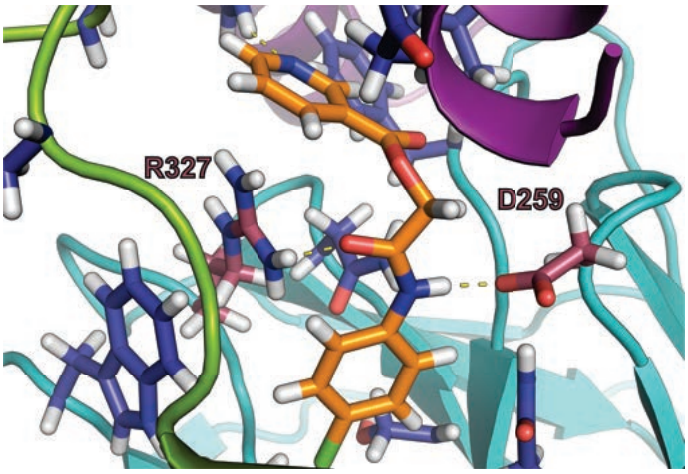


Figure 7.4: Docking pose of an exemplary molecule from the docking calculation of the SDDL library to cavity *M2*. Residues D259 and R327 (highlighted as sticks in berry) are close enough to each other to form ionic interactions. The molecule (orange) is inserted between these two residues, forming polar interactions to both with its amide group. Similar molecule poses were observed frequently from the docking calculation to this cavity. DEP (violet), DHEX-GGL linker (green) and $G_{\beta 5}$ (cyan). Residues from *M2* and in 4 Å distance to the molecule are shown as sticks in violet-blue.

screen against *region 2* did not show any responses in the *in vitro* assays and there are certain disadvantages to this cavity, which might make it seem less interesting as a docking target. However, the SDDL library is an in-house library that our collaborators can easily access, which makes the selection of molecules from the docking calculations of this library against cavity *region 2* to test them *in vitro* conceivable. Molecules selected from docking calculations against cavity *M2* have not been tested so far and, therefore, some molecules should be selected from the SDDL docking screen to be tested in *in vitro* assays.

7.5 Conclusions

The overall goal of this project is to find small molecules binding to the RGS7 complex and modulating its conformation and possibly function. The first step was to predict possible binding sites for small molecules. This could be accomplished by docking of molecule fragments using SEED as well as an additional analysis using the fpocket algorithm and prior knowledge on the overall protein complex. Finally, three possible cavities could be identified that were targeted in docking calculations using two different molecule libraries, the ZINC15 *lead-like* library and the SDDL library. Based on the results of these docking calculations, two cavities were identified as possible small molecule binding sites, cavities *region 2* and *M2*. Molecules from the docking screens against

these cavities now need to be chosen to be subjected to *in vitro* assays to confirm whether molecules might bind to these predicted binding sites.

Although the overall success of this project could not be validated *in vitro* yet, the computational part of the project was quite successful. Results of binding site prediction using the two different methods of fragment docking calculations and the fpocket algorithm complemented each other in most cases making the predicted molecule binding sites more believable. Furthermore, the docking screens themselves proved quite useful as additional evaluation tools. However, further conclusions how successful these approaches really are can only be made after a validation *in vitro* has been done. Since the docking results from the SDDL library are already partially evaluated and the molecule library is an in-house library of the Scripps Research Institute and should therefore be relatively quickly accessible to our collaboration partners, this *in vitro* validation could probably be pursued within a few weeks or months. Anyhow, this was not possible within the scope and time frame of this thesis and will be concluded later on.

7.6 Perspectives

Due to the missing *in vitro* validation of the predictions made in this study, there are still some untied ends which need to be fixed. For that, possible small molecule binders need to be selected from the docking calculations of the SDDL library to the proposed cavities and communicated to our collaborators at the Martemyanov lab to be tested in *in vitro* assays. However, there are also several additional possible options to continue with this project.

As described above, cavity *region 2* is quite shallow and positively charged but might also be an interaction site for other proteins. It might therefore be interesting to design small peptides that could bind to this region. This task can quickly get quite demanding though. The more amino acids the peptide contains, the more possible options can be created by combining the natural amino acids. Additionally a computational selection of potential peptide binders gets more difficult the longer the peptide is, since the flexibility of the peptide increases which complicates prediction of molecule binding poses. However, based on the volume of *region 2* it might be enough to limit a peptide library to di- and tripeptides. Additionally, the library could be further limited by setting the boundary conditions of at least one negatively charged amino acid in the peptide and no positvely charged amino acids to account for the quite positively charged cavity. It would be interesting to dock such a peptide library to cavity *region 2* and select peptides to be tested *in vitro* from these docking calculations. If one of these peptides could be confirmed as a binder, the project could be continued and the analogous peptides based on artificial amino acids could be tested as well, to obtain non-natural molecules binding to the RGS7 complex.

Another approach could be the discovery of covalently binding molecules. Covalent interactions between small molecules and proteins can easily be achieved by formation of disulfide bonds between the molecule and cysteine side chains in the protein. This is especially interesting since this kind of molecule binders has already been discovered for other members of the RGS protein family.[140,141] Although there are no cysteines located directly in the proposed binding sites, some can at least be found in proximity to them: C206 in chain D is in proximity to *region 2* and C200 and C207 in chain A are in proximity to *M2*.

All predictions in this project were made based on the crystal structure of the RGS7 complex. However, it is possible that the complex could take a different conformation, opening up potential small molecule binding sites that are not visible in the crystal structure. This could possibly be explored by applying MD simulations to the protein complex and monitoring interesting structural changes.

8 | General methods

8.1 Computational methods

8.1.1 Structure preparation for docking

To prepare crystal structures for docking, all additional proteins (such as e.g. effector proteins) and molecules as well as additional chains were removed and the structure was inspected closely to fix issues if necessary. Missing side chains of amino acids were added using *Dunbrack* rotamers in UCSF Chimera[233,238] and one conformation of duplicate side chains was deleted if applicable. Chain breaks were treated as non-existent i.e. left uncharged. Histidine protonation, i.e. protonation in δ- or ε-position or in both positions, was assigned according to the local environment, while other protonation was done according to pk_a values. Following the protonation, all hydrogen atoms as well as added side chains were minimised using CHARMM and the CHARMm22 force field,[209] unless otherwise stated.

8.1.2 Re-ranking of docking results

Re-ranking was done by comparing ranking lists from two different docking calculations of the same molecule library to two different receptor structures to each other. Depending on the chosen re-ranking mode either molecules with high ranks in both lists are favoured in the re-ranking list (dual re-ranking) or molecules with higher ranks in one list of a 'target structure' compared to the rank in a list of the 'anti-target structure' (selective re-ranking).

For the dual re-ranking, a previously described approach was used.[239] First, a relative rank (R_{rel}) is calculated for each molecule and both ranking lists by basically dividing the rank r of each molecule by the total number of molecules m in the ranking list (see eq. 8.1).

$$R_{rel,i} = \frac{r_i - 1}{m_i - 1} \tag{8.1}$$

where i is the indicator of each docking. The new re-ranking score D is then calculated using these relative ranks (see eq. 8.2).

$$D = \frac{(R_{rel,1} - R_{rel,2})^2 + R_{rel,1} + R_{rel,2}}{2} \tag{8.2}$$

More details on the calculations can be found in the original publication.[239]

For the selective re-ranking, the new re-ranking score S was obtained by simply calculating the ratio of the ranks of a molecule in both ranking lists (see eq. 8.3).

$$S = \frac{r_2}{r_1} \tag{8.3}$$

where r_1 is the rank of the molecule from the docking calculation to the 'target structure' and r_2 is the rank of the molecule from the docking calculation to the 'anti-target structure'.

8.1.3 SEED docking

SEED is a tool that can be used to dock small- to medium-sized fragments to a protein or specified parts of a protein.[178] For that, the protein structure was used as prepared for DOCK (see also section 8.1.1) including all hydrogens atoms. The region of the protein that should be explored was defined in the input and the default values were used for all other parameters. A fragment library of 160 diverse fragments generated with DAIM[240] was internally available already and used here. Depending on the aim of this analysis, all 160 fragments or a subset of the library were then docked to the specified protein part using SEED (v.3.3.4).[178] The resulting fragment poses were evaluated visually as described in more detail in the chapters for the specific targets.

8.1.4 Selection of ligands for the optimisation of β_2AR docking setups

To evaluate the performance of the docking calculations targeting the β_2AR, i.e. the enrichment of known ligands of the β_2AR in the top ranks of the ranking list, a test set was compiled. Literature known ligands were selected from different resources such as ChEMBL[192] or PubChem.[241] Additionally, functional data of the ligands was compiled to classify them as 'inverse agonist', 'agonist' or 'antagonist'. During the selection process it was made sure that the affinity and functional activity of each molecule was found in at least two sources to confirm their behaviour and account for unspecific assay data. In total, a set with 125 enantiomers of 70 different ligands was compiled. In this set, 8 enantiomers/ 5 ligands were defined as 'inverse agonists', 62/ 37 as 'agonists' and 54/ 28 as 'antagonists'.
To test the enrichment of these ligands in the docking calculations to the β_2AR, 50 decoy molecules were generated for each molecule using DUD-E.[155] In total, 6129 uniq decoys were generated for the ligand set and docked with the ligands to quantify the enrichment of the ligands in the docking calculations to the β_2AR structures using ROC plots.

8.2 Assays

Assays were conducted in two different laboratories at Philipps-University Marburg and at the University of Nottingham in different cell lines and tissue cultures. This section is therefore split into two parts.

8.2.1 Assays performed in Marburg

Cell culture. HEK293T cells expressing the β_2AR endogenously were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 2 mM L-glutamine, 10% fetal bovine serum, 100 U/ml Penicillin and 100 µg/mL Streptomycin at 37°C in 5% CO_2:95% air atmosphere.

Cisbio assay. For an initial characterisation, the HTRF® cAMP G_s dynamic assay (Cisbio Bioassays, France) was used. To obtain the optimal assay window and assay results the standard protocol provided by *Cisbio* was optimised and adapted to the requirements for both agonist and antagonist mode.

The assay was performed with non-adherent HEK293T cells. For that, cells were suspended in modified Tyrode buffer (10 mM HEPES, 1 mM $CaCl_2$, 0.5 mM $MgCl_2$, 4.2 mM KCl, 146 mM NaCl, 5.5 mM Glucose; pH = 7.4) supplemented with 0.5 mM IBMX at a concentration of 1400 cells/µL and 5 µL cell suspension transferred to the wells of a 384-well plate.

Agonist mode. To determine agonism, 5 µL modified Tyrode buffer or compound dilution in modified Tyrode buffer (at two times final concentration) was added to each well containing cell suspension. Cells were then incubated at room temperature for 10 min, before adding both detection reagents as recommended in the standard protocol (5 µL one time dilution in the provided Lysis buffer of each per well). Plates were then incubated 1 h at room temperature in the dark and finally read on a *Tecan Spark 20M* plate reader using the settings recommended by *Cisbio*. Each measurement was conducted in technical duplicates and measurements to account for basal activity and background fluorescence of the detection reagents included for each experiment. As a reference an Isoprenaline dose-response curve was included in each experiment.

Antagonist mode. To determine antagonism, 3 µL modified Tyrode buffer or compound dilution in modified Tyrode buffer (at 3.3 times final concentration) were added to each well containing cell suspension. After incubating the plate for 10 min at room temperature, 2 µL modified Tyrode buffer or 400 nM Isoprenaline in modified Tyrode buffer were added to each well and the plate was incubated for another 10 min at room temperature. The detection reagents were then added as recommended in the standard protocol (5 µL one time dilution in the provided Lysis buffer of each per well) and the plate incubated for 1 h at room temperature in the dark. The plate was then read at a *Tecan Spark 20M* plate reader using the settings recommended by *Cisbio*. Each measurment was conducted in technical duplicates and measurements to account for basal activity and background fluorescence of the detection reagents included for each experiment. A Propranolol dose-response curve was measured in each experiment as a reference.

Data analysis. The resulting data was analysed as described in the *Cisbio* assay protocol by calculating the HTRF ratio according to equation 8.4 and plotted using Graphpad Prism 7.[213] Since measurements were not conducted with a statistically relevant number of replica, no IC_{50}values were calculated.

$$HTRFratio = \frac{Signal_{665\,nm}}{Signal_{620\,nm}} \cdot 10^4 \tag{8.4}$$

8.2.2 Assays performed in Nottingham

Cell culture. CHO cells expressing stably either β_2AR or β_1AR as well as the CRE-SPAP reporter gene were grown in Dulbecco's modified Eagle's medium nutrient mix F12 (DMEM/F12) containing 10% fetal calf serum and 2 mM L-glutamine at 37°C in humidified 5% CO_2:95% air atmosphere.

^{3}H-CGP 12177 whole cell binding assay. After plating the cells to white 96-well plates and growing them to confluence over night, the medium was removed from each well and replaced by 100 µL serum-free media (sfm, DMEM/F12 containing 2 mM L-glutamine) or compound (at twice the final concentration in sfm). Immediately after this, 100 µL ^{3}H-CGP 12177 in sfm (1:2 dilution in wells) was added and the cells incubated for 2 h at 37°C in humidified 5% CO_2:95% air atmosphere. After this incubation step everything was removed from the wells and the cells were washed twice with 200 µL 4°C phosphate-buffered saline. To each well 100 µL Microscint 20 was added and the plates incubated at room temperature in the dark for several hours before being counted on a TopCount.[201] To define non-specific binding 10 µM Propranolol was used in all plates and the final concentration of ^{3}H-CGP 12177 was 0.5-0.9 nM. For each compound 7-point concentration response curves were measured as triplicates in each experiment and a sigmoidal curve was fitted to the data using Graphpad Prism 7.[213] To determine IC_{50} values, the following equation was used:

$$\%\text{ specific binding} = 100 - \frac{100}{1 + 10^{(log(\mathrm{IC_{50}}) - log([\mathrm{A}])}} \tag{8.5}$$

where [A] is the concentration of the competing ligand and IC_{50} is the concentration at which half of the specific binding of ^{3}H-CGP 12177 has been inhibited.
To calculate K_D values the Cheng-Prusoff equation was used with the IC_{50} values and the known concentration of radioligand, [^{3}H-CGP 12177]:

$$\mathrm{K}_D = \frac{\mathrm{IC_{50}}}{1 + \frac{[^3H-CGP\,12177]}{\mathrm{K}_D(^3H-CGP\,12177)}} \tag{8.6}$$

The K_D of ^{3}H-CGP 12177 in these cells was 0.42 nM for the β_1AR and 0.17 nM for the β_2AR.[201]

CRE-SPAP production assay. After plating the cells to clear 96-well plates and growing them to confluence over night, media was removed from all wells and replaced with 100 µL sfm for 24 h to serum starve the cells. The serum was then again removed and replaced by 100 µL fresh sfm or, where used, antagonists diluted in sfm. To each well 10 µL compound in sfm was then added and the cells subsequently incubated for 5 h at 37°C and humidified 5% CO_2:95% air atmosphere. After this incubation, everything was removed from the cells, replaced by 40 µL sfm and the cells incubated for 1 h at 37°C and humidified 5% CO_2:95% air atmosphere before being placed in a 65°C oven for 30 min. After the plates had cooled to room temperature, 100 µL of 5 mM *para*-Nitrophenylphosphate in DEA buffer (1 M diethanolamine, 0.28 M NaCl, 0.5 mM $MgCl_2$, pH 9.85) was added to each well. The plates were then read on an MRX plate reader through a 405 nm filter once the yellow colour had developed.[202] As a positive control 10 µM isoprenaline was used in all plates. For each compound 7-point concentration response curves were measured as triplicates in each experiment and a sigmoidal curve was fitted to the data using Graphpad Prism 7[213] (eq. 8.7).

$$Response = \frac{E_{max}}{1 + 10^{log(EC_{50}) - log([A])}} \tag{8.7}$$

where E_{max} is the maximum response, [A] is the agonist concentration and EC_{50} is the concentration of agonist that produces 50% of the maximal response.

As a control experiment to prove that the agonist responses were occurring via the transfected βAR, the affinity (log K_D value) of CGP 20712A or ICI 118551 was calculated from the rightward shift of the agonist concentration responses in the presence of a fixed concentration of antagonist using the Gaddum equation:

$$DR = 1 + \frac{[B]}{K_D} \tag{8.8}$$

where DR (dose ratio) is the ratio of the agonist concentration required to stimulate an identical response in the presence and absence of the fixed concentration of antagonist [B].

References

[1] Fredriksson, R.; Lagerström, M. C.; Lundin, L.-G.; Schioth, H. B. *Mol Pharmacol* **2003**, *63*, 1256–1272.

[2] Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E. *Nat Rev Drug Discov* **2017**, *16*, 829–842.

[3] Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C.; Motoshima, H.; Fox, B.; Le Trong, I.; Teller, D.; Okada, T.; Stenkamp, R.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739–745.

[4] Lagerstrom, M. C.; Schioth, H. B. *Nat Rev Drug Discov* **2008**, *7*, 339–357.

[5] Shonberg, J.; Kling, R. C.; Gmeiner, P.; Löber, S. *Bioorg Med Chem* **2015**, *23*, 3880–3906.

[6] Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. *Science* **2007**, *318*, 1258–1265.

[7] 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. *Nature* **2007**, *450*, 383–U4.

[8] Rasmussen, S. G. F. et al. *Nature* **2011**, *477*, 549–U311.

[9] Rosenbaum, D. M.; Rasmussen, S. G.; Kobilka, B. K. *Nature* **2009**, *459*, 356–363.

[10] Magnani, F.; Serrano-Vega, M. J.; Shibata, Y.; Abdul-Hussein, S.; Lebon, G.; Miller-Gallacher, J.; Singhal, A.; Strege, A.; Thomas, J. A.; Tate, C. G. *Nat Protoc* **2016**, *11*, 1554–1571.

[11] Weis, W. I.; Kobilka, B. K. *Annu Rev Biochem* **2018**, *87*, 897–919.

[12] Pándy-Szekeres, G.; Munk, C.; Tsonkov, T. M.; Mordalski, S.; Harpsøe, K.; Hauser, A. S.; Bojarski, A. J.; Gloriam, D. E. *Nucleic Acids Res* **2018**, *46*, 440–446.

[13] Vénien-Bryan, C.; Li, Z.; Vuillard, L.; Boutin, J. A. *Acta Crystallogr F Struct Biol Commun* **2017**, *73*, 174–183.

[14] García-Nafría, J.; Tate, C. G. *Mol Cell Endocrinol* **2019**, *488*, 1–13.

[15] García-Nafría, J.; Tate, C. G. *Annu Rev Pharmacol Toxicol* **2020**, *60*, 51–71.

[16] Shoemaker, S. C.; Ando, N. *Biochemistry* **2018**, *57*, 277–285.

[17] Kolakowski, L. J. *Recept Channels* **1994**, *2*, 1–7.

[18] Stevens, R. C.; Cherezov, V.; Katritch, V.; Abagyan, R.; Kuhn, P.; Rosen, H.; Wüthrich, K. *Nat Rev Drug Discov* **2013**, *12*, 25–34.

[19] Ballesteros, J. A.; Weinstein, H. In *Receptor Molecular Biology*; Sealfon, S. C., Ed.; Methods in Neurosciences; Academic Press: San Diego, CA, 1995; Vol. 25; Chapter 19, pp 366–428.

[20] Zhou, Q. et al. *Elife* **2019**, *8:e50279*, 1–31.

[21] Katritch, V.; Fenalti, G.; Abola, E. E.; Roth, B. L.; Cherezov, V.; Stevens, R. C. *Trends Biochem Sci* **2014**, *39*, 233–244.

[22] Gurevich, V. V.; Gurevich, E. V. *Int J Mol Sci* **2017**, *18*, 2519.

[23] Deupi, X.; Kobilka, B. *Adv Protein Chem* **2007**, *74*, 137–166.

[24] Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. *Chem Rev* **2017**, *117*, 139–155.

[25] Manglik, A.; Kim, T. H.; Masureel, M.; Altenbach, C.; Yang, Z.; Hilger, D.; Lerch, M. T.; Kobilka, T. S.; Thian, F. S.; Hubbell, W. L.; Prosser, R. S.; Kobilka, B. K. *Cell* **2015**, *161*, 1101–1111.

[26] Hilger, D.; Masureel, M.; Kobilka, B. K. *Nat Struct Mol Biol* **2018**, *25*, 4–12.

[27] Downes, G. B.; Gautam, N. *Genomics* **1999**, *62*, 544–552.

[28] De Lean, A.; Stadel, J. M.; Lefkowitz, R. J. *J Biol Chem* **1980**, *255*, 7108–7117.

[29] Yao, X. Y.; Ruiz, G. V.; Whorton, M. R.; Rasmussen, S. G.; DeVree, B. T.; Deupi, X.; Sunahara, R. K.; Kobilka, B. *Proc Natl Acad Sci U S A* **2009**, *106*, 9501–9506.

[30] Liu, X.; Xu, X.; Hilger, D.; Aschauer, P.; Tiemann, J. K.; Du, Y.; Liu, H.; Hirata, K.; Sun, X.; Guixà-González, R.; Mathiesen, J. M.; Hildebrand, P. W.; Kobilka, B. K. *Cell* **2019**, *177*, 1243–1251.

[31] Thomsen, A. R. et al. *Cell* **2016**, *166*, 907–919.

[32] Nguyen, A. H. et al. *Nat Struct Mol Biol* **2019**, *26*, 1123–1131.

[33] Luttrell, L. M.; Roudabush, F. L.; Choy, E. W.; Miller, W. E.; Field, M. E.; Pierce, K. L.; Lefkowitz, R. J. *Proc Natl Acad Sci U S A* **2001**, *98*, 2449–2454.

[34] Shenoy, S. K.; Drake, M. T.; Nelson, C. D.; Houtz, D. A.; Xiao, K.; Madabushi, S.; Reiter, E.; Premont, R. T.; Lichtarge, O.; Lefkowitz, R. J. *J Biol Chem* **2006**, *281*, 1261–1273.

[35] Grundmann, M. et al. *Nat Commun* **2018**, *9*, 341.

[36] Luttrell, L. M. et al. *Sci Signal* **2018**, *11*.

[37] Gutkind, J. S.; Kostenis, E. *Nat Rev Mol Cell Biol* **2018**, *19*, 615–616.

[38] Wisler, J. W.; Xiao, K.; Thomsen, A. R.; Lefkowitz, R. J. *Curr Opin Cell Biol* **2014**, *27*, 18–24.

[39] Wisler, J. W.; Rockman, H. A.; Lefkowitz, R. J. *Circulation* **2018**, *137*, 2315–2317.

[40] Gurevich, V. V.; Gurevich, E. V. *Pharmacol Ther* **2020**, *211*, 107540.

[41] Rajagopal, S.; Ahn, S.; Rominger, D. H.; Gowen-MacDonald, W.; Lam, C. M.; DeWire, S. M.; Violin, J. D.; Lefkowitz, R. J. *Mol Pharmacol* **2011**, *80*, 367–377.

[42] Kenakin, T.; Watson, C.; Muniz-Medina, V.; Christopoulos, A.; Novick, S. *ACS Chem Neurosci* **2012**, *3*, 193–203.

[43] Kenakin, T.; Christopoulos, A. *Nat Rev Drug Discov* **2013**, *12*, 205–216.

[44] Langley, J. N. *J Physiol* **1905**, *33*, 374–413.

[45] Dale, H. H. *J Physiol* **1906**, *34*, 163–206.

[46] Ahlquist, R. P. *Am J Physiol* **1948**, *153*, 586–600.

[47] Dale, H. *Transactions of the Faraday Society* **1943**, *39*, 319b–322b.

[48] Ahlquist, R. P. *Perspect Biol Med* **1973**, *17*, 119–122.

[49] Lands, A. M.; Arnold, A.; McAuliff, J. P.; Luduena, F. P.; Brown, T. G. *Nature* **1967**, *214*, 597–598.

[50] Starke, K. *Rev Physiol Biochem Pharmacol* **1981**, *88*, 199–236.

[51] Kobilka, B. K.; Dixont, R. A. F.; Frielle, T.; Dohlman, H. G.; Bolanowski, M. A.; Sigalt, I. S.; Yang-fengt, T. L.; Francket, U. T. A.; Caron, M. G.; Lefkowitz, R. J. *Proc Natl Acad Sci U S A* **1987**, *84*, 46–50.

[52] Lefkowitz, R. J. *Angew Chem Int Ed* **2013**, *52*, 6366–6378.

[53] Lohse, M. J. *J Mol Med* **2015**, *93*, 955–962.

[54] Kobilka, B. *Angew Chem Int Ed* **2013**, *52*, 6380–6388.

[55] Xiao, R.-P.; Ji, X.; Lakatta, E. G. *Mol Pharmacol* **1995**, *47*, 322–329.

[56] Molinoff, P. B. *Drugs* **1984**, *28*, 1–15.

[57] Barnes, P. J. *Life Sci* **1993**, *52*, 2101–2109.

[58] Johnson, M. *J Allergy Clin Immunol* **2006**, *117*, 18–24.

[59] Barnes, P. J. *J Biol Chem* **2011**, *286*, 32899–32905.

[60] Smiley, R. M.; Finster, M. *J Matern Fetal Med* **1996**, *5*, 106–114.

[61] Arrowsmith, S.; Kendrick, A.; Wray, S. *Obstet Gynaecol Reprod Med* **2010**, *20*, 241–247.

[62] do Vale, G. T.; Ceron, C. S.; Gonzaga, N. A.; Simplicio, J. A.; Padovan, J. C. *Curr Hypertens Rev* **2019**, *15*, 22–31.

[63] Wood, A. J. J. *Am Heart J* **1984**, *108*, 1070–1077.

[64] Tsai, M.-C.; Liu, H.-C.; Yeung, C.-Y. *Medicine* **2019**, *98*, e14078.

[65] Thai, T.; Wang, C.-Y.; Chang, C.-Y.; Brown, J. D. *J Clin Med* **2019**, *8*, 268.

[66] Taylor, M. R. *Pharmacogenomics J* **2007**, *7*, 29–37.

[67] Tandale, A.; Joshi, M.; Sengupta, D. *Sci Rep* **2016**, *6*, 1–11.

[68] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res* **2000**, *28*, 235–242.

[69] Wacker, D.; Fenalti, G.; Brown, M. A.; Katritch, V.; Abagyan, R.; Cherezov, V.; Stevens, R. C. *J Am Chem Soc* **2010**, *132*, 11443–11445.

[70] Liu, X.; Ahn, S.; Kahsai, A. W.; Meng, K. C.; Latorraca, N. R.; Pani, B.; Venkatakrishnan, A. J.; Masoudi, A.; Weis, W. I.; Dror, R. O.; Chen, X.; Lefkowitz, R. J.; Kobilka, B. K. *Nature* **2017**, *548*, 480–484.

[71] 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. *Science* **2019**, *364*, 1283–1287.

[72] Liu, X. et al. *Nat Chem Biol* **2020**, *16*, 749–755.

[73] Ishchenko, A. et al. *IUCrJ* **2019**, *6*, 1106–1119.

[74] Dixon, R. A.; Sigal, I. S.; Rands, E.; Register, R. B.; Candelore, M. R.; Blake, A. D.; Strader, C. D. *Nature* **1987**, *326*, 73–77.

[75] Strader, C. D.; Sigal, I. S.; Register, R. B.; Candelore, M. R.; Rands, E.; Dixon, R. A. F. *Proc Natl Acad Sci U S A* **1987**, *84*, 4384–4388.

[76] Strader, C. D.; Sigal, I. S.; Rios Candelore, M.; Rands, E.; Hill, W. S.; Dixon, R. A. *J Biol Chem* **1988**, *263*, 10267–10271.

[77] Wieland, K.; Zuurmond, H. M.; Krasel, C.; Ijzerman, A. P.; Lohse, M. J. *Proc Natl Acad Sci U S A* **1996**, *93*, 9276–9281.

[78] Ring, A. M.; Manglik, A.; Kruse, A. C.; Enos, M. D.; Weis, W. I.; Garcia, K. C.; Kobilka, B. K. *Nature* **2013**, *502*, 575–579.

[79] Strader, C. D.; Candelore, M. R.; Hill, W. S.; Sigal, I. S.; Dixon, R. A. *J Biol Chem* **1989**, *264*, 13572–13578.

[80] Liapakis, G.; Ballesteros, J. A.; Papachristou, S.; Chan, W. C.; Chen, X.; Javitch, J. A. *J Biol Chem* **2000**, *275*, 37779–37788.

[81] Scharf, M. M.; Bünemann, M.; Baker, J. G.; Kolb, P. *Mol Pharmacol* **2019**, *96*, 851–861.

[82] Masureel, M. et al. *Nat Chem Biol* **2018**, *14*, 1059–1066.

[83] Scharf, M. M.; Zimmermann, M.; Wilhelm, F.; Stroe, R.; Waldhoer, M.; Kolb, P. *ChemMedChem* **2020**, *15*, 882–890.

[84] Pardon, E.; Betti, C.; Laeremans, T.; Chevillard, F.; Guillemyn, K.; Kolb, P.; Ballet, S.; Steyaert, J. *Angew Chem Int Ed* **2018**, *57*, 5292–5295.

[85] Munro, S.; Thomas, K. L.; Abu-Shaar, M. *Nature* **1993**, *365*, 61–65.

[86] Pertwee, R. G. *British Journal of Pharmacology* **2006**, *147*.

[87] Mechoulam, R.; Hanuš, L. *Chem Phys Lipids* **2000**, *108*, 1–13.

[88] Lawrence, D. K.; Gill, E. W. *Mol Pharmacol* **1975**, *11*, 595–602.

[89] Devane, W. A.; Dysarz, F. A.; Johnson, M. R.; Melvin, L. S.; Howlett, A. C. *Mol Pharmacol* **1988**, *34*, 605–613.

[90] Matsuda, L. A.; Lolait, S. J.; Brownstein, M. J.; Young, A. C.; Bonner, T. I. *Nature* **1990**, *346*, 561–564.

[91] Kohno, M.; Hasegawa, H.; Inoue, A.; Muraoka, M.; Miyazaki, T.; Oka, K.; Yasukawa, M. *Biochem Biophys Res Commun* **2006**, *347*, 827–832.

[92] Ryberg, E.; Larsson, N.; Sjögren, S.; Hjorth, S.; Hermansson, N. O.; Leonova, J.; Elebring, T.; Nilsson, K.; Drmota, T.; Greasley, P. J. *B J Pharmacol* **2007**, *152*, 1092–1101.

[93] Pertwee, R. G.; Howlett, A. C.; Abood, M. E.; Alexander, S. P. H.; Marzo, V. D.; Elphick, M. R.; Greasley, P. J.; Hansen, H. S.; Kunos, G. *Pharmacol Rev* **2010**, *62*, 588–631.

[94] Console-Bram, L.; Marcu, J.; Abood, M. E. *Prog Neuropsychopharmacol Biol Psychiatry* **2012**, *38*, 4–15.

[95] Howlett, A. C.; Qualy, J. M.; Khachatrian, L. L. *Mol Pharmacol* **1986**, *29*, 307–313.

[96] Bouaboula, M.; Poinot-Chazel, C.; Marchand, J.; Canat, X.; Bourrié, B.; Rinaldi-Carmona, M.; Calandra, B.; Le Fur, G.; Casellas, P. *Eur J Biochem* **1996**, *237*, 704–711.

[97] Sugiura, T.; Kondo, S.; Kishimoto, S.; Miyashita, T.; Nakane, S.; Kodaka, T.; Suhara, Y.; Takayama, H.; Waku, K. *J Biol Chem* **2000**, *275*, 605–612.

[98] Shoemaker, J. L.; Ruckle, M. B.; Mayeux, P. R.; Prather, P. L. *J Pharmacol Exp Ther* **2005**, *315*, 828–838.

[99] Soethoudt, M. et al. *Nat Commun* **2017**, *8*.

[100] Lynn, A. B.; Herkenham, M. *J Pharmacol Exp Ther* **1994**, *268*, 1612–1623.

[101] Galiègue, S.; Mary, S.; Marchand, J.; Dussossoy, D.; Carrière, D.; Carayon, P.; Bouaboula, M.; Shire, D.; LE Fur, G.; Casellas, P. *Eur J Biochem* **1995**, *232*, 54–61.

[102] Núñez, E.; Benito, C.; Pazos, M. R.; Barbachano, A.; Fajardo, O.; González, S.; Tolón, R. M.; Romero, J. *Synapse* **2004**, *53*, 208–213.

[103] Van Sickle, M. D.; Duncan, M.; Kingsley, P. J.; Mouihate, A.; Urbani, P.; Mackie, K.; Stella, N.; Makriyannis, A.; Piomelli, D.; Davison, J. S.; Marnett, L. J.; Di Marzo, V.; Pittman, Q. J.; Patel, K. D.; Sharkey, K. A. *Science* **2005**, *310*, 329–332.

[104] Cabral, G. A.; Raborn, E. S.; Griffin, L.; Dennis, J.; Marciano-Cabral, F. *B J Pharmacol* **2008**, *153*, 240–251.

[105] Cabral, G. A.; Griffin-Thomas, L. T. *Expert Rev Mol Med* **2009**, *11*, 1–25.

[106] Howlett, A. C.; Barth, F.; Bonner, T. I.; Cabral, G.; Casellas, P.; Devane, W. A.; Felder, C. C.; Herkenham, M.; Mackie, K.; Martin, B. R.; Mechoulam, R.; Pertwee, R. G. *Pharmacol Rev* **2002**, *54*, 161–202.

[107] Jhaveri, M. D.; Sagar, D. R.; Elmes, S. J.; Kendall, D. A.; Chapman, V. *Mol Neurobiol* **2007**, *36*, 26–35.

[108] Maccarrone, M.; Bab, I.; Bíró, T.; Cabral, G. A.; Dey, S. K.; Di Marzo, V.; Konje, J. C.; Kunos, G.; Mechoulam, R.; Pacher, P.; Sharkey, K. A.; Zimmer, A. *Trends Pharmacol Sci* **2015**, *36*, 277–296.

[109] Rossi, F.; Tortora, C.; Argenziano, M.; Di Paola, A.; Punzo, F. *Int J Mol Sci* **2020**, *21*, 1–16.

[110] Ashton, J.; Glass, M. *Curr Neuropharmacol* **2007**, *5*, 73–80.

[111] Dhopeshwarkar, A.; Mackie, K. *Mol Pharmacol* **2014**, *86*, 430–437.

[112] Li, X. et al. *Cell* **2019**, *176*, 459–467.e13.

[113] Hua, T. et al. *Cell* **2020**, *180*, 655 – 665.e18.

[114] Xing, C. et al. *Cell* **2020**, *180*, 645–654.e13.

[115] Devane, W. A.; Hanus, L.; Breuer, A.; Pertwee, R. G.; Stevenson, L. A.; Griffin, G.; Gibson, D.; Mandelbaum, A.; Etinger, A.; Mechoulam, R. *Science* **1992**, *258*, 1946–1949.

[116] Mechoulam, R.; Ben-Shabat, S.; Hanus, L.; Ligumsky, M.; Kaminski, N. E.; Schatz, A. R.; Gopher, A.; Almog, S.; Martin, B. R.; Compton, D. R.; Pertwee, R. G.; Griffin, G.; Bayewitch, M.; Barg, J.; Vogel, Z. *Biochem Pharmacol* **1995**, *50*, 83–90.

[117] Zhang, Y.; Xie, Z.; Wang, L.; Schreiter, B.; Lazo, J. S.; Gertsch, J.; Xie, X. Q. *Int Immunopharmacol* **2011**, *11*, 1303–1310.

[118] Lin, S. W.; Sakmar, T. P. *Biochemistry* **1996**, *35*, 11149–11159.

[119] Singh, R.; Hurst, D. P.; Barnett-Norris, J.; Lynch, D. L.; Reggio, P. H.; Guarnieri, F. *J Pept Res* **2002**, *60*, 357–370.

[120] McAllister, S. D.; Hurst, D. P.; Barnett-Norris, J.; Lynch, D.; Reggio, P. H.; Abood, M. E. *J Biol Chem* **2004**, *279*, 48024–48037.

[121] Schwartz, T. W.; Frimurer, T. M.; Holst, B.; Rosenkilde, M. M.; Elling, C. E. *Annu Rev Pharmacol Toxicol* **2006**, *46*, 481–519.

[122] Berman, D. M.; Wilkie, T. M.; Gilman, A. G. *Cell* **1996**, *86*, 445–452.

[123] Watson, N.; Linder, M. E.; Druey, K. M.; Kehrl, J. H.; Blumer, K. J. *Nature* **1996**, *383*, 172–175.

[124] Hunt, T. W.; Fields, T. A.; Casey, P. J.; Peralta, E. G. *Nature* **1996**, *383*, 175–177.

[125] Hepler, J. R.; Berman, D. M.; Gilman, A. G.; Kozasa, T. *Proc Natl Acad Sci U S A* **1997**, *94*, 428–432.

[126] Ross, E. M.; Wilkie, T. M. *Annu Rev Biochem* **2000**, *69*, 795–827.

[127] He, W.; Cowan, C. W.; Wensel, T. G. *Neuron* **1998**, *20*, 95–102.

[128] Arshavsky, V. Y.; Pugh, E. N. *Neuron* **1998**, *20*, 11–14.

[129] Lan, K.-L.; Zhong, H.; Nanamori, M.; Neubig, R. R. *J Biol Chem* **2000**, *275*, 33497–33503.

[130] Neubig, R. R.; Siderovski, D. P. *Nat Rev Drug Discov* **2002**, *1*, 187–197.

[131] Abramow-Newerly, M.; Roy, A. A.; Nunn, C.; Chidiac, P. *Cell Signal* **2006**, *18*, 579–591.

[132] Zheng, B.; De Vries, L.; Gist Farquhar, M. *Trends Biochem Sci* **1999**, *24*, 411–414.

[133] Patil, D. N.; Rangarajan, E. S.; Novick, S. J.; Pascal, B. D.; Kojetin, D. J.; Griffin, P. R.; Izard, T.; Martemyanov, K. A. *Elife* **2018**, *7*, e42150.

[134] Snow, B. E.; Krumins, A. M.; Brothers, G. M.; Lee, S.-F.; Wall, M. A.; Chung, S.; Mangion, J.; Arya, S.; Gilman, A. G.; Siderovski, D. P. *Proc Natl Acad Sci U S A* **1998**, *95*, 13307–13312.

[135] Anderson, G. R.; Posokhova, E.; Martemyanov, K. A. *Cell Biochem Biophys* **2009**, *54*, 33–46.

[136] Hu, G.; Wensel, T. G. *Proc Natl Acad Sci U S A* **2002**, *99*, 9755–9760.

[137] Martemyanov, K. A.; Yoo, P. J.; Skiba, N. P.; Arshavsky, V. Y. *J Biol Chem* **2005**, *280*, 5133–5136.

[138] Posner, B. A.; Gilman, A. G.; Harris, B. A. *J Biol Chem* **1999**, *274*, 31087–31093.

[139] Hooks, S. B.; Waldo, G. L.; Corbitt, J.; Bodor, E. T.; Krumins, A. M.; Harden, T. K. *J Biol Chem* **2003**, *278*, 10087–10093.

[140] O'Brien, J. B.; Wilkinson, J. C.; Roman, D. L. *J Biol Chem* **2019**, *294*, 18571–18585.

[141] Salaga, M.; Storr, M.; Martemyanov, K. A.; Fichna, J. *Bioessays* **2016**, *38*, 344–354.

[142] Ostrovskaya, O.; Xie, K.; Masuho, I.; Fajardo-Serrano, A.; Lujan, R.; Wickman, K.; Martemyanov, K. A. *Elife* **2014**, *3*, e02053.

[143] Zhou, H.; Chisari, M.; Raehal, K. M.; Kaltenbronn, K. M.; Bohn, L. M.; Mennerick, S. J.; Blumer, K. J. *Proc Natl Acad Sci U S A* **2012**, *109*, 19977–19982.

[144] Cheever, M. L.; Snyder, J. T.; Gershburg, S.; Siderovski, D. P.; Harden, T. K.; Sondek, J. *Nat Struct Mol Biol* **2008**, *15*, 155–162.

[145] Brooijmans, N.; Kuntz, I. D. *Annu Rev Biophys Biomol Struct* **2003**, *32*, 335–373.

[146] Torres, P. H.; Sodero, A. C.; Jofily, P.; Silva-Jr, F. P. *Int J Mol Sci* **2019**, *20*, 4574.

[147] Pinzi, L.; Rastelli, G. *Int J Mol Sci* **2019**, *20*, 4331.

[148] Fischer, E. *Ber Dtsch Chem Ges* **1894**, *27*, 2985–2993.

[149] Koshland, D. E. *Proc Natl Acad Sci U S A* **1958**, *44*, 98–104.

[150] Monod, J.; Wyman, J.; Changeux, J. P. *J Mol Biol* **1965**, *12*, 88–118.

[151] Changeux, J. P.; Edelstein, S. *F1000 Biol Rep* **2011**, *3*, 1–15.

[152] Salmaso, V.; Moro, S. *Front Pharmacol* **2018**, *9*, 1–16.

[153] Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. *Biophys Rev* **2014**, *6*, 75–87.

[154] Kuntz, I. D.; Meng, E. C.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J Mol Biol* **1982**, *161*, 269–288.

[155] Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. *J Med Chem* **2012**, *55*, 6582–6594.

[156] Mysinger, M. M.; Shoichet, B. K. *J Chem Inf Model* **2010**, *50*, 1561–1573.

[157] Ertl, P. *J Chem Inf Comput Sci* **2003**, *43*, 374–380.

[158] Gorse, A.-D. *Curr Top Med Chem* **2006**, *6*, 3–18.

[159] Van Hilten, N.; Chevillard, F.; Kolb, P. *J Chem Inf Model* **2019**, *59*, 644–651.

[160] Opassi, G.; Gesù, A.; Massarotti, A. *Drug Discov Today* **2018**, *23*, 565–574.

[161] Irwin, J. J.; Shoichet, B. K. *J Chem Inf Model* **2005**, *45*, 177–182.

[162] Chevillard, F.; Kolb, P. *J Chem Inf Model* **2015**, *55*, 1824–1835.

[163] Chevillard, F.; Rimmer, H.; Betti, C.; Pardon, E.; Ballet, S.; van Hilten, N.; Steyaert, J.; Diederich, W. E.; Kolb, P. *J Med Chem* **2018**, *61*, 1118–1129.

[164] Ertl, P.; Altmann, E.; Mckenna, J. M. *J Med Chem* **2020**, *63*, 8408–8418.

[165] Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. *J Chem Inf Model* **2012**, *52*, 1757–1768.

[166] Sterling, T.; Irwin, J. J. *J Chem Inf Model* **2015**, *55*, 2324–2337.

[167] Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. *J Comput Aided Mol Des* **2012**, *26*, 749–773.

[168] Allen, W. J.; Fochtman, B. C.; Balius, T. E.; Rizzo, R. C. *J Comput Chem* **2017**, *38*, 2641–2663.

[169] Shoichet, B. K.; Bodian, D. L.; Kuntz, I. D. *J Comput Chem* **1992**, *13*, 380–397.

[170] Shoichet, B. K.; Kuntz, I. D. *Protein Eng* **1993**, *6*, 723–732.

[171] Coleman, R. G.; Carchia, M.; Sterling, T.; Irwin, J. J.; Shoichet, B. K. *PLoS One* **2013**, *8*, e75992.

[172] Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J Comput Chem* **1992**, *13*, 505–524.

[173] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J Am Chem Soc* **1984**, *106*, 765–784.

[174] Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins* **1999**, *34*, 4–16.

[175] Lorber, D. M.; Shoichet, B. K. *Protein Sci* **1998**, *7*, 938–950.

[176] Lorber, D.; Shoichet, B. *Curr Top Med Chem* **2005**, *5*, 739–749.

[177] Scarsi, M.; Apostolakis, J.; Caflisch, A. *J Phys Chem A* **1997**, *101*, 8098–8106.

[178] Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. *Proteins* **1999**, *37*, 88–105.

[179] Majeux, N.; Scarsi, M.; Caflisch, A. *Proteins* **2001**, *42*, 256–268.

[180] McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.

[181] Torrens-Fontanals, M.; Stepniewski, T. M.; Aranda-García, D.; Morales-Pastor, A.; Medel-Lacruz, B.; Selent, J. *Int J Mol Sci* **2020**, *21*, 5933.

[182] Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. *J Med Chem* **2004**, *47*, 1879–1881.

[183] Durrant, J. D.; McCammon, J. A. *BMC Biol* **2011**, *9*, 71.

[184] Rechlin, C.; Scheer, F.; Terwesten, F.; Wulsdorf, T.; Pol, E.; Fridh, V.; Toth, P.; Diederich, W. E.; Heine, A.; Klebe, G. *ACS Chem Biol* **2017**, *12*, 1397–1415.

[185] Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. *Proc Natl Acad Sci U S A* **2011**, *108*, 18684–18689.

[186] Van Gunsteren, W. F. et al. *Angew Chem Int Ed* **2006**, *45*, 4064–4092.

[187] Schlick, T. In *Molecular Modelling and Simulation: An Interdisciplinary Guide*, 2nd ed.; Antman, S. S., Marsden, J. E., Sirovich, L., Eds.; Springer Science & Business Media, 2010; pp 425–462.

[188] Leach, A. R. *Molecular Modelling: Principles and Applications*; Pearson Education Limited, 1996; pp 313–370.

[189] Harvey, M. J.; Giupponi, G.; De Fabritiis, G. *J Chem Theory Comput* **2009**, *5*, 1632–1639.

[190] Kolb, P.; Rosenbaum, D. M.; Irwin, J. J.; Fung, J. J.; Kobilka, B. K.; Shoichet, B. K. *Proc Natl Acad Sci U S A* **2009**, *106*, 6843–6848.

[191] Rogers, D.; Hahn, M. *J Chem Inf Model* **2010**, *50*, 742–754.

[192] Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. *Nucleic Acids Res* **2012**, *40*, D1100–D1107.

[193] Sabio, M.; Jones, K.; Topiol, S. *Bioorg Med Chem Lett* **2008**, *18*, 5391–5395.

[194] Weiss, D. R.; Ahn, S.; Sassano, M. F.; Kleist, A.; Zhu, X.; Strachan, R.; Roth, B. L.; Lefkowitz, R. J.; Shoichet, B. K. *ACS Chem Biol* **2013**, *8*, 1018–1026.

[195] Schmidt, D.; Gunera, J.; Baker, J. G.; Kolb, P. *ACS Med Chem Lett* **2017**, *8*, 481–485.

[196] Chevillard, F.; Stotani, S.; Karawajczyk, A.; Hristeva, S.; Pardon, E.; Steyaert, J.; Tzalis, D.; Kolb, P. *Proc Natl Acad Sci U S A* **2019**, *116*, 11496–11501.

[197] Carlsson, J.; Coleman, R. G.; Setola, V.; Irwin, J. J.; Fan, H.; Schlessinger, A.; Sali, A.; Roth, B. L.; Shoichet, B. K. *Nat Chem Biol* **2011**, *7*, 769–778.

[198] Kruse, A. C.; Weiss, D. R.; Rossi, M.; Hu, J.; Hu, K.; Eitel, K.; Gmeiner, P.; Wess, J.; Kobilka, B. K.; Shoichet, B. K. *Mol Pharmacol* **2013**, *84*, 528–540.

[199] Kuhn, B.; Gilberg, E.; Taylor, R.; Cole, J.; Korb, O. *J Med Chem* **2019**, *62*, 10441–10455.

[200] Kolb, P.; Phan, K.; Gao, Z.-G.; Marko, A. C.; Šali, A.; Jacobson, K. A. *PLoS One* **2012**, *7*, e49910.

[201] Baker, J. G. *Br J Pharmacol* **2005**, *144*, 317–322.

[202] Baker, J. G.; Proudman, R. G. W.; Hill, S. J. *Mol Pharmacol* **2014**, *85*, 811–829.

[203] Baker, J. G.; Hall, I. P.; Hill, S. J. *Mol Pharmacol* **2004**, *65*, 986–998.

[204] Baker, J. G. *J Pharmacol Exp Ther* **2005**, *313*, 1163–1171.

[205] Zheng, Z.; Huang, X.-p. P.; Mangano, T. J.; Zou, R.; Chen, X.; Zaidi, S. A.; Roth, B. L.; Stevens, R. C.; Katritch, V. *J Med Chem* **2017**, *60*, 3070–3081.

[206] Wawer, M.; Bajorath, J. *J Chem Inf Model* **2010**, *50*, 1395–1409.

[207] Baker, J. G. *B J Pharmacol* **2010**, *160*, 1048–1061.

[208] Miljuš, T. The Molecular Basis for Biased Signalling by Human Cannabinoid CB2 Receptor. Ph.D. thesis, ETH Zürich, 2018.

[209] Momany, F. A.; Rone, R. *J Comput Chem* **1992**, *13*, 888–900.

[210] OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com, 2013.

[211] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2020.

[212] Beglov, D.; Roux, B. *J Phys Chem B* **1997**, *101*, 7821–7826.

[213] GraphPad Software, San Diego, California USA, www.graphpad.com,.

[214] Motulsky, H. J.; Mahan, L. C. *Mol Pharmacol* **1984**, *25*, 1–9.

[215] Fenalti, G.; Giguere, P. M.; Katritch, V.; Huang, X.-P.; Thompson, A. A.; Cherezov, V.; Roth, B. L.; Stevens, R. C. *Nature* **2014**, *506*, 191–196.

[216] Huang, W. et al. *Nature* **2015**, *524*, 315–321.

[217] Hua, T. et al. *Nature* **2017**, *547*, 468–471.

[218] Wang, S.; Wacker, D.; Levit, A.; Che, T.; Betz, R. M.; McCorvy, J. D.; Venkatakrishnan, A. J.; Huang, X. P.; Dror, R. O.; Shoichet, B. K.; Roth, B. L. *Science* **2017**, *358*, 381–386.

[219] Staus, D. P.; Hu, H.; Robertson, M. J.; Kleinhenz, A. L.; Wingler, L. M.; Capel, W. D.; Latorraca, N. R.; Lefkowitz, R. J.; Skiniotis, G. *Nature* **2020**, *579*, 297–302.

[220] Šali, A.; Blundell, T. L. **1993**, *234*, 779–815.

[221] Shao, Z.; Yan, W.; Chapman, K.; Ramesh, K.; Ferrell, A. J.; Yin, J.; Wang, X.; Xu, Q.; Rosenbaum, D. M. *Nat Chem Biol* **2019**, *15*, 1199–11205.

[222] Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J Comput Chem* **2008**, *29*, 1859–1865.

[223] Jo, S.; Kim, T.; Im, W. *PLoS Med* **2007**, *2*, e880.

[224] Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. *Biophys J* **2009**, *97*, 50–58.

[225] Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Conteras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. *J Comput Chem* **2014**, *35*, 1997–2004.

[226] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr, A. D. *J Comput Chem* **2010**, *31*, 671–690.

[227] Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell Jr, A. D. *J Comput Chem* **2012**, *33*, 2451–2468.

[228] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J Comput Chem* **2005**, *26*, 1781–1802.

[229] Huang, J.; MacKerell, A. D. *J Comput Chem* **2013**, *34*, 2135–2145.

[230] Lee, J. et al. *J Chem Theory Comput* **2016**, *12*, 405–413.

[231] Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graph* **1996**, *14*, 33–38.

[232] Roe, D. R.; Cheatham III, T. E. *J Chem Theory Comput* **2013**, *9*, 3084–3095.

[233] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J Comput Chem* **2004**, *25*, 1605–1612.

[234] Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Huang, C. C.; Ferrin, T. E. *BMC Bioinformatics* **2006**, *7*, 339.

[235] Huang, W.; Masureel, M.; Qu, Q.; Janetzko, J.; Inoue, A.; Kato, H. E.; Robertson, M. J.; Nguyen, K. C.; Glenn, J. S.; Skiniotis, G.; Kobilka, B. K. *Nature* **2020**, *579*, 303–308.

[236] Guilloux, V. L.; Schmidtke, P.; Tuffery, P. *BMC Bioinformatics* **2009**, *10*, 168.

[237] Schmidtke, P.; Barril, X. *J Med Chem* **2010**, *53*, 5958–5867.

[238] Shapovalov, M. V.; Dunbrack, R. L. *Structure* **2011**, *19*, 844–858.

[239] Schmidt, D.; Bernat, V.; Brox, R.; Tschammer, N.; Kolb, P. *ACS Chem Biol* **2015**, *10*, 715–724.

[240] Kolb, P.; Caflisch, A. *J Med Chem* **2006**, *49*, 7384–7392.

[241] Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. *Nucleic Acids Res* **2019**, *47*, D1102–D1109.

Acknowledgements

Aus Gründen des Persönlichkeitsschutzes wird von der elektronischen Veröffentlichung der Danksagung abgesehen.

Due to protection of personal privacy, the acknowledgements are not included in the digital version of the dissertation.

Curriculum vitae

Aus Gründen des Persönlichkeitsschutzes wird von der elektronischen Veröffentlichung des Lebenslaufs abgesehen.

Due to protection of personal privacy, the Curriculum vitae is not included in the digital version of the dissertation.

Publications

4 Analyzing kinase similarity in small molecule and protein structural space to explore the limits of multi-target screening.
Denis Schmidt, Magdalena M. Scharf, Dominique Sydow, Eva Aßmann, Maria Marti-Solano, Marina Keul, Andrea Volkamer, Peter Kolb
Molecules, submitted.

3 A Focus on Unusual ECL2 Interactions Yields β_2-Adrenergic Receptor Antagonists with Unprecedented Scaffolds.
Magdalena M. Scharf†, Mirjam Zimmermann†, Florian Wilhelm, Raimond Stroe, Maria Waldhoer, Peter Kolb
ChemMedChem **2020**, 15, 882-890.

2 Comparative docking to distinct G protein-coupled receptor conformations exclusively yields ligands with agonist efficacy.
Magdalena M. Scharf, Moritz Bünemann, Jillian G. Baker, Peter Kolb
Mol. Pharmacol. **2019**, 96, 851-861.
Molecular Pharmacology Highlighted Trainee Author, December 2019

1 Solvation by aqueous solutions of imidazole-based ionic liquids: 2-A comparison between alkyl and alkoxy side-chains.
Jéssica C. de Jesus, Paulo A. R. Pires, Magdalena Scharf, Omar A. El Seuod
Fluid Phase Equilibria **2017**, 451, 48-56

Patent Applications

1 Waldhoer, M., Zenone, L., Wilhelm, F., Zimmermann, M., Ostermaier, M. K., Scharf, M. M., Kolb, P.
Beta Adrenergic Receptor Antagonists.
EPO 20162433.5 - 1109 (2020)

Conferences

03 – 05 Nov. 2019 **15th German Conference on Cheminformatics 2019**, Mainz, Germany
Poster presentation: *Comparative docking to distinct G protein-coupled receptor conformations exclusively yields ligands with agonist efficacy.*

09 – 16 Feb. 2019 **GRS & GRC Molecular Pharmacology 2019**, Ventura, USA
Poster presentation: *Prediction of novel agonists for the β_2-adrenergic receptor using multi-conformation docking.*

12 – 14 March 2018 **32nd Molecular Modelling Workshop**, Erlangen, Germany
Oral presentation: *Computer-aided design of ligands with tailored efficacies for the β_2-adrenergic receptor.*

14 – 17 March 2017 **Spring School in Computational Chemistry**, Espoo, Finland
Poster presentation: *Ligand activity prediction for the β_2-adrenergic receptor.*

Erklärung

Hiermit versichere ich, dass ich meine Dissertation mit dem Titel

Investigations toward the rational modulation of G protein-coupled receptor signalling pathways using *in silico* methods

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den _____

_____

Magdalena M. Scharf