DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Structural Basis of Ligand Recognition at Free Fatty Acid Receptors "

verfasst von / submitted by

Paul Schubert

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Pharmazie (Mag.pharm.)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 449
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Diplomstudium Pharmazie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Gerhard Ecker

The present paper was prepared from April until September 2018 in collaboration with the Institute of Pharmacy at the free University of Berlin under the direction of Prof. Dr. Gerhard Wolber.

Acknowledgements

First of all, I want to thank my supervisor at the Institute of Pharmacy of The Free University of Berlin, Dr. Gerhard Wolber as well as my supervisor at the University of Vienna, Univ.-Prof. Dr. Gerhard Ecker, for giving me the opportunity to work on this project.

Furthermore, I would like to express sincere gratitude to Dr. Marcel Bermúdez and Apotheker Trung Ngoc Nguyen for sharing the professional guidance, helpfulness and constructive criticism during the work on this paper. Additionally, I thank the entire Prof Wolber group for the welcoming and friendly environment they provided.

I also want to thank my boy Rokko for his support and motivation over the last years.

Above all, I want to thank my parents for supporting me and giving me the possibility to study abroad.

Abstract

This paper's aim is to lead to a deeper understanding of structural and functional properties of Free Fatty Acid Receptors. On the basis of the latest state of knowledge of this receptor-family, common computational methods were used to gain new insights. Homology modeling made it possible to generate models of all receptor subtypes based on an existing crystal structure of subtype FFAR1. On the basis of these models, further experiments were conducted. A comparative analysis between FFAR1 and FFAR2 resulted in an explanation for the carbon-chain-length-selectivity of each receptor. Molecular docking of both endogenous and synthetic ligands led to the elucidation of the according binding modes, including the dual binding mode of orthosteric and allosteric ligands at FFAR2. Concluding molecular dynamic simulations revealed additional details in a dynamic context.

German abstract – Kurzfassung

Diese Arbeit verfolgt das Ziel, ein tieferes Verständnis für strukturelle und funktionelle Eigenschaften der Free Fatty Acid Rezeptoren zu erlangen. Basierend auf dem aktuellen Wissensstand über diese Rezeptorfamilie wurde mit der Hilfe von gängigen computergestützten Methoden versucht neue Erkenntnisse zu gewinnen. Homologiemodellierung ermöglichte es fundierend auf einer bereits ermittelten Kristallstruktur des Subtyps FFAR1 weitere Modelle der Subtypen FFAR2, FFAR3 und FFAR4 zu generieren. Auf der Grundlage dieser Modelle konnten weitere Experimente durchgeführt werden. Eine vergleichende Analyse zwischen FFAR1 und FFAR2 konnte eine Erklärung für die jeweilige Selektivität, welche von der Kettenlänge der dazugehörigen Fettsäure abhängt, liefern. Docking von sowohl endogenen als auch synthetischen Bindungspartnern führte zur Aufklärung der jeweiligen Bindungsmodi, einschließlich dem dualen Bindungsmodus von orthosterischen und allosterischen Liganden mit FFAR2. Abschließende Molekulardynamiksimulationen konnten weitere Details in einem dynamischen Kontext darstellen.

Table of contents

1 Introduction

G protein-coupled receptors (GPCR) are receptors in the cell membrane which transmit extracellular signals to intracellular responses mainly through heterotrimeric guanine nucleotide-binding proteins (G proteins), but also through G protein-coupled receptor kinases (GRKs) and arrestins [1]. A multitude of ligand types including ions, hormones, neurotransmitters, peptides and proteins interact with GPCRs [2]. They picture the biggest family of transmembrane receptors and are involved in a multitude of physiological and pathophysiological processes [3]. Approximately 50% of all marketed drugs are targeting GPCRs [4,5]. The following sections will give you an introduction on GPCRs and a deeper look at the free fatty acid receptors (FFAR) – a GPCR family that binds free fatty acids as their endogenous ligands.

1.1 GPCR Classification

More than 800 different human genes are coding for functional GPCRs. By comparison of their primary sequence three main families with no detectable shared sequence homology among themselves were classified: A, B, C or alternative 1, 2, 3. Class A represents the largest family and includes nearly 85% of the GPCR genes [6]. Beside the ABC/123 classification also other systems have been developed [7].

Kolakowski [8] classified all GPCRs of both vertebrates and invertebrates in classes A-F. These are the Class A Rhodopsin-like, which account for over 80% of all GPCRs, Class B Secretin-like, Class C Metabotropic glutamate receptors, Class D Pheromone receptors, Class E cAMP receptors and the Class F Frizzled/smoothened family. There are 286 human non-olfactory Class A receptors, the majority of which bind endogenous peptides, amines or lipid-like substances [7,9].

Derived from their phylogenetic distance the GRAFS classification system grouped the GPCRs into five main families: Glutamate, Rhodopsin, Adhesion, Frizzled/Taste and Secretin [7]. FFARs belong to the Rhodopsin family.

1.2 Structure of GPCRs

On the basis of their structure GPCRs belong to the superfamily of the hepta-helical trans-membrane (TM) receptors [10,11]. Seven transmembrane domains are connected through intra- and extracellular loops which are enabling interactions between the receptors and their ligands and can vary in structure, sequence and length [12,13]. Since the discovery of different intracellular binding partners of GPCRs [14,15], the term seven-transmembrane do-main (7-TMD) receptor gets more and more established [2].

A   B
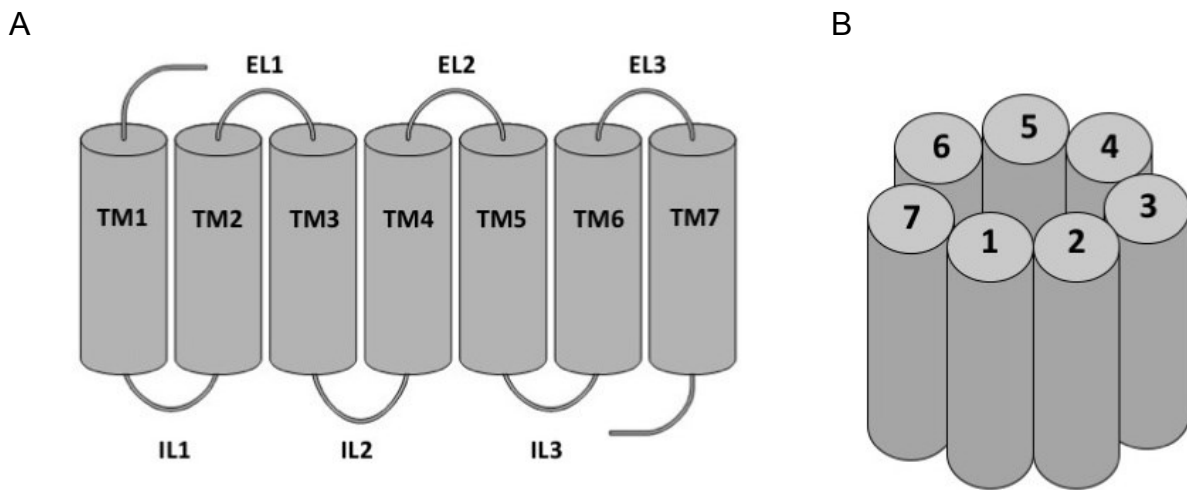
Figure 1: Topology of a GPCR. A GPCR has an extracellular N-terminus and an intracellular C-terminus. Seven membrane spanning helical domains are linked through three extracellular and three intracellular loops (A). Viewed from outside the cell the TM domains are arranged in a counterclockwise fashion (B).

The seven membrane-spanning helical domains of GPCRs are responsible for the anchoring in the cell membrane. They are counterclockwise organized when considered from the extracellular side (Figure 1). In contrast to the variable extracellular and intracellular domains, the tertiary structure of TMDs are highly conserved within the GPCR superfamily. Specific amino acid sequences (motifs) are found throughout the GPCR families. For example, the E/DRY motif in TM3 and the NPxxY motif in TM7, which play an important role for the receptor activation, can be found in almost all Rhodopsin-like receptors [16].

The orthosteric binding sites are usually found in extracellular or transmembrane regions. Receptors, which have their binding sites located in the N-terminal domain, are characterized by a very long N-terminal amino acid sequence (up to 2800 amino acids). A short N-terminal amino acid sequence indicates that the receptor's binding site is located on another region [16,17]. The intracellular regions interact with G-proteins and other signaling molecules. Intracellular loop (IL) 2, IL 3 and the intracellular C-terminal domain are strongly involved on G-protein binding [18].

Ballesteros and Weinstein[19] established a numbering system for GPCRs, that assigns the helix number and the number 50 to the highest conserved residue within each TMD. As a result, all the remaining transmembrane residue positions are getting defined as a sequence-intern distance from that amino acid.
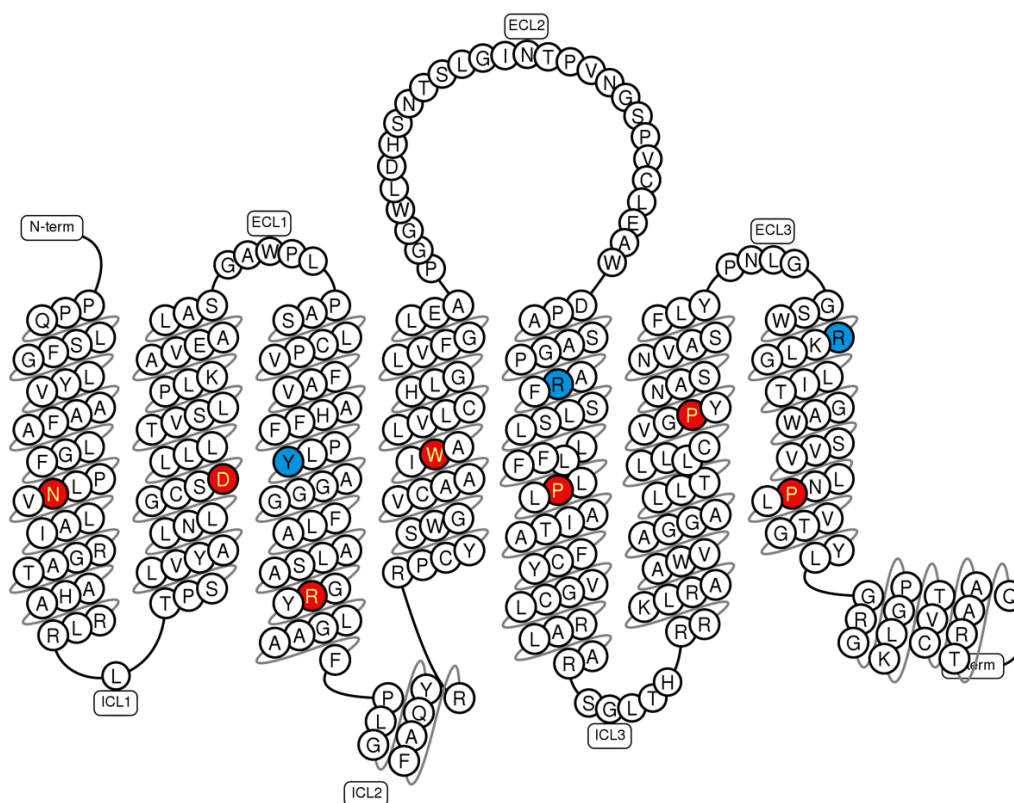

Figure 2: Snake plot of FFAR1. Most conserved residues of each TMD colored in red. Key residues for orthosteric ligand binding are shown in blue. The figure was created by using the snake plot tool from gpcrdb.com.

TMD1 is characterized by an asparagine in the GNxxV-motif. Mutational studies of the α1b-adrenergic receptor showed, that this residue has influence on the activation state [15]. The most conserved amino acid in TMD2 is Asp2.50. Arg3.50 is part of the DRY-motif or rather the ARG-cage [9]. While Arg3.50 plays an important role for the signal transduction to the g-protein [2], the whole DRY-motif is attributed to have an important function for stabilizing the receptor in an inactive conformation. The role of Trp4.50 and Pro5.50 is not yet clarified. Pro6.50 creates kinks in the helix that are important for conformational changes from inactive to active states. Pro7.50 plays an important role as part of the NPxxY motif which is essential for receptor activation [20]. Pro5.50 is the only exception in this numbering system, because technically Tyr5.58 is the most conserved residue in TMD5. However, Pro5.50, which is often associated with Phe5.47, represents a better marking position [16]. Besides the mentioned residues which are most conserved within all GPCRs, also those residues that are responsible or involved in ligand binding obtain special attention in GPCR research (Figure 2).

1.3 Crystallization and structure determination of GPCRs

The crystallization of GPCRs was highly challenging due to the stabilizing effect of the membrane on the protein structure as well as the highly dynamic nature of the receptor itself. The crystal structure of bovine rhodopsin was the first GPCR resolved in 2000 [21]. Over the next seven years 14 crystal structures were solved and all of them were non-human rhodopsin receptors. In 2007 the first human GPCR structure (human β2 receptor)[22] was resolved and since then X crystal structures for human GPCRs are available.

The development of new methods in crystallography and protein engineering was acquired to achieve the aim of structure enlightenment and mechanistical understanding of GPCRs. Important approaches have been the insertion of thermostabilizing mutations and insertion of the T4-lysozyme for stabilizing IL3. The use of anti- or nanobodies as well as the use of ligands that show a high affinity to the binding side contributed to the achievement of GPCR structure determination [23]. By combination of these techniques it became even possible to crystallize active-like receptor conformations.

1.4 Pharmacology of GPCRs

Classically GPCRs have been pharmacologically exploited through the endogenous binding site. But it has been shown that different binding sites can also be targeted to modulate GPCR signaling. We differentiate between orthosteric sites, which interact with the endogenous ligands, and allosteric sites, which are located distinct from the orthosteric site and also offer opportunities for ligand interaction [2].

GPCR ligands can be classified by their ability to shift the equilibrium between active and inactive conformation with regard to basal activity. Full agonists activate the receptor and produce the maximum response a receptor can elicit, whereas partial agonists increase the activity between basal and maximum response. Antagonists competitive block the binding of agonists and do not shift the basal activity. Inverse agonists competitive prevent agonist binding and lead to decrease of the basal activity [24].

Bock A. et al[25] described GPCRs as integrative and highly dynamic signaling machines because of their ability to trigger several signaling mechanisms. Beside G protein activation they can also result in G protein independent signaling pathways like the activation of arrestins or different types of G protein-coupled receptor kinases (GRK) (Figure 4) [26]. G protein activation starts with the binding of a ligand that stabilizes the active receptor conformation, which then binds a G protein on its intracellular surface. In its trimeric ground state, the G protein is bound to a guanosindiphosphat (GDP) molecule. The formation of a GPCR-G protein complex leads to the exchange of GDP with guanosintriphosphat (GTP) and dissociation of the trimer into its α & βγ subunits. Various pathways are getting activated, depending on the G protein type: Gαs activates adenylate cyclase and thereby the production of the second messenger cAMP, while Gαi,o inhibits the same pathway. Gαq,o activates phospholipase Cβ and thus calcium channels indirectly. Altogether there are 20 different Gα subunits [27,28] which can be divided in four distinct G protein families (Gs, Gi/0, Gq/11, G12/13) [29]. Furthermore, six different β and twelve γ subunits exist [30].
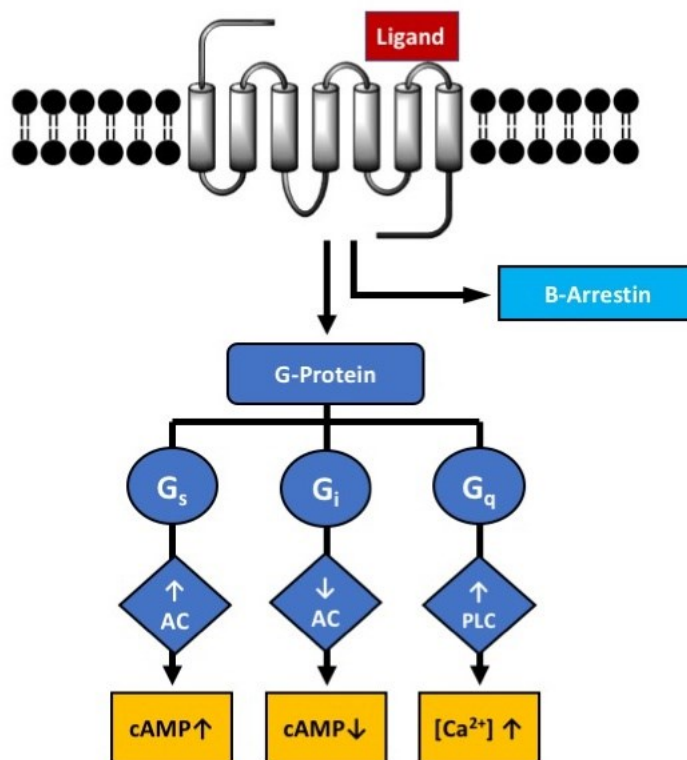
Figure 3: *GPCR signaling. Depending on the G-protein sub-class G-protein-mediated signaling causes different intracellular responses. GPCRs may also signal through G-protein-independent mechanisms like β-Arrestin activation.*

Certain ligands have the ability to trigger distinct signaling profiles at the same receptor through stabilizing favorable conformations. This phenomenon is described as biased signaling or functional selectivity [31]. Biased signaling can occur between different G-protein subtypes but also between those and G-protein independent signaling mechanisms, where focus has been mainly on arrestin-interactions [32]. Signaling pathways lead to different effects that can be both beneficial and adverse. The interest in biased ligands is increasing over the last years because of their potential to act as drugs that show only the desired positive effects with less side effects [33].

1.5 Free fatty acids and free fatty acid receptors

Fatty acids (FA) consist of a carboxylic acid and an either saturated or unsaturated aliphatic chain. As long as they are not ester-bound, they are called non-esterified or ''free'' fatty acids (FFA). Fatty acids can be categorized by the length of their aliphatic tails as short-chain fatty acids (SCFA; ≤5 carbons) [34], medium-chain fat-ty acids (MCFA; 6-12 carbons) [35] and long-chain fatty acids (LCFA; 13-21 carbons) [36].

Consumption of dietary fat and dietary fiber leads to the synthesis of FAs mainly by the liver. FAs mostly bind to glycerol and form triglycerides. Through lipolysis of triglycerides FFAs get into circulating plasma [37,38]. FFAs fulfill many important physiological roles. Besides proteins and carbohydrates, they act as important energy sources in almost all body tissues. They also have a big role in crucial functions like receptor signaling, gene expression and regulation of energy homeostasis [39]. An unbalanced diet can result in elevated blood levels of lipids or lipoproteins and therefore cause related diseases like metabolic syndrome, obesity, type 2 diabetes, cardiovascular diseases and fatty liver [40,41].

The free fatty acid receptors (FFAR) are activated by free fatty acids and consist of four different receptor subtypes: FFAR1 – 4 [42]. Because of their potential for the treatment of metabolic and inflammatory diseases this receptor family obtains growing interest as novel drug targets [2]. FFAR1 & FFAR4 are activated by saturated and unsaturated LCFAs like myristic acid (C14:0), palmitic acid (C16:0), oleic acid (C18:1), linoleic acid (C18:2), α-linoleic acid (C18:3), arachidonic acid (C20:4), EPA (C20:5, n-3) and docosahexaenoic acid (C22:6, n-3) [42]. However, SCFAs like acetic acid (C2), propanoic acid (C3), butyric acid (C4) and pentanoic acid (C5) activate FFAR2 as well as FFAR3 [43,44]. The fact that several FFAs can activate the same FFAR, whereas one FFA can activate several FFARs, highlights the complexity of FFAR pharmacology. Binding of FFARs and their endogenous ligands leads to the activation of various intracellular signal transduction pathways. While mostly g-protein depended signaling is triggered, FFAR activation can also result in g-protein independent signaling mediated by beta-arrestin (Figure 4). FFARs are expressed in different tissues such as pancreatic beta cells, white adipocytes and the central nervous system and take part in the regulation of metabolism and inflammation next to other important physiological functions. [45].
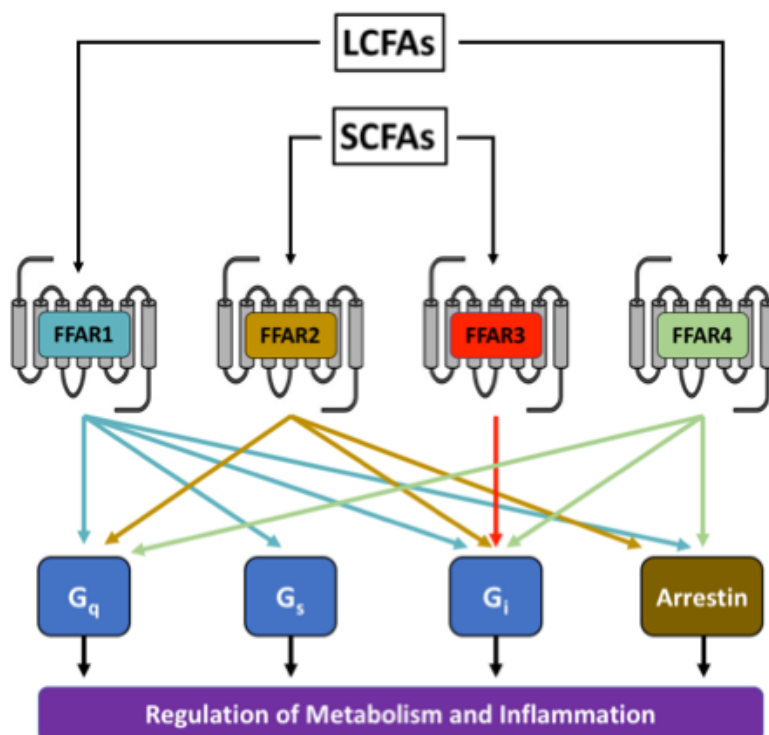
Figure 4: *FFAR signaling. Activation of FFARs can lead to the activation of a variety of signaling pathways. G-protein-mediated, as well as Arrestin-mediated signaling can be triggered.*

1.6 FFAR1 Crystal structures

Srivastava et al[46] solved the first crystal structure of FFAR1 (Figure 3), also known as human GPR40 (hGPR40), in 2014. This structure serves as an important starting point for the ongoing studies in this thesis. The receptor is bound to TAK-875, a potent and selective partial agonist of FFAR1 which reached phase III clinical trials as a potential treatment of type-2 diabetes mellitus under the name Fasiglifam [47]. With a resolution of 2.3A the structure was determined by X-ray crystallography. The structure is declared with the four-letter code 4PHU and available at the RCSB protein databank.

Since the revealing of the first FFAR1 crystal structure by Srivastava, three more have been published so far. Two years later Ho et. all were able to crystallize FFAR1 bound to the synthetic full agonist compound 1 [48]. The latest structures were revealed in 2017 when Lu et. all determined FFAR1 structures in complex with the partial agonist MK-8666 as well as in complex with MK-8666 and the agonistic and positive allosteric modulating ligand (AgoPAM) AP8 [49].

2 Methods

2.1 Homology Modeling

Homology modeling (HM) is a method for tertiary structure prediction of proteins with no available experimentally determined structure. By aligning the sequence of the target (the protein with an unknown 3D STRUCTURE) with the template (a related homologous protein) this approach delivers structural models from the protein of interest [50]. The quality of the model is depending on the sequence identity between target and template. A similarity of at least 70% results in models that are almost identical to experimentally determined structures. Below a similarity of 20% the obtained models can vary greatly from experimentally determined structures [51]. Although their quite low sequence similarities GPCRs are suitable proteins for HM because of the similar overall TMD-architecture [52]. In this thesis homology models of FFAR2, FFAR3 and FFAR4 were built using a FFAR1 crystal structure as template.

For every FFAR subtype model the crystal structure of FFAR1 bound to the allosteric agonist TAK-875 (PDB: 4PHU), that represents the receptor in an inactive-like state, was used as template. This structure was chosen as template because of its best resolution (2.33A) compared to the other available FFAR1 structures. MOE (Molecular Operating Environment) software was used for the whole HM-process. The sequence alignment, which can be considered as the most critical part [53], was made under main attention of a rational alignment of the TMD sequences. The standard settings were used and 10 main chain models each with 3 side chain samples were built using the OPLS-AA force field at a temperature of 300 K. As a result, we obtained 30 intermediate models with an RMS gradient of 1. Refinement protocols with an RMS gradient of 0.5 were applied to build the final homology model. A manual model refinement was accomplished by optimizing the protein geometry considering dihedral angel distributions and atom clashes.

2.2 Ligand selection and preparation

Selection and preparation of ligands are essential steps before we are able to perform molecular docking experiments. Main criterions for the selection were receptor-selectivity and potency. Selective synthetic ligands have been reported for each FFAR subtype, while endogenous ligands can interact with two receptors likewise (LCFAs with FFAR1 & FFAR4; SCFAs with FFAR2 / FFAR3). The selected ligands were created and energy minimized in MOE (*Molecular Operating Environment (MOE)*, 2014.09; Chemical Computing Group Inc.).

2.3 Molecular Docking

Molecular docking is a method that allows binding mode analysis by flexibly fitting small molecules into a binding pocket. It is one of the most frequently used in silico methods in pharmaceutical research because it enables virtual analysis of interactions between big quantities of diverse small molecules with a protein target. The process is based on the known chemical and spatial structure of the starting molecules. Search algorithms fit the ligand into the target protein's binding site, which is mostly assumed to be rigid [54].

To measure the fit of a ligand's pose into its active site scoring functions are used. Scoring involves energy calculations (electrostatic, hydrogen bonds, van der Waals,…) to evaluate the strength of intermolecular interactions between receptor and ligand [55].

All molecular docking experiments were performed using the automated docking software GOLD (Genetic Optimization for Ligand Docking, 5.22). Standard settings were used, and diverse solutions option was selected. For fitness function Goldscore was selected and 10 GA runs for each ligand were performed. The orthosteric binding site for FFAR1, FFAR2 and FFAR3 was defined as a point between the conserved $ARG^{5.40}$ and $ARG^{7.34}$ within a 10A radius. For FFA4 the binding site was defined by $ARG99^{2.64}$ within a 10A radius as well. Additional allosteric docking at FFAR2 was performed where $TRP253^{7.32}$ was used to define the binding site within 10A. The generated poses were analyzed using LigandScout Version 4.09 [56,57] and minimized with the MMFF94 force field. Docking poses with the best interactions and spatial conformation were selected.

2.4 Molecular Dynamics Simulations and Dynophores

Molecular dynamic (MD) simulations calculate the physical movements of atoms and molecules to sample different conformations of a rigid macromolecular system. MD trajectory calculations are subject to the Newtonian equations of motion. Force-fields describe the interactions between atoms and their potential energy contains covalent interactions (bond stretching, bond angle bending, dihedral angle bending) and non-covalent interactions (van der Waals forces, Coulomb force) [58,59]. Many MD software suites have been established, including AMBER [60], GROMACS [61] and Desmond [62].

For this thesis all MD simulations were prepared with Maestro [63]. To mimic the physiological state 0.15 M salt was added to the system. Transmembrane regions were taken from UniProt [64] and imbedded in a preequilibrated palmitoyl-oleoyl-phosphatidyl-cholin (POPC) bilayer membrane. The system Pressure and temperature were kept at atmospheric pressure and 300K. Standard settings were used for all other parameters. MD simulations of 100ns were carried out in Desmond [62].

A pharmacophore consists of the molecule's features that are responsible for its pharmacological effects. All steric and electronic features that are necessary to trigger interactions with a target structure are considered (table 1) [65]. Pharmacophores play an important role in modern cheminformatics, i.e. for screening databases of chemical compounds for potential effective drugs. Dynophores are an extension of classical pharmacophores that allow the investigation of protein-ligand interactions in a dynamic representation. The pharmacophore feature information is extracted and collected from every single step of the MD-simulation and combined to a dynophore. The resulting three-dimensional volumetric feature density clouds enable the depiction of the spatial occurrence of a feature, while barcode plots facilitate a statistical evaluation of the feature's occurrence during the time period according to the MD-simulation [66,67]. For the depiction of pharmacophores and dynophores we used LigandScout Version 4.09.

Depiction in LigandScout	Pharmacophore Feature	Depiction in LigandScout	Pharmacophore Feature
	Positive Ionizable Area		Negative Ionizable Area
	Hydrogen Bond Donor		Hydrogen Bond Acceptor
	Aromatic Ring		Hydrophobic Inter-actions

Table 1: *Selection of pharmacophore features in LigandScout.*

3 Results

3.1 Comparative Investigations based on Homology Models
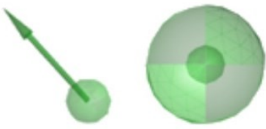
The starting point for the experimental section of this thesis consists of the crystal structure of FFAR1 and the created homology models of FFAR2, FFAR3 and FFAR4. This section is aimed to reveal insights about similarities and differences of all FFAR subtypes by structural and sequential comparing. For the validation of the models we perform a MD simulation of each subtype in its apo form and take a look at the receptors' RMSD plots.

3.1.1 FFAR2 and FFAR3 show high similarity and identity

The sequence identity plot (Figure 6: A) values range from 12.7% to 42.5%. The individual values reveal that FFAR1 and FFAR4 show the lowest, while FFAR2 and FFAR3 show the highest identity between all subtypes. Furthermore, FFAR1 shows similar identities with FFAR2 (29.2%) and FFAR3 (32.1%). Finally, the identity between FFAR2 and FFAR4 amounts to 14.9%. We can observe that all of the lowest scores include the comparison with FFAR4, so we assume that this subtype differs the most from the others. The highest identity between FFAR2 and FFAR3 supports the fact that both of them bind to the same type of ligands (SCFAs).
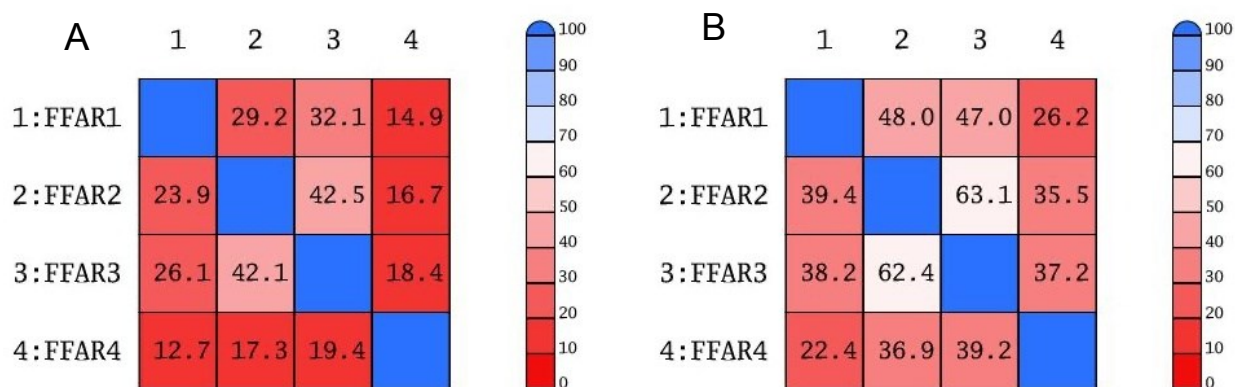

Figure 5: Identity (A) and similarity (B) plots of FFAR subtypes 1-4. The values correspond to percentage rates. A legend for the used color code is depicted.

Sequence similarities between the four subtypes range from 22.4% to 63.1% (figure 6: B). As in the identity plot we observe the highest values between FFAR2 and FFAR3 (63.1%). Surprisingly we achieve the lowest values for sequence similarity between FFAR1 and FFAR4 (22.4%), although they both interact with LCFAs.

3.1.2 Transmembrane regions of FFARs are highly conserved

The alignment of all receptors shows structural similarities and differences (figure 6). We can observe that the TMDs are almost identical except of one big and a few small differences in the length of their α-helices. The intracellular terminal of FFAR4's TMD5 is nine amino acids longer than the TMD5 of FFAR1, while there are a few negligible length varieties of one to three amino acids between all subtypes. Big differences occur between the ECL and ICL regions. These regions diversify in their lengths as well as in their spatial structure. The structural differences can be explained by their general flexibility, given by not being anchored in the cell membrane.

■ : FFAR1   ■ : FFAR2   ■ : FFAR3   ■ : FFAR4
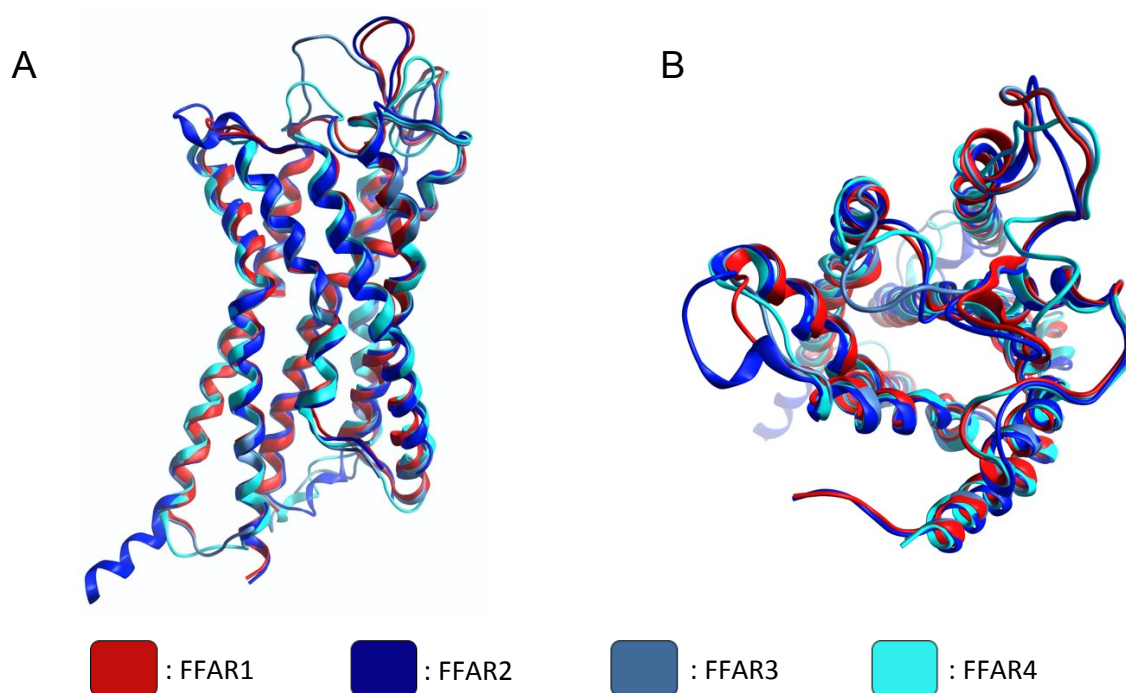
Figure 6*: Superposition of FFAR1 crystal structure and FFAR2-4 homology models. Membrane (A) and extracellular (B) view. FFAR1 is colored in red, FFAR2-4 in different shades of blue (legend depicted)*

The RMSD plots of the apo simulations show the same typical trend for each receptor (figure 8). We observe a rapid rising course until the systems attain an equilibrium state. These states are located on different RMSD levels. FFAR1 shows its equilibrium at the lowest RMSD level, according to the fact that this structure was received by crystallization and therefore is more stable than the homology models.
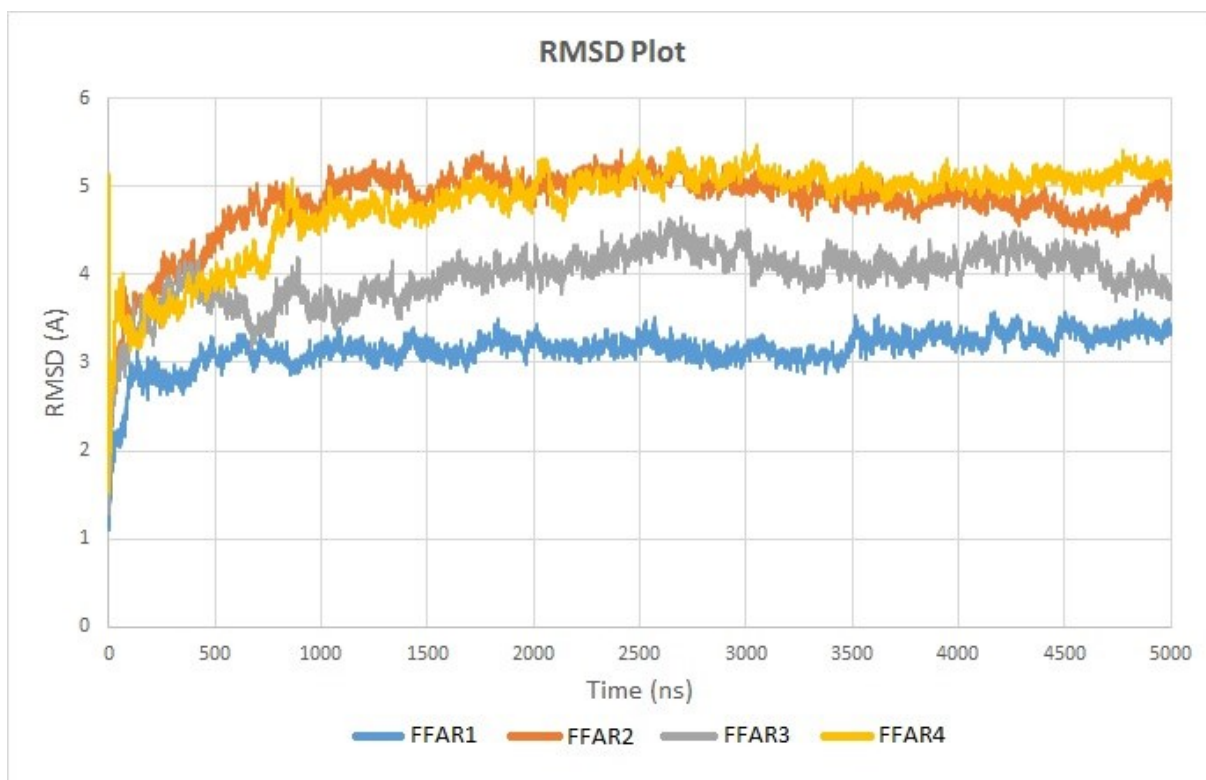
Figure 7: *RMSD Plots of FFAR1-4 apo simulations (legend depicted).*

16

3.2 Binding Modes and Selectivity

After completion and validation of the FFAR2, FFAR3 and FFAR4 homology models we were able to start the molecular docking experiments. The aims of this section are to describe the binding modes of reported selective ligands for each receptor subtype and to explain the binding-selectivity of LCFAs and SCFAs for their corresponding receptor subtypes. Therefore, we first analyzed existing experimental data in terms of binding modes. Then we selected useful reported ligands with which we performed docking experiments.

3.2.1 Crystal structure of FFAR1 reveals orthosteric binding mode

To get an idea of how ligands interact with FFARs we analyzed the crystal structure of FFAR1 in complex with TAK-875 (figure 8). The ligand enters the receptor between TM3 and TM4 through a tube-shaped cavity that is formed by lipophilic amino acids from TM3 (Val81^{3x27}, Ala83^{3x29}, Val84^{3x30}, Phe87^{3x33}), TM4 (Leu138^{4x57}, Phe142^{4x61}) and ECL2 (Trp174^{ECL2}). Lipophilic interactions between these residues and the lipophilic features of TAK-875 (aromatic rings, methyl groups) stabilize the ligands conformation. Arg183^{5x30} and Arg258^{7x34} anchor the ligand by its carboxylate in the core of the receptor and serve as key residues. By acting as H-bond donor they show the strongest interactions involved in this binding mode. Another H-bond interaction with the carboxylate comes from Tyr91^{3x37} on TM3. The polar residue Asn244^{6x55} at TM6 might also play an important role although it doesn't interact directly with TAK-875 but forms an H-bond with Arg258^{7x34} and thus stabilizes its orientation close to the ligand's carboxylate moiety.

From the described observations we assume that ligands of FFAR2, FFAR3 and FFAR4 interact in a similar way. By comparing the structure of endogenous ligands with the structure of reported synthetic ligands we notice that they all share a carboxylic acid or other acidic functional groups (e.g. thiazolidinedione at Rosiglitazone, figure 10). We assume that an interaction between the acidic group of a ligand with a positive charged amino acid of the receptor defines the binding pose for the other FFAR subtypes as well. By comparing the crystal structure and the receptor models, we discovered that the observed
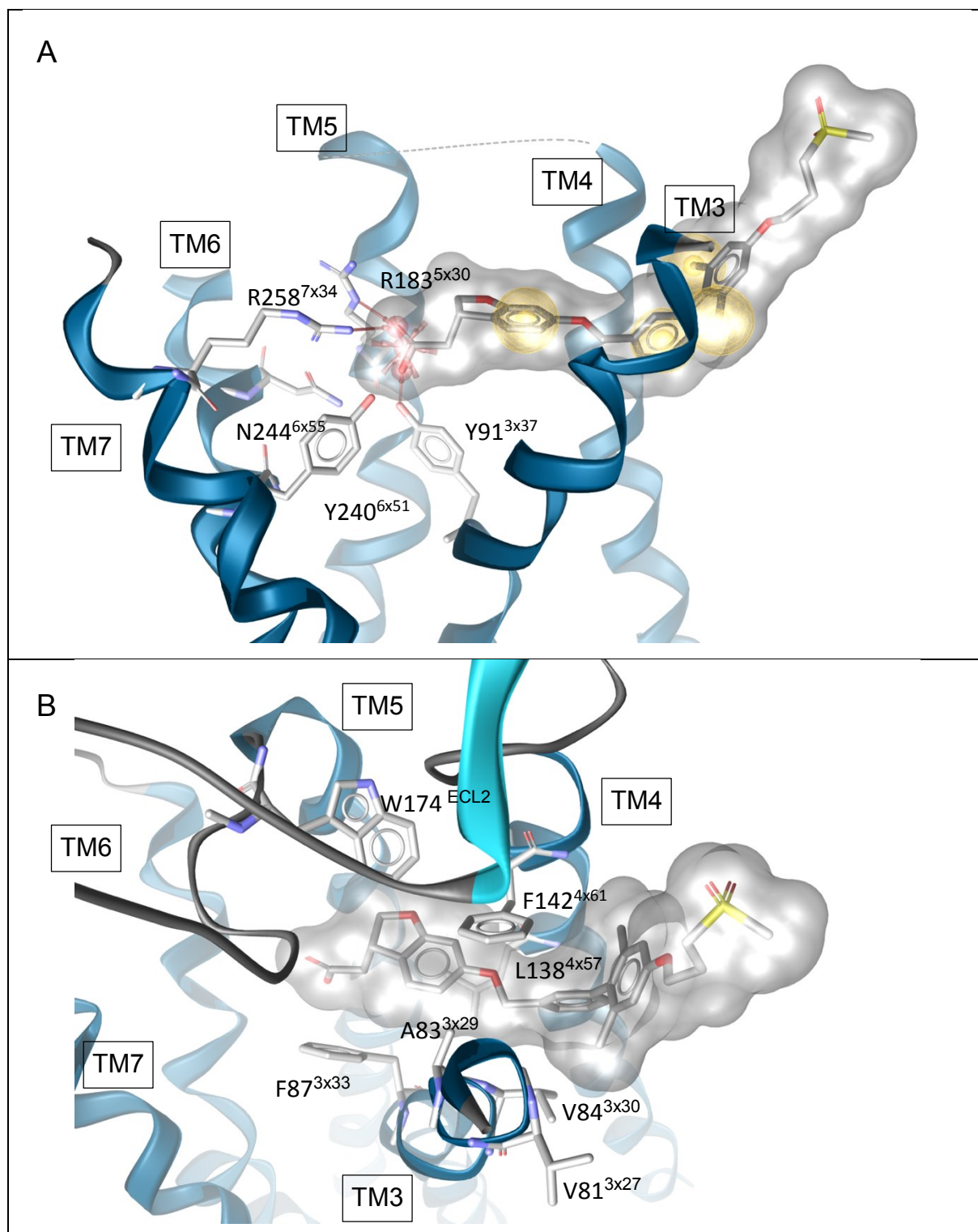
Figure 8: *Crystal structure of FFAR1 in complex with TAK-875. Receptor's TMDs are shown as blue ribbons while TAK-875 and side chains are represented as stick model. Molecular Surface of the ligand shown in transparent grey. A: Transmembrane view of the complex. Positive charged and polar key residues coordinate the carboxylate moiety. Pharmacophore reveals that the binding mode includes hydrogen bonds and hydrophobic interactions. B: Extracellular view. Lipophilic amino acids form a tube-shaped cavity and stabilize the ligand.*

key residues Arg183^{5x40} and Arg258^{7x34} from FFAR1 crystal structure are also conserved at FFAR2 and FFAR3, but not at FFAR4. Therefore, we hypothesize that Arg^{5x40} and Arg^{7x34} are the key residues for ligand binding at FFAR2 and FFAR3, while the anchoring site for carboxylates at FFAR4 must be distinct. By checking the amino acid sequence of FFAR4 we find Arg99^{2x64} as a possible key residue. The putative binding site could be located in the receptor's core, involving TM2, TM3, TM5, TM6 and TM7. Mutation studies of R99^{2x64}A show a loss of ligand affinity for the receptor [68–70]. Based on that knowledge, we assume that Arg99^{2x64} could act as a key residue for ligand binding and use that residue as starting point for our molecular docking experiments with the FFAR4 model.

3.2.2 FFARs bind saturated and unsaturated fatty acids

Endogenous FFAR1/FFAR4 ligands are LCFAs that can vary in chain length and extent of saturation. The selected ligands include saturated, mono unsaturated and poly unsaturated fatty acids with chain lengths of 14 to 22 carbons. Monocarboxylic acids that consist of two, three and four carbons are representative endogenous ligands of FFAR2 & FFAR3 (figure 9).

FFAR1 / FFAR4   **FFAR2 / FFAR3**

Figure 9: Selected endogenous ligands of FFAR1/FFAR4 and FFAR2/FFAR3.

3.2.3 Selective synthetic ligands have been reported for all FFARs

GW9508 (3-(4-{[(3-phenoxyphenyl)methyl]amino}phenyl)propanoic acid) is one of the first reported synthetic FFAR1 agonists. This ligand comes from a series based on an N-substituted 3-(4-aminophenyl)-propanoic acid template [71]. GW9508 is able to activate FFAR4 and FFAR1 but displays 70-fold higher potency at FFAR1 [72]. Rosiglitazone is an antidiabetic thiazolidinedione drug that was remarked to activate FFAR1 [73]. We can observe that it doesn't implicitly need a carboxylic group for an interaction with the receptor

but a functional group with acidic properties, like the ligand's thiazolidinedione. New ligands were identified by combining the GW9508 series with structure-activity relationship exploration. One of the resulting ligands is TAK875 (2-[(3S)-6-({3-[4-(3-methanesulfonylpropoxy)-2,6-dimethylphenyl]phenyl}methoxy)-2,3-dihydro-1-benzofuran-3-yl]acetic acid), the first selective FFAR1 agonist that was brought into clinical trials by Takeda [47]. Christiansen et al explored TUG770 (3-(4-{2-[2-(cyanomethyl)phenyl]ethynyl}-2-fluorophenyl)propanoic acid), a selective and potent FFAR1 agonist, by optimizing a series of 4-phenethynyldihydrocinnamic acid (figure 10) [74,75].

Figure 10: Selective synthetic FFAR1 ligands.

Because FFAR1 and FFAR4 are both activated by LCFAs, synthetic ligands for one subtype assess potential activity for the other. By modification of PPARy active ligands, NCG21 (4-(4-{2-[phenyl(pyridin-2-yl)amino]ethoxy}phenyl)butanoic acid), an agonist with modest selectivity for FFAR4 over FFAR1, was developed [76]. From modifying a series of dihydrocinnamic acid-based chemicals, which had shown FFAR1 activity, the first highly potent and selective FFAR4 agonists were identified. TUG-891 (3-(4-{[5-fluoro-2-(4-methylphenyl)phenyl]methoxy}phenyl)propanoic acid) is one of these synthetic ligands and shows a more than 1000-fold selectivity over FFAR1 in an arrestin-recruitment assay. Metabolex (now Cymabay Therapeutics) developed a few selective FFAR4 ligands. The most potent of them is Compound B [77]. Other companies developed FFAR4 agonists too. Compound A (2-[3-[2-chloro-5-(trifluoromethoxy)phenyl]-3-azaspiro[5.5]undecan-9-

yl]acetic acid) from Merck shows strong potency and a high selectivity for FFAR4 over FFAR1 (figure 11) [78].

Figure 11: Selective orthosteric FFAR4 agonists.

Orthosteric FFAR2 ligands based on a 4-oxobutanoic acid backbone have been reported. Compound 1 (3-{cyclopropyl[4-(2,5-dichlorophenyl)-1,3-thiazol-2-yl]carbamoyl}-4-phenylbutanoic acid) and Compound 2 were synthesized by Hudson et al. and act as potent agonists (figure 12)[79]. A pair of phenylacetamides were the first synthetic allosteric ligands of FFAR2. One of those was 4-CMTB (4-chloro-α-(1-methylethyl)-N-2-thiazolylbenzeneacetamide) which was proven to be an allosteric agonist that is able to activate the receptor directly, as well as to increase the potency of SCFAs when added at the same time. For AZ1729 the same effects were observed (figure 12). In addition, Bolognini et al. reported AZ1729 to act as a Gi-biased ligand [80].
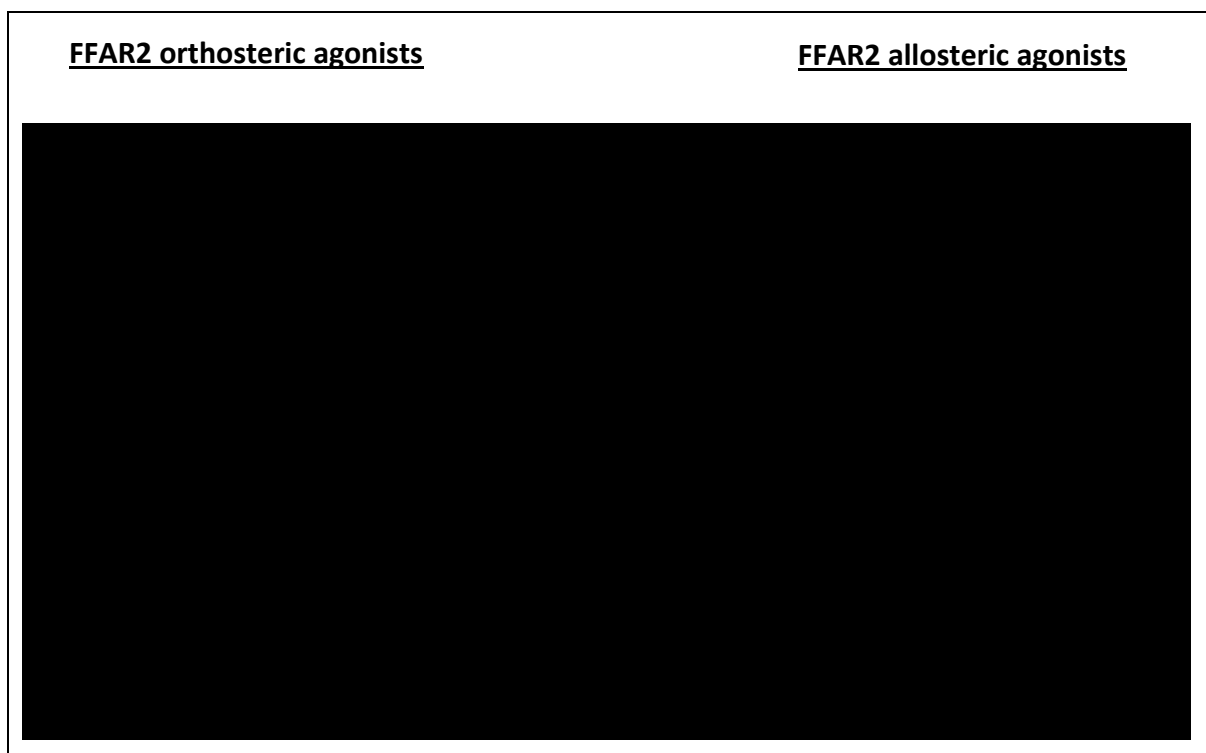
FFAR2 orthosteric agonists   **FFAR2 allosteric agonists**

Figure 12*: Orthosteric and Allosteric FFAR2 ligands.*

There are no high selective synthetic orthosteric FFAR3 ligands reported yet. To explore selectivity between FFAR2 and FFAR3 Schmidt et al. examined a library of small carboxylic acids and enforced a structure-activity relationship analysis. Bulkier and substituted sp3-hybridized α-carbons prefer FFAR3 activation. At least two modest selective FFAR3 ligands could be identified: Compound 7 (1-methylcyclopropanecarboxylic acid) and Compound 12 (3-pentenoic acid) (figure 13)[81].
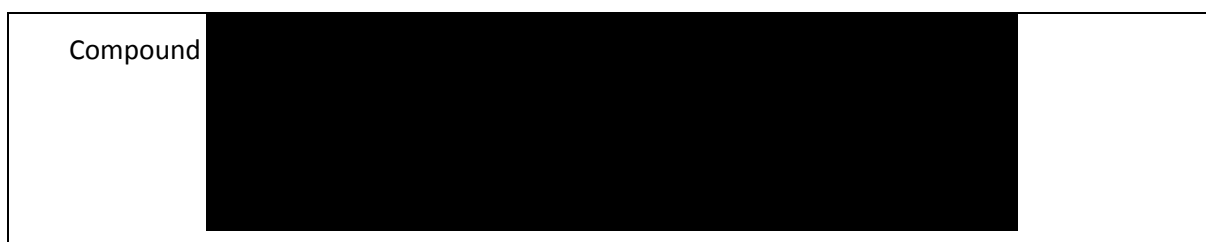
Compound

Figure 13*: Selective orthosteric FFAR3 Ligands.*

3.2.4 LCFAs bind at FFAR1 and FFAR4

Docking of LCFAs alpha-linolenic acid and docosahexaenoic acid into the FFAR1 crystal structure resulted in our assumed orthosteric binding mode (figure 14). Both acids enter the receptor's core between TM3 and TM4. The carboxylates get anchored by the positive charged ARG183^{5x30} and ARG258^{7x34}. Additionally, the carboxylate of alpha-linolenic acid shows an H-bond interaction with TYR91^{3x37}. The remaining interactions have lipophilic character. Each double bond acts as a partner for lipophilic interactions with appropriate amino acids of the receptor. According to that, docosahexaenoic acid shows more interactions with the receptor, including interactions with LEU171^{ECL2} and TRP174^{ECL2}. Both ligands show a quite common binding pose but the hydrophobic tails have a slightly different orientation. That is why the methyl groups on the end of their tails have lipophilic interactions with different FFAR1 amino acids (alpha-linolenic acid with VAL81^{3x27} and docosahexaenoic acid with LEU158 ^{ECL2}).

By docking the LCFAs alpha-linolenic acid and myristic acid, we wanted to check our assumption that ARG99^{2x64}, which is located on the extracellular end of TM2, plays the key role for ligand binding at FFAR4. The experiments delivered two docking poses which show the same interactions between the carboxylates and ARG99^{2x64} (figure 15). In addition, the carboxylate of myristic acid shows an H-bond interaction with THR195^{ECL2}. The lipophilic tails have different orientations. While the saturated tail of myristic acid is orientated straight in intracellular direction, the tail of the unsaturated alpha-linolenic acid makes a kink at the half of the chain that makes the end of the ligand point towards extracellular direction. Both ligands enter the receptor through a cavity between TM4 and TM5. According to the unsaturated structure, alpha-linolenic acid shows more lipophilic interactions with the receptor. PHE115^{3x29}, MET118^{3x32}, THR119^{3x33}, LEU173^{4x61}, LEU196^{ECL2} and ILE300^{6x55} form the lipophilic cavity and stabilize the lipophilic tail in its docking pose. However, myristic acid has only one property to interact with the receptor. The methyl group on the end of the ligand's tail interacts with the neighboring PHE211^{5x43} and VAL212^{5x44} on TM5.

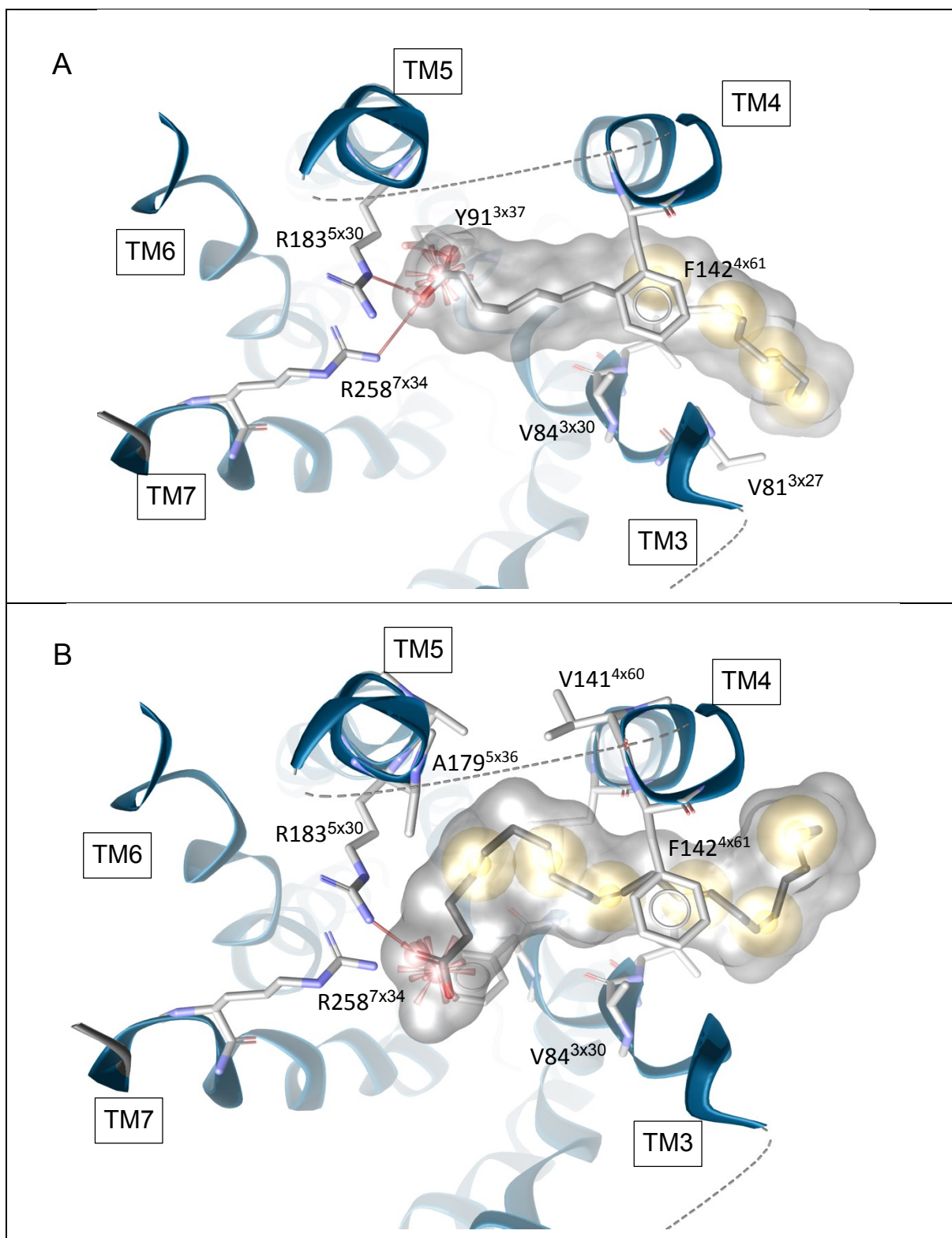*Figure 14*: Docking of endogenous ligands at FFAR1 in extracellular view. Receptor's TMDs are shown as blue ribbons while ligands and side chains are represented as stick model. Molecular Surface of the ligands shown in transparent grey. Pharmacophores reveal that the binding mode includes hydrogen bonds and lipophilic interactions. A: Docking pose of alpha-linolenic acid. B: Docking pose of Docosahexoic acid.

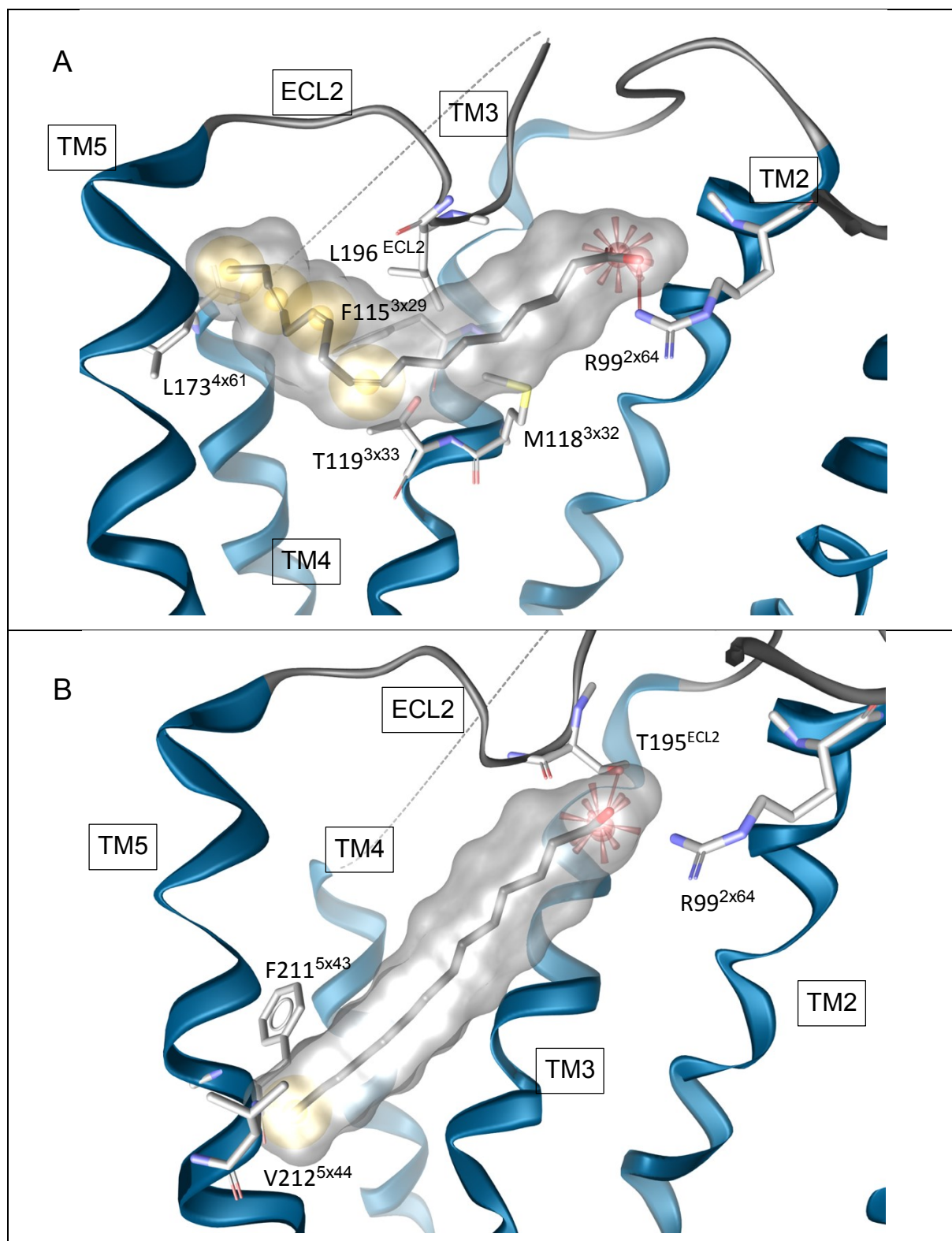***Figure 15***: *Docking of endogenous ligands at FFAR4 in transmembrane view. Receptor's TMDs are shown as blue ribbons while ligands and side chains are represented as stick model. Molecular Surface of the ligands shown in transparent grey. Pharmacophores reveal that the binding mode includes hydrogen bonds and lipophilic interactions. A: Docking pose of alpha-linolenic acid. B: Docking pose of myristic acid.*

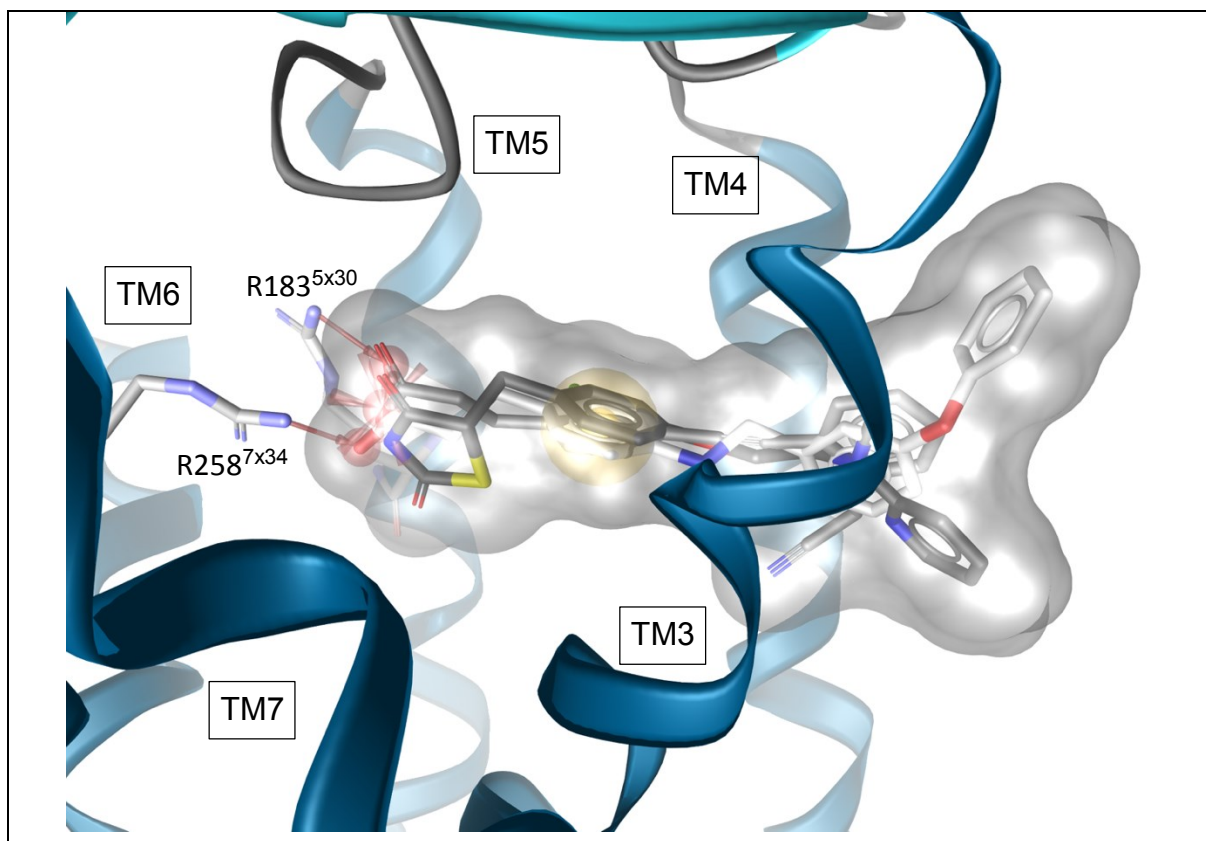### 3.2.5 Synthetic FFAR1 Ligands interact with orthosteric key residues

Figure 16: Transmembrane view of synthetic ligands docked into FFAR1. Receptor's TMDs are shown as blue ribbons while the aligned ligands and side chains are represented as stick model. Molecular Surface of the ligands shown in transparent grey.

An alignment of all three docked synthetic FFAR1 ligands (GW9508, Rosiglitazone, TUG770) shows that the individual binding modes are almost identical, except of small structure-depended differences in the orientation of their lipophilic tail (figure 16). As expected ARG183^{5x30} and ARG258^{7x34} act as key residues by anchoring the ligands by their acidic features (carboxylate at GW9508 and TUG-770; thiazolidinedione at rosiglitazone). TYR91^{3x37} also helps for the anchoring of the acidic end of the ligand by acting as an H-bond donor. Lipophilic amino acids shape the binding pocket and stabilize the ligands in their binding pose. PHE87^{3x33}, PHE142^{4x61}, LEU171^{ECL2} and TRP174^{ECL2} are important residues which are involved in the docking pose of each ligand. PHE87^{3x33} delimits the binding pocket from the intracellular side, while PHE142^{4x61}, LEU171^{ECL2} and TRP174^{ECL2} form the pocket's limit to the extracellular direction. TM3, TM4 and TM5 limit the ligand's space in transmembrane direction.

3.2.6 Synthetic FFAR4 ligands interact with ARG99[2x64]

When we take a look at the selected selective FFAR4 agonists we discovered that they do not share much similarities, despite a carboxylate on the one end, and quite bulky structures on the other end (compared to the endogenous ligands). The carboxylates of Metabolex compound B and Merck compound A are both anchored by ARG99[2x64] and show an additional H-bond interaction with THR195[ECL2] (figure 17). Metabolex compound B shows a multitude of lipophilic interactions with residues from TM2, TM3, TM4, TM5, TM6, TM7 and ECL2 which might stabilize the ligand's position in the binding pocket. However, Merck compound A shows less lipophilic features and accordingly interacts with less FFAR4 amino acids. PHE115[3x29], THR119[3x33], LEU196[ECL2] and PHE211[5x43] interact with the ligand's three lipophilic structures, which include a benzyl ring and it's two p-substituted chloride and trifluoromethane.
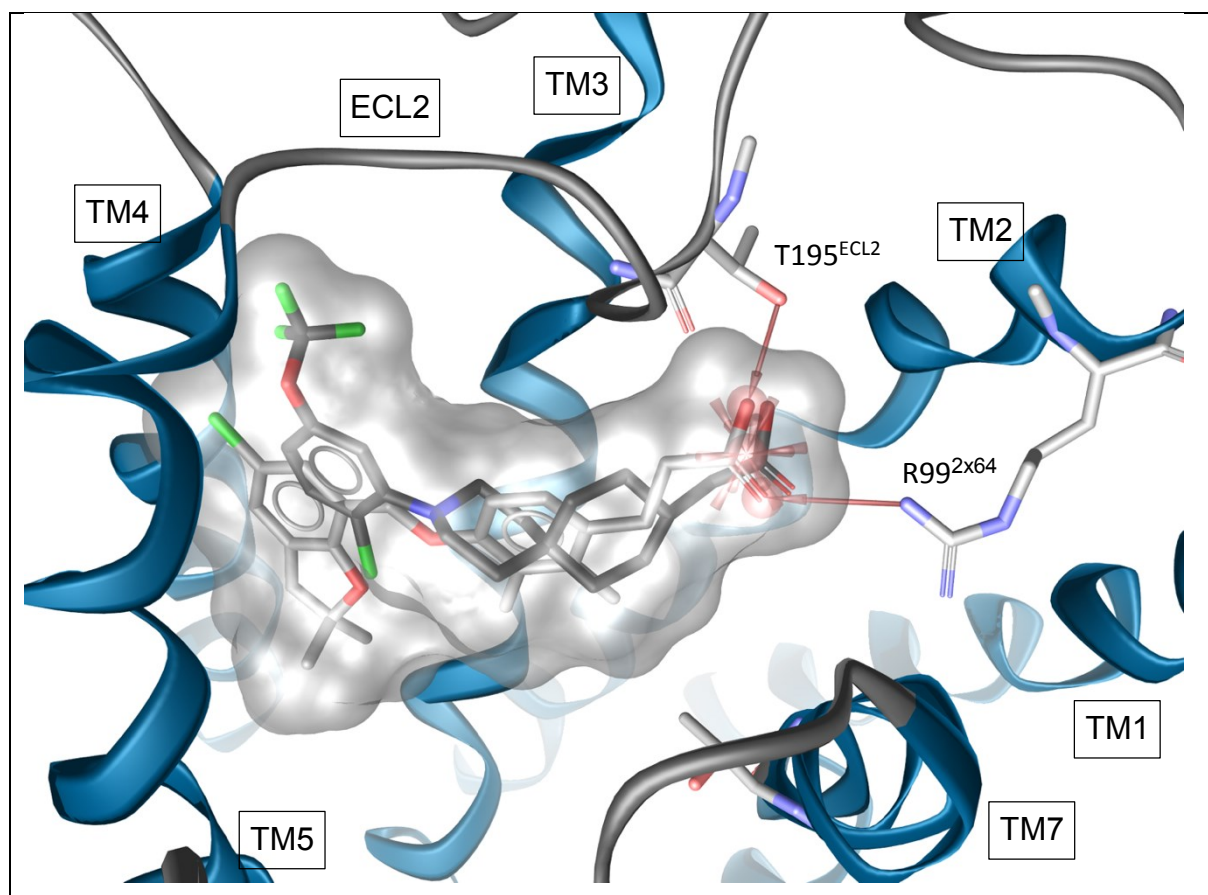

Figure 17: *Docking of Metabolex compound B and Merck compound A at FFAR4. Receptor's TMDs are shown as blue ribbons while the aligned ligands and side chains are represented as stick model in dark grey (Merck compound A) and light grey (Metabolex compound B). Molecular Surface of the ligands shown in transparent grey.*

Compared to the described interactions of FFAR4 with Metabolex ompound B and Merck compound A, the synthetic ligands NCG21 and TUG891 are missing the H-bond interaction between the ligands' carboxylate and THR195^{ECL2}. So just ARG99^{2x64} interacts with the carboxylate and anchors the ligand by its acidic end. Besides that, the ligands show lipophilic features that interact with the receptor (figure 18). The three aromatic rings of NCG21 interact with residues of TM3 (MET118^{3x32}, THR119^{3x33}), TM5 (VAL212^{5x44}), TM6 (TRP293^{6x48}, ILE297^{6x52}, ILE300^{6x55}), TM7 (PHE319^{7x34}, VAL323^{7x38}) and ECL2 (LEU196^{ECL2}), that shape the binding pocket and stabilize the ligand and its position. TUG-891 has also three aromatic rings that interact in a similar way with the same residues as NCG21 does. Additionally, this ligand has a methyl group that interacts with PHE115^{3x29}, THR119^{3x33} and LEU196^{ECL2}. Both ligands might enter the receptor from the extracellular side and their docking poses are entirely positioned in the core of the receptor.
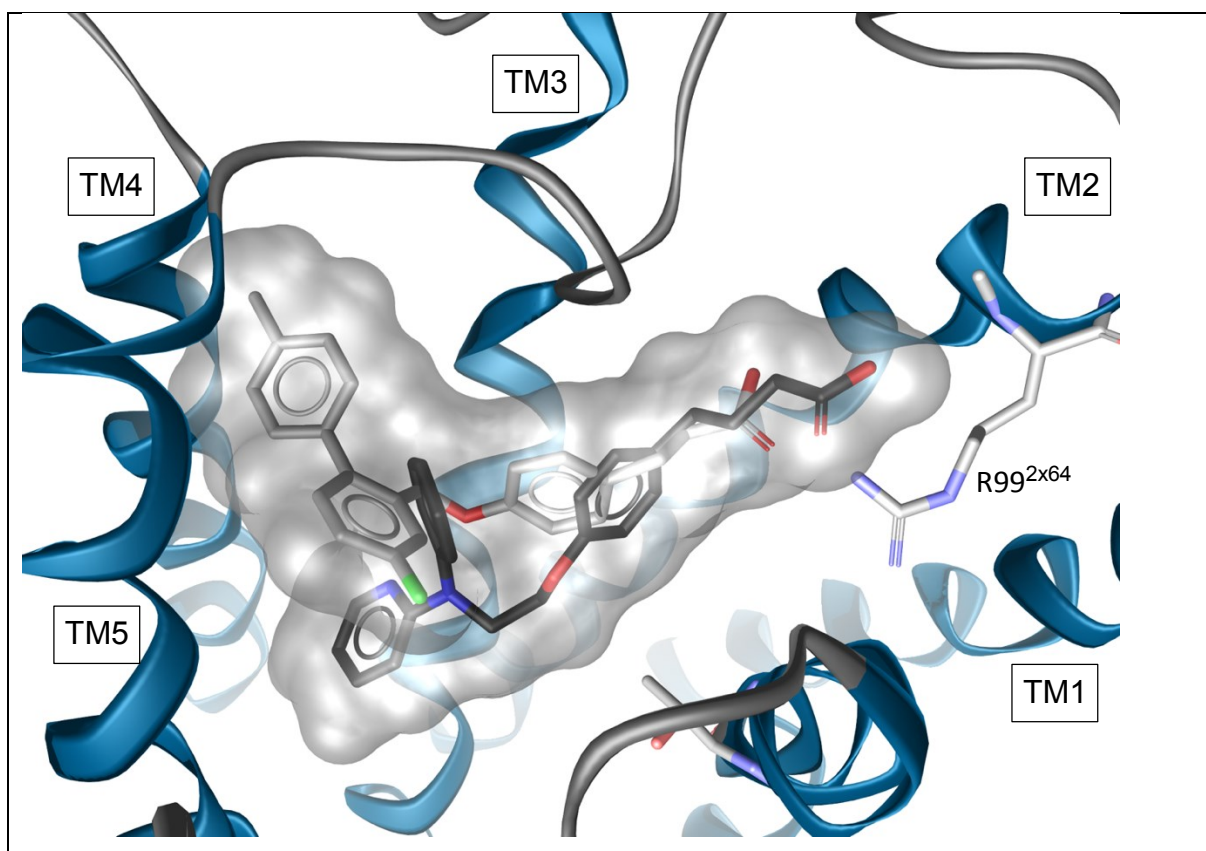

***Figure 18**: Docking of NCG21 and TUG891 at FFAR4. Receptor's TMDs are shown as blue ribbons while the aligned ligands and side chains are represented as stick model in dark grey (NCG21) and light grey (TUG891). Molecular Surface of the ligands shown in transparent grey.*

3.2.7   FFAR2 and FFAR3 interact with SCFAs and small synthetic Ligands

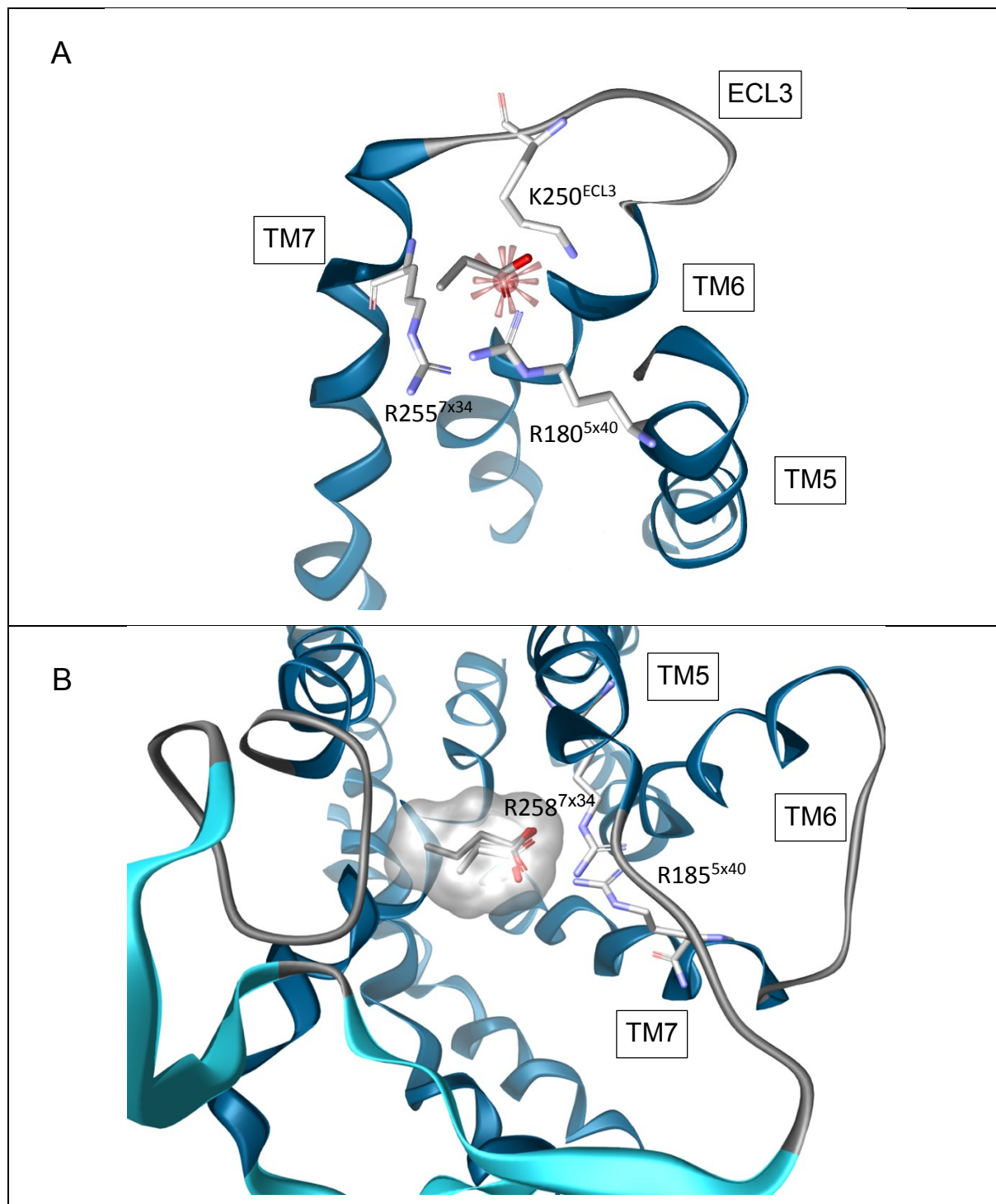*Figure 19*: *Extracellular view of SCFAs docked into FFAR2 and FFAR3. Receptor's TMDs are shown as blue ribbons while the aligned ligands and side chains are represented as stick model. Molecular Surface of the ligands shown in transparent grey. A: Docking pose of propionic acid at FFAR2. B: Alignment of docking poses of acetic acid (light grey), propionic acid (medium grey) and butyric acid (dark grey) at FFAR3.*

Due to their small size and simple structure it is no difficulty for SCFAs to enter the assumed FFAR2 binding pocket from the extracellular side and get in direct proximity to the key residues. The binding mode is as simple as the structure itself (figure 19: A). H-Bond interactions between the ligand's carboxylate and positive charged FFAR key residues ARG180^{5x40} and ARG255^{7x34} are responsible for the ligand-receptor interaction. Additionally, LYS250^{ECL3} shows an impact on the observed docking pose. Superposition of acetic, propionic and butyric acid shows that the position of their carboxylates is almost identical. Depending on the length of their aliphatic chain, the lipophilic tail interacts with the extracellular side more or less.

Docking experiments with SCFAs at FFAR3 resulted in almost identical binding poses (figure 19: B). The ligands carboxylates are placed approximate to ARG185^{5x40} and ARG258^{7x34} where they form hydrogen bonds with those positive charged residues. These residues correspond to the observed conserved key residues at FFAR1 and FFAR2. In contrast to the observed binding mode at FFAR2 there is no interaction with a residue located on ECL3, but there are lipophilic interactions between the aliphatic tail of butyric acid and PHE96^{3x33}, PHE155^{ECL2} and THR168^{ECL2}.

Synthetic FFAR3-selective ligands compound 7 and compound 12 interact with the receptor in a similar way as we observed with the endogenous ligands (figure 20). However, there are a few mentionable differences. Because of the bulkier lipophilic tails, the synthetic ligands are not able to dock into the receptor's core as deep as the SCFAs do. As a result, the distance between the ligand's carboxylate and ARG258^{7x34} is too big to form an H-bond. While compound 7 shows no interaction with ARG258^{7x34} at all, compound 12 shows at least an electrostatic interaction with the positive charged residue. Compound 7 shows lipophilic interactions between its cyclopropane and PHE96^{3x33} and TYR100^{3x37}. Lipophilic interactions at compound 12 are formed between the double-bond and PHE96^{3x33} and VAL184^{5x39}.
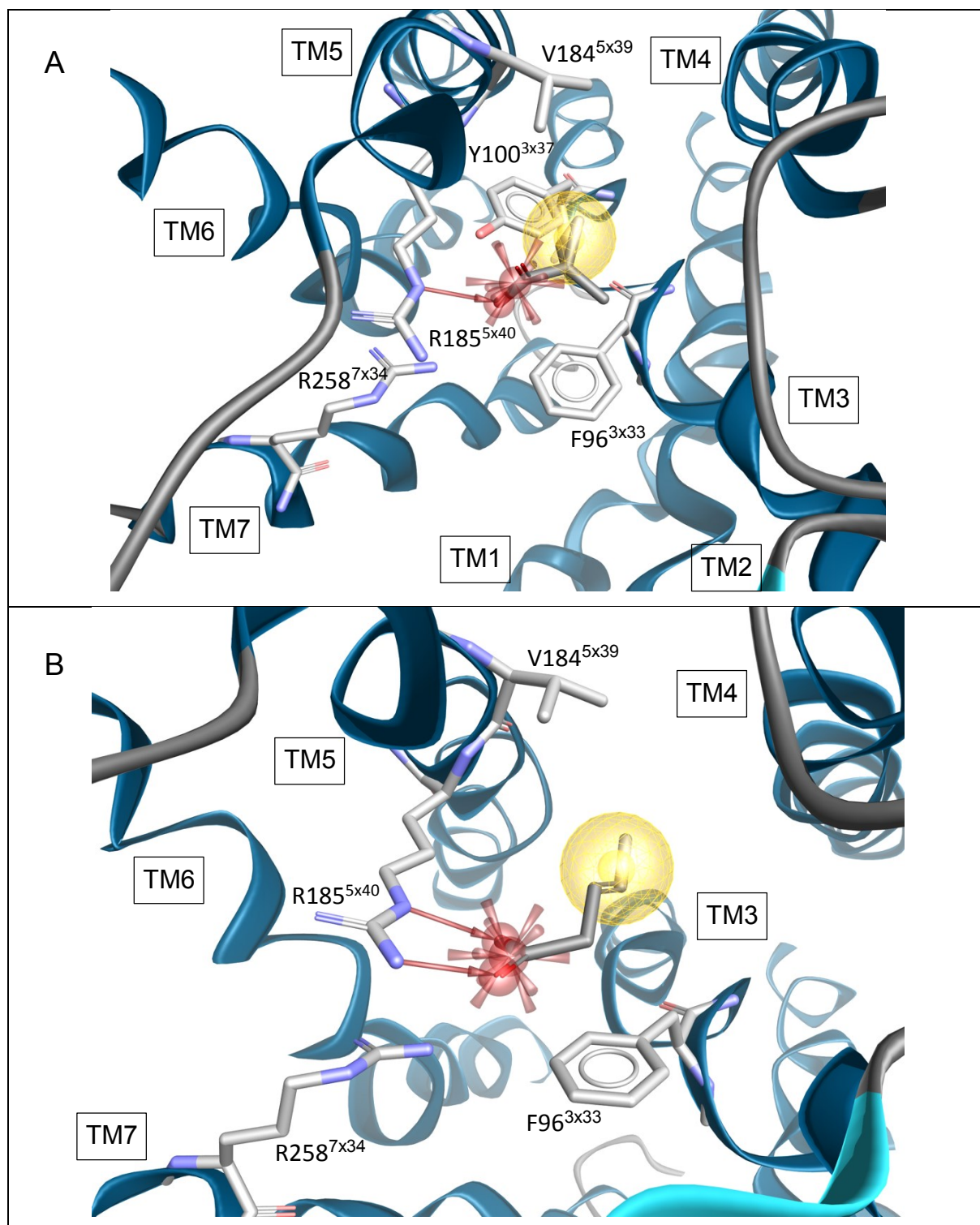
Figure 20: *Docking of synthetic ligands at FFAR3. Receptor's TMDs are shown as blue ribbons while the aligned ligands and side chains are represented as stick model. Molecular Surface of the ligands shown in transparent grey. A: Docking of compound 7. B: Docking of compound 12.*

3.2.8 FFARs are chain length selective

We started the attempt by finding an explanation for SC/LC selectivity between FFAR subtypes with the determination of all residues that are involved in the binding of LC endogenous ligand docosahexoic acid at FFAR1 and therefore form the FFAR1 binding pocket. Then we compared these residues with the corresponding conserved residues at the SC selective FFAR2.

Our earlier docking studies (section 3.2.4 / section 3.2.7) already revealed that both subtypes show identical conserved key residues for ligand binding (ARG183^{5x30} & ARG258^{7x34} at FFAR1 / ARG180^{5x30} & ARG255^{7x34} at FFAR2). When we compare the residues that form the tube shaved cavity which enables access to the core of FFAR1 (VAL81^{3x27}, ALA83^{3x29}, VAL84^{3x30}) with the corresponding residues at FFAR2 (LEU84^{3x27}, SER86^{3x29}, PHE87^{3x30}) we notice that the FFAR2 residues are bulkier than the residues at FFAR1. As a consequence, the binding pocket at FFAR2 doesn't provide as much space as the FFAR1 binding pocket, what leads to preferred binding of bigger ligands (LCFAs) at FFAR1 and of smaller ligands (SCFAs) at FFAR2. Between VAL84^{3x30} (FFAR1) and PHE87^{3x30} (FFAR2) we find the biggest difference in size. The bulky PHE87^{3x30} acts as a block that prevents the possible ligand entrance between TM3 and TM4 at FFAR2 (figure 21: A & B). After the residue comparison under the aspect of molecular size we took a look at differences in their hydrophilic/lipophilic properties. We found two polar residues at FFAR2 (SER86^{3x29}, TYR90^{3x33}) on positions, in which nonpolar residues stabilize the lipophilic tails of LCFAs at FFAR1 (figure 21: C, D). Therefore, it is obvious that binding of LCFAs to FFAR2 is less favorable due to the polar surrounding and their aliphatic tails. The described polar residues at FFAR2 don't have a negative influence on the binding of SCFAs, because their short aliphatic tails are not orientated towards the receptors core, but towards the extracellular side and the agile extracellular loop regions (see section 3.2.7).

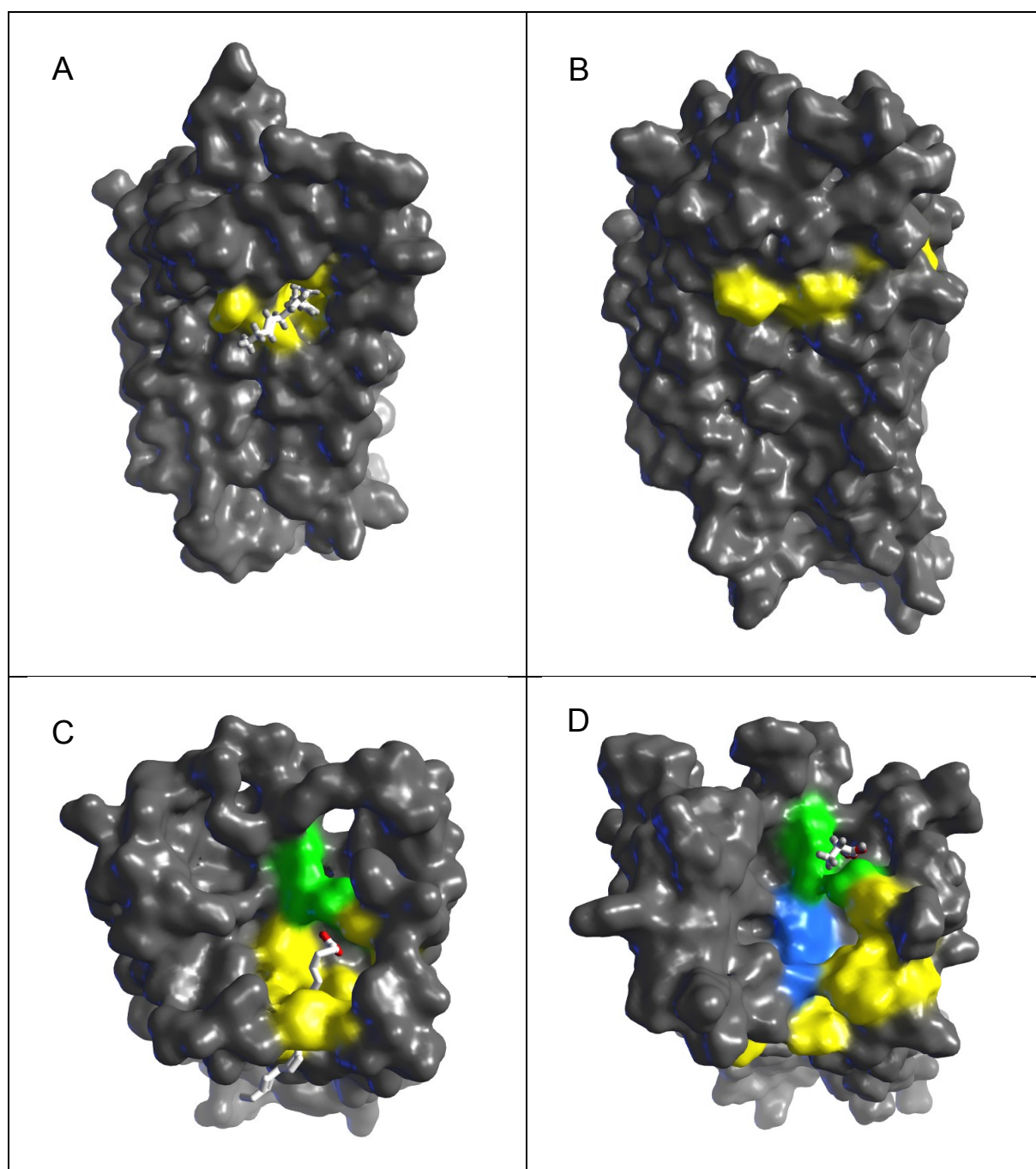*Figure 21*: *FFARs are chain length selective. Receptors molecular surface shown in grey. Residues of the binding pocket are colored by their polar properties (lipophile: yellow, hydrophile: blue, positive charged: green). Endogenous ligands are represented as stick model. A: Transmembrane view of FFAR1. B: Transmembrane view of FFAR2. C: Extracellular view of FFAR1. D: Extracellular view of FFAR2.*

3.2.9 Allosteric modulators bind adjacent to orthosteric ligands

The orthosteric binding mode for synthetic ligands at FFAR2 shows the same key residues that we observed for the binding of endogenous FFAR2 ligands (section 3.2.7). The carboxylate of Compound 1 gets anchored by ARG180^{5x40} and ARG255^{7x34}. Electrostatic interactions between the contrary charged features are responsible for the fixation of the ligand. While the phenyl ring is interacting with VAL259^{7x38} at TM7, the rest of the interactions are located on extracellular regions. The last residue that is involved in the binding mode PHE73^{ECL1}, which shows lipophilic interactions with the cyclopropyl and the thiazol feature of Compound 1 (figure 22).
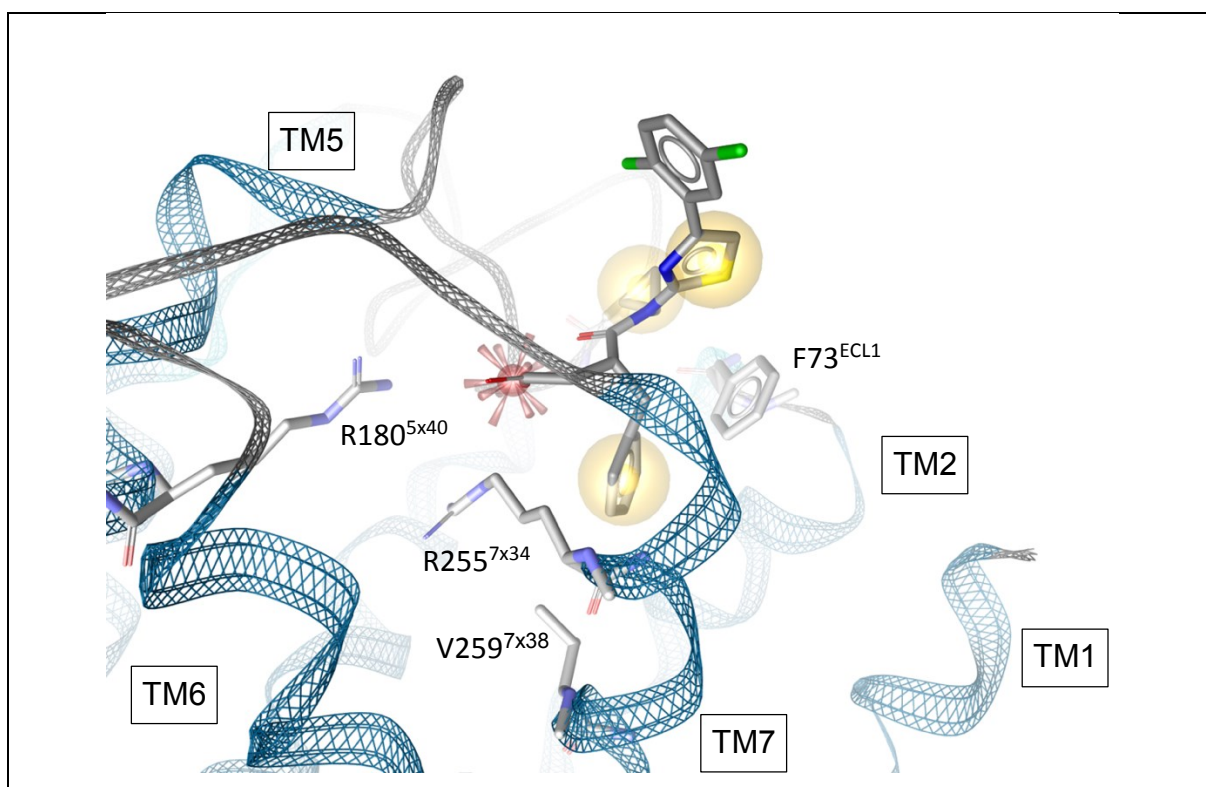
Figure 22: *Orthosteric binding of Compound1 at FFAR2. Receptor's TMDs are shown as blue snake while ligands are represented as stick model.*

Our attempts of docking orthosteric and allosteric ligands simultaneously resulted in a binding mode in which both ligands bind in a direct surrounding (figure 23: A, B). The proposed allosteric binding pocket is located next to the orthosteric binding pocket and is formed by residues that are located on the extracellular endings of TM1, TM2, TM7 and
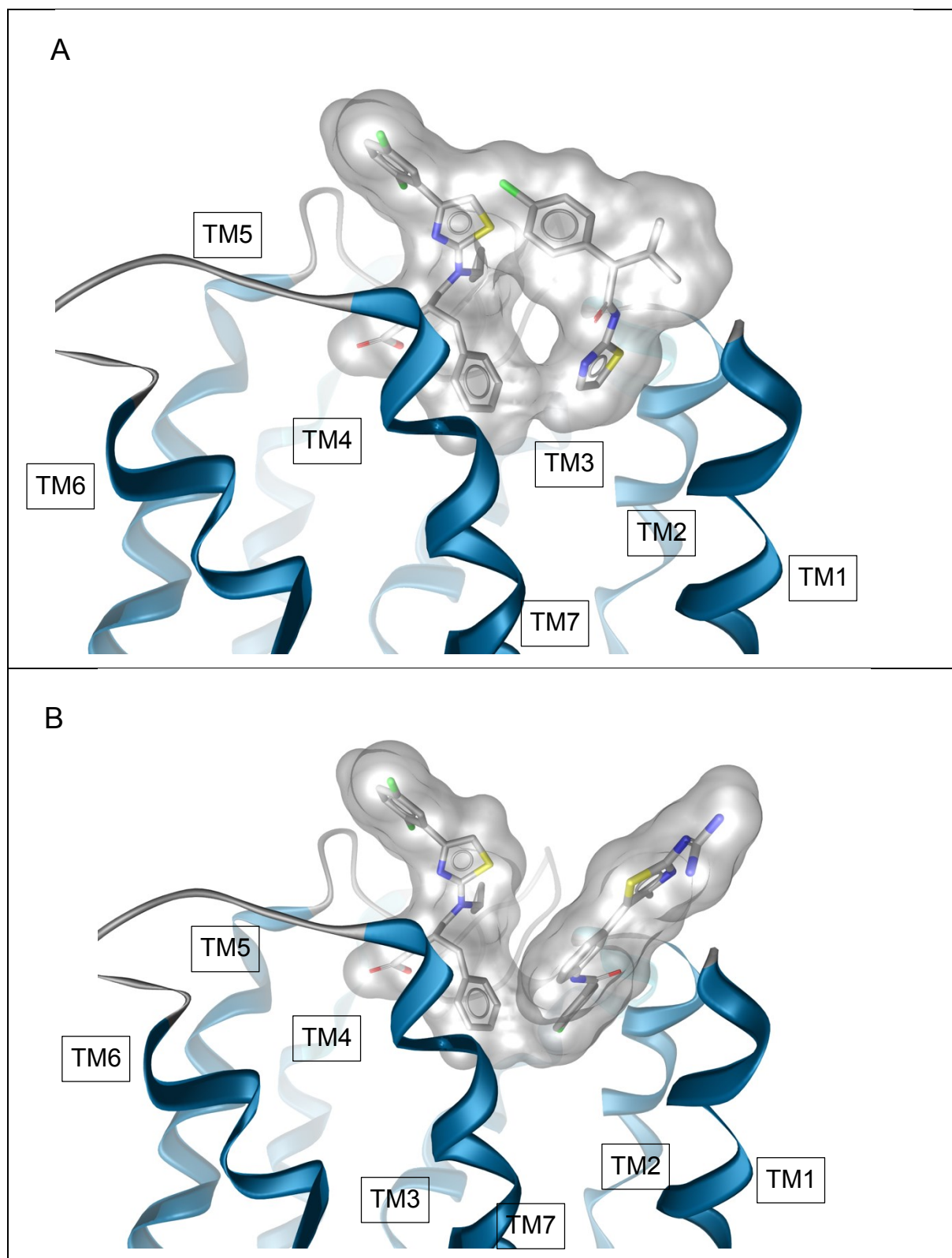
Figure 23: *Simultaneous binding of orthosteric ligands and allosteric modulators at FFAR2. Receptor's TMDs are shown as blue ribbons while ligands are represented as stick model. Molecular Surface of the ligands shown in transparent grey. A: Simultaneous binding of Compound 1 and 4CMTB. B: Simultaneous binding of Compound 1 and AZ1729.*

residues at ECL1. Although the orthosteric and allosteric ligands are located in immediate proximity, the docking shows no direct interaction between both ligands. 4CMTB docks with its thiazol feature into the receptors core, while the chloro-benzene side points towards extracellular direction. Lipophilic interactions and one ionic interaction between FFAR2 and 4CMTB are responsible for the binding mode. ILE10^{1x35} and ALA69^{2x64} interact with the ligand's thiazol. PHE73^{ECL1} and TRP253^{7x32} both show lipophilic interactions with the benzene ring, while only TRP253^{7x32} also interacts with chlorine. Besides that, there is also an ionic interaction between the negative charged ASN72^{ECL1} and the basic secondary amine of 4CMTB which has the biggest impact on the allosteric binding at FFAR2 (figure 24).
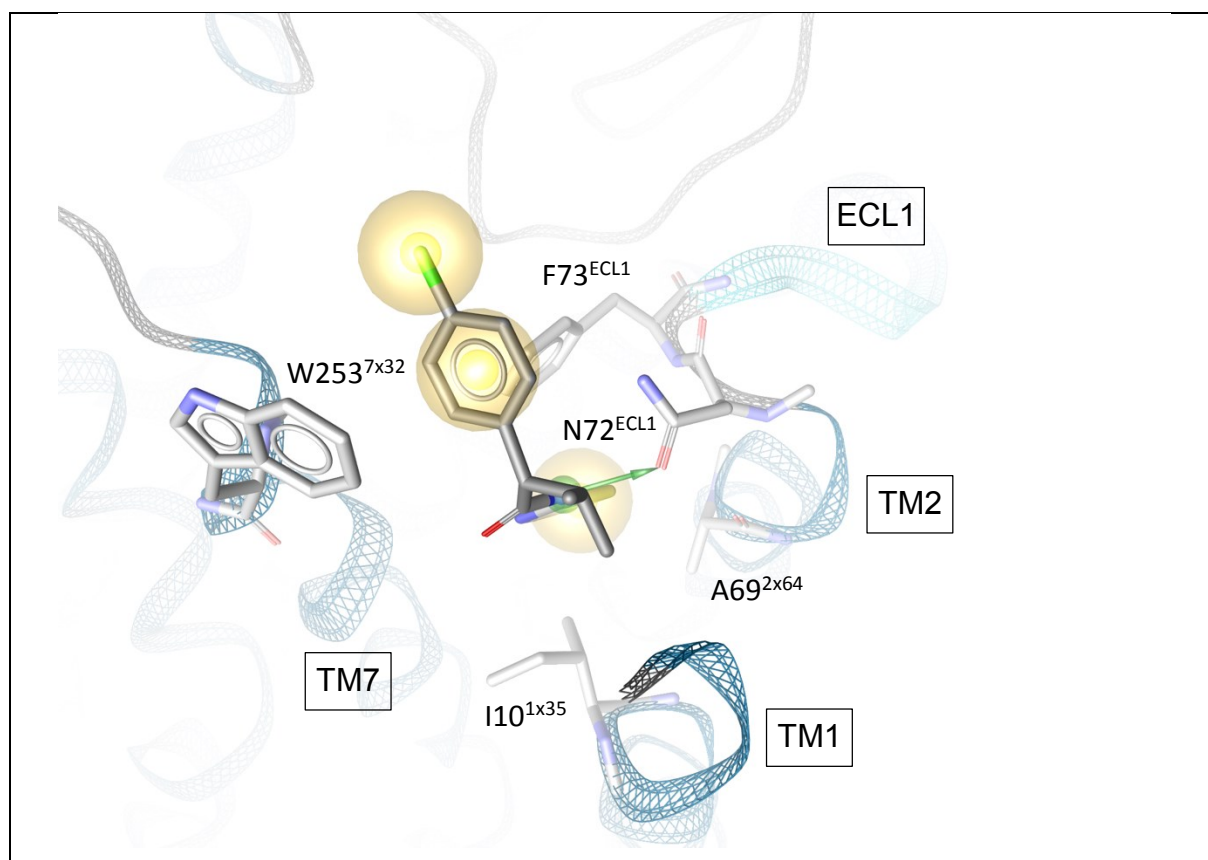

Figure 24: *Allosteric binding of 4CMTB at FFAR2. Receptor's TMDs are shown as blue snake while ligands are represented as stick model.*

Docking of AZ1729 results in a quite similar binding mode as described for 4CMTB. Four out of five residues involved in the binding of 4CMTB (ILE10[1x35], ALA69[2x64], PHE73[ECL1], TRP253[7x32]) are also responsible for the allosteric binding mode of AZ1729 at FFAR2. ILE10[1x35], ALA69[2x64] and PHE73[ECL1] interact with the lipophilic benzenes. PHE73[ECL1] not only interacts with the benzene, it also shows lipophilic interactions with fluorine. The ligand's thiazol interacts with TRP253[7x32] on TM7. In contrast to the observed binding mode of FFAR2 and 4CMTB, ASN72[ECL1] at ECL1 is not involved in the interaction of FFAR2 and AZ1729. Instead, we find TYR14[1x39] to have a big impact on the allosteric binding mode. Thy hydroxyl group of TYR14[1x39] shows ionic interactions with the ligand's fluorine, and therefore forms the strongest interaction that is involved in allosteric binding of AZ1729 at FFAR2 (figure 25).
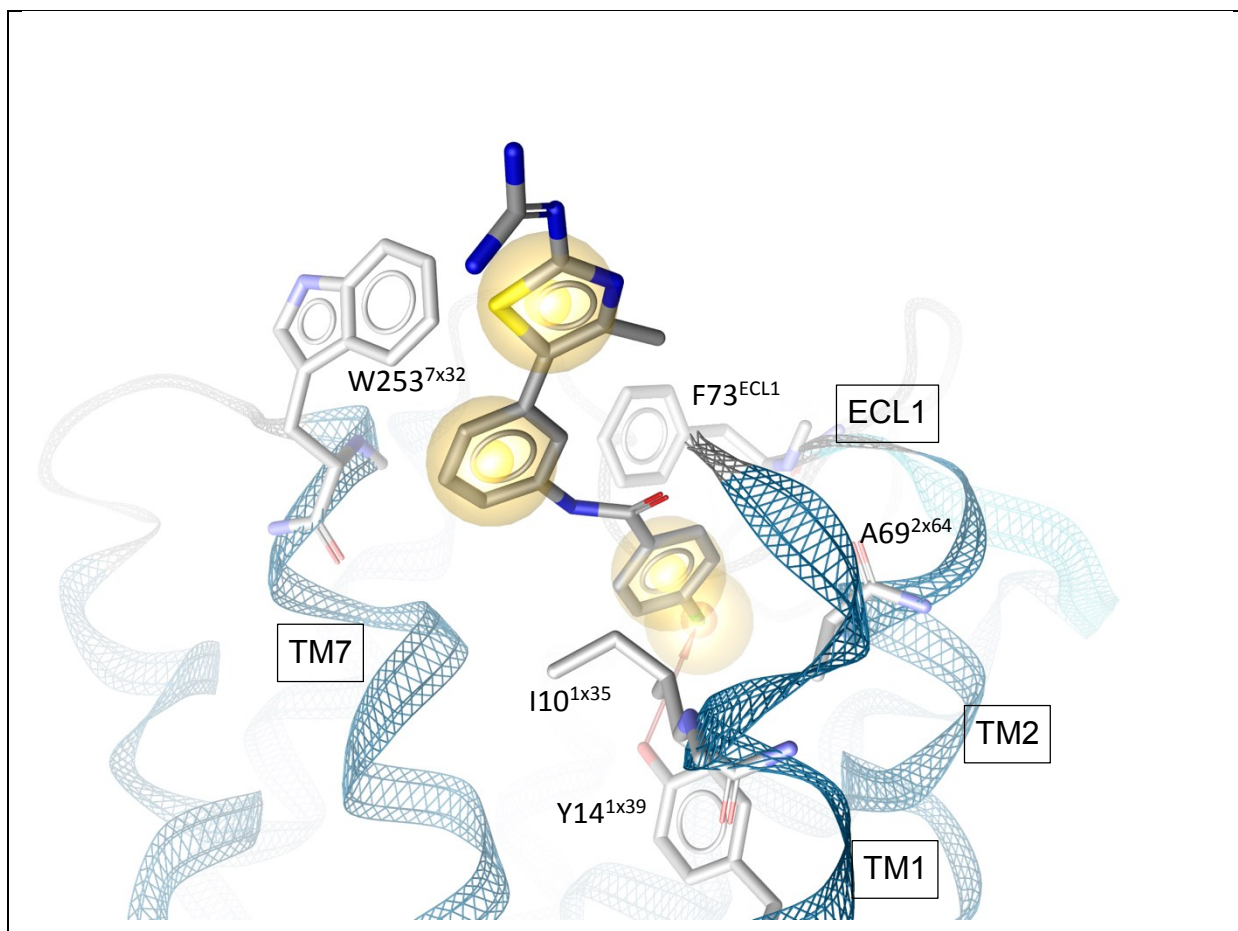

Figure 25: Allosteric binding of AZ1729 at FFAR2. Receptor's TMDs are shown as blue snake while ligands are represented as stick model.

3.3 Molecular Dynamic Simulations

After completing our docking studies, we wanted to take a deeper look on the protein-ligand interactions. Therefore, we performed a series of MD-simulations, which give us the opportunity to carry out investigations on the dynamic behavior of proteins and ligands, either alone or in complex.

3.3.1 ECL2 shows the highest variability of the FFAR1 binding pocket

With the use of the VMD-Tool we observed FFAR1 over a period of 200ns and superposed the receptor conformations after each 50ns (figure 26). This attempt served to show conformational changes of the whole receptor, while our main interest lies on the observed FFAR1 binding pocket (see section 3.2.1). At the regions where ligands get anchored by the key residues (ARG183^{5x30} and ARG258^{7x34}) we barely see any motion. The extracellular regions of TM3 and TM4, which are part of the ligand's access to the receptor's show more flexibility. ECL2 is the most variable element of the binding pocket and is directly involved in the orthosteric binding mode. Based on those observations we assume that the access of the binding pocket has an important role for ligand binding.
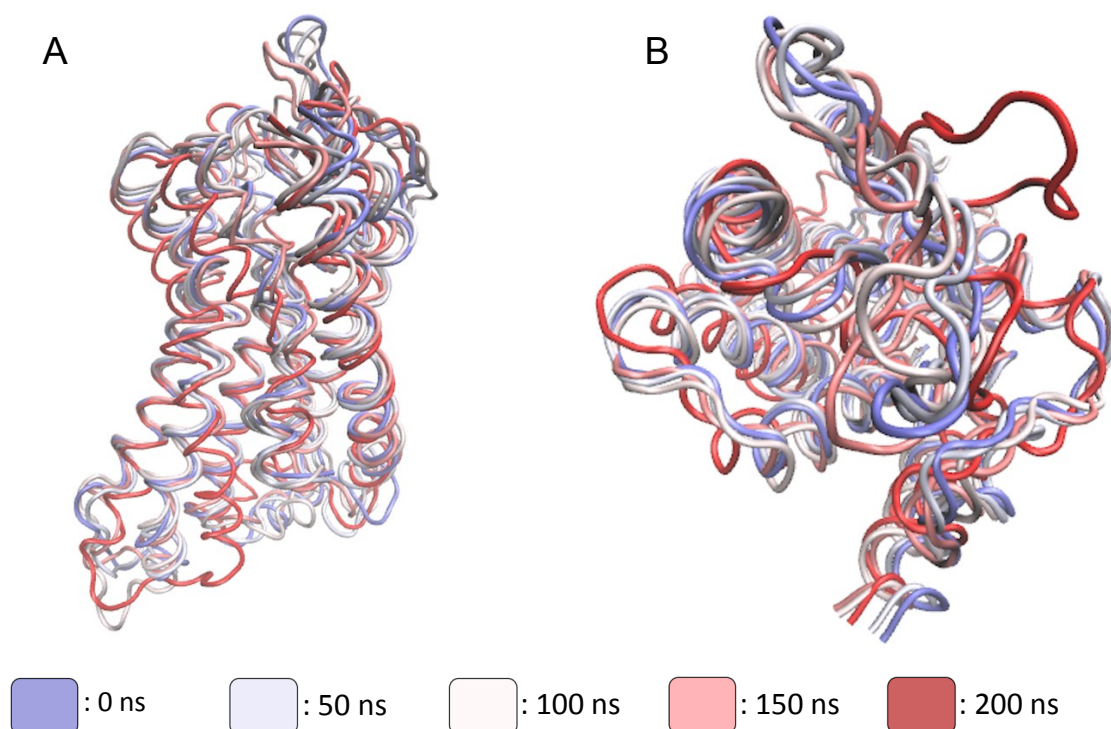

Figure 26: *MD-simulation of FFAR1. The different receptor conformations are aligned and colored according to each timestep (legend depicted). A: Transmembrane view. B: Extracellular view.*

3.3.2 Molecular dynamic simulation reveals additional features

Comparing of the static pharmacophore from only the first frame (figure 27: A) with the dynamic dynophore of all frames (figure 27: B) reveals new details about the interaction of TAK875 with FFAR1. The lipophilic features appear to be quite stable and don't show a lot of motion. Interestingly a fourth lipophilic interaction between a ligand's methyl group and PHE142[4x61] can be found at the dynophore, while it does not occur at the first frame of the MD-simulation. The barcode plot reveals that this feature occurs at 92,4% of the trajectory and therefore seems to be relevant for the binding mode. Also, other features can be observed that only occur at thy dynophore. These include an aromatic interaction between the ligand's aromatic ring and PHE142[4x61] as well as an H-bond interaction between the ligand's sulfonyl and TRP150[ECL2]. According to their occurrence over the trajectory, which is 0,1% for the H-bond feature and 0,7% for the aromatic feature, these interactions seem to be an artifact and play no important role for the binding mode.
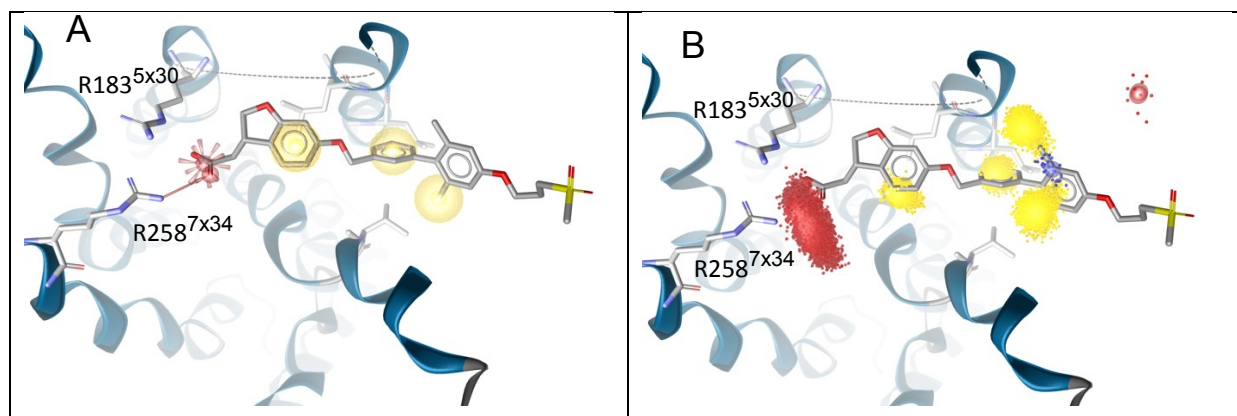

Figure 27*: Static and dynamic interactions of TAK875 at FFAR1. Receptor's TMDs are shown as blue ribbons while ligands and side chains are represented as stick model. A: Static pharmacophore. B: Dynamic dynophore.*

While the lipophilic features are almost stable, we find a lot of motion at the ligand's carboxylate, which is responsible for the anchoring of the ligand in the receptor's core (see section 3.2.1). The negative ionizable feature as well as the H-bond interaction appear as a broad cloud in the dynophore. To find out if this feature-cloud is a result of the ligand's movement alone or if the carboxylate and the according residues (ARG183[5x30] and ARG258[7x34]) move as an ensemble, we measured the interaction distance between ARG258[7x34] and the central carbon of the carboxylate. The outcome shows that the interaction switches occasionally between two different poses, where the distance is either ~ 3,9 Å or ~4,8 Å (figure 28).
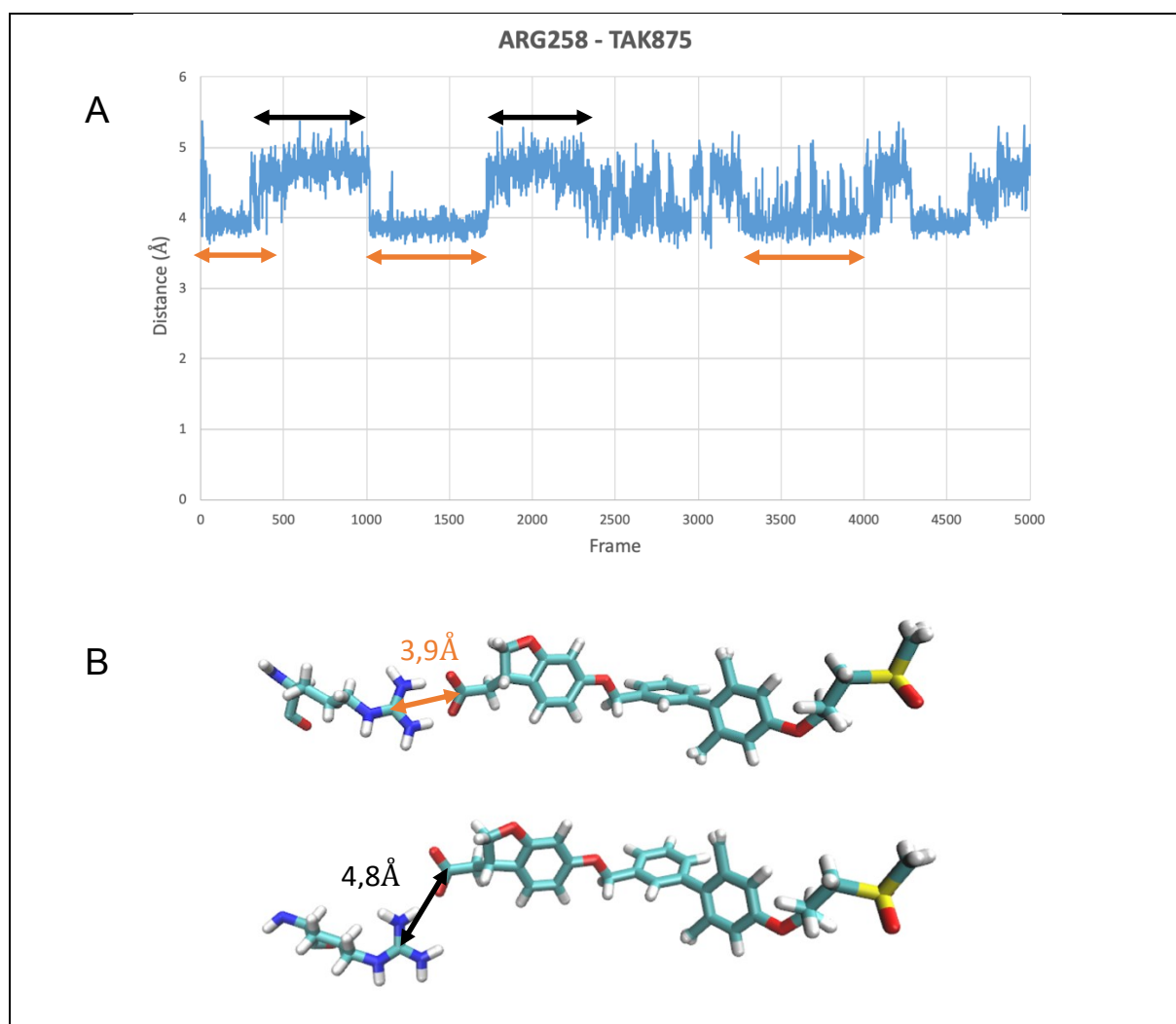

Figure 28*: Distance between ARG258[7x34] and the carboxylate of TAK875 A: Distance plot. B: Binding poses of FFAR1 and TAK875.*

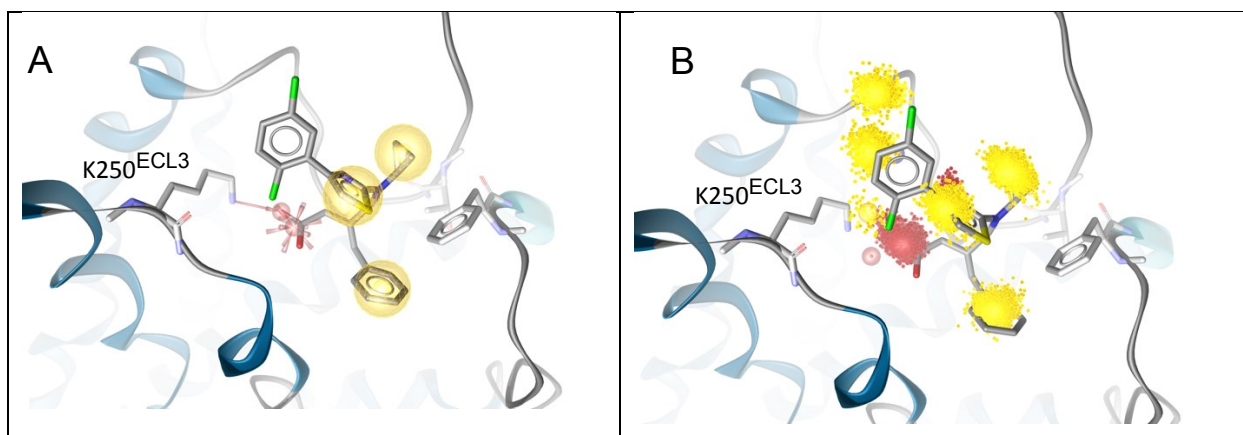### 3.3.3 Shifted dichlorophenyl enables three additional interactions

Figure 29: *Static and dynamic interactions of Compound 1 at FFAR2. Receptor's TMDs are shown as blue ribbons while ligands and side chains are represented as stick model. A: Static pharmacophore. B: Dynamic dynophore.*

Observations based on dynophores reveal that there occur more interactions between Compound 1 and FFAR2 than the static pharmacophore depicts (figure 29). Three more lipophilic interactions can be noticed at the dynophore. The additional lipophilic interactions involve the ligand's aromatic ring and its two chlorines as well as the receptor's LEU173[5x33]. Therefore, the ligand shifted its dichlorophenyl group closer to the extracellular end of TM5 and stays almost stable in this conformation. The negative ionizable feature of the ligand's carboxylate is 100% persistent and interacts with LYS250[ECL3] over the whole trajectory. LYS250[ECL3] also shows an additional H-bond interaction with the carboxylate's oxygen which only occurs during 21,1% of the MD-simulation. Other features that can be seen on the dynophore occur very rarely and are therefore not relevant for the binding of Compound 1 at FFAR2.

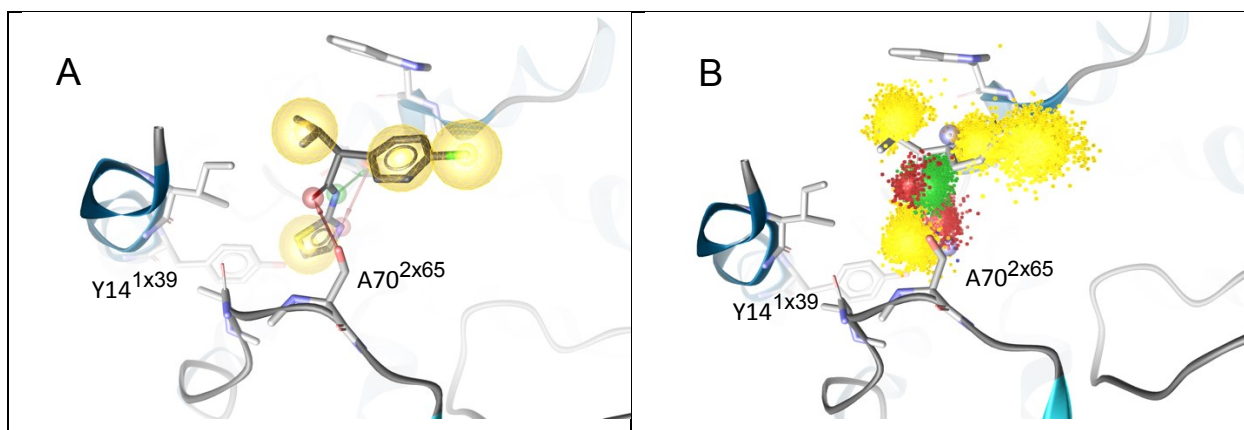### 3.3.4 Allosteric binding mode of 4CMTB is very stable

Figure 30: *Static and dynamic interactions of 4CMTB at FFAR2. Receptor's TMDs are shown as blue ribbons while ligands and side chains are represented as stick model. A: Static pharmacophore. B: Dynamic dynophore.*

The allosteric binding mode of 4CMTB at FFAR2 seems to be very stable and consistent figure 30). No additional features occur during the simulated time period of 200ns. All interactions, including four lipophilic and three H-bond interactions, stay more or less at their position. Therefore, the ligand seems to be very well fixed in its allosteric binding mode. The barcode plots reveal that more residues are involved in the binding mode than can be assumed by pharmacophore investigations. SER256^{7x35} on TM7 shows H-bond interactions during 81% of the trajectory and therefore must have an important influence on the binding mode. Additionally, TYR14^{1x39} and ALA70^{2x65} are involved in lipophilic interactions with the ligand's thiazol during the simulated time period.

4 Summary and discussion

The fact that approximately 50% of all marketed drugs target GPCRs highlights the importance of GPCRs as a drug target [5]. A deeper understanding of these complex signaling machines is crucial for related pharmaceutical research such as the designing of drugs that target this receptor type specifically. The discovery of first GPCR crystal structures enabled first insights about static conformations, while the understanding of dynamic behavior is a widely discussed topic in current GPCR research. FFARs are widespread in almost all types of body tissue and fulfill a variety of physiological functions. As a result, this receptor family is a potential target for the treatment of metabolic and inflammation related diseases, such as type 2 diabetes and cardiovascular diseases. Using an experimental setup that included computational methods such as homology modelling, molecular docking and MD-simulations, it was possible to gain a deeper insight into structural and functional properties of FFARs.

Based on an existing crystal structure of FFAR1, we were able to create models of all other FFAR subtypes. Since GPCRs share the architecture of their seven alpha-helical TM regions, those regions did not show any irregularities in all generated models. The highest sequence similarity was observed between FFAR2 and FFAR3, both subtypes that bind SCFAs as their endogenous ligands. While the TM regions showed almost no variety, we were able to observe varieties in the ECL and ICL regions. The variability of these regions increased correlating to their sequence length. Modelling of the loop regions can be a very challenging process. Different techniques, that can potentially lead to different results, have established. For the modelling of ECL2, which is involved in ligand binding and therefore very important for the experiments we conducted, we used the ECL2 region of the FFAR1 crystal structure as a template for the sequence alignment. We considered this to be the most rational method although ECL regions of FFAR2, FFAR3 and FFAR4 differ in sequence-length. After completing the models, MD-simulations of 200ns were performed on each model. The resulting RMSD plots showed an expected trend and can therefore be seen as a validation tool. The successful performance of the following experiments that were conducted can also be considered to be an overall validation.

Molecular docking experiments were performed to gain insights into receptor-ligand interaction and in order to determine key residues. The results showed that FFAR1, FFAR2 and FFAR3 own the same key residues located on positions 5x40 and 7x34. While observing the crystal structure of FFAR1 in complex with TAK875, we also found two conserved arginines at the same position on TM5 and TM7. ARG183^{5x40} and ARG258^{7x34} act as ideal binding partners for acidic functional groups and therefore can anchor potential ligands in the receptor's core. In addition to their carboxylates, FFAs have an aliphatic tail, that can be either saturated or unsaturated and vary in chain length. SCFAs can easily enter the receptor from the extracellular side and access the binding pocket. However, LCFAs are much bulkier and therefore require appropriate spaces to enter the receptor. A tube-shaped cavity between TM3 and TM4 allows the ligands to reach the key residues in the receptor's core. The lipophilic residues that shape this cave create the perfect surrounding for aliphatic chains of endogenous ligands, as well as for lipophilic tails of synthetic ligands. While FFAR1, FFAR2 and FFAR3 share the common key residues, FFAR4 does not have any corresponding residues located at 5x40 and 7x34 that could indicate a similar binding mode. We examined the receptor's amino acid sequence for residues that could potential anchor the characteristic acidic groups of its ligands. After a visual inspection, we deemed ARG99^{2x64} to be a potential key residue. The following docking experiments resulted in a plausible binding mode and revealed further details of the orthosteric binding mode at FFAR4.

One of the main objectives of this thesis was to find an explanation for the chain-length-dependent-selectivity between FFAR subtypes. We therefore compared FFAR1, which only binds LCFAs, with FFAR2, which binds SCFAs exclusively. Our experiment began with the determination of all the residues involved in ligand binding at FFAR1 and that therefore participate in forming the orthosteric FFAR1 binding pocket. Comparing of these residues with the corresponding residues at FFAR2 lead to observations that account for the preferred binding of each preferred ligand type. We were able to determine two aspects that lead to selectivity – polarity and steric conditions. Lipophilic amino acids at the FFAR1 binding pocket form an ideal surrounding for lipophilic tails of LCFAs, while polar residues on the other hand would reject aliphatic carbon chains that were too long. In addition, the FFAR2 binding pocket does not provide the necessary space for LCFAs.

Furthermore, the bulky PHE87^{3x30} acts as a block that denies the entrance of any ligand between TM3 and TM4.

It was very challenging to find a legitimate theory for the simultaneously binding of orthosteric and allosteric ligands at FFAR2, because there was no experimental data available that provided any suspicious key residues. Our series of docking experiments lead to the described binding mode where allosteric ligands bind in an extracellular vestibule next to the orthosteric binding pocket. The fact that a comparable allosteric binding mode was assumed at the human M2 muscarinic acetylcholine receptor [82] supports our theory.

While 3D-Pharmacophores only depict a static snapshot of interactions, the MD-simulation experiments we carried out gave us an insight into ligand binding in a dynamic context. Since ligand-receptor interactions are dynamic processes, conformational changes during those interactions can lead to additional features. Those features can be present during the entire process, but they can also occur with a certain occurrence frequency due to the flexibility of the protein and the ligand. The combination of an MD-simulation with distance measuring between FFAR1 key residues and TAK875's carboxylate, revealed that the ligand and ARG258^{7x34} move as an ensemble and switch collectively between two conformations.

Computational methods are a great tool to visualize molecular mechanisms and can be used very efficiently to generate scientific theories. Nonetheless, it is important to keep in mind that all the provided results have to be validated by experiments in order to exhibit a sufficient level of confidence. Therefore, mutational studies could be performed to check the importance of key residues that we identified in our docking studies. Furthermore, it would be interesting to extend our approaches with experiments that focus on the intracellular response in order to gain a deeper understanding of the complex universe of GPCR signaling. Considering the development of constantly improving methods in GPCR research, such as the structure determination by crystallization, we are confident that new and exciting outcomes will be presented soon.

5 German summary and discussion – Zusammenfassung und Diskussion

Die Tatsache, dass etwa 50% aller vermarkteten Arzneimittel an GPCRs angreifen unterstreicht den Stellenwert von GPCRs als Arzneistofftarget[5]. Ein tieferes Verständnis dieser komplexen Rezeptorengruppe ist fundamental für weiterführende pharmazeutische Forschung und das Design von GPCR-spezifischen Wirkstoffen. Neueste Kristallisationstechniken in Kombination mit modernen Strukturaufklärungsmethoden konnten bereits strukturelle Einblicke auf statischer Ebene liefern. Das dynamische Verhalten stellt ein breit diskutiertes Thema in der aktuellen GPCR-Forschung dar. FFARs befinden sich auf fast allen Körpergeweben. Da sie eine Vielzahl von physiologischen Funktionen erfüllen, stellen sie ein potentielles Target für die Behandlung von metabolischen und entzündungsbedingten Erkrankungen, wie z.B. Typ 2 Diabetes Mellitus und kardiovaskulären Erkrankungen, dar. Unser Versuchsaufbau, bestehend aus computergestützten Methoden wie Homologiemodellierung, Docking und Molekulardynamiksimulationen, ermöglichte es tiefere strukturelle und funktionelle Einblicke über FFARs zu erhalten.

Basierend auf existierenden Kristallstrukturen von FFAR1 konnten wir Modelle von allen weiteren Subtypen erstellen. Da die transmembranären Regionen von GPCRs immer als α-Helices vorliegen, konnten in diesen Bereichen keine nennenswerten Unterschiede zwischen den Modellen festgestellt werden. Die höchste Sequenz-Ähnlichkeit teilen sich FFAR2 und FFAR3 - jene Subtypen welche SCFAs als endogene Liganden binden. Während die TM-Regionen keine auffälligen Abweichungen aufweisen, konnten wir Unterschiede in den ECL und ICL Regionen feststellen. Die strukturelle Variabilität dieser Regionen nimmt in Korrelation zu ihrer Sequenzlänge zu. Das Modellieren der beweglichen Schleifenregionen stellt einen sehr herausfordernden Prozess dar. Verschiedene Techniken, welche potenziell zu unterschiedlichen Ergebnissen führen können, haben sich etabliert. Für die Modellierung von ECL2, welcher an der Bindung der Liganden beteiligt und daher für unseren Versuchsaufbau sehr wichtig ist, haben wir die ECL2-Region der FFAR1-Kristallstruktur als Vorlage für die Sequenzanordnung verwendet. Wir erachteten dies für die zweckmäßigste Methode, obwohl sich die ECL-Regionen von FFAR2, FFAR3 und FFAR4 in der Sequenzlänge unterscheiden. Nach Fertigstellung der Modelle wurden

MD-Simulationen von 200ns an jedem Modell durchgeführt. Die resultierenden RMSD-Plots zeigten einen erwarteten Trend und können daher als Validierung betrachtet werden. Die erfolgreiche Durchführung der anschließend durchgeführten Experimente kann ebenfalls als eine Gesamtvalidierung der Homolgiemodelle angesehen werden.

Docking-Experimente wurden durchgeführt, um Einblicke in Rezeptor-Ligand-Interaktionen zu gewinnen und um jene Aminosäuren zu identifizieren, die als Schlüsselrollen fungieren. Die Ergebnisse zeigten, dass FFAR1, FFAR2 und FFAR3 zwei Arginine, welche für die Ligandenbindung hauptverantwortlich sind, an den Positionen 5x40 und 7x34 besitzen. ARG183^{5x40} und ARG258^{7x34} fungieren aufgrund ihrer Basizität als ideale Interaktionspartner für saure funktionelle Gruppen und können daher potentielle Liganden im Kern des Rezeptors verankern. Neben ihren Carbonsäuren bestehen FFAs aus einem aliphatischen Schwanz, der gesättigt oder ungesättigt ist und in der Kettenlänge variieren kann. SCFAs können von der extrazellulären Seite in den Rezeptor eindringen und die Bindungstasche erreichen. LCFAs sind jedoch viel voluminöser und benötigen dementsprechende räumliche Gegebenheiten, um den Kern eines Rezeptors betreten zu können. Ein röhrenförmiger Hohlraum zwischen TM3 und TM4 ermöglicht es dem Liganden, die besagten Arginine im Kern des Rezeptors zu erreichen. Die lipophilen Reste, welche diese Höhle formen, bilden die perfekte Umgebung für aliphatische Ketten endogener Liganden sowie für lipophile Reste synthetischer Liganden. Im Gegensatz zu FFAR1-3 weist FFAR4 keine basischen Aminosäuren an den Positionen 5x40 und 7x34 auf. Daraufhin untersuchten wir die Aminosäuresequenz des Rezeptors auf Aminosäuren, welche in der Lage wären die charakteristischen Säuregruppen seiner Liganden zu verankern. Nach einer visuellen Inspektion verdächtigten wir ARG99^{2x64} am orthosterischen Bindungsmodus maßgeblich beteiligt zu sein. Die folgenden Docking-Experimente resultierten in einem plausiblen Bindungsmodus und enthüllten weitere Details des orthosterischen Bindungsmodus von FFAR4.

Eines der Hauptziele dieser Arbeit war es, eine Erklärung für die kettenlängenabhängige Liganden-Selektivität der FFAR-Subtypen zu finden. Wir verglichen daher FFAR1, welcher nur LCFAs bindet, mit FFAR2, welcher ausschließlich SCFAs bindet. Unser Experiment begann mit der Identifizierung aller Aminosäuren, welche an der Ligandenbindung an FFAR1 beteiligt und Bestandteil der orthosterischen FFAR1-Bindungstasche sind. Aus

dem Vergleich dieser Aminosäuren mit den entsprechenden Aminosäuren auf FFAR2 lassen sich Rückschlüsse ziehen, welche die Bindung des jeweils bevorzugten Liganden-Typs erklären. Wir konnten zwei Aspekte bestimmen, die zur Selektivität beitragen - Polarität und sterische Bedingungen. Lipophile Aminosäuren der FFAR1-Bindungstasche bilden eine ideale Umgebung für lipophile Schwänze von LCFAs. Polare Aminosäuren würden an dieser Position zu lange aliphatische Kohlenstoffketten ablehnen. Ferner bietet die FFAR2-Bindungstasche nicht den notwendigen Raum für LCFAs. Darüber hinaus fungiert die sperrige Seitenkette von PHE87^{3x30} als Absperrung, welche den Eintritt eines Liganden zwischen TM3 und TM4 verhindert.

Da bisher keine experimentellen Daten vorliegen die relevante Anhaltspunkte liefern, war es schwierig, eine legitime Theorie für die gleichzeitige Bindung von orthosterischen und allosterischen Liganden an FFAR2 zu finden. Unsere Serie von Docking-Experimenten führte zu dem beschriebenen Bindungsmodus, bei dem allosterische Liganden in einer extrazellulären Einbuchtung neben der orthosterischen Bindungstasche binden. Die Tatsache, dass ein vergleichbarer allosterischer Bindungsmodus bereits am humanen M2-Muskarin-Acetylcholin-Rezeptor angenommen wurde [82], unterstützt unsere Theorie.

Während 3D-Pharmacophore nur eine statische Momentaufnahme von Interaktionen darstellen, lieferten die durchgeführten MD-Simulationen nähere Erkenntnisse über die Ligandenbindung in einem dynamischen Kontext. Da Ligand-Rezeptor-Wechselwirkungen dynamische Prozesse sind, können Konformationsänderungen während der Ligandenbindung zu zusätzlich auftretenden Interaktionen führen. Diese können während des gesamten Prozesses vorhanden sein, oder aufgrund der Flexibilität des Proteins und des Liganden mit einer bestimmten Häufigkeit auftreten. Die Kopplung einer MD-Simulation mit einer Abstandsmessung zwischen den beiden Schlüsselaminosäuren ARG183^{5x40} und ARG258^{7x34} und der Säurefunktion von TAK875 zeigte, dass sich der Ligand und ARG258^{7x34} als Ensemble bewegen und gemeinsam zwischen zwei Konformationen wechseln.

Computergestützte Methoden sind ein hervorragendes Tool zur Visualisierung molekularer Mechanismen und können sehr effizient zur Erstellung wissenschaftlicher Theorien eingesetzt werden. Nichtsdestotrotz ist es wichtig sich vor Augen zu halten, dass alle gelieferten Ergebnisse durch Experimente validiert werden müssen, um ihre Aussagekraft zu bestätigen. Daher könnten im Anschluss auf diese Arbeit Mutationsstudien durchgeführt werden, um die Bedeutung jener Aminosäuren zu überprüfen, die wir in dieser Arbeit als essentiell für die Ligandenbindung identifiziert haben. Darüber hinaus wäre es interessant, unsere Ergebnisse mit Experimenten zu erweitern, welche sich auf die intrazelluläre Ebene konzentrieren, um ein tieferes Verständnis über das komplexe Universum der GPCR-Signalübertragung zu gewinnen. In Anbetracht der Entwicklung von sich ständig verbessernden Methoden in der GPCR-Forschung, wie z.B. der Strukturbestimmung durch Kristallisation, sind wir zuversichtlich, dass bald neue und aufregende Ergebnisse präsentiert werden.

6 References

1. Glukhova, A. *et al.* Rules of Engagement: GPCRs and G Proteins. *ACS Pharmacol. Transl. Sci.* (2018) doi:10.1021/acsptsci.8b00026.

2. Milligan, G., Shimpukade, B., Ulven, T. & Hudson, B. D. Complex pharmacology of free fatty acid receptors. *Chemical Reviews* (2017) doi:10.1021/acs.chemrev.6b00056.

3. Kimura, I., Ichimura, A., Ohue-Kitano, R. & Igarashi, M. Free fatty acid receptors in health and disease. *Physiol. Rev.* (2020) doi:10.1152/physrev.00041.2018.

4. Flower, D. R. Modelling G-protein-coupled receptors for drug design. *Biochimica et Biophysica Acta - Reviews on Biomembranes* (1999) doi:10.1016/S0304-4157(99)00006-4.

5. Klabunde, T. & Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* (2002) doi:10.1002/1439-7633(20021004)3:10<928::AID-CBIC928>3.0.CO;2-5.

6. Bjarnadóttir, T. K. *et al.* Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* (2006) doi:10.1016/j.ygeno.2006.04.001.

7. Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* (2003) doi:10.1124/mol.63.6.1256.

8. Kolakowski, L. F. GCRDb: A G-protein-coupled receptor database. *Receptors and Channels* (1994).

9. Davies, M. N. *et al.* On the hierarchical classification of G protein-coupled receptors. *Bioinformatics* (2007) doi:10.1093/bioinformatics/btm506.

10. Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* (2013) doi:10.1038/nature11896.

11. Hollenstein, K. *et al.* Insights into the structure of class B GPCRs. *Trends in Pharmacological Sciences* (2014) doi:10.1016/j.tips.2013.11.001.

12. Schiöth, H. B. & Fredriksson, R. The GRAFS classification system of G-protein coupled receptors in comparative perspective. in *General and Comparative*

Endocrinology (2005). doi:10.1016/j.ygcen.2004.12.018.

13. Levoye, A. *et al.* The orphan GPR50 receptor specifically inhibits MT1 melatonin receptor function through heterodimerization. *EMBO J.* (2006) doi:10.1038/sj.emboj.7601193.

14. Reiter, E., Ahn, S., Shukla, A. K. & Lefkowitz, R. J. Molecular Mechanism of β-Arrestin-Biased Agonism at Seven-Transmembrane Receptors. *Annu. Rev. Pharmacol. Toxicol.* (2012) doi:10.1146/annurev.pharmtox.010909.105800.

15. S. Flordellis, C. The Plasticity of the 7TMR Signaling Machinery and the Search for Pharmacological Selectivity. *Curr. Pharm. Des.* (2012) doi:10.2174/138161212799040556.

16. Ballesteros, J. & Palczewski, K. G protein-coupled receptor drug discovery: Implications from the crystal structure of rhodopsin. *Current Opinion in Drug Discovery and Development* (2001).

17. Ballesteros, J. *et al.* Functional Microdomains in G-protein-coupled Receptors. *J. Biol. Chem.* (1998) doi:10.1074/jbc.273.17.10445.

18. Qin, K., Dong, C., Wu, G. & Lambert, N. A. Inactive-state preassembly of Gq-coupled receptors and G q heterotrimers. *Nat. Chem. Biol.* (2011) doi:10.1038/nchembio.642.

19. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* (1995) doi:10.1016/S1043-9471(05)80049-7.

20. Ragnarsson, L., Andersson, Å., Thomas, W. G. & Lewis, R. J. Mutations in the NPxxY motif stabilize pharmacologically distinct conformational states of the α 1B - and β 2 -adrenoceptors. *Sci. Signal.* (2019) doi:10.1126/scisignal.aas9485.

21. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science (80-.).* (2000) doi:10.1126/science.289.5480.739.

22. Rasmussen, S. G. F. *et al.* Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* (2007) doi:10.1038/nature06325.

23. Shonberg, J., Kling, R. C., Gmeiner, P. & Löber, S. GPCR crystal structures: Medicinal chemistry in the pocket. *Bioorganic Med. Chem.* (2015) doi:10.1016/j.bmc.2014.12.034.

24. Rao, V. R., Cohen, G. B. & Oprian, D. D. Rhodopsin mutation G90D and a

molecular mechanism for congenital night blindness. *Nature* (1994) doi:10.1038/367639a0.

25. Bock, A., Kostenis, E., Tränkle, C., Lohse, M. J. & Mohr, K. Pilot the pulse: Controlling the multiplicity of receptor dynamics. *Trends in Pharmacological Sciences* (2014) doi:10.1016/j.tips.2014.10.002.

26. Gurevich, V. V. & Gurevich, E. V. Molecular mechanisms of GPCR signaling: A structural perspective. *International Journal of Molecular Sciences* (2017) doi:10.3390/ijms18122519.

27. Conklin, B. R. & Bourne, H. R. Structural elements of G alpha subunits that interact with G beta gamma, receptors, and effectors. *Cell* **73**, 631–641 (1993).

28. Rens-Domiano, S. & Hamm, H. E. Structural and functional relationships of heterotrimeric G-proteins. *FASEB J.* (1995) doi:10.1096/fasebj.9.11.7649405.

29. Cabrera-Vera, T. M. *et al.* Insights into G Protein Structure, Function, and Regulation. *Endocrine Reviews* (2003) doi:10.1210/er.2000-0026.

30. Clapham, D. E. & Neer, E. J. G protein beta gamma subunits. *Annu. Rev. Pharmacol. Toxicol.* (1997) doi:10.1146/annurev.pharmtox.37.1.167.

31. Urban, J. D. *et al.* Functional selectivity and classical concepts of quantitative pharmacology. *Journal of Pharmacology and Experimental Therapeutics* (2007) doi:10.1124/jpet.106.104463.

32. Shukla, A. K., Xiao, K. & Lefkowitz, R. J. Emerging paradigms of β-arrestin-dependent seven transmembrane receptor signaling. *Trends in Biochemical Sciences* (2011) doi:10.1016/j.tibs.2011.06.003.

33. Correll, C. C. & McKittrick, B. A. Biased ligand modulation of seven transmembrane receptors (7TMRs): Functional implications for drug discovery. *Journal of Medicinal Chemistry* (2014) doi:10.1021/jm401677g.

34. Cifuentes, A. *Foodomics: Advanced Mass Spectrometry in Modern Food Science and Nutrition. Foodomics: Advanced Mass Spectrometry in Modern Food Science and Nutrition* (2013). doi:10.1002/9781118537282.

35. Jameson, E. & Walter, J. H. Medium-chain acyl-CoA dehydrogenase deficiency. *Paediatr. Child Health (Oxford).* **29**, 123–126 (2019).

36. Beermann, C. *et al.* Short term effects of dietary medium-chain fatty acids and n-3 long-chain polyunsaturated fatty acids on the fat metabolism of healthy volunteers. *Lipids Health Dis.* (2003) doi:10.1186/1476-511X-2-10.

37. Zechner, R., Strauss, J. G., Haemmerle, G., Lass, A. & Zimmermann, R. Lipolysis: Pathway under construction. *Current Opinion in Lipidology* (2005) doi:10.1097/01.mol.0000169354.20395.1c.

38. Hillgartner, F. B., Salati, L. M. & Goodridge, A. G. Physiological and molecular mechanisms involved in nutritional regulation of fatty acid synthesis. *Physiological Reviews* (1995) doi:10.1152/physrev.1995.75.1.47.

39. Hara, T., Kimura, I., Inoue, D., Ichimura, A. & Hirasawa, A. Free fatty acid receptors and their role in regulation of energy metabolism. *Rev. Physiol. Biochem. Pharmacol.* (2013) doi:10.1007/112_2013_13.

40. Kahn, S. E., Hull, R. L. & Utzschneider, K. M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* (2006) doi:10.1038/nature05482.

41. Sanz, Y., Santacruz, A. & Gauffin, P. Gut microbiota in obesity and metabolic disorders. in *Proceedings of the Nutrition Society* (2010). doi:10.1017/S0029665110001813.

42. Briscoe, C. *et al.* Free fatty acid receptors (version 2019.4) in the IUPHAR/BPS Guide to Pharmacology Database. *IUPHAR/BPS Guid. to Pharmacol. CITE* **2019**, (2019).

43. Brown, A. J. *et al.* The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J. Biol. Chem.* (2003) doi:10.1074/jbc.M211609200.

44. Le Poul, E. *et al.* Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. *J. Biol. Chem.* (2003) doi:10.1074/jbc.M301403200.

45. Offermanns, S. Free Fatty Acid (FFA) and Hydroxy Carboxylic Acid (HCA) Receptors. *Annu. Rev. Pharmacol. Toxicol.* (2014) doi:10.1146/annurev-pharmtox-011613-135945.

46. Srivastava, A. *et al.* High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* (2014) doi:10.1038/nature13494.

47. Negoro, N. *et al.* Discovery of TAK-875: A potent, selective, and orally bioavailable GPR40 agonist. *ACS Med. Chem. Lett.* (2010) doi:10.1021/ml1000855.

48. Ho, J. D. *et al.* Structural basis for GPR40 allosteric agonism and incretin stimulation. *Nat. Commun.* (2018) doi:10.1038/s41467-017-01240-w.

49. Lu, J. *et al.* Structural basis for the cooperative allosteric activation of the free fatty

acid receptor GPR40. *Nat. Struct. Mol. Biol.* (2017) doi:10.1038/nsmb.3417.

50. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* (1986) doi:10.1002/j.1460-2075.1986.tb04288.x.

51. Chung, S. Y. & Subbiah, S. A structural explanation for the twilight zone of protein sequence homology. *Structure* (1996) doi:10.1016/S0969-2126(96)00119-0.

52. Costanzi, S. Homology modeling of class a g protein-coupled receptors. *Methods Mol. Biol.* (2012) doi:10.1007/978-1-61779-588-6_11.

53. Saxena, A., Sangwan, R. S. & Mishra, S. Fundamentals of Homology Modeling Steps and Comparison among Important Bioinformatics Tools: An Overview. *Sci. Int.* (2013) doi:10.17311/sciintl.2013.237.252.

54. Guedes, I. A., de Magalhães, C. S. & Dardenne, L. E. Receptor-ligand molecular docking. *Biophysical Reviews* (2014) doi:10.1007/s12551-013-0130-2.

55. Lensink, M. F., Méndez, R. & Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd Edition. in *Proteins: Structure, Function and Genetics* (2007). doi:10.1002/prot.21804.

56. Wolber, G. & Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* (2005) doi:10.1021/ci049885e.

57. Wolber, G., Dornhofer, A. A. & Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput. Aided. Mol. Des.* (2006) doi:10.1007/s10822-006-9078-7.

58. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational methods in drug discovery. *Pharmacological Reviews* (2014) doi:10.1124/pr.112.007336.

59. Bermudez, M. & Wolber, G. Structure versus function - The impact of computational methods on the discovery of specific GPCR-ligands. *Bioorganic Med. Chem.* (2015) doi:10.1016/j.bmc.2015.03.026.

60. Case, D. A. *et al.* The Amber biomolecular simulation programs. *Journal of Computational Chemistry* (2005) doi:10.1002/jcc.20290.

61. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* (1995) doi:10.1016/0010-4655(95)00042-E.

62. Bowers, K. J. *et al.* Scalable algorithms for molecular dynamics simulations on commodity clusters. in *Proceedings of the 2006 ACM/IEEE Conference on*

Supercomputing, SC'06 (2006). doi:10.1145/1188455.1188544.

63. Schrödinger Release 2017-3. Maestro. (2017).

64. Bateman, A. *et al.* UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkw1099.

65. Wermuth, C. G., Ganellin, C. R., Lindberg, P. & Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* (1998) doi:10.1351/pac199870051129.

66. Dominique S. Dynophores: Novel Dynamic Pharmacophores. Implementation of pharmacophore generation based on molecular dynamics trajectories and their graphical representation. (Humbold University Berlin).

67. Bock, A. *et al.* Ligand binding ensembles determine graded agonist efficacies at a G protein-coupled receptor. *J. Biol. Chem.* (2016) doi:10.1074/jbc.M116.735431.

68. Hudson, B. D. *et al.* The pharmacology of TUG-891, a potent and selective agonist of the free fatty acid receptor 4 (FFA4/GPR120), demonstrates both potential opportunity and possible challenges to therapeutic agonism. *Mol. Pharmacol.* (2013) doi:10.1124/mol.113.087783.

69. Watson, S. J., Brown, A. J. H. & Holliday, N. D. Differential signaling by splice variants of the human free fatty acid receptor GPR120. *Mol. Pharmacol.* (2012) doi:10.1124/mol.111.077388.

70. Hudson, B. D., Shimpukade, B., Milligan, G. & Ulven, T. The molecular basis of ligand interaction at free fatty acid receptor 4 (FFA4/GPR120). *J. Biol. Chem.* (2014) doi:10.1074/jbc.M114.561449.

71. Garrido, D. M. *et al.* Synthesis and activity of small molecule GPR40 agonists. *Bioorganic Med. Chem. Lett.* (2006) doi:10.1016/j.bmcl.2006.01.007.

72. Briscoe, C. P. *et al.* Pharmacological regulation of insulin secretion in MIN6 cells through the fatty acid receptor GPR40: Identification of agonist and antagonist small molecules. *Br. J. Pharmacol.* (2006) doi:10.1038/sj.bjp.0706770.

73. Kotarsky, K., Nilsson, N. E., Flodgren, E., Owman, C. & Olde, B. A human cell surface receptor activated by free fatty acids and thiazolidinedione drugs. *Biochem. Biophys. Res. Commun.* (2003) doi:10.1016/S0006-291X(02)03064-4.

74. Christiansen, E. *et al.* Discovery of potent and selective agonists for the free fatty acid receptor 1 (FFA1/GPR40), a potential target for the treatment of type II diabetes. *J. Med. Chem.* (2008) doi:10.1021/jm8010178.

75. Christiansen, E. *et al.* Discovery of TUG-770: A highly potent free fatty acid receptor 1 (FFA1/GPR40) agonist for treatment of type 2 diabetes. *ACS Med. Chem. Lett.* (2013) doi:10.1021/ml4000673.

76. Suzuki, T. *et al.* Identification of G protein-coupled receptor 120-selective agonists derived from PPARγ agonists. *J. Med. Chem.* (2008) doi:10.1021/jm800970b.

77. Shimpukade, B., Hudson, B. D., Hovgaard, C. K., Milligan, G. & Ulven, T. Discovery of a potent and selective GPR120 agonist. *J. Med. Chem.* (2012) doi:10.1021/jm300215x.

78. Calder, P. C. Marine omega-3 fatty acids and inflammatory processes: Effects, mechanisms and clinical relevance. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* (2015) doi:10.1016/j.bbalip.2014.08.010.

79. Brown, A. J. *et al.* Pharmacological properties of acid N-thiazolylamide FFA2 agonists. *Pharmacol. Res. Perspect.* (2015) doi:10.1002/prp2.141.

80. Bolognini, D. *et al.* A novel allosteric activator of free fatty acid 2 receptor displays unique Gi-functional bias. *J. Biol. Chem.* (2016) doi:10.1074/jbc.M116.736157.

81. Schmidt, J. *et al.* Selective orthosteric free fatty acid receptor 2 (FFA2) agonists: Identification of the structural and chemical requirements for selective activation of FFA2 versus FFA3. *J. Biol. Chem.* (2011) doi:10.1074/jbc.M110.210872.

82. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* (2013) doi:10.1038/nature12735.

7   List of Figures