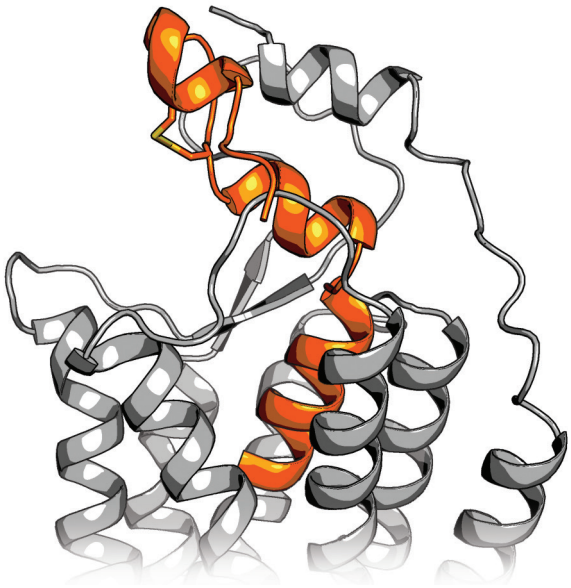
DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

67/2018

LASSE KARHU

Computational Analysis of Orexin Receptors and Their Interactions with Natural and Synthetic Ligands



DIVISION OF PHARMACEUTICAL CHEMISTRY AND TECHNOLOGY
FACULTY OF PHARMACY
DOCTORAL PROGRAMME IN DRUG RESEARCH
UNIVERSITY OF HELSINKI

Division of Pharmaceutical Chemistry and Technology

Faculty of Pharmacy

University of Helsinki

Computational analysis of orexin receptors and their interactions with natural and synthetic ligands

Lasse Karhu

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Pharmacy, University of Helsinki, for public examination in the Auditorium 1041, Biocenter 2 in Viikki, on 26th of October 2018, at 12 noon.

Helsinki 2018

Supervisors Docent Henri Xhaard, Ph.D.
Division of Pharmaceutical Chemistry and Technology
Faculty of Pharmacy
University of Helsinki, Finland

Docent Alex Bunker, Ph.D.
Division of Pharmaceutical Biosciences
Faculty of Pharmacy
University of Helsinki, Finland

Reviewers Associate Professor Jana Selent, Ph.D.
Hospital del Mar Medical Research Institute
Pompeu Fabra University
Barcelona, Spain

Docent Hugo Gutiérrez-de-Terán, Ph.D.
Department of Cell and Molecular Biology
Uppsala University, Sweden

Opponent Professor Peter Kolb, Ph.D.
Department of Pharmaceutical Chemistry
Philipps-University Marburg, Germany

ISBN 978-951-51-4576-5 (paperback)
ISBN 978-951-51-4577-2 (PDF, http://ethesis.helsinki.fi)

Published in the DSHealth series *Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis*
ISSN 2342-3161 (print)
ISSN 2342-317X (online)

Hansaprint, Vantaa, Finland, 2018

Abstract

The orexinergic system is a key regulator of the sleep-wake cycle, and as such, presents a prominent target for drug development against ailments such as insomnia and narcolepsy. The system comprises two G protein-coupled receptors (GPCR), OX_1 and OX_2, and two neuropeptides, orexin-A and orexin-B. In the beginning of the study presented here, several antagonists (blockers) of the receptors were available but drug-like agonists (activators) were not. The search for the latter was hampered by the poor understanding how the endogenous ligands, the orexin peptides, activate their receptors.

The main objective for the thesis research was to elucidate the binding mode of orexin peptides at their cognate receptors, along with the activation determinants, using both computational and traditional experimental methods.

We produced homology models for the OX_1 receptor based on related GPCRs, and subsequently adopted the orexin receptor structures reported during the study. Peptide binding mode was probed through rigid-body docking, which resulted in two alternative binding modes. These were followed up by extensive molecular dynamics simulations within membrane environment, accompanied with simulations of small-molecule binding. Deriving from the simulations, we proposed a single, well-defined binding mode for the orexin peptide C-terminus within the canonical GPCR binding site. In addition, we observed that the small-molecular antagonist was remarkably stable within the binding site, whereas the recently reported agonist Nag26 was more mobile. The pool of simulations allowed us to observe differences between the agonists and the antagonist, leading to suggestions on determinants of agonist and antagonist binding.

To assess the bioactive conformation of orexin peptides, we produced conformationally constrained orexin peptide variants. These showed that the stabilization of the straight α-helical conformation of the orexin-A is detrimental to potency, but not necessary to efficacy, at least with the utilized stapled peptides and α-aminoisobutyric acid insertions at the tested sites. We assume that the modifications were directly incompatible with the binding interactions, or the stabilized conformation was sub-optimal.

The literature review focuses on the functions and characteristics of GPCRs and the orexinergic system, and provides insight into computational tools used in the study.

Tiivistelmä

Oreksiinijärjestelmä on tärkeä uni–valverytmin säätelijä ja näin ollen kiinnostava lääkekehityskohde muun muassa nukahtamis- ja narkolepsialääkkeille. Järjestelmään kuuluu kaksi G-proteiinikytkentäistä reseptoria (GPCR), OX_1 ja OX_2, sekä kaksi neuropeptidiä, oreksiini-A ja -B. Väitöskirjassa esitetyn tutkimuksen alussa tunnettiin useita reseptoriantagonisteja (salpaajia), mutta lääkkeenkaltaisia agonisteja (aktivaattoreita) ei tunnettu. Endogeenisten agonistien, eli oreksiini-peptidien, sitoutumistapaa reseptoreihin ei tunnettu, mikä haittasi etenkin agonistien etsintää.

Väitöstutkimuksen päätavoite oli selvittää oreksiinipeptidien sitoutumistapa oreksiinireseptoreihin ja reseptorin aktivoitumiselle oleelliset vuorovaikutukset. Tähän käytettiin sekä tietokoneavusteisia että perinteisiä kokeellisia menetelmiä.

Tuotimme OX_1 reseptorista malleja ensin läheisiin reseptorirakenteisiin pohjautuen, ja kun oreksiinireseptorien rakenteita julkaistiin, hyödynsimme niitä. Peptidien sitoutumista selvitettiin telakoimalla. Tämän pohjalta esitimme kaksi vaihtoehtoista sitoutumistapaa, joille ajoimme kattavia molekyylidynaamisia simulaatioita. Lisäksi simuloimme pienmolekyylien sitoutumista. Simulaatioiden pohjalta esitimme oreksiinipeptidien C-terminaalille yhden tarkasti muotoillun sitoutumistavan reseptorin sitoutumistaskuun. Simulaatiot osoittivat myös, että antagonist suvoreksantin sitoutuminen on hyvin vakaata, mikä puolestaan ei pitänyt paikkaansa hiljattain julkaistun pienmolekyyliagonistin Nag26 kohdalla. Simulaatioiden vertailun pohjalta ehdotimme reseptorin sammuttamisen ja aktivoimisen kannalta oleellisia vuorovaikutuksia.

Lisäksi tuotimme konformaatiostabiloituja oreksiinipeptideitä tutkiaksemme peptidien bioaktiivista konformaatiota. Oreksiini-A:n suoran α-heliksirakenteen vakauttaminen, ainakin käyttämillämme menetelmillä, alentaa peptidien voimakkuutta, muttei välttämättä tehokkuutta. Oletamme, että tuottamamme muutokset ovat suoraan yhteensopimattomia oleellisten vuorovaikutusten kanssa tai vaihtoehtoisesti vakautettu konformaatio on epäsopiva.

Kirjallisuuskatsaus käsittelee G-proteiinikytkentäisten reseptorien ja oreksiini-järjestelmän ominaisuuksia ja toimintaa sekä esittelee tutkimuksessa käytetyt tietokoneavusteiset menetelmät.

Acknowledgements

While there is only one name in the cover of the book, many more have contributed towards its completion. Here, I would like thanks those people and organizations.

First and foremost, I will express my gratitude to my supervisors Dr. Henri Xhaard and Dr. Alex Bunker. Henri has been guiding my work for eight years, starting with my Master's thesis and culminating here to my doctoral thesis. I thank you for the countless discussions and rounds of comments, especially the legible ones! While at times frustrating, your drive towards perfection has left a clear mark on my work. I thank Alex for opening the black box of computational tools for me, and for being there in times of dire need, providing a new pathway for my research.

I am grateful for my custos Prof. Jari Yli-Kauhaluoma, who opened the first door on my journey to become a scientist. Your enthusiasm towards pharmaceutical chemistry is a beacon capable of illuminating even the darkest of days.

I offer my thanks to my pre-examiners, Prof. Jana Selent and Dr. Gutiérrez-de-Terán, for finding time to review my thesis and for their kind and motivating comments. I am also thankful for Prof. Peter Kolb for agreeing to participate in the dissertation as an opponent, and for running the Glisten network for European GPCR researchers.

I wish to thank Prof. Jyrki Kukkonen, Dr. Erik Wallén, Dr. Teppo Leino and Dr. Aniket Magarkar for scientific collaboration. Jyrki deserves a special note for his direct communication, which is quite refreshing after the initial shock, and of course for his expertise within the field of cell-based assays. Erik has been the driving force concerning the stapled peptides, and his visits dealing with all aspects of life have been welcome breaks from the computer screen. Teppo has brought his expertise and compounds to the project, which has pushed the research project into novel paths. Aniket has guided me into the world of MD simulations, and been there for me every time I got lost in the parameters and settings.

I will offer a special thanks to Dr. Ainoleena Turku, both for her scientific collaboration and personal support through my studies. I believe it is not common for a tutor to stick around for 12 years to help her tutee. Perhaps there was something in the first few weeks that made her feel indebted to me.

I thank the CDD research group, both current and former members, for all the discussions, breaks, conference trips, barbeque events, sauna and ice swimming

excursions etc. You were always there to turn to when facing problems or to toast with in times of success (or if it was not November – or especially if it was).

Concerning practical matters, I am grateful for the University of Helsinki Research Foundation, the Finnish Cultural Foundation, the Finnish Pharmaceutical Society, the Orion Research Foundation and the doctoral programs FinPharma Doctoral Program and Doctoral Programme in Drug Research for funding my research. I will offer the most sincere thanks to education planning officer Marjukka Laakso for her flexibility concerning certain paperwork and deadlines.

My family has been an inexhaustible source for support and repetitive jokes that have carried me through my doctoral studies. I am grateful to my parents for creating an environment that has supported learning since the first grade, for my twin brother for sparring throughout school and for my older brothers for setting up examples.

Finally, I will express my deepest gratitude to my wife Elina for her love and support, and to my daughter Emma who enlightens my days and reminds me of what is truly important.

Helsinki, October 2018

Lasse Karhu

Table of Contents

List of original publications

This dissertation is based on the following publications referred to in the text by the Roman numerals I–III.

I **Karhu L**, Turku A, Xhaard H: Modeling of the OX_1R–orexin-A complex suggests two alternative binding modes. *BMC Structural Biology* **2015**, 15:9

II **Karhu L**, Weisell J, Turunen PM, Leino TO, Pätsi H, Xhaard H, Kukkonen JP, Wallén EAA: Stapled truncated orexin peptides as orexin receptor agonists. *Peptides* **2018**, 102: 54-60

III **Karhu L**, Magarkar A, Bunker A, Xhaard H: Determinants of orexin receptor binding and activation — a molecular dynamics study. *Manuscript*

Personal contributions

I Study design, homology modeling, peptide docking, data analysis, figure preparation and writing the manuscript.

II Study design, planning and modeling the peptide staple locations, data analysis, figure preparation and writing the manuscript.

III Study design, system preparation, molecular dynamics simulations, data analysis, figure preparation and writing the manuscript.

Additional publications

IV Turku A, Borrel A, Leino TO, **Karhu L**, Kukkonen JP, Xhaard H: Pharmacophore Model To Discover OX_1 and OX_2 Orexin Receptor Ligands. *Journal of Medicinal Chemistry* **2016**, 59: 8263–8275

V Turku A, Leino TO, Karhu L, Yli-Kauhaluoma J, Kukkonen JP, Wallén EAA, Xhaard H: Azulene-based effectors of the orexin receptors – the value of shape and electrostatics. *Manuscript*

Abbreviations

Aib	α-Aminoisobutyric acid
AMP	Adenosine monophosphate
CG	Coarse grained
CHOL	Cholesterol
CNS	Central nervous system
EC_{50}	Half maximal effective concentration
ECL1–ELC3	Extracellular loop 1–3
GDP	Guanosine diphosphate
GPCR	G protein-coupled receptor
GRK	G protein-coupled receptor kinase
GTP	Guanosine triphosphate
ICL1–ILC3	Intracellular loop 1–3
IP_3	Inositol trisphosphate
MD	Molecular dynamics
MM	Molecular mechanics
NMR	Nuclear magnetic resonance
PME	Particle mesh Ewald
POPC	1-Palmitoyl-2-oleoylphosphatidylcholine
QM	Quantum mechanics
RESP	Restrained electrostatic potential
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SDM	Site-directed mutagenesis
TM1–TM7	Transmembrane helix 1–7
WT	Wild-type

Amino acids

Alanine	Ala	A	Leucine	Leu	L	Glycine	Gly	G
Arginine	Arg	R	Lysine	Lys	K	Histidine	His	H
Asparagine	Asn	N	Methionine	Met	M	Isoleucine	Ile	I
Aspartate	Asp	D	Phenylalanine	Phe	F	Tryptophan	Trp	W
Cysteine	Cys	C	Proline	Pro	P	Tyrosine	Tyr	Y
Glutamate	Glu	E	Serine	Ser	S	Valine	Val	V
Glutamine	Gln	Q	Threonine	Thr	T			

1 Introduction

The orexinergic system comprises two G protein-coupled receptors OX_1 and OX_2, and two endogenous peptide agonists, orexin-A and -B. These components are mainly expressed within the central nervous system, where they participate in the regulation of various systems, the main physiological function being the regulation of the sleep–wake cycle. The activation of the system promotes alertness, whereas decreased signaling causes sleepiness. The onset of narcolepsy is closely linked with the destruction or malfunctioning of the orexinergic system.

At the beginning of the research presented in this thesis, the amino acid sequences of the orexin receptors were available, but the crystal structures were still to be solved. For the peptide ligands, the sequences and a collection of NMR-derived 3D structures in aqueous solution had been published, but it remained unclear whether the bioactive conformation was among them. Studies had shown that the C-terminus of the peptides was the key to biological activity and highlighted several amino acids both in the receptors and in the peptides that were important. The pharmaceutical industry had synthesized and reported a handful of well-characterized small-molecular antagonists and a plethora of their analogs, linked with the race for the orexin antagonist hypnotics. In contrast, the endogenous peptides and their analogs were the only available agonists. A patent had been filed, reporting the first small molecular agonist, but the patent or the application had not been published yet.

We set out to elucidate the binding mode of orexin peptides at their cognate receptors, with the contingent aim to replicate the identified binding interactions with a small molecule.

This dissertation comprises two peer-review publications and one manuscript (Publications I–III). In Publication I, we describe two alternative binding mode options for the orexin-A peptide at the OX_1 receptor, achieved through homology modeling and peptide docking. These binding modes are taken up in the Publication III and subjected to extensive molecular dynamics simulations to address the stability of the predicted interactions. Additionally, we simulated two small molecular ligands, an agonist and antagonist to compare the differences in binding modes. The simulations allowed us to propose a single, fine-tuned binding mode for orexin peptides and to suggest determinants of agonist and antagonist binding.

In Publication II, we produced conformationally constrained orexin peptide analogs with the aim of elucidating the bioactive conformation. A suitable constraint

might have also enabled the production of shorter orexin peptides that retained bioactivity. Many of the conformationally constrained peptides were active, but to our disappointment, with markedly decreased potency. However, this provided new insight into the bioactive conformation and the peptide recognition within the receptor binding site.

2 Review of the literature

2.1 G protein-coupled receptors

G protein-coupled receptors (GPCR) is the largest family of transmembrane proteins with approximately 800 members in human, ~350 of them non-sensory (smell, taste, or vision), recognizing a vast array of different ligands and even photons through covalently bound chromophore "ligands".[1] They reside at cell membranes, facilitating signal transduction across membranes. GPCRs are among the most successful drug targets. Approximately one third of all approved drugs act on a GPCR, and GPCRs constitute 12% of the protein targets for the approved drugs.[2,3] Similar numbers are seen with molecules in clinical trials.

2.1.1 Classification

There are two main classification schemes for GPCRs, which overlap to a significant degree. The first wide-spread system was the A–F classification, which was based on sequence and functional similarities.[4,5] The classes were A for rhodopsin-like and olfactory/taste receptors, B for secretin receptors, C for glutamate receptors, D for fungal mating pheromone receptors, E for cyclic AMP receptors and F for frizzled/smoothened receptors. Classes D and E are not found in vertebrates. A more recent GRAFS classification[1] defines five groups of human GPCRs: glutamate (G), rhodopsin-like (R), adhesion (A), frizzled/taste2 (F) and secretin (S). Of these, rhodopsin-like group is by far the largest, comprising approximately 700 members, including the estimated 460 olfactory receptors. The rhodopsin-like receptors are subdivided into branches α, β, γ, and δ. Orexin receptors belong to the β-branch, along with 33 other peptide-binding receptors.

2.1.2 Structure and function

GPCRs share a common fold, consisting of seven transmembrane helices (TMs 1–7) connected by three extracellular and three intracellular loops, an extracellular N-terminus and an intracellular C-terminus (Figure 1).[6] The helices pack into a bundle with a crevice between the extracellular ends of the TMs 2–7. This crevice is the most common binding site for ligands in the rhodopsin-like group.[7] Small molecular ligands often penetrate deep into the receptor, well within the cell membrane. For large ligands such as peptides, the extracellular loops (ECL1–3) also participate in the ligand binding. Some GPCRs, from groups other than the

rhodopsin-like, are activated by interactions formed by the N-terminus alone. In these cases, the ligand might be a protein of the extracellular matrix, for example.

In this thesis, I consider only the rhodopsin-like GPCRs. While there are many similarities in function and signaling between the groups, many details in the following chapters might not be true for the other subfamilies.

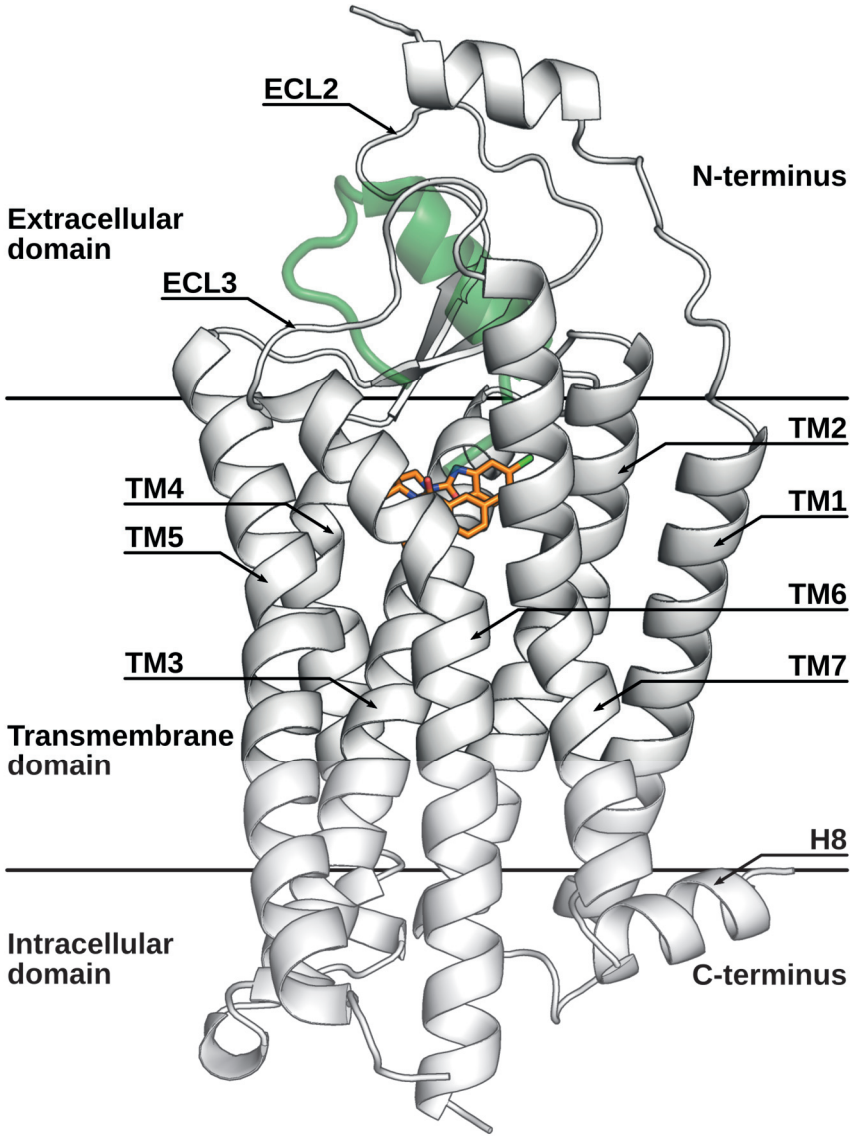


Figure 1. The conserved GPCR fold, illustrated with the OX_1 structure 4ZJ8[8]. The horizontal lines show the approximate membrane location, suvorexant in orange highlights the canonical small-molecule binding pocket, while an overlaid endothelin[9] (in green, transparent for clarity) shows that peptide binding often includes also the extracellular domain.

2.1.3 GPCR activation

Long before the discovery of GPCRs, an abstract "receptor" was suggested to bind acetylcholine at the neuromuscular junction, resulting in a conformational change from an inactive receptor to an active one.[10] Few decades later, and after the discovery of GPCRs, this view was updated to a so-called two-state model, which states that the active and inactive conformations of the receptor are in equilibrium in the absence of ligands.[11] Ligands could prefer binding to a specific state (agonist to active and inverse-agonists to inactive) and shift the equilibrium through stabilization, or bind without preference or effect on the equilibrium (antagonist). This model had the advantage of explaining the basal activity of some receptors.

Parallel, it was observed that an unidentified, guanosine-related membrane component X took part in the ligand–receptor interaction.[12] The coined ternary complex model described a low-affinity ligand–receptor interaction followed by the X component binding, and a high-affinity ligand binding to a precoupled receptor–X complex. The ternary complex of ligand, receptor and the component X could bind a guanosine nucleotide, freeing the X component to activate downstream effectors. The component X was identified as the G protein by the research group of A. Gilman,[13] earning him a shared Nobel Prize in 1994. The ternary complex model was later updated with the finding that the formation of the ternary complex was separable from the G protein activation, and that the receptor needs to transition from an inactive state R to an active state R* in order for the guanosine to bind.[14]

The Figure 2 shows a simplified view of the GPCR activation cycle. The important steps are ligand binding, G protein binding and receptor activation (R–R*). It remains unclear, which binds first, the ligand or the G protein, or if both are biologically relevant options. Receptor activation requires the bound G protein, but the ligand presence is not necessary, at least not in all receptors.

Further studies have unveiled additional conformational states and elucidated the exchanges between them.[15–17] The current understanding is that unliganded GPCRs exist in an equilibrium of states, most of them different inactive states, but some also active, which mirrors the basal activity of some GPCRs. An antagonist binds all the states with little preference and its binding should not disturb this equilibration. An inverse agonist prefers to bind to the inactive conformational states, and the binding-induced stabilization shifts the equilibrium towards the inactive receptor. Vice versa, an agonist shifts the equilibrium towards the pool of activated conformations by preferentially binding the active states and stabilizing them. It is

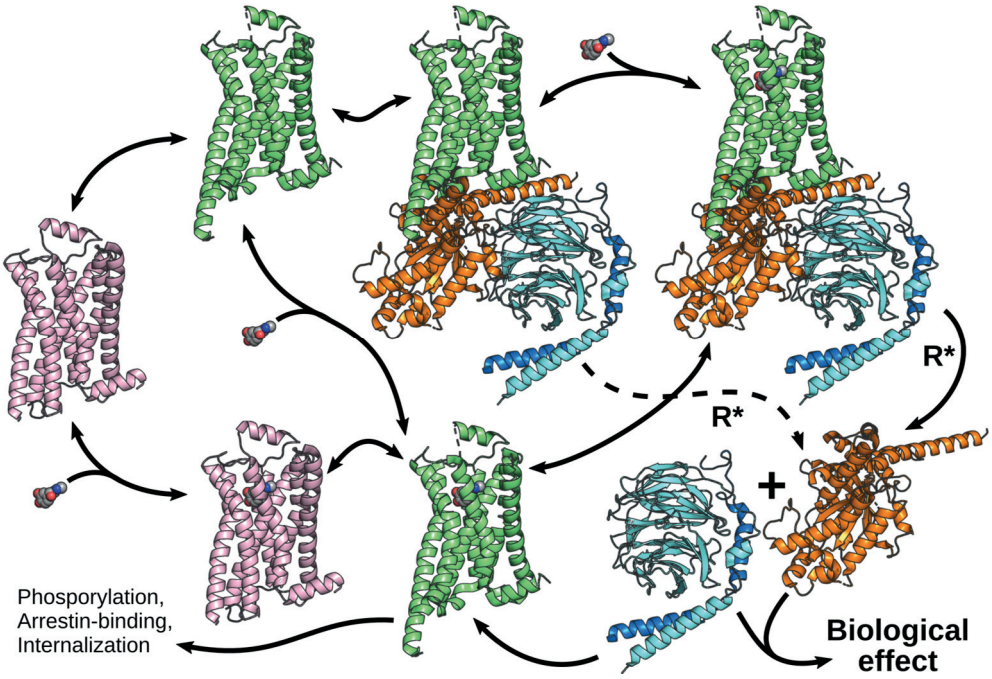


Figure 2. The functional cycle of a GPCR. Pink: Receptor in the inactive conformation. Green: Receptor with the G protein binding site open. Orange, light blue and blue: G_α, G_β and G_γ subunits, respectively. Spheres: Ligand; R: receptor activation event, i.e. the receptor-induced GDP→GTP transfer at the G protein.*

also possible for the ligands to bind "unfavorable" conformational states and induce a transformation into their preferred state. However, no ligand is likely to impose the receptor to adapt a single, rigid conformation, as that would be highly unfavorable in terms of entropy. The identification of multiple states illustrates the receptor activation pathway, and has helped in explaining the observation that certain ligands can selectively induce only one pathway through a receptor that couples to multiple pathways. This mechanism, which lies outside the scope of this thesis, is called biased agonism.

2.1.4 Downstream signaling

As the name G protein-coupled receptor suggests, the G proteins have been traditionally considered the main effector protein, the secondary messenger, of GPCRs. The G protein consists of three domains, namely α, β, and γ.[18] The α-subunit hosts a binding site for guanine nucleotides and GTPase activity. While the α-subunit joins and leaves the complex during the signaling cycle, the β- and γ-subunits form an inseparable dimer. There are at least 20 different α-subtypes, falling into four

classes (α_s, $\alpha_{i/o}$, α_q, $\alpha_{12/13}$), each class with different effects on the downstream effector proteins.[19] For the β-subunit, there are five subtypes and twelve for the γ-subunit, but most combinations have similar downstream effects.

The complete G protein (αβγ-trimer) with a bound guanosine diphosphate (GDP) is able to bind a GPCR which has an open intracellular binding site, whether ligand-binding-induced or spontaneously formed.[18] The GPCR activation induces a conformational change in the α-subunit, which leads to GDP dissociation and the binding of GTP (guanosine triphosphate) from the cytosol, which in turn leads to the separation of the α-subunit from the βγ-dimer and both units leaving the GPCR. While the receptor is free to activate the next G protein, the α-subunit and the βγ-dimer pass on to activate further effector proteins. As the α-subunit has GTPase activity, the bound GTP is eventually cleaved to GDP, rendering the subunit inactive. It then recruits the βγ-dimer, cutting off its signaling too, and thus regenerating the inactive GDP-bound G protein trimer for the next activation event.

In addition to G proteins, GPCRs bind also G protein-coupled receptor kinases (GRKs) and arrestins.[17] Initially, it was thought that these proteins merely served to desensitize the GPCRs by GRK-mediated phosphorylation and subsequent arrestin binding to block G protein/transducing binding[20] (hence the name arrestin,[21] as it was observed to arrest the phosphodiesterase activity after rhodopsin activation). However, it was shown that instead of merely blocking G protein signaling, arrestins had downstream signaling of their own. For endogenous ligands, the traditional view still holds (prolonged activity leads to arrestin-mediated desensitization), but not long ago synthetic ligands have been found to provide continuous G protein activity without arrestin-mediated desensitization,[22] or in contrast only arrestin-mediated signaling without the expected G protein signals.[23] This is called biased signaling, and the pharmacological application thereof are under intense research. Some GPCRs have also been shown to form functional dimers.[24] On top, many aspects of GPCR signaling appear to be dependent on the expression system and the relative abundances of different proteins. For example, biased signaling observed *in vitro* may be driven by an abundance of a certain G protein subtype over another instead of a biased agonist.

2.1.5 Determinants of activation

As described above, there are multiple inactive and active conformations (the active conformation in this context refers to the conformation capable of G protein

binding, not necessarily to the event that triggers nucleotide change within the G protein). However, in the light of available GPCR crystal structures, there are certain hallmarks to both (Figure 3).[25,26] The most notable change to take place upon GPCR activation is the outswing of the intracellular end of TM6 (Figure 3A), which opens up the G protein binding site.[27] This is of course not an isolated event, but accompanied with distinct changes in the interhelical interactions at the intracellular side and within the receptor core. Comparison of active and inactive structures has highlighted a series of interactions present consistently only in either of the groups.[26] Close interactions of the pairs 3x46–6x37, 1x53–7x53, 7x53–8x50, and 7x54–8x51 are seen in all inactive GPCRs, while these interactions are broken upon the opening of the G protein binding site to give rise to interactions between 6x41–5x55 and 7x53–3x46 (Figure 3D–E). The conserved tyrosines at 7x53 and 5x58 are also observed to reorganize to form a water-mediated interaction upon activation with Arg^{3x50} hydrogen-bonding to Tyr^{5x58}, which is related to the TM6 moving away and TM5–TM7 distance diminishing (Figure 3B).[25] In a subset of GPCRs, an "ionic lock" between Arg^{3x50} and Glu^{6x30} stabilizes the inactive conformations and reorganizes to form other interactions upon activation,[28] but in orexin receptors an arginine is present at the 6x30, and a suitable "replacement" acidic residue is not found within the intracellular end of the TM6. Another feature linked with activation is the "core triad" below the orthosteric binding site, formed by Pro^{5x50}, Phe^{6x44} and a hydrophobic residue at 3x40 (often isoleucine).[29] The observation is that in the inactive state, the hydrophobic residue at 3x40 lies in-between Pro^{5x50} and Phe^{6x44} (Figure 3C). An activation-linked inward movement of the extracellular end of the TM5 would shift or push the hydrophobic residue at 3x40 aside, thus allowing Phe^{6x44} to move closer to TM5 and Pro^{5x50}, facilitating the outward swing of the intracellular end of TM6. The caveat here is that some crystallized receptors, such as the M2, feature a smaller valine instead of the bulky isoleucine at 3x40, and fail to show marked differences between the active and inactive core triad conformations, perhaps owing to less restriction on the conformation of Phe^{6x44}.[29–31]

The residue at site 6x48, most often Trp^{6x48}, has been called the "toggle switch" and thought to be linked with GPCR activation, as spectroscopic studies on rhodopsin indicated that the Trp^{6x48} would change rotamers during activation.[32,33] While this hypothesis was undermined by the active rhodopsin, β_2 and M2 crystal structures showing inactive-like vertical conformation to Trp^{6x48} side chain,[31,34,35] NTS_1 has been observed with a horizontal rotamer for the tryptophan side chain.[36] In any case,

the conserved aromatic residue at 6x48 might well play a role in the activation cascade.

At the binding site level, the determinants of activation are more diverse, as befits the vast diversity of GPCR ligands. It is suggested that a common activation

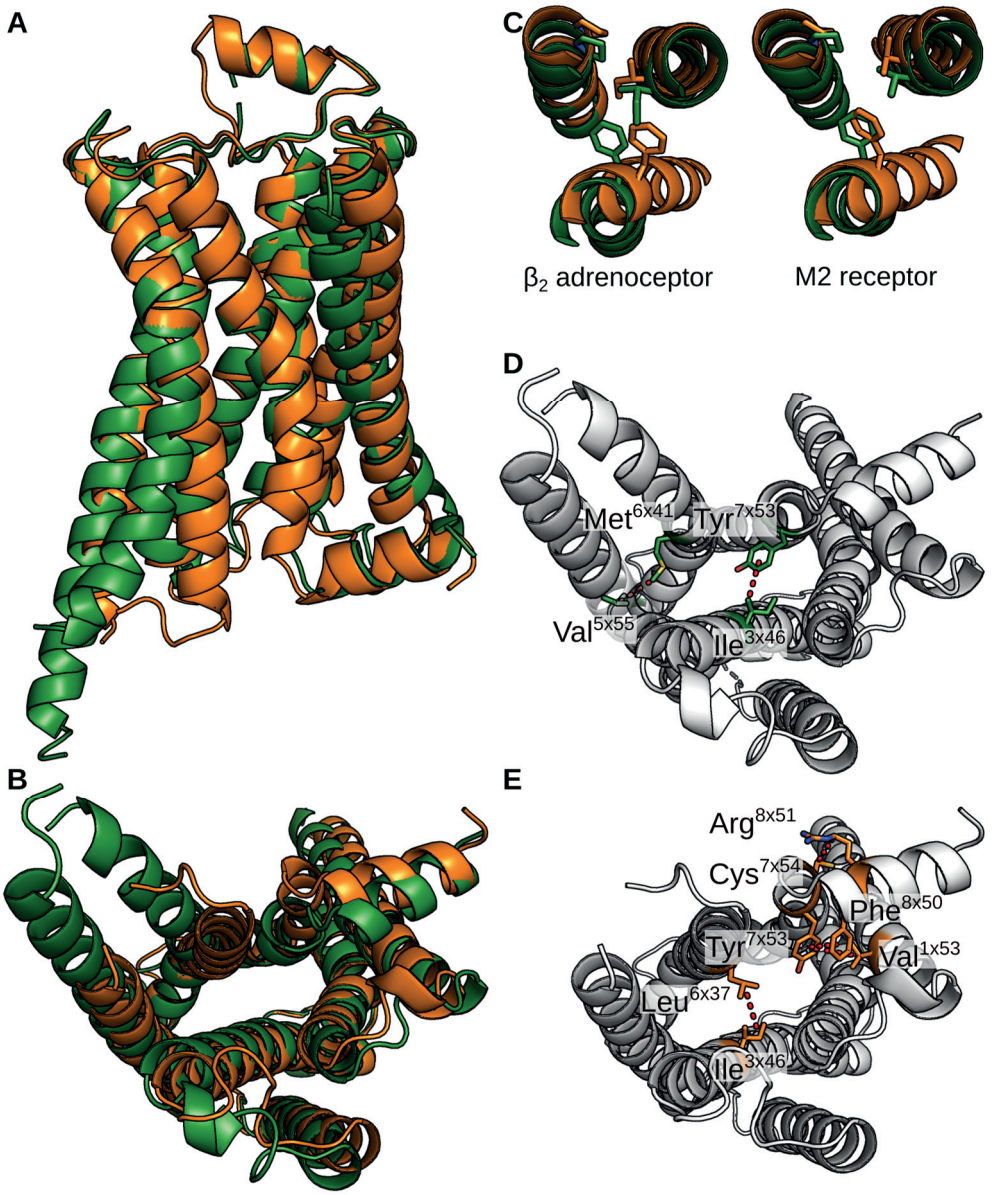


Figure 3. Inactive and active GPCR conformation, as seen with the β2-adrenoceptor (PDB id: 2RH1[37] and 3SN6[27]). A) View from TM6–7; B) Intracellular view; C) The core triad (M2 receptor: PDB id: 3UON[30] and 4MQS[31]); D–E) Common interactions in active (D) and inactive (E) GPCR crystal structures. Green: Active; Orange: Inactive.

mechanism would be the contraction of the ligand binding cavity.[25] Adrenoceptors, for example, are thought to be activated by an inward motion of TM5 induced by the ligand binding between TM5 and TMs 3 and 7, whereas the peptide-binding NTS_1 shows an inward tilt of TMs 6 and 7 towards the ECL2.[38]

2.1.6 Ligand binding

2.1.6.1 Binding site

The GPCRs have a canonical orthosteric binding site within the extracellular ends of the transmembrane helices. The ligand-binding depth for small molecules is similar across receptors, but peptides bind at different depths.[7] The Figure 4 displays the binding depth for co-crystallized peptides, and the depth of our simulation-derived peptide binding.

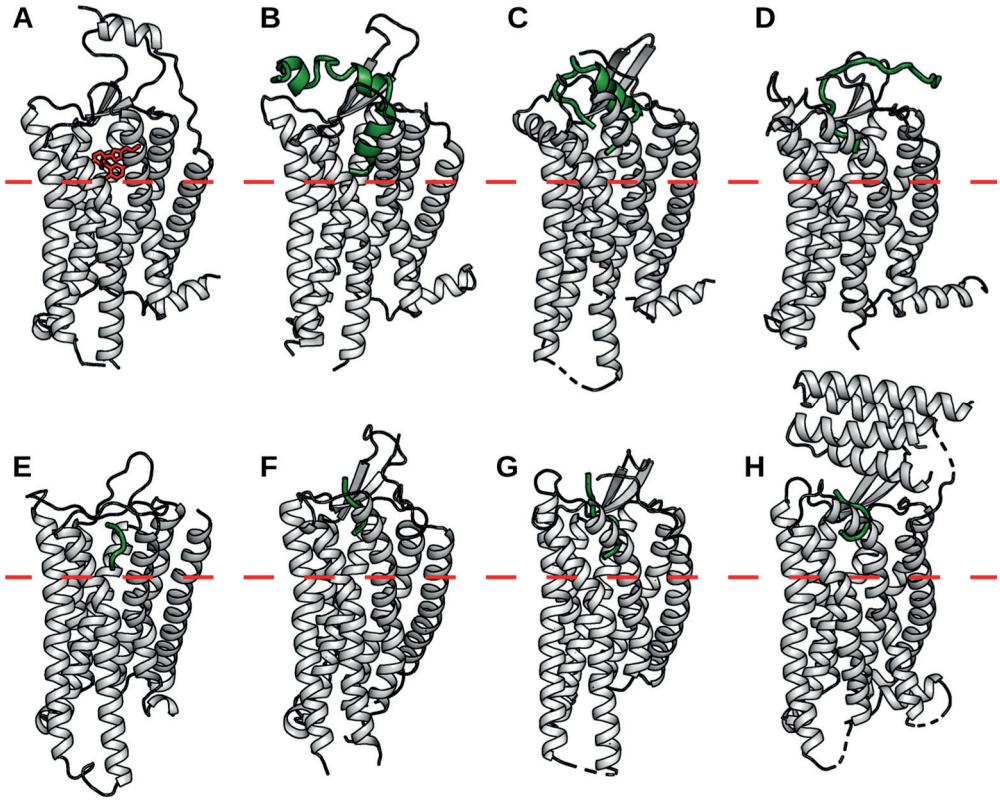


Figure 4. Ligand binding depth. A) Suvorexant in OX_1[8]; B) Orexin-A after MD simulation; C) Endothelin-1 in ET_B[9]; D) Apelin in apelin receptor[39]; E) an agonist peptide DAMGO in μ receptor[40]; F) $NT_{8–13}$ in NTS_1[38]; G) Modified angiotensin II in AT_2[41]; H) an antagonist peptide PMX53 in C5a (complement) receptor[42]. The red dashed line highlights the small-molecule binding site depth.

2.1.6.2 The binding event in terms of thermodynamics

Ligand binding is a complex event. The ligand needs to diffuse into the receptor binding site, which is sometimes assisted by the receptor extracellular loops.[43] Along the process, the ligand adopts a suitable conformation. Within the extracellular fluid, the ligand is surrounded by water molecules, which it needs to shed upon binding. The same goes for the binding site; it is not an empty cavity prior ligand binding but filled with water, which needs to diffuse out of the cavity. Only then can the ligand–receptor interactions take place.

Irreversible ligands aside, ligands bind through non-covalent interactions: electrostatic attraction between opposite (partial) charges, hydrogen bonds, π-electron interactions and van der Waals interactions. While there is no actual "hydrophobic interaction", the term is useful in describing an observed effect of non-polar moieties packing together. This rises from the fact that it is more favorable for water to interact with polar groups (including other water molecules), and the packing of hydrophobic groups reduces the surface area of the less favorable polar–non-polar interface.

There are two main components in thermodynamics: entropy and enthalpy. Entropy stands for the disorder of the system, and enthalpy describes the potential energy of the system. Formation of chemical bonds decreases enthalpy, while the breakage of bonds and different kinds of bond strains increase enthalpy. These are often combined to calculate a change in the Gibbs free energy of binding with the following equation: $\Delta G = \Delta H - T\Delta S$. H stands for enthalpy, T for temperature and S for entropy. This change has to be negative in order for an event to take place spontaneously. The effects of various ligand-binding sub-events on these terms are summarized in the Table 1.

Table 1. Thermodynamics of ligand binding.

Event	Decreased enthalpy	Increased enthalpy	Increased entropy	Decreased entropy
Adoption of bioactive conformation		Possible deviation from the lowest energy structure		Restriction of ligand conformation
Binding site desolvation	Water–water interactions form	Broken water–binding site interactions	Water is released into bulk solvent	
Ligand desolvation	Water–water interactions form	Broken water–ligand interactions	Water is released into bulk solvent	
Ligand binding	Ligand–receptor interactions form			Restriction of ligand position

Often, ligand binding is though only in terms of binding interactions. However, the conformational constraining of the ligand upon binding also plays a major role. For example, it is beneficial to construct a rigid ligand, which is in a biologically active conformation, instead of a flexible molecule that can adopt an active conformation. Even if the active conformation is among the low-energy conformations, the binding would result in decrease of conformational freedom and thus in decrease of entropy. Solvation effects are also important. It is not enough for the ligand to form favorable interactions with the binding site; the interactions need to be superior to the interactions formed by water, both with the binding site and with the ligand, otherwise the total enthalpic effect remains unfavorable.

2.2 Computational methods

The focus of the thesis research was on the use of computational tools to predict orexin peptide binding into orexin receptors. The chapters below will offer a peak into the toolbox and review the scientific discoveries others have reached with similar methods.

2.2.1 Molecular mechanics and force field

Molecular mechanics (MM) stands for the modeling principle where classical mechanics are used to mathematically model a molecular system. Generally, atoms are treated as balls with a fixed radius, mass, and charge. Bonds and three-atom angles are modeled as springs, or harmonic potentials, and bond rotation is expressed in terms of a periodic function with multiple minima. Short-range non-bonded interactions are usually calculated in a combination of Lennard-Jones potential[44] and Coulomb interaction, while long-range electrostatic interactions are often treated with Particle-mesh Ewald (PME) method.[45]

Force field stands for a collection of mathematical functions and numerical parameters that can be used to perform MM calculations. These include the atom radii and partial charges, the ideal bond lengths and angles along with the spring constants, and the form and parameters for the bond rotation.

Using MM with its assumptions is a trade-off between accuracy and efficiency. Several problems rise from the fact that electrons are not considered: bonds cannot form or break and charges on the atoms remain fixed. Also, a force field cannot treat an atom or a molecule which it is not parametrized for. However, an alternative method would be to resort to quantum mechanics (QM) calculations, which rely on

calculation of the electron density. This of course provides superior accuracy, but the cost on calculation time is such that QM is unfeasible for most anything larger than a small molecule in a single conformation, whereas MM can deal with large systems and timescales up to milliseconds. Also, bearing in mind the limitations on MM, computational methods can often reach the limit of experimental error when one sticks to well-parametrized systems.[46]

Molecular mechanics is used with both static and dynamic systems. For static complexes, such as receptor modeling, small molecule conformation generation or for docking purposes, MM can be used to calculate the potential energy of the system. This can lead the selection of low-energy conformations for both small molecules and proteins or the ranking of docking poses. In contrast to static systems, where MM is used to calculate the energies of pre-generated systems, the dynamic use of MM allows the relocation of particles to yield new conformations. This is useful in energy minimization, for example, which is often linked with conformation generation, and especially in molecular dynamics (MD) simulations.

2.2.2 Molecular dynamics simulations

Molecular dynamics simulation stands for the use of molecular mechanics to derive forces, which are then applied to the system of atoms using classical mechanics. An MD simulation is essentially a series of steps, called frames. Each frame consists of coordinates and speeds for all individual particles of the system. From the coordinates and force-field-derived parameters, the simulation engine calculates the forces acting on each atom. These forces are then used to calculate the changes in the particle speeds, and each particle is allowed to move for a short period of time, typically in the ballpark of few femtoseconds. With the new set of coordinates and speeds, the process is repeated as many times as requested by the user.

MD simulations enable the observation of a biological system in the atomic level. The simulation allows for the examination of interatomic interactions over time, as well as the large-scale movements within the system. Given long enough simulation times, one could most likely observe ligand binding events, receptor conformation changes, transporter protein fluctuations, ion channel function, or protein–protein interactions.

Conventional or classical MD simulations “seek” the lowest energy conformations. This is expected behavior, if one wishes to study, say, the stability of

predicted interactions. However, if the goal is to observe transitions, this poses problems, because even if both "ends" would be low-energy conformations, the transition from one to another usually requires the system to adopt one or more high-energy conformations along the way. While it is possible for the conventional MD simulations to cross barriers of higher energy, it might require significantly longer simulations due to the improbability of such events. As transitions are of paramount importance in biology, several methods have been implemented.

Accelerated molecular dynamics introduces a "boost" potential, which increases the energy of conformations that would fall below a set threshold. This effectively makes deep energy-wells shallower, allowing for easier escape over the neighboring energy barriers.[47,48] Metadynamics, on the other hand, introduces similar destabilizing potential component to simulation states that have already been sampled.[49] This requires the setup of geometrical criteria to differentiate between the states.

2.2.2.1 MD simulations on GPCRs

GPCRs have been subjected to countless MD simulations to probe their ligand-binding interactions and activation cascades.[50] For example, massive classical simulation efforts have elucidated the route and mechanism of small-molecule ligand binding into adrenoceptors[43] and the binding mechanism of allosteric modulators of the M2 receptor.[51] Through metadynamics, similar results have been obtained with a fraction of the computational cost.[52] For the geometrical criteria, the study used simply the distance between a ligand and a residue at the bottom of the binding site in a direction perpendicular to the membrane. In addition, the researchers placed an inverted "cup" on top of the receptor, which applied a repulsive force to the ligand if it was about to diffuse away from the vicinity of the receptor. These ligand-binding simulations offer insight into the pathway of ligand entry and the events that take place. In addition, they allow the estimation of binding affinity, which is often of great interest to medicinal chemists.

As the GPCR activation takes place over millisecond timeframes[53], the transition from inactive to active has eluded even the longest unbiased simulations up to date.[54,55] However, the reverse event has been caught by conventional MD simulations[54]. Relying on an assumption that the inactivation of the receptor takes the same steps backwards as the receptor activation, this allowed for detailed suggestions on the molecular mechanics within the activation cascade. Building on these

simulations, an accelerated MD simulation was able to observe the activation event of the M2 receptor.[56]

2.2.3 Homology modeling

Homology modeling refers to the methods, where a protein with a defined 3D structure is taken as a template for the building of a model for the protein of interest ("target"). The template and the target proteins should be close homologs for the resulting model to be reliable, preferably from the same protein family. Additionally, an ideal template structure should have high resolution, no crystallization-induced defects and a high sequence identity for the target protein. The class A GPCRs do not always display high sequence identity as the group is quite large and versatile, but the common tertiary fold compensates this to a certain degree, at least for the transmembrane bundle.[57–59] Successful homology modeling rests on two cornerstones: sequence alignment and template selection. Sequence alignment stands for the assignment of homologous amino acids between the template and the target; together with the template they form a map for the modeling program where to place each residue.

2.2.4 Peptide docking

Peptide docking is a challenging task due to the large number of atoms and the inherent flexibility of the peptides.[60,61] The prediction of very short peptides is possible with the tools intended for small molecules,[62] but longer peptides require specialized software. The currently available peptide docking tools were recently reviewed by Ciemny and co-workers.[60] However, most peptide-docking software is benchmarked with peptides of 15 amino acids or fewer, most of which are not helical.[63,64]. Also buried binding sites have been problematic.[65–67]

A tempting, and often the only, option in docking longer peptides is to resort to protein–protein docking software such as the Schrödinger PIPER[68] or ZDOCK[69]. Both perform rigid-body docking and score based on shape complementarity, electrostatics and solvation effects.

2.3 Orexinergic system

The orexinergic system was discovered in 1998 by two independent research groups. Early in January, an article in PNAS[70] reported that an mRNA sequence

expressed in the rat hypothalamus encodes a putative peptide precursor that could be cleaved to yield two C-terminally amidated peptides, sharing sequence similarity with each other and secretin. Conservation in mouse was shown, along with the localization of the mRNA only in the hypothalamus within the brain, while the predicted protein product was observed enclosed in vesicles within cells projecting to various areas of the brain. The 28-amino-acid peptide was also found to be neuroexcitatory *in vitro*, while the other peptide was not synthesized as the starting position could not be deduced from the mRNA. By comparison to secretin, the group suggested that the peptides could act through two (yet unidentified) GPCRs to activate adenylyl cyclases, and based on the neuronal projections, the peptides could serve as regulators of nutritional homeostasis. By the hypothalamic origin and the resemblance to incretins, the group named the peptides hypocretins and the precursor preprohypocretin.

Some six weeks later, an article in Cell[71] outlined an extensive study aimed at deorphanization of GPCRs by screening various tissue extracts against a panel of cell lines expressing orphan GPCRs. The group identified a brain extract from rat which produced a robust Ca^{2+}-elevation through an orphan GPCR named HFGAN72 from the human brain. As the response could be obliterated by protease pre-treatment, the active component was likely a peptide. Three active components were purified and identified. The main activity was assigned with a 33-amino-acid peptide with two intramolecular disulfide bridges, an N-terminal pyroglutamyl residue, and C-terminal amidation. An identical peptide was purified from a bovine hypothalamus extract. The group termed the peptide orexin-A. Two other components were identified as a linear, C-terminally amidated 28-amino-acid peptide and the N-terminally truncated 3–28 fragment thereof. The former was named orexin-B and the latter orexin-$B_{3–28}$ as it was unclear whether the shorter peptide was biologically relevant or an artifact from the extraction and purification. Working backwards from orexin-A, the group obtained the rat cDNA responsible for the peptide precursor prepro-orexin, and subsequently the corresponding mouse and human genomic fragments, to learn that orexin-A for these species was identical and orexin-B different by two amino acids from the rodent orexin-B. The HFGAN72 receptor was confirmed as the receptor for both orexin peptides, but orexin-A was 2–3-fold more potent than orexin-B. Through a BLAST search of the GenBank database, the group identified a gene for another receptor, which, when cloned and expressed, turned out to bind both orexin-A and -B with high affinity. The deorphanized receptor was labeled as the OX_1 receptor, and the GenBank-derived was named the OX_2 receptor. Both the peptides and the

receptors were predominantly expressed in the brain, which befit the hypothesis of neuropeptides. The name orexin (the Greek word for appetite: ὄρεξις, orexis) was selected because the peptides increased food intake of rats, and the mRNA for the precursor peptide was upregulated by fasting.

Quickly after their publication, the latter group lead by Masashi Yanagisawa, noticed[72] that their orexin peptides[71] were identical to hypocretins[70] discovered by the group of J. Gregor Sutcliffe.

2.3.1 Signaling and physiological functions

The main downstream signaling pathway for orexin receptors appears to be the G_q-mediated activation of phospholipase C, which produces IP_3 and induces the intake and release of Ca^{2+}, resulting in the robust elevation of intracellular Ca^{2+}-concentration. However, depending on the cell line or tissue, also $G_{i/o}$- and G_s-mediated regulation of adenylyl cyclase is observed. At the cellular level, orexinergic signaling is neuroexcitatory through membrane depolarization.[73] Orexin receptors are mainly expressed within the CNS, where they participate in the regulation of the sleep–wake cycle, energy homeostasis, stress and the reward system.[73] Orexin receptors are also found in several peripheral tissues, but the clinical significance of these remains unclear, especially as the expression of orexin peptides is limited to only a few sites throughout the body. Distribution by circulation has not been shown to our knowledge, but pharmacokinetic parameters have been established experimentally. Orexin-B degrades rapidly in blood,[74] while orexin-A exhibits a half-life of approximately 20–30 minutes,[74–76] possibly due to peptidase protection offered by the disulfide bridges. The complex details of orexin receptor downstream signaling and cellular effects[77] are outside the scope of this thesis.

2.3.2 Therapeutic potential

As the orexinergic systems participates in many physiological functions, there are also multiple potential therapeutic areas. However, as the main physiological function appears to be the modulation of sleep and alertness, the focus of the pharmaceutical industry has been on the development of orexin antagonists as hypnotics.[78] In addition to inducing sleep, the orexinergic system could also be targeted to reduce sleepiness, which would be key to successful narcolepsy treatment, especially as malfunctions of the orexinergic system contribute significantly to the onset of narcolepsy.[79,80] Orexin receptors have also been located in certain cancer cell

lines, and the activation of these receptors directed the cells to apoptosis, raising interest as orexin receptors as potential cancer medication targets.[81,82]

2.3.3 Orexin receptors

The orexin receptors OX_1 and OX_2 are quite similar at sequence level (full length identity 64%, the TM bundle excluding ICL3 80% identical). Closest relatives by sequence identity in human are the NPFF1 and NPFF2 receptors at 37% and 35% identities for the TM bundle, respectively. The crystal structures for both orexin receptors have been elucidated quite recently.[8,83] The seven transmembrane helices and the H8 at the C-terminus pack into the canonical GPCR fold, along with the short loops ICL1, ECL1, ICL2 and ICL4. The ECL2 adopts a β-hairpin fold similar to other peptide-binding GPCRs. The loop is stabilized by the conserved disulfide bridge between the ECL2 and the extracellular end of the TM3. In the first orexin receptor crystal structure, 4S0V for the OX_2,[83] the conformation for the receptor N-terminus could not be solved, but subsequent structures of both receptor subtypes[8,84] have shown a two-turn amphipathic α-helix nine residues upstream of the TM1. In the OX_1 structures, the N-terminal helix packs against the ELC2 hairpin, parallel to the membrane plane, whereas the later OX_2 structures show the helix facing away from the receptor, again parallel to the membrane (Figure 5). Assays with orexin-A and N-terminal deletion constructs of OX_1 and OX_2 showed abolished binding and activation of receptors,[8] strongly suggesting a vital role for the N-terminus in orexin peptide binding. The authors suggest that the amphipathic helix could recruit the orexin peptide from the solution and guide it into the binding site formed by the N-terminus, the ECL2 and the canonical GPCR binding pocket.

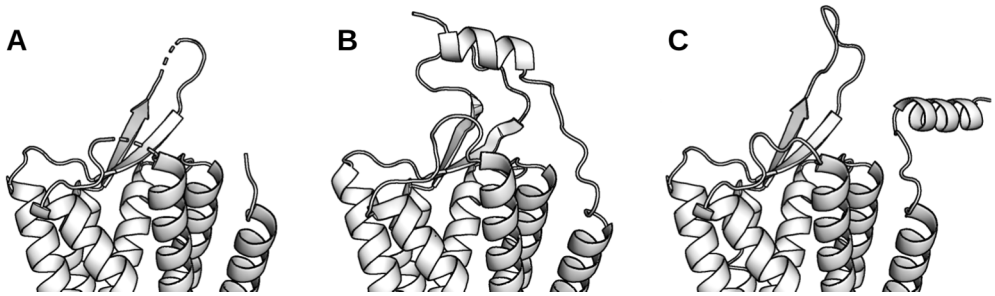


Figure 5. The extracellular domain of orexin receptors. A) OX_2 (PDB id: 4S0V); B) OX_1 (PDB id: 4ZJ8); C) OX_2 (PDB id: 5WQC)

The orexin receptors have been subjected to site-directed mutagenesis (SDM) with the intention to locate important residues for ligand binding and receptor function.[8,85–88] The mutations and their effects are listed in the Tables 2 and 3, and illustrated in the Figure 6. Three mutations have a drastic effect on the receptor function across a wide selection of ligands. The mutation of Tyr215/223^{5x39} to alanine for example drastically lowers the orexin peptides' binding affinity and potency, and also abolishes antagonist binding.[87,88] However, the receptor does express, and localizes to the plasma membrane. Similar, but not quite as disruptive effect is seen with alanine mutations of Trp206/214^{45x54} and Phe219/227^{5x43}. Interestingly, all these residues are aromatic, and closely situated. Trp^{45x54} and Tyr^{5x39} pack closely together at the junction of ECL2 and TM5, facing the TM4 and not so much the binding site. Phe^{5x43} is one helical turn below Tyr^{5x39}, facing the TM3 as much as the binding site. If the packing of these residues is critical for the local folding of the binding site, these mutations could cause a deformation which would explain the observed effects.

Orexin peptide binding or potency are also affected by mutations of 2x60, 45x51, 5x47, 6x48, 6x55, and to smaller extent by 3x32, 3x36, 7x34, 7x38, and 7x42. Thr^{2x60}, Asp^{45x51}, and Asn^{6x55} offer polar interaction sites to the binding site, and it is not surprising to find them contributing to the peptide-binding interactions. The aromatic residues Tyr^{5x47} and Tyr^{6x48} are side-by-side at the bottom of the binding cavity. As described above, the site 6x48 has been linked with the receptor activation cascade. While Tyr^{5x47} faces the TM6 and not the binding cavity, it could be linked to Tyr^{6x48} conformation or motions during the activation cascade, or to the large-scale helical reorganization. Interestingly, alanine mutations of Tyr^{5x47} and Tyr^{6x48} also produced a marked decrease in the efficacy of orexin-A, which supports the hypothesis of an impaired activation cascade.

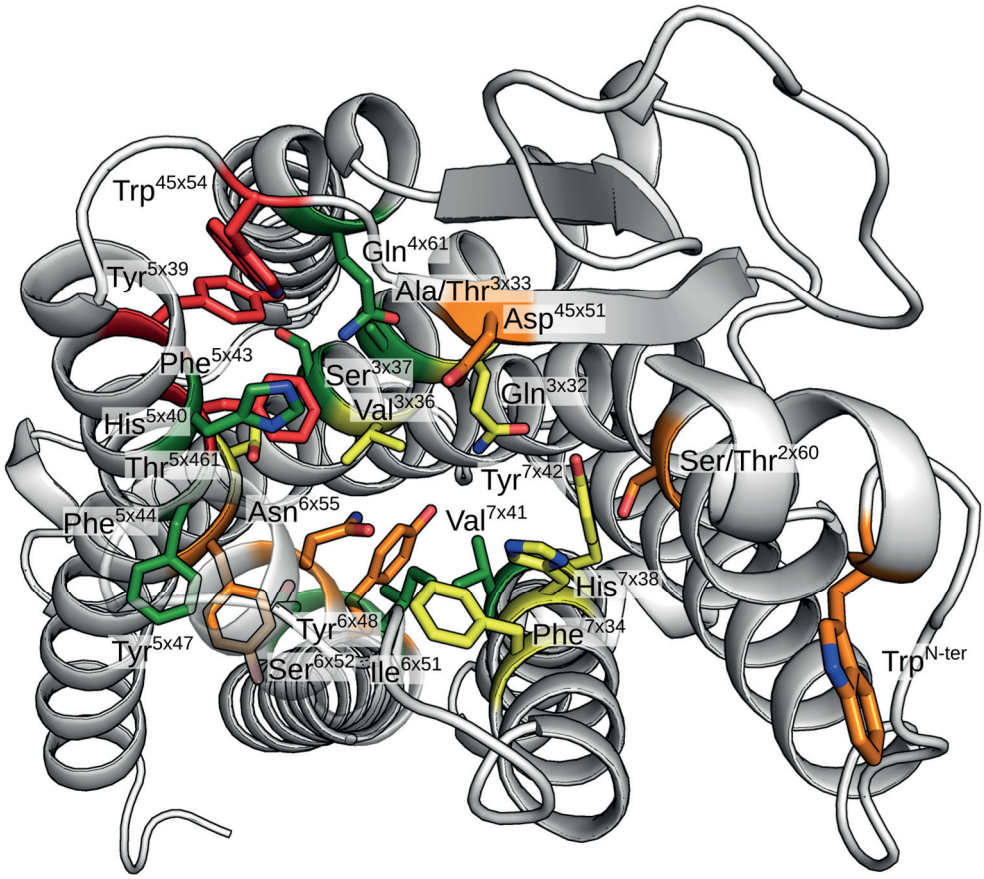


Figure 6. Amino acids in the orexin receptors subjected to site-directed mutagenesis. Coloring indicates the effect of an alanine mutation: Red, orange, and yellow for decreasing deleterious effect, green for no effect.

Table 2. Effects of site-directed mutations on the OX_1 receptor.

Mutation	Site	Orexin-A EC_{50}	Orexin-A E_{max}[a]	Orexin-B EC_{50}[c]	Almorexant IC_{50}	Almorexant binding[a]	SB-649868 IC_{50}[c]	SB-674042 IC_{50}[a]	SB-674042 binding[a]	SB-334867 IC_{50}[c]
W36N	N-ter	↓↓[d]								
Q126A	3x32	=[a]	↓↓		↓↓↓[a]	NDB		↓↓	↓↓↓	
A127T	3x33	=[a]	=		↓↓↓[a]	NDB		↓↓	↓↓	
V130A	3x36	↓↓[a]	=		=[a]	=		=	=	
Q179A	4x61	↑↑[c]		↑↑↑	=[c]		↓			
Q179H	4x61				=[c]		↓↓			=
Q179N	4x61				=[c]		=			=
D203A	45x51	↓↓↓[a]	=			NDB			=	
W206A	45x54	↓↓↓[a]	↓↓			NDB			NDB	
Y215A	5x39	↓↓↓[a]	↓			NDB			NDB	
F219A	5x43	↓↓↓[a]	=		↓↓↓[a]	NDB		↓↓↓	NDB	
Y224A	5x47	↓↓[a]	=		=[a]	NDB		=	=	
Y311A	6x48	↓↓↓[a,c]	↓↓	↓↓↓	↓↓↓[a]	NDB		=	↓↓	
Y311F	6x48	=[c]		=	=[c]		=			=
N318A	6x55	↓↓[d]								
H344A	7x38	↓↓↓[a]	↓		↓↓[a]	NDB		↓↓	↓↓	
Y348A	7x42	=[a]	↓		↓↓[a]	NDB		=	↓	

NDB = no detectable (specific) binding, in direct saturation assay; a, Malherbe et al. 2010; c, Heifetz et al. 2013(?); d, Yin et al. 2016

Symbol	Fold effect on peptide EC_{50} or binding affinity, or antagonist IC_{50}	Fold effect on antagonist binding affinity
↑↑↑	> 100	> 50
↑↑	20–100	10–50
↑	10–20	3–10
=	0.5–10	0.5–3
↓	0.1–0.5	0.1–0.5
↓↓	0.05–0.1	0.05–0.1
↓↓↓	< 0.05	< 0.05

Table 3. Effects of site-directed mutations on the OX_2 receptor.

Mutation	Site	Orexin-A EC_{50}	Orexin-A binding[b]	Orexin-A E_{max}[a]	Orexin-B EC_{50}	Orexin-B binding[b]	Almorexant IC_{50}	Almorexant binding[a]	SB-649868 IC_{50}[c]	EMPA IC_{50}	EMPA binding	JNJ IC_{50}[b]	JNJ binding[b]	SB-674042 IC_{50}[b]	SB-674042 binding[b]
W44N	N-ter	↓↓↓[d]													
T111A	2x60	↓↓↓[a]		=			= [a]	=		↓↓[a]	↓[a]				
Q134A	3x32	↓↓[a] / ↓[b,c]	↓	↓	↓[b] / = [c]	↓	= [a] / ↑[c]	↓↓	↑	= [a,b]	= [a,b]	↓↓	↓↓	=	=
T135A	3x33	= [a,b]	=	=	= [b]	↓↓	= [a]	=		↓↓↓[a,b]	NDB[a] / ↓↓↓[b]	↓↓↓	↓	↓↓	↑
V138A	3x36	= [a]		=			= [a]	=		↓↓[a]	↓↓[a]				
S139A	3x37	= [a,b]	=	=	= [b]	=	= [a]	=		= [a,b]	= [a,b]	=	=	=	=
Q187A	4x61	= [c]			= [c]		= [c]		=						
D211A	45x51	↓↓↓[a]		=			= [a]	=		↓[a]	↓[a]				
W214A	45x54	↓↓[a]		↓			↓↓↓[a]	NDB		↓↓↓[a]	NDB[a]				
Y223A	5x39	↓↓↓[a,b]	NA	↓	↓↓↓[b]	NA	↓↓↓[a]	NDB		↓↓↓[a] / NAB[b]	NDB[a] / NA[b]	NAB	NA	NAB	NA
H224A	5x40	= [b]	=		= [b]	=				= [b]	= [b]	=	=	=	=
F227A	5x43	↓↓↓[a] / ↓↓[b]	NA	↓	↓↓[b]	NA	↓↓↓[a]	NDB		↓↓↓[a,b]	NDB[a] / NA[b]	↓↓↓	NA	↓↓	NA
F227W	5x43	↓↓[c]			↓↓[c]				↑↑						
F228A	5x44	= [a,b]	=	↓	= [b]	=	= [a]	=		= [a,b]	= [a,b]	=	=	=	=
T231A	5x461	↓[b]	NA		= [b]	NA				= [b]	NA[b]	=	NA	=	NA
Y232A	5x47	↓↓[a]		↓↓			= [a]	↓↓		↓↓[a]	NDB[a]				
Y232F	5x47						↑[c]		↑						
Y317A	6x48	↓[a]		↓↓			↓↓[a]	NDB		↓↓[a]	NDB[a]				
Y317F	6x48	= [a]		↓			= [a]	=		= [a]	↓[a]				
I320A	6x51	= [a,b]	=	=	= [b]	=	= [a]	=		↓↓↓[a], = [b]	NDB[a] / ↓↓[b]	=	↓	=	=
S321A	6x52	= [b]	NA		= [b]	NA				= [b]	NA[b]	=	NA	↓	NA
N324A	6x55	↓↓[b] / ↓[d]	NA		↓[b]	NA				↓[b]	NA[b]	=	NA	=	NA
F346A	7x34	↓↓[a,b]	NA	↓	= [b]	NA	= [a]	↑		= [a,b]	= [a], NA[b]	=	NA	=	NA
H350A	7x38	↓↓[a] / ↓[b]	NA	↓	= [b]	NA	↓↓[a]	NDB		↓↓↓[a] / ↓[b]	NDB[a] / NA[b]	↓	NA	=	NA
V353A	7x41	= [a]		=			= [a]	=		= [a]	= [a]				
Y354A	7x42	= [a,b]	↓	↓	= [b]	↓	= [a]	=		↓↓↓[a] / ↓↓[b]	NDB[a] / ↓↓[b]	=	↓	=	↓

NDB = no detectable (specific) binding, in direct saturation assay; NA = radiotracer did not bind (for displacement assay); NAB = no agonist binding (for antagonist assay); a, Malherbe et al. 2010; b, Tran et al. 2011; c, Heifetz et al. 2013(?); d, Yin et al. 2016. Symbols as in Table 2.

The antagonists differ slightly in their binding modes, at least in the light of the effects on the site-directed mutagenesis.[86–88] However, most are affected by mutations of Gln^{3x32}, Tyr^{6x48} and the aromatic residues in the TM7. An interesting site for an alanine mutation is the $Thr135^{3x33}$ in OX_2, as the corresponding wild-type residue in OX_1 is alanine. As expected, the mutation does not affect orexin-A, which binds OX_1 and OX_2 with similar affinity, but reduces the affinity of orexin-B, which favors the OX_2. The mutation also impairs the binding and activity of OX_2-specific antagonists EMPA and JNJ-10397049. One would expect the OX_1-specific antagonist SB-674042 to gain affinity and potency, but interestingly a moderate decrease in potency is observed, even though there is a small gain in affinity. The reverse mutation $A127^{3x33}T$ in OX_1 produces the expected loss of SB-674042 affinity and potency, but here also almorexant affinity and potency are heavily decreased, even though it is an antagonist for both receptors, and the mutation $T135^{3x33}A$ in the OX_2 had no effect on almorexant.

Mutations of 3x37, 5x40, 5x44, and 7x41 have no significant effect neither on orexin peptide affinity or potency, nor affinity or potency of the tested small molecular antagonists.

Chimeras between the OX_1 and the OX_2 showed that ligand selectivity is mostly conferred by the TM3.[88,89] Two studies, employing different cut points, found that either the region ECL1-TM3-ICL2-TM4 or the region TM2-ELC1-TM3 carries the largest effect on the preference of the orexin-B for the OX_2 receptor, as well as the subtype selectivity for the antagonists. However, also combinatorial effects were observed; for instance, the introduction of either the OX_1 N-terminal and TM1, or OX_1 TM7 and C-terminal into OX_2 carries no effect on the potency of an OX_1-specific antagonist, but a dual substitution of both induces a ten-fold increase in potency.

2.3.4 Orexin peptides

There are two orexin peptides, orexin-A and orexin-B. Orexin-A is equipotent towards both receptors.[71,90,91] Orexin-B is as potent as orexin-A at OX_2, but 10-fold less potent in OX_1.[71,90,92] As described above, orexin-A is a C-terminally amidated 33-amino-acid peptide with N-terminal pyroglutamate and two intracellular disulfide bridges, and orexin-B is linear C-terminally amidated 28-amino-acid peptide (Figure 7A). The peptides are conserved across vertebrates and are also similar to each other in terms of C-terminal sequence; 11 out of 15 amino acids are identical. The structure

of both peptides in aqueous solution has been elucidated by NMR.[93–95] Orexin-B shows two helical segments, Leu7–Asn20 (helix I) and Ala22–Met28 (helix II), bent 90° 1.5 helical turns from the C-terminus (Figure 7B).[94] In orexin-A, there are three helical sections, Leu16–Ala23 (helix I), Asn25–Leu33 (helix II), and Asp5–Lys10 (helix III). The disulfide bridges stabilize a turn between the helices III and I (Figure 7). One study reports 30 conformation for orexin-A,[93] falling into two categories; 5 conformations are straight (Figure 7C), joining the helices I and II, while 25 conformation are bent two helical turns from the C-terminus, in the opposite direction than orexin-B (Figure 7E). Perhaps depending on the experimental conditions, another study describes orexin-A in a 90°-bent conformation where the "hinge" is similarly located, but the orientation of the N-terminus is oriented in another direction by 90° (Figure 7D).[95]

Truncation studies on the orexin have shown that the conserved C-terminus of the peptides is vital for biological activity,[90] as suggested by the conservation of the C-termini between the peptides. Even one-amino-acid C-terminal truncation yields an inactive peptide, and the switch from amide the carboxyl has the same effect.[90] In contrast, N-terminal truncations are tolerable.[90–92,96,97] Successive truncations display reductions in potency, until peptides shorter than 19 amino acids fail to reach maximal receptor activation. It is noteworthy that the entire N-terminal section with the short helix and the disulfide-bridge-stabilized turn can be removed from orexin-A while retaining the biological activity.

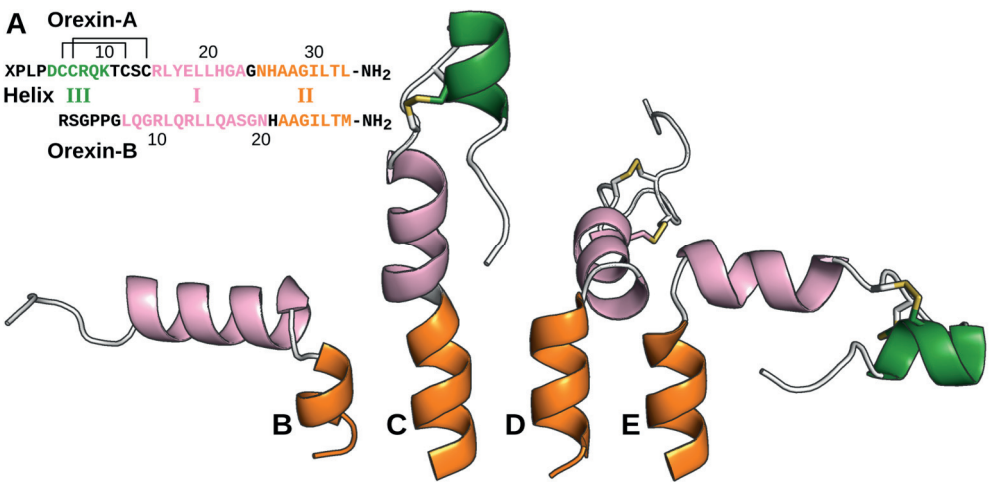


Figure 7. Orexin peptides. A) Sequences, disulfide bridges and locations for helices; B–E) Solution NMR structures of Orexin-A (C–E) and Orexin-B (B).

Within the helices I and II of the orexin peptides, hydrophobic residues are grouped on one side, while the other is governed by polar residues (Figure 8B). This gives the peptides a strong amphipathic nature. It has been postulated that this would be a requirement for receptor interactions.[93] The amphipathicity also hints towards membrane interactions, as the membrane surface would provide a suitable water–lipid interface for the orexin peptides. This has implications for ligand-binding kinetics; a peptide "seeking" for a receptor could first settle on top of the membrane, and diffuse along the two-dimensional space in the search for the receptor. This is beneficial over the direct approach from the solvent, where there are more decrees of freedom for the peptide movement and thus the binding would introduce a larger decrease to entropy. The pre-coupling with the membrane would also reduce the number of water molecules the peptide needs to shed upon entering the receptor. Also, from the statistical point of view, in search of a membrane-bound receptor, it makes sense to limit the search to the membrane.

Several mutation "scans" have been performed on the orexin peptides (Tables 4 and 2006, Figure 8A).[90–92,96,97] Alanine scan, where each residue in turn is replaced by alanine, is the most common, and for both peptides, there are two individual studies reporting such scans. Orexin-A has been used as 14–33[91] or 15–33 fragment,[96] and orexin-B either in full[92] or as 6–28 fragment.[90] In addition, orexin-B has been subjected to a full D-amino-acid scan[92] and a proline scan,[90] while only the two C-terminal residues of orexin-A have been switched to D-amino acids.[96] German and co-workers[97] have carried out site-directed mutagenesis on orexin-$A_{17–33}$. Apparently this shorter fragment is more sensitive to mutations, as the results are incomparable to the body of other data. Especially the potency of orexin-A at OX_2 was very low, and the effect of most mutations was significantly larger than in other studies. Thus, these results are not included in the tables.

Taken together, the site-directed mutagenesis studies follow the trend set by the truncation studies; the C-terminus is the most important section for activity (Tables 4 and 2006). The most pronounced effect is tied with five C-terminal amino acids. There, any mutation is highly detrimental to biological activity, except for L26 in orexin-B, which is tolerated by both receptors. On mutations T27A and M28A on orexin-B, two studies disagree; one shows drastic drops for both mutations on both receptor subtypes,[92] whereas another reports only slight loss of potency for T27A at OX_2 and for M28A at both receptor subtypes.[90] Apart from the C-terminus, Leu20/15 is vital for bioactivity. An alanine mutation results in moderate-to-large decrease in

potency for the OX_1, and in small-to-moderate decrease in OX_2. A proline mutation is also intolerable, which is not surprising as the residue is in the middle of a helical segment. However, a D-Leu mutation here has no effect on the biological activity. Except for Leu20/15 and Leu19/14, amino acids upstream of His26/21 appear to tolerate alanine (or alanine to glycine) and D-amino-acid mutations quite well. Some proline mutations are less tolerated.

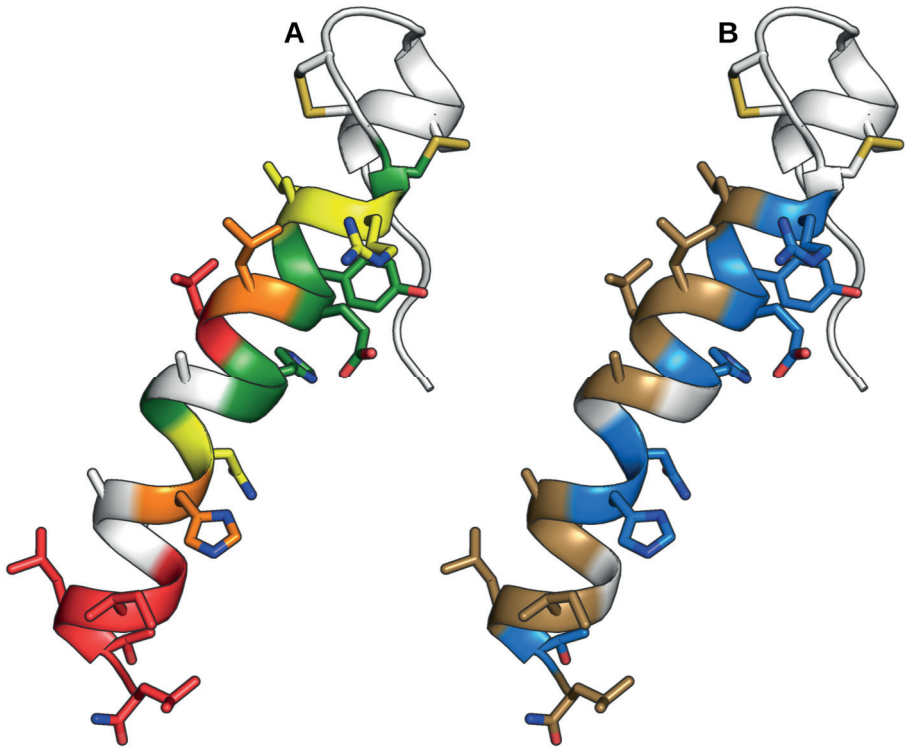


Figure 8. The important amino acids and the amphipathic nature of orexin-A. A) Amino acids color-coded on the tolerability of an alanine mutation. Red, orange, and yellow for decreasing deleterious effect, green for no effect. B) Amino acids with hydrophobic and hydrophilic side chains in brown and blue, respectively.

Table 4. Mutations on orexin-A.

		Ala		D-WT
Site	WT	OX_1	OX_2[a]	OX_1[b]
14	C	=[a]	=	
15	R	=[a,b]	=	
16	L	=[a] / ↓[b]	=	
17	Y	=[a,b]	=	
18	E	=[a] / ↑[b]	=	
19	L	=[a] / ↓↓[b]	=	
20	L	↓↓[a,b]	↓	
21	H	=[a,b]	=	
22	G	=[a,b]	=	
23	A			
24	G	=[a,b]	=	
25	N	=[a,b]	=	
26	H	↓[a] / ↓↓[b]	=	
27	A			
28	A			
29	G	↓↓[b]		
30	I	↓↓↓[b]		
31	L	↓↓[b]		
32	T	↓↓[b]		↓↓↓
33	L	↓↓[b]		↓↓

a, Ammoun et al. 2003
b, Darker et al. 2001

Table 5. Mutations on orexin-B

		Ala / Gly for WT Ala		D-WT / D-Ala for WT Gly		Pro	
Site	WT	OX_1	OX_2	OX_1[a]	OX_2[a]	OX_1[b]	OX_2[b]
1	R	=[a]	=[a]	=	=		
2	S	=[a]	=[a]	=	=		
3	G	=[a]	=[a]	=	=		
4	P	=[a]	=[a]	=	=		
5	P	=[a]	=[a]	=	=		
6	G	=[a,b]	=[a,b]	=	=	↓	=
7	L	=[a] / ↓[b]	=[a] / ↓[b]	=	=	=	↑
8	Q	=[a,b]	↓[a] / =[b]	=	=	=	=
9	G	=[a,b]	=[a,b]	=	=	↓	=
10	R	↓[a,b]	↓[a,b]	↓	↓	↓	=
11	L	↓[a,b]	=[a,b]	↓	↓	↓↓↓	↓
12	Q	=[a,b]	=[a,b]	=	=	↓	=
13	R	=[a] / ↓[b]	=[a,b]	=	↓	↓↓	↓
14	L	=[a] / ↓[b]	=[a] / ↓[b]	↓↓	↓	↓↓↓	↓
15	L	↓↓[a] / ↓↓↓[b]	↓↓[a] / ↓[b]	=	=	↓↓↓	↓↓↓
16	Q	=[a,b]	=[a,b]	↓	↓	↓	↓
17	A	↑[a]	=[a]	=	=	↓	=
18	S	↓[a] / =[b]	↓[a] / =[b]	↓	↓	=	↓
19	G	↓[a] / =[b]	↓[a] / =[b]	=	=	=	↓
20	N	↓↓[a] / =[b]	↓↓[a] / =[b]	=	=	↓	=
21	H	↓[a] / =[b]	↓[a] / =[b]	↓	↓	=	↓
22	A	↓	=[a]	↓↓	↓↓	↓↓↓	↓↓
23	A	↓↓	↓	↓↓↓	↓↓↓	↓↓↓	↓↓
24	G	↓↓↓[a,b]	↓↓↓[a] / ↓↓[b]	↓↓↓	↓↓↓	↓↓↓	↓↓
25	I	↓↓↓[a,b]	↓↓↓[a,b]	↓↓↓	↓↓↓	↓↓↓	↓↓↓
26	L	↓↓↓[a,b]	↓↓↓[a,b]	↓↓↓	↓↓↓	↓	=
27	T	↓↓↓[a,b]	↓↓↓[a] / ↓[b]	↓↓↓	↓↓↓	↓↓↓	↓↓
28	M	↓↓↓[a] / ↓[b]	↓↓↓[a] / ↓[b]	↓↓↓	↓↓↓	↓↓↓	↓↓

a, Asahi et al. 2003
b, Lang et al. 2004

Symbol	Fold effect on peptide EC_{50}
↑↑↑	> 100
↑↑	20–100
↑	5–20
=	0.5–5
↓	< 0.5

Through mutations, few OX_2-selective peptides have been discovered. Orexin-B with L11A and L15D-Leu is 400-fold more potent at the OX_2 over the OX_1[92], while orexin-$B_{10–28}$ is even more selective (at least 1750-fold)[90]. However, [Ala[11],D-Leu[15]]orexin-B is ten-fold more potent than orexin-$B_{10–28}$. OX_1 selectivity appears to be harder to achieve, as the only peptide favoring the OX_1 is the orexin-$A_{2–33}$ with unclosed disulfide bridges with a modest three-fold preference.[90] Interestingly, German and co-workers[97] found that L20D-Leu, which corresponds to the L15D-Leu mutation in orexin-B, actually rescues some of the poor potency on orexin-$A_{17–33}$ towards OX_2, while it is detrimental to potency for OX_1.

2.3.4.1 Predictions on orexin peptide binding

In addition to the work presented here, to our knowledge only one group has attempted to elucidate the binding mode for orexin peptides. Heifetz and colleagues[85] built homology models of the orexin receptors with the D_3 receptor as a template and subjected the models to brief MD simulations. From the simulations, they picked a pool of models for both subtypes. The article is not explicit on the matter, but from the figures it appears that they elected to use orexin-$A_{15–33}$ and orexin-$B_{10–28}$ in lieu of the full peptides. The fragments were docked into the pool of models through ensemble docking, which allows the use of multiple targets. They suggest a docking pose where the peptide C-terminus lies at the level of Tyr311/317[6x48], H26 side chain faces the TM4-side of the TM5, the peptide hinge is bent 90° and the middle helix of the peptide points towards the TM7. Due to the use of D_3 as a template, which has a markedly shorter ECL2 loop than the orexin receptors, the models feature an ECL2 collapsed atop the binding site. As there were multiple peptide-binding receptor structures available at the time of their study, all displaying the β-hairpin ECL2, the choice of D_3 seems peculiar, and serves to highlight the importance of the modeling template selection. Also, while collecting several target models from an MD simulation could alleviate the potential problems linked with selecting a single model, the simulation was equilibrated only for 200 ps, which is quite short. Either due to short equilibration, or the use of the bent peptide conformations for docking, TM7 of the OX_1 model appears to feature a sharp bend, allowing the deep binding of the bent peptide. The collapsed ELC2 might also force a deep binding mode and affect the TM7. The authors do not discuss compatibility with full peptides, but it appears that especially orexin-A, with the N-terminal bulk, would most likely be incompatible with the presented binding mode.

2.3.5 Small molecular ligands

Several small molecular ligands for orexin receptors have been developed. The pharmaceutical industry has mainly been interested in antagonists, which have been researched as hypnotics. The first, and this far only, antagonist suvorexant reached the market in August 2014. Suvorexant has also been the molecule of choice for the crystallographers, being the co-crystallized antagonist in the first crystal structures of both orexin receptor subtypes, and therefore its binding interactions are well known. The binding mode of suvorexant is discussed in detail later. Suvorexant is a dual antagonist, binding both orexin receptor subtypes. Other dual antagonists have also been developed and tested in clinical trials. There are also subtype-selective antagonists for both orexin receptors, two of which have been crystallized with their targeted receptors: EMPA with OX_2 and SB-674042 with OX_1.

Concerning agonists, the academic community has been more active than the pharmaceutical industry. Most notably, the research group of Masashi Yanagisawa, who were also the first to identify orexin receptors, has patented two series of orexin receptor agonists[98,99], and published a detailed description of the work leading to a full OX_2-agonist[100]. Our research group has produced and reported azulene-based orexin receptor agonists (Additional publication V).

3 Aims of the study

The main goal of the study was to understand the binding interactions of orexin receptor ligands — foremost the natural peptides, but also synthetic small molecules, and to deduce determinants for receptor activation. To this end, we carried out peptide docking and extensive molecular dynamics simulations. To probe the bioactive conformation of orexin peptides, we produced and tested conformationally constrained orexin peptides.

The specific aims of the thesis were:

- To understand the molecular mechanism by which orexin peptides bind to their cognate receptors, and how receptor activation is triggered. (Publications I and III)
- To assess the bioactive conformation of the orexin peptides. (Publication II)
- To discover or design novel chemical compounds or (modified) peptides that bind to and affect orexin receptor signaling. (Publication II and additional publications IV and V)

4 Materials and Methods

4.1 Numbering of GPCR residues

As GPCRs share a common overall fold, it is often reasonable to compare homologous amino acid residues from different receptors. To facilitate this, Ballesteros and Weinstein suggested a unified labeling convention more than two decades ago.[101] They proposed that since all transmembrane helices have conserved regions, in each helix the most conserved amino acid would be denoted n.50, where n stands for the number of the transmembrane helix from the N-terminus. Other residues on the helix would be numbered consecutively using this midpoint as a reference. This convention was widely accepted and remains still in active use. However, the Ballosteros–Weinstein system considers only the protein primary structure with predicted locations for the helices. The recent surge in resolved GPCR 3D structures has shown that, on closer inspection, the conformation is not identical between receptors; the helices often show local constrictions and bulges as results of amino acid deletions and insertions. Should the Ballosteros–Weinstein convention be used, a homologous amino acid would receive a different number in two receptors if, in either of them, there was an insertion or deletion between the amino acid in question and the denoted central amino acid. To take this structural aspect into account, the GPCRdb project has introduced an "updated" numbering scheme[102], where the central amino acids remain the same as in Ballosteros–Weinstein scheme, but other amino acids are numbered based on their superposition on an ideal α-helix. In case of a constriction, the "missing" residue number is simply skipped, whereas in the case of a bulge, the "extra" residue is given the number of the previous residue appended with 1, e.g. 5x461 for the conserved bulge after the residue 5x46.

4.2 Homology models of the orexin receptors

As the orexin receptor crystal structures[8,83] were solved during the research project, we initially had to rely on homology models. Also, when the first crystal structure was published, it was of OX_2, while we were currently working with OX_1, and additionally, few amino acids were missing from the extracellular loops, so homology modeling remained necessary throughout the research project.

4.2.1 Template selection and sequence alignment

In publication I, we derived a sequence alignment for all 19 class A GPCRs* that had a published 3D structure by superimposing the 3D structures and visually inspecting which amino acids corresponded between the receptors. This served two purposes; we could identify structurally conserved regions (both in terms of sequence and secondary structure) among the receptors to guide the following sequence alignment of OX_1 to the sequences of the crystallized receptors, and we were able to find locations where intrahelical amino acid insertions and deletions had taken place, in case we would need to introduce such gaps in our alignment of OX_1 and the template to be selected. We added the sequence of the OX_1 based on the conserved residues in each transmembrane helix. For the ECL2, we noticed that all published structures of peptide-binding GPCRs (CXCR4, NTS_1, and opioid μ, κ, δ and NOP)[38,103–107] featured a β-hairpin structure between TMs 4 and 5, which was constrained above TM3 by the conserved disulfide bridge. We made sure that the alignment of the OX_1 to the crystallized receptors had no gaps within the expected β-strands.

For templates we chose the rat NTS_1[38] and the CXCR4[107]. The NTS_1 was the first choice, as it lies in the β-branch of the GPCR phylogenetic tree with the orexin receptors.[1] However, while NTS_1 is also activated by a peptide ligand, neurotensin is smaller than orexin peptides (13 amino acids in neurotensin versus 28 and 33 in orexin peptides). As the NTS_1 binding cavity is constricted by an inward tilt of the extracellular end of the TM6, we had doubts whether the α-helical orexin-A would fit in. Therefore, we selected the CXCR4 as a secondary template, since the structure shows a more open binding cavity, as befits a receptor whose cognate ligand is a small protein. We also constructed a hybrid template by rotating the TM6 in the NTS_1

*Class A GPCRs available at the time, and PDB identifiers thereof:

Rat NTS_1 receptor, 4GRV[38], Mouse μ receptor, 4DKL[103], κ receptor, 4DJH[105], Mouse δ receptor, 4EJ4[104], NOP receptor, 4EA3[106], CXCR4, 3ODU[107], $5\text{-}HT_{1B}$ receptor, 4IAR[108], $5\text{-}HT_{2B}$ receptor, 4IB4[109], β_2-adrenoceptor, 2RH1[37], Turkey β_1-adrenoceptor, 2VT4[110], D_3 receptor, 3PBL[111], H_1 receptor, 3RZE[112], M_2 receptor, 3UON[30], Rat M_3 receptor, 4DAJ[113], A_{2A} receptor, 3EML[114], PAR1, 3VW7[115], Squid rhodopsin, 2Z73[116], Bovine rhodopsin, 1U19[117], $S1P_1$ receptor, 3V2Y[118]

structure to match the TM6 in the CXCR4 structure. Both template structures display 23.6% sequence identity to OX_1, which would often be considered poor for homology modeling. However, we were confident that the conservation of the overall fold would compensate for this.

As the OX_2 structure[83] was published during the peer-review process of Publication I, we included that in the study. The sequence alignment was of course straightforward in this case, as the sequence identity of the TM bundles of the OX_1 and OX_2 is close to 80%, and nothing suggested that there would be any insertions or deletions. In the Publication III, we used homology modeling to fill in the few missing residues in the OX_2 structure. Here, the sequence alignment was trivial, as we used the OX_2 as both target and template.

4.2.2 Model production and evaluation

For the Publication I, we produced ten OX_1 models from each of the templates, consisting of residues Tyr41^{1x27}–Gln246^{5x69} and Arg291^{6x28}–Cys375^{8x60}, and in the Publication III, 30 models of the OX_2 comprising Pro50^{1x28}–Gln254^{5x69} and Lys294^{6x25}–Cys381^{8x60}. We did not include the receptor termini or the ICL3, as suitable templates were not available at the time. The C-terminus and ICL3 would also have had little impact in the ligand binding interactions, which was the main focus. We used MODELLER[119] 9v8 and 9.14 with default settings for model construction. In Publication I, we evaluated the models on the basis of ECL2 and ECL3 conformations, as these were the main variable sections. We sought models where the turn of the ECL2 hairpin resembled the crystallized peptide-binding GPCRs, and the ECL3 did not constrict the binding site entry. From each template, we selected one model. In the Publication III, we selected the model with the smallest RMSD in comparison to the template structure 4S0V[83].

4.3 Peptide docking

For all publications, we used the straight conformation of the orexin-A[93]. The bent conformation is dominant in the published NMR structures, but our preliminary docking suggested that it would not have been possible for the bent peptide to fully reach the predicted binding site, unless the receptor structure deviated from the crystallized peptide-binding GPCRs. However, modeling such receptor conformation without a suitable template would not have been feasible.

In Publication I, we decided to use the orexin-$A_{16–33}$ fragment to avoid the N-terminal disulfide-bridge-stabilized hook from "colliding" with the receptor extracellular loops and thus limiting the sampling of the peptide C-terminus within the binding cavity. In Publication III, we used the orexin-$A_{2–33}$ in docking, since the best poses from Publication I, which we sought to replicate, were achievable also for the full-length peptide. However, we were forced to omit the first residue, pyroglutamate, as the docking protocol did not include the required parameters. For the simulation, we selected two high-scoring docking poses that represented the two docking modes discussed in Publication I.

We used ZDOCK[69] and RDOCK[120] with default settings to carry out the peptide docking and refining. These tools are initially built for protein–protein docking, but they have found use also in peptide docking, since purpose-built peptide docking software has been available mainly for short, non-helical peptides and exposed binding sites. ZDOCK performs an exhaustive rigid-body docking and performs pose clustering. We excluded the poses where the peptide C-terminus did not interact with the receptor, or the peptide traversed between the transmembrane helices into the membrane space. The rest of the poses were passed to RDOCK for interaction optimization and re-scoring. We used Discovery Studio[121] as an interface for ZDOCK and RDOCK, and for the visual examination of the results.

4.4 Analysis of peptide location and orientation

For publication I, we sought to characterize the preferred binding area and space available for the orexin peptides. We approached this through an RMSD-based clustering of the docking poses. For the pooled docking poses, we calculated a full pairwise Cα-RMSD matrix with Matlab[122]. Then, we extracted modelwise clusters with an algorithm[123], where the pose with the largest number of close neighbors (RMSD < 3 Å) — the cluster seed — is extracted from the pool with all its neighbors to create the a cluster. This is repeated until no two poses are within the cutoff. Each cluster was assigned the median RDOCK score of its members. For visualization of the cluster locations, we used the cluster seeds as representative poses. To illuminate the cluster packing and distribution, we produced a 2D representation of the RMSD matrix via multidimensional scaling.

In Publication I, we also assessed the solvent accessible surface area of the docked peptides with Naccess[124], and measured their binding depth as a distance along the z-axis (set to be perpendicular to the membrane plane) between the peptide

L33 Cα and the Cα-plane of Thr223^{5x461}, Tyr311^{6x48} and Tyr348^{7x42}. Direction of the peptide H26 side chain, calculated as a vector in xy-plane from A28 Cα to H26 Cα, served as a marker for the peptide rotational orientation.

4.5 Analysis of binding interactions

In Publication I, we considered any distance below 4 Å between non-hydrogen atoms in the receptor and in the peptide as a contact between the respective amino acid residues. For a group of similar docking poses, the interactions were collected in heatmaps to identify frequent interactions, and a representative pose was selected, in turn, by checking which of group member best reproduced the frequent interactions. The analysis was implemented entirely in Matlab. From the representative poses, we identified atom-to-atom interactions visually.

In Publication III, we carried out a more detailed analysis of binding interactions. We began with a similar distance-based mapping as in Publication I, again with a 4 Å cutoff, using the Gromacs tool *gmx mindist*. This served to highlight peptide side-chain orientations and hydrophobic contacts. In addition, we analyzed both direct and water-mediated hydrogen bonding with the Gromacs tool *gmx hbond* with default settings (donor–acceptor distance <3.5 Å, donor–hydrogen–acceptor angle <30°). Direct hydrogen bonding was trivial as the tool is intended for the very task, but for the water-mediated interactions, we devised an in-house Matlab-script. The script functions as follows:

1. Search all hydrogen bonds between water and the receptor–ligand complex (*gmx hbond*).
2. Discard water molecules that never have more than one hydrogen-bonding partner at a time.
3. For each water molecule, identify the simulation frames where multiple bonding takes place.
4. For each frame, identify the bonding partners and record the existence of the bridged interaction at this frame.
5. Convert atom-to-atom bridges to residue-to-residue bridges.

In a molecular dynamics simulation, the water molecules rarely remain in the same area through the simulation, but exchange rapidly with the bulk solvent. Therefore, the script described above does not consider the bridging water molecule

to be a part of the bridged interaction, but records only the amino acid heads of the bridge.

We treated the small molecules in Publication III similarly, but instead of treating the whole molecule as one entity, we divided them into fragments for the analysis. As far as the computer programs were concerned, these fragments were treated just as amino acids were.

The entire simulation trajectory was analyzed for the direct interactions, while frames at 3 ns intervals were used for the water-mediated interactions.

4.6 Molecular dynamics simulations

4.6.1 Overview

In Publication III, we produced a total of 36 µs of MD simulations on different ligands in combination with the OX_2 receptor in two different membrane compositions. These are summarized in Table 6. For the system setup and analysis we employed Gromacs 5.1[125], and for the MD simulations Gromacs 4.6.7[126].

Table 6. Molecular dynamics simulations

Ligand	Membrane	Length (µs)
Apo	POPC	3
Suvorexant	POPC	3
Nag26	POPC	3
Nag26	POPC	1
Orexin-A (TM5)	POPC	3
Orexin-A (TM5)	POPC	1
Orexin-A (TM7)	POPC	3
Orexin-A (TM7)	POPC	1
Apo	POPC-CHOL (25%)	3
Suvorexant	POPC-CHOL (25%)	3
Nag26	POPC-CHOL (25%)	3
Nag26	POPC-CHOL (25%)	1
Orexin-A (TM5)	POPC-CHOL (25%)	3
Orexin-A (TM5)	POPC-CHOL (25%)	1
Orexin-A (TM7)	POPC-CHOL (25%)	3
Orexin-A (TM7)	POPC-CHOL (25%)	1

4.6.2 Force field parametrization and simulation protocol

For the protein and peptide, we selected the Amber99sb-ildn force field,[127] as it was the most recent Amber force field available. For the membrane lipids we used the Splipids parameters,[128] which were developed to be used in combination with Amber force fields. As Amber does not contain parameters for the pyroglutamate (the first residue in orexin-A), we created the connectivity manually, assigned regular carbonyl atom types for the lactam carbonyl, and used partial charges from the Tinker molecular modeling package.[129]

For suvorexant, we were able to produce bond parameters directly from Antechamber[130] with Gaussian-derived[131] RESP charges. Nag26 parametrization was not as straightforward. Antechamber provided a good template, but the parameters for many of the "basic" bonds, such as the aromatic C–C bond, appeared to be slightly different from regular Amber parameters. As Nag26 features mostly quite common bonds, we were able to mainly use native Amber bond and angle parameters, and resorted to OPLS/AA if suitable Amber parameter was not found. This mixing is possible as Amber and OPLS/AA are very similar. For dihedrals, we used mainly the parameters from Antechamber, as they were mostly identical with native Amber or differed only slightly. For sulfonamide, we substituted OPLS/AA parameters, as they favored a 90° angle for the N-S-C-C torsion and 90°/270° angle for the X-N-S-X torsion, which is supported by crystal structures.[132] Also for the torsions of the aromatic amine, we used OPLS/AA parameters as Amber and Antechamber provided none. For partial charges, we used AM1-BCC.[133]

We used periodic boundary conditions, and selected a hexagonal prism for the simulation box, as it offers a 14.4% reduction in volume in respect to a rectangular prism with similar periodic distance. Time-step was 2 fs, center-of-mass translation was removed every 10 steps (solvent treated separately from the membrane with the embedded components) and Verlet pair-list was updated every 20 steps. We used the physiological temperature of 310 K (Nose–Hoover thermostat for solvent, membrane, and protein complex separately), and pressure of 1 bar independently for the membrane xy-plane and z-dimension (Parrinnello–Rahman). Short-range interactions were cut off at 1 nm, dispersion correction applied for long range pressure and energy and PME with default settings (cubic, 0.12 nm Fourier spacing) for long-range electrostatics. LINCS constraints were used for all bonds.

4.6.3 System setup, equilibration and production

The molecular dynamics simulation systems in Publication III comprised the OX_2 receptor model, an optional ligand, membrane model, solvent and ions. The process of receptor modeling and peptide docking is described above.

4.6.3.1 Small molecule placement

Coordinates for suvorexant were extracted from the crystal structure 4S0V[83] after superimposition to the OX_2 receptor model. Nag26 was docked with Glide induced-fit into the binding site of the crystal structure 4S0V, where suvorexant, but not waters, had been deleted (Additional publication IV). Poses were then ranked by their similarity to suvorexant binding. To maintain consistency, we did not use the induced-fit-treated receptor from the docking, but again placed the Nag26 into the OX_2 model.

4.6.3.2 Membrane and system assembly

We constructed two hexagonal membranes for the simulations with Charmm-GUI.[134] In the pure POPC (*1-Palmitoyl-2-oleoylphosphatidylcholine*) membrane, there were 150 lipids per leaflet, and in the mixed POPC-cholesterol membrane, we used 126 POPC molecules and 42 cholesterol molecules per leaflet to yield 25% cholesterol content. Both membranes were solvated into 12 nm high box of TIP3p water.[135]

As cholesterol is shown to favor certain locations around GPCRs, we carried out a coarse-grained (CG) simulation to decide initial cholesterol locations. In a CG simulation, groups of atoms are replaced with single particles. Less particles means fewer calculations per time frame, so CG simulations can achieve remarkably longer time scales. The OX_2 receptor was embedded into three POPC-cholesterol membranes (10, 25 and 50%) and simulated for 10 μs in each membrane. The simulations revealed eight preferred locations for cholesterol molecules around the receptor. From a suitable frame, we back-mapped the CG receptor and the surrounding lipids into an all-atom system.

To embed the receptor complex into the membrane, we used the Gromacs tool *membed*. The tool shrinks the embeddable complex with a given factor (we used 10% of the original size in the membrane plane, and original size along the z-axis), superimposes the shrunk structure on the membrane, deletes the overlapping lipids,

and gradually "inflates" the complex to its original size, pushing membrane lipids out of the way. For the POPC membrane, we aimed at a symmetrical deletion of three lipids per leaflet, but the slightly different center-of-mass at the peptide-bound receptors caused the shrunk complex to be translated approximately 0.4 Å, resulting in the deletion of an additional lipid from the intracellular leaflet. In the case of the cholesterol-containing membrane, each receptor was first joined with the close lipids extracted from the CG simulation. As this complex is larger and the close lipids were not symmetrically distributed – five cholesterol molecules in the extracellular leaflet and three on the inside, along with 15 POPC in both leaflets, and the target is a multicomponent membrane, we ensured symmetrical outcome by pre-creating a circular hole in a suitable place at the membrane by deleting 18 POPC molecules from both leaflets, along with 6+4 cholesterol molecules. We then embedded the receptor–lipid(–ligand) complex into the pre-generated hole with same settings as above.

Finally, we neutralized the system charge and added 100 nmol/l of NaCl. The final systems contained the receptor, optionally either orexin-A, suvorexant or Nag26, 293–294 POPC or 246 POPC + 82 CHOL lipids, 25 100–25 200 water molecules, 75 Na ions and 88–89 Cl ions (the orexin-A carries a positive charge).

4.6.3.3 Equilibration and production

The equilibration scheme is presented in Table 7. Each phase was simulated for 10 ns. The equilibrated systems were then simulated for 3 µs. Additionally, we started 1 µs replica simulations with the same files for the simulations with agonist ligands (see Table 6).

Table 7. Positional restraints during the equilibration.

Phase	Restrained atoms	Strength kJ/(mol x nm^2)
1	Protein, ligand and cholesterol heavy atoms	1000
2	Protein and ligand heavy atoms	800
3	Protein and ligand heavy atoms	600
4	Protein and ligand heavy atoms	400
5	Protein and ligand heavy atoms	200
6	All Cα and small molecule heavy atoms	200
7	Helical or peptide Cα, and small molecule heavy atoms	200

For the equilibration and production simulations, we used the Sisu supercomputer from the CSC – IT Center for Science. Each simulation was distributed over 336 CPU cores, averaging approximately 130 ns/day, and thus totaling a little over nine months of wall-clock calculation time and 255 years' worth of CPU time. The system setup and analysis were carried out on a typical desktop computer.

4.6.4 Analysis

The analysis on the receptor and the ligand-binding interactions was conducted with Gromacs tools[125], VMD[136] and Matlab[122]. For analysis, we centered the protein complex and packed the membrane and the solvent into a hexagonal prism. Then, we aligned the simulation frames based on the receptor Cα with such settings that the outcome was a smooth, continuous trajectory with no jumps. Most analysis was carried out on these trajectories, with the following exception: due to limitations of the analysis software and the physical RAM, the analysis concerning water, and the preferred cholesterol locations were done with frames every 3 ns. In the analysis of membrane thickness and area per lipid, we used unaligned coordinates, as the receptor-based alignment tends to tilt the membrane and throw off the analysis.

We analyzed several aspects of the receptor and its interactions, as well as few membrane properties. RMSD (root-mean-square deviation) was calculated both for the receptor and all ligands. An RMSD value describes the similarity of two conformations of the same structure. First, the comparable conformations are superimposed, then a square-root of the sum of all squared pairwise distances is calculated. For amino acids, we considered only the Cα, as it appeared to reflect the conformational differences as well as all-atom or heavy atom calculations. For small molecules, we took into account only heavy atoms. The alignment preceding the calculation has a large effect on the value. Should a rigid ligand move around in the receptor binding site, two outcomes are possible for the calculation: if the alignment is based on the receptor, RMSD for the ligand atoms reflects the ligand motions, but if the alignment considers only the ligand, RMSD would be close to zero, as the ligand conformation remains close to the initial conformation. Thus, we produced two RMSD values for the ligands: "external" RMSD stands for receptor-based alignment and therefore describes the stability of ligand position, rotation and conformation, and "internal" for the ligand-only alignment, which describes the

conformational stability. For the receptor and orexin-A, we also calculated RMSF root-mean-square fluctuation), which highlights the most mobile regions.

The analysis of ligand-binding interactions is described above. As the interactions between receptor residues are also of interest, we used the same methods to map receptor–receptor hydrogen bonds and water-mediated interactions. At the "core triad" area, we also followed the distance and angle of the residues Val142^{3x40}, Pro235^{5x50} and Phe313^{6x44}, and close to the G protein binding site, the interaction distances Leu310^{6x41}–Val240^{5x55} and Tyr364^{7x53}–Ile148^{3x46}, which are conserved in the activated GPCRs structures, and the distances Ile148^{3x46}–Leu306^{6x37}, Val75^{1x53}–Tyr364^{7x53}, Tyr364^{7x53}–Phe371^{8x50} and Asn365^{7x54}–Arg372^{8x51} conserved in the inactive GPCRs. In addition, we monitored the salt bridges observed in the orexin receptor crystal structures, namely Asp211^{45x51}–Arg328^{6x59}, Glu212^{45x52}–His224^{5x40}, Arg339^{7x27}–Glu118^{2x67}, and Asp115^{2x64}–His350^{7x38}.

To estimate stable locations for water cholesterol molecules through the simulations, we used the VMD tool *Volmap* with default settings.

For the membrane, we calculated the area per lipid and membrane thickness with GridMAT-MD[137]. As the tool assumes a rectangular box, we dealt with the edge-problems by surrounding the simulation box with periodic images in the xy-plane, and considered only the lipids in the "central" image for the analysis. As an estimate for the membrane thickness, we calculated an area-weighted mean of the P–P distance for the POPC headgroups. The order parameters for the POPC tails we calculated with *gmx order*, treating the double bond as described by Pluhackova et al.[138]

4.7 Design of stapled peptides

In Publication II, we report a series of conformationally constrained peptides, where a hydrocarbon linker is placed between two amino acids one helical turn apart to force an α-helical conformation. This process is called stapling.[139] For the ease of synthesis, we used orexin-$A_{15–33}$ instead of the full peptide. The usage of a truncated peptide might also allow us to observe if we could "rescue" some of the potency lost with the truncation.

As our goal was to stabilize the straight α-helical conformation of orexin-A, our initial stapling location was the pair of alanine residues A23 and A27 (Figure 9). A staple here spans the flexible linker between the C-terminal and central helix and

should “lock” the straight helix, as well as provide overall stability to the helix. We also placed staples at A27–L31 close to the C-terminus, and at L19–A23 at the N-terminus of the fragment we used. A staple near the end of a helix might provide stability by stopping the “fraying” of the helix. The three positions all reside at the hydrophobic face of the orexin peptide, which is suitable for a hydrophobic staple. However, we decided to include a control position on the reverse side of the peptide, which is predominantly polar. For this, we selected the site N25–G29.

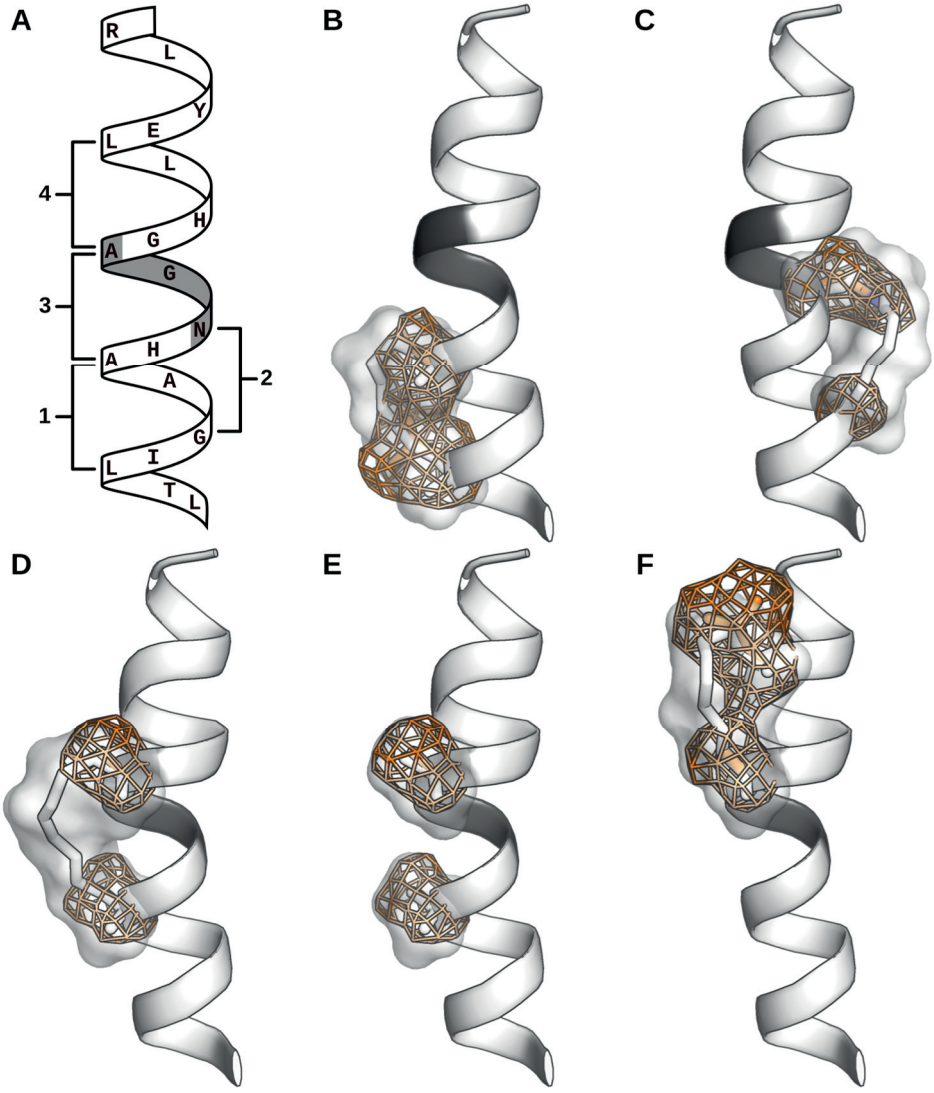


Figure 9. Peptide stapling sites on the orexin-$A_{15–33}$. A) Schematic of the stapling sites. B–F) Comparison of modifications to wild-type side chains in terms of volume. E shows the Aib-variant, others the canonical eight-atom staple. The C-terminus faces down, the flexible hinge is in gray, the wild-type side chains and their volume in orange sticks and mesh, the modifications in white sticks and white volumes.

The peptides were assayed for their ability to increase intracellular Ca^{2+}-concentration via the orexin receptors, and also for their ability to block the Ca^{2+}-elevation induced by sub-EC_{50} concentrations of orexin-A.

5 Results and Discussion

5.1 Orexin peptide binding

We approached orexin peptide binding through homology modeling of the OX_1 receptor, using mainly the rat NTS_1 and human CXCR4 as templates, until the crystal structure of the OX_2 was released, at which point we adopted it as the docking target and the template for OX_1 modeling. For the ligand, we opted to use orexin-A, first as the orexin-$A_{15–33}$ fragment, then as full-length peptide. For the conformation, we selected the straight α-helical conformation of the orexin-A, as it seemed improbable that the bent conformations shown by the NMR structures would penetrate deep enough to interact with residues such as the Tyr311/317^{6x48}, Val130/138^{3x36}, and Tyr224/232^{5x47}. The validity of this hypothesis for the bioactive conformation is discussed in detail below (5.2.2). However, as we have concentrated our binding interaction analysis efforts at the peptide C-terminus, the results should not be heavily influenced by the hinge conformation and the interactions towards the peptide N-terminus.

5.1.1 Peptide docking

5.1.1.1 Homology models

In the Publication I, we initially constructed three OX_1 models based on three template structures: NTS_1, CXCR4, and a hybrid between the two, NTS_1 helices and side chains but with CXCR4-sized binding site. During peer review, the crystal structure of OX_2 was released, so the final article mainly considered the NTS_1-based model in addition to an OX_2-based model, while the two other models were demoted to "secondary models". The NTS_1-based model was quite accurate in terms of overall fold; extracellular ends of the TMs 5 and 6 lean slightly inwards and the ECL3 and the top of ECL2 differ in comparison to the OX_2 crystal structure. At side-chain level, there were multiple differences in rotamers, most notably Gln126^{3x32}, His216^{5x40}, Tyr337^{7x31}, Phe340^{7x34} and His344^{7x38}.

5.1.1.2 Docking

We used orexin-$A_{15–33}$ instead of the full peptide. The docking protocol treated both ligand and peptide as rigid bodies, and we anticipated that possible collisions of the peptide N-terminal hook into the receptor extracellular loops would restrict the

mapping of the C-terminus, which was our main interest. In reality, the receptor loops and the peptide N-terminus would be flexible and would thus allow C-terminal docking poses that the rigid-body-docking would miss.

Either due to the differences in the TMs 5 and 6 of the models, or the side-chain rotamers, the two models (OX_2- and NTS_1-based) produced and favored differing pools of docking poses. The docking of orexin-$A_{15–33}$ into the OX_2-based model produced a tight bouquet of docking poses, where the high-scoring poses were dominantly upright or leaned slightly towards the TM5. In contrast, the NTS_1-based model produced a fan-shaped pool of docking poses, most likely because of the narrower binding cavity due to the inward tilt of the TM6. The top-scoring clusters showed a similar fan-like distribution, but the top-scoring individual poses were again upright in the middle of the binding site. In both models, a good docking score was tied to deep binding, but in reverse, deep binding did not guarantee a good score.

An examination of the high-scoring individual poses from both models revealed that there were two main clusters of docking poses, separated by a rotation of ~100° around the helical axis. We figured the reason for this division was that the orexin-A C-terminus has two bulky residues, His26 and Ile30, facing the same direction, and there are limited pockets available at the binding site large enough to house this bulk. We selected the His26 side chain direction as a rotational marker and noticed that it faced the TM5 in one of the main clusters and the TM7 in the other. Therefore, we termed the two clusters TM5 and TM7 binding mode, accordingly. Both models produced high-scoring poses with both binding modes, but the OX_2-based model favored the TM7 binding mode over the TM5 binding mode, whereas in the NTS_1-based model, the preference was reversed. Within the binding mode clusters, there are also small differences in peptide tilting. In the TM5 binding mode cluster, the NTS_1-based model shows one bundle of poses that lean slightly towards the TM5 and another bundle, where the peptide is slightly deeper and upright, whereas all poses from the OX_2-based model share the position of the latter group. In the TM7 binding mode cluster, both models produce only one bundle of poses each, both leaning slightly towards the TM1, with the poses from the OX_2-based model a bit deeper.

We mapped common interaction patterns for both clusters and compared them to the list of receptor amino acids that have been highlighted by SDM studies, and also assessed the complementarity of hydrophobic and hydrophilic regions between the peptide and the receptor model. On these accounts, both binding modes had

merits, and as both binding modes appeared also to be compatible with full-length orexin-A, we reported both binding modes as equal alternatives. For detailed descriptions, see Publication I.

5.1.2 Molecular dynamics simulations on bound orexin-A

In Publication III, we set out to assess the stability of the binding modes identified in the Publication I. As the binding modes had turned out to be compatible with the full-length peptides, we begun by docking the orexin-A peptide into an OX_2-receptor model, only this time as a full-length peptide. From the resulting poses, we selected two high-scoring poses that were close matches to the previously reported binding modes. As there were sub-groups within the TM5 binding mode cluster in the Publication I, we used the poses derived from the OX_2-based model as references when selecting the full-length docking pose (Figure 10). The bound peptides were then simulated in two different membrane environments for a combined total of 16 μs.

The initial straight conformation of the orexin-A peptide did not appear to be stable without the external support of the peptide N-terminus binding to the receptor ELC2. The bending of the peptide, whether bound or free of the ECL2, dominated the RMSD calculations, masking the events at the interesting C-terminus. Thus, we analyzed only the peptide C-terminus.

In the TM5 binding mode, the peptide C-terminus was stable throughout the simulations, but in the cholesterol-membrane simulations, it moves slightly in the extracellular direction. In contrast, the binding of the peptide in the TM7 binding

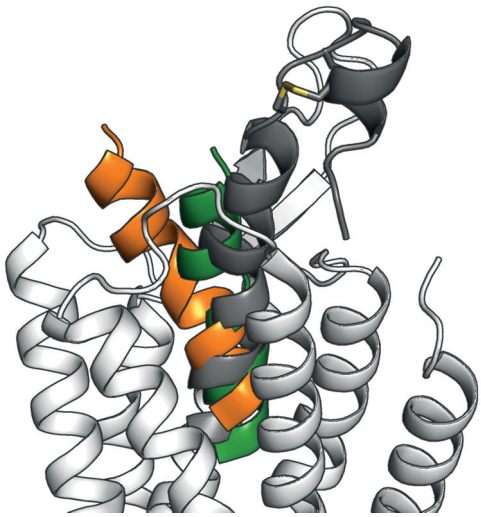


Figure 10. Selection of TM5-mode docking pose for the simulations. Two "middle" poses from Publication I, one for the NTS_1-based model, the other for the OX_2-based model, shown in orange and green, respectively. The selected starting pose for simulations in gray.

mode was unstable in the simulations. Both the peptide conformation and the location were less stable. Often, the hinge region would collapse towards the TM1, and in the 3 µs simulation in the cholesterol-containing membrane, also the helical structure of the C-terminus melts within the binding site. Due to this significant difference in the C-terminal stability, we feel that the TM5 binding mode is likely a more relevant candidate for the orexin peptide binding mode. Therefore, the detailed binding interactions are presented only for the TM5 binding mode. For further description of the TM7 binding mode through the simulations, please refer to the Publication III. However, as many peptides bind their GPCRs in an extended or partially extended conformation (Figure 4), our preference for the stable α-helix may later turn out to be misplaced.

As the starting conformation for the TM5-binding-mode simulation in Publication III was selected to recreate the TM5 binding mode from the Publication I, and the peptide C-terminus remained stable through the simulations, it is not surprising that the pattern of binding interactions is similar between the two studies. However, the simulation had more freedom in the movement of amino acids and in the optimization of interactions than the rigid-body docking. Also, the docking protocol treats solvent implicitly, whereas the simulation featured an all-atom solvent, allowing for the identification of water-mediated interactions. The binding interactions observed with the TM5 binding mode simulation are presented in the Table 8 and Figure 11.

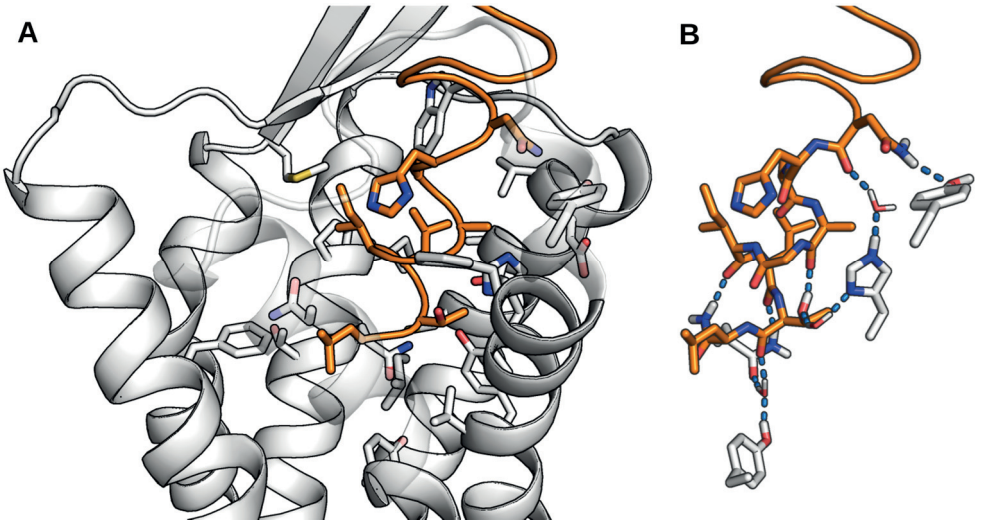


Figure 11. The TM5 binding mode. A) Overview of the binding pose. Part of the TM6 and the ECL3 are transparent for clarity. B) Hydrogen-bonding between the peptide and the binding site is often mediated by water.

Table 8. The binding interactions in the TM5 binding mode.

Residue	Interactions
Asn25	Tyr^{7x31} (Often H-bond between side chains, sometimes also to Asn25 main-chain carbonyl.) Direct or water-mediated H-bonding with Asp^{2x62} and/or Glu^{2x67}. One simulation maintains an interaction from the starting conformation: H-π bond from Asn25 NH_2 to Phe^{7x34}
His26	Packing or H-bonding with ECL2 (Asp^{45x51} and Glu^{45x52}) or ECL3 (Lys^{6x58} and Phe^{7x34}). Cholesterol-membrane simulations show local tightening of the helix, relocating the His26.
Ala27	Faces ECL1 but makes no defined interactions.
Ala28	In a pocket lined by His^{7x38}, Thr^{2x60}, Val^{2x63}, and the aliphatic chain of Asp^{2x64}.
Gly29	Allows for close packing of His^{7x38} against the side of the peptide and the A28–G29 amide bond.
Ile30	Between Met^{4x65} and Pro^{3x29}. Main-chain carbonyl binds with the C-terminal amide.
Leu31	Side chain in a pocket lined by Pro^{3x29}, Ile^{3x28}, Trp^{23x50}, and Thr/Ser^{2x60}. Main-chain carbonyl H-bound with Gln^{3x32}.
Thr32	Side chain: Tyr^{7x42}, Val^{7x41}, His^{7x38}, Ile^{6x51}, and Thr^{2x60}. H-bond mostly to Ala28 main-chain carbonyl, unless His^{7x38} flips upward, exposing a suitable binding partner (in cholesterol-membrane simulations). Main-chain carbonyl: a single water mediates interactions to Gln^{3x32} side-chain carbonyl and Tyr^{6x48} hydroxyl.
Leu33	Side chain in a pocket lined by Ile^{6x51}, Asn^{6x55} and Phe^{5x43}. Main-chain carbonyl interacts transiently with water beneath the peptide.
NH_2	Mostly binds to Gly29 or Ile30 main-chain carbonyls, transient interactions with water beneath the peptide.

A keen eye will notice that the interactions listed in the Table 8 are somewhat different from the Table 1 and Figure 9 in Publication I. This is mostly because the interactions in the Publication I were defined from a single representative pose. As the TM5 binding mode was more popular among the docking poses from the NTS_1-based model, the representative pose rose within those poses. Further, the representative pose shows the slightly higher binding along with the TM5-facing tilt. In contrast, the starting conformation for the simulation was selected to represent the TM5 binding mode adopted by the docking poses from the OX_2-based model, which were deeper and upright. This reflects on Asn25 and His26, which are closer to TM5 in the interaction listing in the Publication I, and on some of the C-terminal residues, which are not as deep in the Publication I. In hindsight, it would have been prudent to select the representative pose for the TM5 binding mode among the deep-and-

upright cluster, which was present in both models, instead of the NTS_1-based-model-specific TM5-leaning cluster.

5.2 Orexin peptide bioactive conformation

The NMR-derived 3D structures of the orexin peptides highlight the flexibility of the peptide especially at the hinge region between the C-terminal helix and the middle/N-terminal helix, some 1.5–2 helical turns from the C-terminus. For orexin-A, three different hinge conformations have been reported, and orexin-B adopts a fourth. It is likely, although not yet shown, that the orexin peptides would bind to their receptors in a conformation similar to the conformations seen in the solution. This is discussed further in 5.2.2.

5.2.1 Modified peptides

In Publication II, we produced conformationally stabilized variants of orexin-$A_{15–33}$ through helical stapling and aminoisobutyric acid (Aib) replacements to probe the bioactive conformation (Figure 9). Our hypothesis was that the peptide would bind in the straight α-helical conformation, as that allowed for the docking of the peptide deep enough to interact with residues shown to be important by the site-directed mutagenesis. We selected four sites for modification in search of a site where the helix-stabilizing modification would be compatible with the peptide–receptor interactions. Three of the sites were on the hydrophobic side of the peptide, as the hydrocarbon staple is hydrophobic, and one was on the hydrophilic side, mainly intended as a negative control (Figure 9 for the staple placement, Figure 8 for the hydrophobic/hydrophilic side). Initially, we introduced a "standard" α-helix-stabilizing eight-atom staple in each site, labelled "nA", where n stands for the site.

Starting from the peptide C-terminus, 1A and 2A were (near) inactive up to a concentration of 20 μM (Figure 12). 1A was extremely weak at OX_1 and inactive at OX_2, while 2A was inactive at both subtypes. This was not surprising. Site 1 is at the vital C-terminus and the staple replaces A27 and L31, which do not tolerate mutations well (Tables 4 and 2006). The mutation L31A, or the corresponding L26A in orexin-B, is destructive for the activity. The mutation L26D-Leu in orexin-B is intolerable, and a D-alanine mutation of A22 in orexin-B, corresponding to A27, also results in a moderate drop in potency. Concerning site 2 (N25, G29), the mutation N25A in orexin-A seems to be tolerable, but one study reports a moderate drop in potency for the corresponding mutation N20A for both receptors. The G29 and the

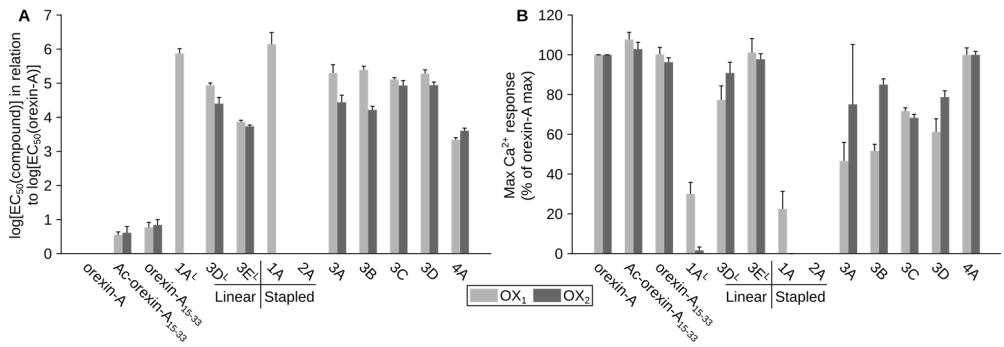


Figure 12. Activity of modified peptides. A) Potency, and B) Maximal response in relation to orexin-A.

corresponding G24 show moderate to large decreases in potency with alanine or D-alanine substitutions. Put together, the site 2 seems to fulfil our expectations of a negative control.

Peptides 3A and 4A were more potent but still showed marked drops in potency with respect to orexin-$A_{15–33}$. 4A reached full receptor activation with near three orders of magnitude lower potency, and while 3A produced a robust response, it did not reach response saturation within the tested concentration range (up to 20 μM). Assuming full efficacy, the decrease in potency for 3A would be three-to-four orders of magnitude. In contrast to the sites 1 and 2, the amino acids at sites 3 and 3 (A23 & A27 and L19 & A23, respectively) should have tolerated the mutations better. The effect of the mutations L19A (orexin-A) and L14A (orexin-B) varies between no effect to moderate decrease in potency between studies and receptor subtypes. There are no mutants for A23 and A27 in orexin-A, but the corresponding S18 and A22 in orexin-B have been mutated. The S18 tolerated both alanine and D-serine mutations with only a small decrease in potency. The A22G mutation is tolerable, but a D-alanine mutation shows a moderate decrease in potency. However, the loss of potency for the stapled peptides 3A and 4A was markedly greater than expected based on the SDM data. There were three likely causes for the observations; the staple-stabilized conformation was wrong *per se*, conformational flexibility is required upon receptor binding, or the modifications were directly incompatible with ligand–receptor interactions.

We approached the putative problem of a wrong conformation by synthesizing a series of peptides (3B, 3C, 3D) with varying staples at site 3, which spans the hinge region. We hoped that these peptides could adopt a more suitable conformation. Their activity was, however, similar to 3A, so either the conformation was still wrong, or the problem lay elsewhere. Next, we tested two linear peptides ($1A^L$ and $3D^L$), where

the unnatural stapling residues were present, but not cross-linked. The rationale was to differentiate between the effects of the amino acid replacement and the cross-linking. As these were slightly more potent than their ring-closed counterparts, it indeed appears that the cross-linking is detrimental to potency. However, the linear precursors were still far less potent than the "wild-type" orexin-$A_{15–33}$. Therefore, the modified residues themselves were the root of the problem.

The amino acids used for the stapling have an alkane side chain with a vinyl group at the end, and an additional methyl group at the Cα. The chiral configuration used in our study is *S*, i.e. corresponding to the L-amino acids. We sought to separate the effects of side-chain modification and Cα-methylation by incorporating α-aminoisobutyric acid (Aib) at site 3 instead of the stapling residues ($3E^L$). Aib is in essence a Cα-methylated alanine, so it was an ideal replacement for A23 and A27 at site 3. Aib is also helix-stabilizing, as the Cα-methylation limits the conformational flexibility of the peptide. The compound $3E^L$ was 5–12-fold more potent than the linear $3D^L$, and the most potent peptide carrying a modification at site 3, but less potent than 4A (3.5-fold in OX_1 and 1.5-fold in OX_2) and far inferior in comparison to unmodified orexin-$A_{15–33}$.

None of the modified ligands was able to block orexin-A-induced Ca^{2+}-elevation, which indicates that the modifications most likely impair the peptides' ability to bind to the orexin receptors, not only their ability to activate the receptor. The purified quantities of the stapled peptides were not sufficient for proper binding assays.

Given that the Aib-substitution introduces only two additional Cα-methyl groups, and carries almost as large a potency decrease as the helical stapling, it seems likely that the problem for all our modifications (at site 3 at least) lies with these two methyl groups, which are also present in the amino acid used for stapling. The root of the problem is likely either the stabilization of the helical conformation, which might be incompatible with the binding site or the events of the peptide entering the binding site, or the methyl groups directly blocking the peptide–receptor interactions which should take place between the receptor and the peptide backbone, which is exposed upon helical bending. A joint effect is also possible, where it is not the methyl groups but the induced helical stabilization which blocks the interactions.

Assuming that the modifications we produced do stabilize the helical conformation (in hindsight, we should have utilized circular dichroism spectroscopy or NMR to verify the increased helicity, as successful stapling does not guarantee

α-helical conformation[140]), our peptides show that the orexin receptors can indeed be activated by all-helical peptides with sub-micromolar concentrations (4A and $3E^L$). However, it might also be that it is the portion of the peptide which is not all-helical, despite the stabilization, which triggers the receptor activation.

5.2.2 Insights into the bioactive conformation

We have presumed that the orexin peptides bind to their receptors in a conformation which is similar to those observed in aqueous solution by NMR. In essence this means α-helical conformation for the peptide C-terminus. There are few solid reasons for this assumption. First, from the thermodynamics' point of view, it is beneficial for the ligand to prefer the bioactive conformation already in the solvent; this way, the binding decreases entropy only by constricting the ligand position, not so much the conformation. It would also reduce the likelihood of ligand binding, or at least introduce a delay, if the peptide needed to cross an energy barrier to reach the bioactive conformation or unfold upon binding. Secondly, truncation of orexin peptides is tolerated down to 19 amino acids, whereas shorter peptides show drastic loss in potency and efficacy. This coincides with the length of peptide which is generally regarded necessary for the adoption of stable α-helical fold, unless specific stabilizing factors such as repeated charged residues are present.[141–144] Also, as the residues lost with these further truncations are not particularly important in the light of the SDM data, it stands to reason to suggest that the reduction of potency would be linked with decreasing helicity.

When considering the orexin peptide bioactive conformation, it is of course prudent to take into account the conformation the orexin receptors. After all, a peptide conformation can hardly be bioactive, if it does not correspond to the requirements of the receptor binding site. The recent crystal structures of OX_1[8] showed a short amphipathic helix at the extracellular receptor N-terminus, close to the ECL2, parallel to the membrane plane. Through mutagenesis, the authors show that the short helix is instrumental in orexin-peptide-mediated receptor activation in both receptor subtypes. If the N-terminal helix does indeed reside close to the ECL2, and not face away of the receptor, as suggested by another study[84], it would "cap" the binding site, thus preventing the binding of a straight helical full-length orexin-A and explaining why the peptide needs to bend at the hinge region. Although the orexin-$A_{15–33}$ used in the Publication II would fit underneath the N-terminal helix in the straight helical

conformation, the straight helix would of course miss many of the interaction sites that a bent peptide would find.

Our MD simulations (Publication III) did not include the newly discovered N-terminal helix, and we began the simulations with the straight conformation of the full-length orexin-A. Across the simulations, there were two main paths for the large-scale conformation for the peptide. If the peptide N-terminal domain found (charge-assisted) hydrogen-bonding partners from the receptor ECL2, the peptide remained upright and maintained the straight conformation at the hinge region. However, if such bonding did not occur, the peptide showed a tendency to bend more or less at the direction of H26. In the TM5 binding mode, this means that the helix I in the middle of the peptide and the N-terminal domain come to rest horizontally over the binding site, atop the latter β-strand and the extracellular end of the TM5. The resulting binding mode and location is strikingly compatible with the location of the receptor N-terminal helix (Figure 13B). The conformation of the peptide does not closely represent any of the published NMR-structures for the orexin peptides (Figure 13C). The direction of the bend is similar to orexin-B, but the hinge is one helical turn higher than in the orexin-B NMR-structure. Hinge location is identical to orexin-A solution NMR-structures, but the direction of the bend is different to both published bent conformations.

A synthesis of the available data suggests that the bioactive conformation for the orexin peptides could be the bent conformation. Our modified peptides show that stabilizing the peptide in the straight helical conformation is detrimental to bioactivity, the NMR-data suggests that the peptide is more likely to be found in the bent conformation, at least in solution, and the recently discovered N-terminal helix caps the binding site in a way that discourages the binding of the straight helical orexin peptides. Our simulations show that the straight conformation is likely not stable without interactions between the peptide N-terminus and the receptor ECL2, which in turn would be blocked by the receptor N-terminal helix. Further, in the light of the receptor N-terminal helix, it is difficult to imagine where else the peptide N-terminus (the N-terminal hook and the middle helix in orexin-A, or the N-terminal helix in orexin-B) would lie than atop the ELC2, towards TM5. But what about the hinge direction?

The solution NMR studies highlight the flexibility of the hinge, showing three possible bending “modes” (Figure 13C). Our simulations show a fourth mode. A crude manual docking of orexin-A as seen in solution[93], not considering the

C-terminal interactions, would place the middle amphipathic helix atop the ECL2 with hydrophobic residues facing down towards the binding site and polar residues facing up towards the receptor N-terminal helix (Figure 13A). The latter would bring together polar residues from both parties, and at a first though, burying hydrophobic residues between the peptide and the receptor would make sense. However, the binding site is rimmed with polar amino acids in the ECL2 and ECL3, and this binding pose would bring these in contact with the hydrophobic peptide residues, which is hardly favorable. In contrast, our simulation-derived binding mode shows the peptide middle helix in a rotational state which presents the polar amino acids from the peptide towards the polar rim of the binding site, and hydrophobic peptide amino acids towards the receptor N-terminal helix. This would bring hydrophobic peptide residues together with the polar residues of the receptor N-terminus, again an unfavorable interaction. However, the rotational state of the amphipathic N-terminal

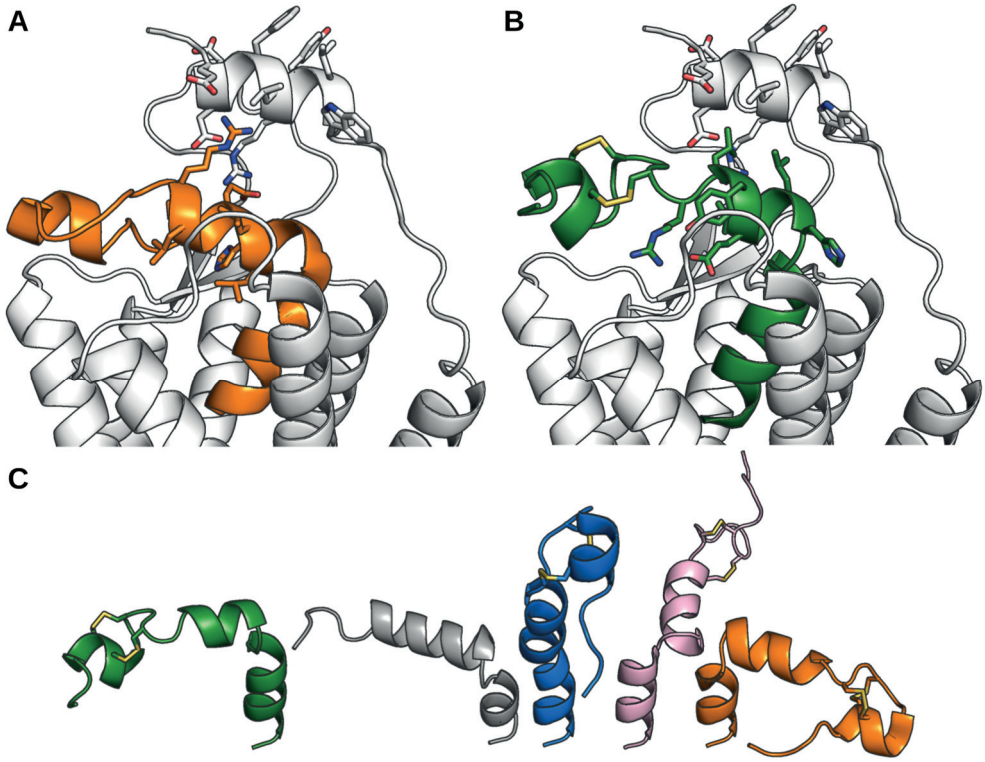


Figure 13. Orexin-A conformation and compatibility with the receptor N-terminal helix. A) Manual insertion of the orexin-A solution-NMR structure; B) Simulation-derived orexin-A conformation; C) Comparison of orexin peptide conformations. Green: Simulation-derived; Gray: Orexin-B (NMR); Blue: Straight conformation of orexin-A (NMR); Pink and orange: two different bent conformations for orexin-A (NMR).

helix of the receptor is peculiar in the crystal structure; the hydrophobic residues face away from the receptor, towards the solvent. As this might reflect the crystallization conditions and packing instead of the biological conformation, we hypothesize that a roll of ~180° would bring the hydrophobic residues from the receptor N-terminal helix in contact with the hydrophobic face of the peptide helix, burying both from the solvent, while the polar side of the receptor N-terminal helix would become exposed the solvent.

On the other hand, and despite the evidence discussed above, the straight α-helical conformation might still turn out to be the bioactive conformation. As mentioned, there are two alternative conformations for the receptor N-terminal helix,[8,84] and only one of them caps the binding site. The other displays the helix facing away from the receptor. However, as mutations in the N-terminal helix impair the peptide-induced receptor activation, it seems likely that the N-terminus would be in contact with the helix. It is possible that the N-terminal conformation seen in the crystal structures is relevant only for the small-molecule-bound receptor, and that the receptor N-terminus might adopt a third conformation upon peptide-binding, which would permit the binding of the straight α-helical peptide. Concerning our modified peptides, which were indisputably weaker than the corresponding wild-type peptide, the reason might have been a direct incompatibility of the modifications with the binding site, not the stabilized all-helical conformation.

5.3 Small molecular ligand binding

5.3.1 Antagonists

The crystallized antagonists suvorexant, EMPA, and SB-674042 share the same binding location (Figure 14).[8,83,84] SB-674042 and Suvorexant are similar molecules, and they also adopt a similar, horse-shoe-like conformation. EMPA differs chemically but also folds similarly and finds the same subpockets for binding (Figure 14).

Concerning large-scale conformation, all molecules feature two sandwich-packed aromatic groups. Suvorexant folds on a 1,4-substituted diazepane ring to achieve this, and SB-674042 via 1,2-substituted pyrrolidine. These rings also overlap in the middle of the binding site, facing Phe^{5x43} and His^{5x40}. The rings are next to Ala/Thr^{3x33}. EMPA produces similar packing through a sulfonamide moiety, which occupies a different space, while an N-ethyl group reaches the site occupied by the

N-heterocycles from other antagonists. Next to this "central" moiety, all antagonists feature an amide carbonyl, pointing roughly in the directions of Asn^{6x55}. A crystallographic water often links the carbonyl with His^{7x38}.

The two closely packed aromatic moieties are chlorinated benzoxazole and toluene in suvorexant, benzene-linked oxadiazole and methylated thiazole in SB-674042, and toluene and methoxypyridine in EMPA. The former stacks next to Pro^{3x29}, occupying a pocket lined by Cys^{2x56}, Ala^{2x59}, Ser/Thr^{2x60}, Val^{2x63}, Trp^{23x50}, Ile^{3x28}, and Gln^{3x32}. The latter is positioned with His^{7x38}, Val^{7x41} and Tyr^{7x42} in TM7, and flanked by Gln^{3x32} and the other aromatic moiety on the other side. A third

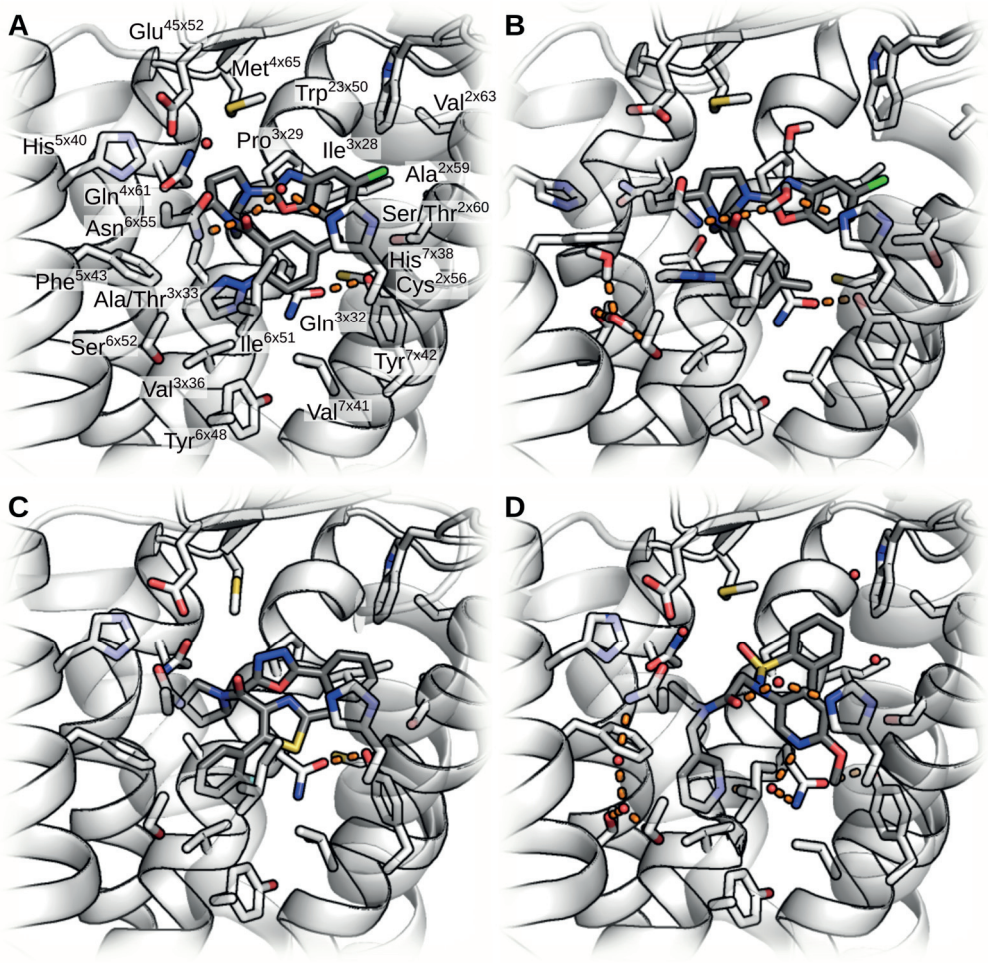


Figure 14. Antagonist binding. A) Suvorexant in OX$_2$. The OX$_1$-binding is identical. B) Suvorexant in OX$_2$ after a 3 µs MD simulation; C) SB-674042 in OX$_1$; D) EMPA in OX$_2$. As the viewpoints and selections of binding site amino acids macth, labels are only shown in panel A. Thr^{5x461} is additionally shown in panels B and D, as it serves as an anchor point for the water-binding network.

aromatic group reaches towards the TM5, into the pocket lined by Tyr^{6x48}, Val^{3x36}, Ile^{6x51} and Phe^{5x43}. This aromatic moiety is a triazole in suvorexant, a fluorobenzene in SB-674042, and pyridine in EMPA. The group might play a part in trapping few water molecules between the TMs5 and 6.

Concerning the selectivity between the orexin receptor subtypes, there are two sites of interest, 2x60 and 3x33, which were also highlighted by SDM studies.[87,88] The OX_1 has Ser^{2x60} and Ala^{3x33}, while OX_2 has Thr^{2x60} and Thr^{3x33}. SB-674042 is likely OX_1-selective because the pyrrolidine produces a steric bulk incompatible with Thr^{3x33} in OX_2, but the Ala^{3x33} in OX_1 leaves enough space for the moiety. For EMPA, there is not a single obvious reason for the selectivity. It might gain OX_2 selectivity by not reaching as close to TM2 and 2x60, where the OX_2 has a larger amino acid than the OX_1; then again, suvorexant is closer to TM2 than EMPA and binds equally to both subtypes. Same goes for the turn structure: SB-674042 protrudes close to TM3 and Ala^{3x33}, EMPA resides further from the Thr^{3x33}, thanks to the sulfonamide-based turn structure, but suvorexant shows that a distance much shorter than EMPA´s is still acceptable. It would be tempting to account the sulfonamide carbonyls for the selectivity, especially with the Thr^{3x33} close by for hydrogen-bonding. However, the crystal structures show the Thr^{3x33} hydroxyl facing the other way and hydrogen-bonding intrahelically. Rotamer change would bring the hydroxyl closer, but the distance and geometry would remain unfavorable. Perhaps a water-mediated interaction takes place, or the Thr^{3x33} hydroxyl plays a role in repositioning Gln^{3x32} to interact with the sulfonamide carbonyls. However, the comparison to suvorexant might be misleading. Although suvorexant is a dual antagonist, it is more potent than EMPA towards the OX_2; the OX_2-selectivity of EMPA arises more from the poor OX_1-binding than exceptional OX_2-binding. The case might simply be that EMPA keeps its distance from both 2x60 and 3x33, and the larger residues in the OX_2 still reach to interact with EMPA to a degree, whereas the smaller residues of the OX_1 do not.

In the Publication III, we simulated suvorexant bound with the OX_2. We began the simulation with the small molecule location and interactions from the crystal structure 4S0V, and the antagonist maintained its interactions throughout both 3 µs simulations (Figure 14B).

5.3.1.1 Mechanism of antagonism

To begin with, it is important to notice that an antagonist merely needs to block agonist binding, and as such, there cannot be an actual (conserved) mechanism of antagonism. However, there are few sites of interest in the binding site, that all crystallized orexin receptor antagonists seem to share.

All orexin receptor crystal structures have been obtained with a bound antagonist, and all show hydrogen-bonding link of Gln^{3x32}–Tyr^{7x42}. Consistently, our simulations with suvorexant maintains this link, while none of the other simulations do so, at least not to the same extent. We hypothesize that this interaction could be a hallmark for an inactive state of orexin receptors.

5.3.2 Agonist Nag26

We simulated the Nag26 agonist in four simulations for a total of 8 µs (Table 6). Unlike the antagonist suvorexant, which remained remarkably stable through its simulations, the Nag26 mobile to a degree that it is difficult to compile a single binding mode (Figure 15). The initial docking pose featured a horseshoe-like conformation similar to suvorexant, but in some simulations, the ligand extends within the binding site. The dimethylamide group, which lies deep in the binding site in the initial conformation, retains its position in the pocket lined by Thr^{5x461}, Tyr^{6x48},

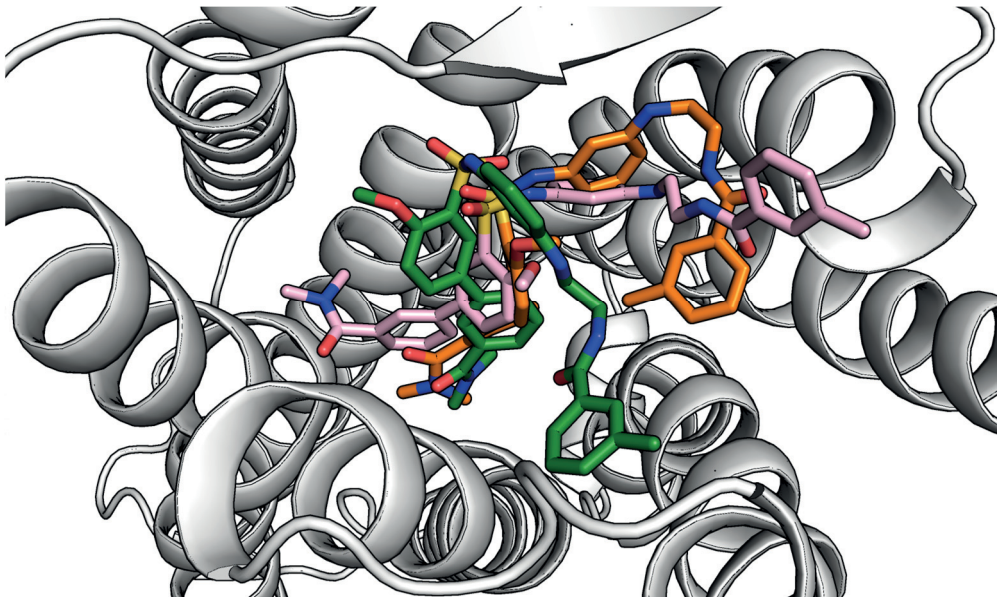


Figure 15. The binding mode of the Nag26 agonist. As the ligand is flexible throughout the simulations, a collection of the most populated poses is shown.

Val^{3x36} and Ser^{6x52} in the cholesterol-membrane simulations, forming water-mediated interactions with the Thr^{5x461}. However, the POPC-membrane simulations show the dimethylamide retreating higher in the binding site, further from the TM5, and then plunging in-between the TMs4 and 5. This pushes on the side chains of the Thr^{5x461} and Phe^{5x43}, resulting in a counterclockwise turn of the extracellular end of the TM5. This is perhaps facilitated by the bulge at Thr^{5x461}–Tyr^{5x47}, which tightens in the process. The rise of the amide group brings the carbonyl to water-mediated interaction with the Asn^{6x55} or Arg^{6x59}. Regardless of the dimethylamide, the adjacent benzene ring remains packed next to Ile^{6x51}, and the linked methoxybenzene is in the level of Asn^{6x55} in the middle of the binding site. The methoxy group participates transiently to interactions with the polar residues from the ECL2 and ECL3, and the benzene ring packs with the toluenyl group at the other end of the ligand, in the case of the folded ligand. The sulfonamide moiety is located similarly to the corresponding structure in the antagonist EMPA (Figure 14), next to the Pro^{3x29}. However, unlike EMPA in the crystal structure, the Nag26 sulfonamide forms extensive interactions with polar groups in the TM3 and ECL2–3. Depending on the simulation, it forms several transient direct or water-mediated interactions with the ECL2 and, again depending on the simulation, hydrogen-bonds with the Arg^{6x59} or Lys^{6x58}. In turn, these residues are often bound to the acidic residues Asp^{45x51} and Glu^{45x52} in the ECL2. This participation in the cross-binding site bridges seems to tilt the ECL3 towards the ECL2, but it is difficult to say if this is normal fluctuation or a ligand-induced or a ligand-stabilized conformation. In the 3 µs simulation in the cholesterol membrane, the ligand retains a deep binding location, which allows direct or water-mediated hydrogen-bonding between the sulfonamide carbonyls and Gln^{3x32}. The remainder of the molecule (the central benzene ring, the linker chain and the toluenyl ring) is quite mobile. In some simulations, it retains or readopts the packed conformation, which brings the toluenyl ring close to the ECL3 and TM7, especially the Phe^{7x34}. In contrast, the extended conformation brings the group in contact with the TM2 and ECL1, and the linker chain under the ECL2, where it forms both direct and water-mediated hydrogen bonds with the carbonyl of the Cys^{45x50}. Depending on the "flip" of the toluenyl group, the amide carbonyl is also able to reach Thr^{2x60}.

5.4 Mechanisms of orexin receptor activation

There are multiple links in the chain of GPCR activation. The cascade results in the opening of the G protein binding site, but as the activation needs to be triggered

by versatile ligands binding different receptors, the cascades in different receptors need to differentiate at some point upstream of the G-protein-binding pocket. For the rhodopsin-like GPCRs, it is suggested that the G protein binding site opening is mirrored in a defined set of interaction changes upon activation. In Publication III, we examined our simulations in terms of the conserved inactive and active interactions at the G protein binding site.[26] All simulations maintained the inactive interactions and, in accordance, none showed the formation of the active interactions. As we begun the simulations with the receptor in the inactive conformation, the simulations were perhaps too short for the receptor to undergo a large conformational change.[53–55] Another likely possibility is that the receptor activation event is energetically slightly unfavorable until the binding of a G protein is able to stabilize it, and our simulation lengths did not provide a thorough enough sampling to catch such a "rare" event.

Following the activation cascade upstream, another suggested common event is the reorganization of the three amino acids called the core triad.[29] In the orexin receptors, the triplet is Phe^{6x44}, Pro^{5x50} and Val^{3x40}. The core triad usually features a bulkier residue at 3x40, often isoleucine, and from the Pro^{5x50} viewpoint, the Phe^{6x44} lies behind the 3x40 amino acid in the inactive conformation. Upon activation, the residues reorganize into an equilateral triangle. This reorganization is tied with the outward movement of the intracellular end of the TM6. In our simulations, the core triad remains close to the starting (inactive) conformation regardless of the ligand. The discussion above on the sampling is also relevant here. It is also possible that the core triad function is not conserved across class A GPCRs. For example, the M2 receptor also features a valine at 3x40, and it shows near identical core triad conformations between the active and the inactive crystal structures.[30,31] The smaller residue at 3x40 could pose less of a barrier to the movement of Phe^{6x44}, thus erasing the distinct inactive and active conformations.

Within the binding site, it is difficult to draw solid conclusions on the activation determinants, and comparison with other receptors is less fruitful due to the intrinsic differences between GPCRs. However, one common pattern among activated GPCRs is the observation that interhelical interactions are often changed. The ligand might bridge a connection between two helices, which happens in adrenoceptors for example, or it might break or stabilize direct helix–helix interactions. As discussed above, the antagonist suvorexant maintains a hydrogen bond between the Gln^{3x32} and Tyr^{7x42}. In contrast, the simulations with orexin-A in

the TM5 binding mode consistently lack this hydrogen bond. Instead, the Tyr^{7x42} seems to favor binding with the Thr^{2x60}. If the Gln^{3x32}–Tyr^{7x42} interaction is important for the receptor inactive state, the alternative binding promoted by orexin-A could serve to tip the scales more in the favor of receptor activation. The Nag26 agonist does not, however, appear to induce such a strong change in the Tyr^{7x42} interaction pattern, but the Gln^{3x32}–Tyr^{7x42} interaction is less stable nonetheless, in the benefit of the Tyr^{7x42}–Thr^{2x60} interaction.

We monitored the polar interactions across the binding site with hopes of identifying distinct differences between the unliganded, antagonist-bound and agonist-bound receptors. The simulations did show differences in these bridges, but they were not correlated with the bound ligand (or a lack thereof). However, the crystallized orexin antagonists show a distinct lack of interactions with the polar residues at the rim of the binding site, and our simulations with suvorexant agree with this. The peptide, on the other hand, traverses between the ECL2 and ECL3 and perhaps bends to lay atop the receptor, producing extensive interactions with the polar residues from both loops. Our simulations also show the Nag26 binding a bit higher than the antagonists, which allows the polar upward facing moieties to form direct and water-mediated interactions with the ECL2 backbone and acidic side chains. In addition, the Nag26 agonist interferes in the ECL2–ECL3 salt bridge. These interactions could serve to pull the extracellular end of the TM6 towards the binding site, which in turn could pivot the entire transmembrane helix, producing the canonical outward swing of the intracellular end. The simulations also show Nag26 stably interacting with the Thr^{5x461}, a site of vital importance to adrenoceptor activation. In two simulations, we observe the Nag26 agonist penetrating between the TMs 4 and 5, pushing the Thr^{5x461} in between the helices, inducing a tightening of the helix bulge at 5x461 and rotating the extracellular end of the TM5. In addition to affecting the binding-site-exposure of many amino acids, this counterclockwise rotation also changes the residues present for interhelical interactions with the TMs 4 and 6, which might serve to destabilize interactions that keep the receptor inactive, thus facilitating activation.

6 Conclusions and perspectives

Through peptide docking and molecular dynamics simulations, we produced a suggested binding mode for the orexin peptides. The binding mode features a water-bridged network at the bottom of the binding site, linking the exposed main-chain carbonyls with Gln^{3x32} and Tyr^{6x48}, both of which are key activation residues among GPCRs. Leu33 side chain lies in a pocket formed by TMs 5 and 6, Thr32 faces the TM7, Leu31 is under the ECL1, Ile30 in front of TMs 3 and 4 under the ECL2. Gly29 packs close to His^{7x38}, Ala28 faces TMs 2 and 7 and Ala27 the ECL1. His26 faces the TM5 and interacts with ECL2 or ECL3, and Asn25 interacts with the extracellular residues of the TM7. We suggest that the rest of the peptide bends to rest on top of the ECL2 and TM5, and that the receptor N-terminal amphipathic helix lies on top of the peptide ligand, shielding the hydrophobic face of the peptide helix II. Alternatively, the peptide remains upright, and the receptor N-terminal helix adopts a position parallel to the peptide.

The constraining of the peptide conformation was detrimental to potency. In accordance to site-directed mutagenesis data, modifications at the C-terminus were intolerable, whereas modifications at the peptide hinge and at the helix II produced bioactive peptides. As the Aib-modification carried almost as large effect as the introduction of the far bulkier staple, it appears that either the binding site is very confined, the dimehtylated Cα blocks key interactions or the modifications stabilize a conformation, which is incompatible with the receptor.

It would be of high interest to repeat the peptide binding simulations with the recently discovered N-terminal receptor helix. Such simulation could offer insight into the functional role of the helix. With hypothetical unlimited resources, an intriguing setup would also be to construct a simulation, where the peptide was initially unbound from the binding site, but in interaction with the N-terminal helix.

The suggested binding mode should be validated by introducing new mutations into the receptors. Especially the ECL3 and the extracellular end of the TM7 would be important additions to the pool of existing mutants.

7 References

(1) 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**, *63* (6), 1256–1272.

(2) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; et al. A Comprehensive Map of Molecular Drug Targets. *Nat. Rev. Drug Discov.* **2016**, *16*, 19.

(3) Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E. Trends in GPCR Drug Discovery: New Agents, Targets and Indications. *Nat. Rev. Drug Discov.* **2017**, *16*, 829–842.

(4) Kolakowski, L. J. GCRDb: A G-Protein-Coupled Receptor Database. *Receptors Channels* **1994**, *2* (1), 1–7.

(5) Attwood, T. K.; Findlay, J. B. C. Fingerprinting G-Protein-Coupled Receptors. *Protein Eng. Des. Sel.* **1994**, *7* (2), 195–203.

(6) Hanson, M. A.; Stevens, R. C. Discovery of New GPCR Biology: One Receptor Structure at a Time. *Structure* **2009**, *17* (1), 8–14.

(7) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **2013**, *494* (7436), 185–194.

(8) Yin, J.; Babaoglu, K.; Brautigam, C. A.; Clark, L.; Shao, Z.; Scheuermann, T. H.; Harrell, C. M.; Gotter, A. L.; Roecker, A. J.; Winrow, C. J.; et al. Structure and Ligand-Binding Mechanism of the Human OX_1 and OX_2 Orexin Receptors. *Nat. Struct. Mol. Biol.* **2016**, *23* (4), 293–299.

(9) Shihoya, W.; Nishizawa, T.; Okuta, A.; Tani, K.; Dohmae, N.; Fujiyoshi, Y.; Nureki, O.; Doi, T. Activation Mechanism of Endothelin ETB Receptor by Endothelin-1. *Nature* **2016**, *537* (7620), 363–368.

(10) del Castillo, J.; Katz, B. F. R. S. Interaction at End-Plate Receptors between Different Choline Derivatives. *Proc. R. Soc. London. Ser. B - Biol. Sci.* **1957**, *146* (924), 369–381.

(11) Leff, P. The Two-State Model of Receptor Activation. *Trends Pharmacol. Sci.* **1995**, *16* (3), 89–97.

(12) De Lean, A.; Stadel, J. M.; Lefkowitz, R. J. Ternary Complex Model Explains the Agonist-Specific Binding Properties of the Adenylate Cyclase-Coupled β-Adrenergic Receptor. *J. Biol. Chem.* **1980**, *255* (15), 7108–7117.

(13) Northup, J. K.; Sternweis, P. C.; Smigel, M. D.; Schleifer, L. S.; Ross, E. M.; Gilman, A. G. Purification of the Regulatory Component of Adenylate Cyclase. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77* (11), 6516–6520.

(14) Samama, P.; Cotecchia, S.; Costa, T.; Lefkowitzs, R. J. A Mutation-Induced Activated State of the β_2-Adrenergic Receptor. *J. Biol. Chem.* **1993**, *268* (7), 4625–4636.

(15) Kobilka, B. K.; Deupi, X. Conformational Complexity of G-Protein-Coupled Receptors. *Trends Pharmacol. Sci.* **2007**, *28* (8), 397–406.

(16) Deupi, X.; Kobilka, B. K. Energy Landscapes as a Tool to Integrate GPCR Structure, Dynamics, and Function. *Physiology* **2010**, *25* (5), 293–303.

(17) Hilger, D.; Masureel, M.; Kobilka, B. K. Structure and Dynamics of GPCR Signaling Complexes. *Nat. Struct. Mol. Biol.* **2018**, *25* (1), 4–12.

(18) Gudermann, T.; Schöneberg, T.; Schultz, G. Functional and Structural Complexity of Signal Transduction via G-Protein-Coupled Receptors. *Annu. Rev. Neurosci.* **1997**, *20* (1), 399–427.

(19) Offermanns, S. G-Proteins as Transducers in Transmembrane Signalling. *Prog. Biophys. Mol. Biol.* **2003**, *83* (2), 101–130.

(20) Lefkowitz, R. J.; Shenoy, S. K. Transduction of Receptor Signals by β-Arrestins. *Science* **2005**, *308* (5721), 512–517.

(21) Zuckerman, R.; Buzdygon, B.; Philp, N.; Liebman, P.; Sitaramayya, A. Arrestin: An ATP/ADP Exchange Protein That Regulates CGMP Phosphodiesterase Activity in Retinal Rod Disk Membranes (RDM). *Biophys. J.* **1985**, *47*, 37a.

(22) Manglik, A.; Lin, H.; Aryal, D. K.; McCorvy, J. D.; Dengler, D.; Corder, G.; Levit, A.; Kling, R. C.; Bernat, V.; Hübner, H.; et al. Structure-Based Discovery of Opioid Analgesics with Reduced Side Effects. *Nature* **2016**, *537* (7619), 185.

(23) Azzi, M.; Charest, P. G.; Angers, S.; Rousseau, G.; Kohout, T.; Bouvier, M.; Piñeyro, G. β-Arrestin-Mediated Activation of MAPK by Inverse Agonists Reveals Distinct Active Conformations for G Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (20), 11406–11411.

(24) Gomes, I.; Jordan, B. A.; Gupta, A.; Rios, C.; Trapaidze, N.; Devi, L. A. G Protein Coupled Receptor Dimerization: Implications in Modulating Receptor Function. *J. Mol. Med.* **2001**, *79* (5), 226–242.

(25) Manglik, A.; Kruse, A. C. Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry* **2017**, *56* (42), 5628–5634.

(26) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Heydenreich, F. M.; Flock, T.; Miljus, T.; Balaji, S.; Bouvier, M.; Veprintsev, D. B.; Tate, C. G.; et al. Diverse Activation Pathways in Class A GPCRs Converge near the G-Protein-Coupling Region. *Nature* **2016**, *536* (7617), 484–487.

(27) Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; et al. Crystal Structure of the β_2 Adrenergic Receptor–Gs Protein Complex. *Nature* **2011**, *477* (7366), 549–555.

(28) Preininger, A. M.; Meiler, J.; Hamm, H. E. Conformational Flexibility and Structural Dynamics in GPCR-Mediated G Protein Activation: A Perspective. *J. Mol. Biol.* **2013**, *425* (13), 2288–2298.

(29) Huang, W.; Manglik, A.; Venkatakrishnan, A. J.; Laeremans, T.; Feinberg, E. N.; Sanborn, A. L.; Kato, H. E.; Livingston, K. E.; Thorsen, T. S.; Kling, R. C.; et al. Structural Insights into μ-Opioid Receptor Activation. *Nature* **2015**, *524* (7565), 315–321.

(30) Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; et al. Structure of the Human M2 Muscarinic Acetylcholine Receptor Bound to an Antagonist. *Nature* **2012**, *482* (7386), 547–551.

(31) Kruse, A. C.; Ring, A. M.; Manglik, A.; Hu, J.; Hu, K.; Eitel, K.; Hübner, H.; Pardon, E.; Valant, C.; Sexton, P. M.; et al. Activation and Allosteric Modulation of a Muscarinic Acetylcholine Receptor. *Nature* **2013**, *504* (7478), 101–106.

(32) Pope, A.; Eilers, M.; Reeves, P. J.; Smith, S. O. Amino Acid Conservation and Interactions in Rhodopsin: Probing Receptor Activation by NMR Spectroscopy. *Biochim. Biophys. Acta - Bioenerg.* **2014**, *1837* (5), 683–693.

(33) Lin, S. W.; Sakmar, T. P. Specific Tryptophan UV-Absorbance Changes Are Probes of the Transition of Rhodopsin to Its Active State. *Biochemistry* **1996**, *35* (34), 11149–11159.

(34) Rasmussen, S. G. F.; Choi, H.-J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; DeVree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; et al. Structure of a Nanobody-Stabilized Active State of the β_2 Adrenoceptor. *Nature* **2011**, *469* (7329), 175–180.

(35) Standfuss, J.; Edwards, P. C.; D'Antona, A.; Fransen, M.; Xie, G.; Oprian, D. D.; Schertler, G. F. X. The Structural Basis of Agonist-Induced Activation in Constitutively Active Rhodopsin. *Nature* **2011**, *471* (7340), 656–660.

(36) Krumm, B. E.; White, J. F.; Shah, P.; Grisshammer, R. Structural Prerequisites for G-Protein Activation by the Neurotensin Receptor. *Nat. Commun.* **2015**, *6*, 7895.

(37) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. a; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; et al. High-Resolution Crystal Structure of an Engineered Human β_2-Adrenergic G Protein–coupled Receptor. *Science* **2007**, *318* (5854), 1258–1265.

(38) White, J. F.; Noinaj, N.; Shibata, Y.; Love, J.; Kloss, B.; Xu, F.; Gvozdenovic-Jeremic, J.; Shah, P.; Shiloach, J.; Tate, C. G.; et al. Structure of the Agonist-Bound Neurotensin Receptor. *Nature* **2012**, *490* (7421), 508–513.

(39) Ma, Y.; Yue, Y.; Ma, Y.; Zhang, Q.; Zhou, Q.; Song, Y.; Shen, Y.; Li, X.; Ma, X.; Li, C.; et al. Structural Basis for Apelin Control of the Human Apelin Receptor. *Structure* **2017**, *25* (6), 858–866.

(40) Koehl, A.; Hu, H.; Maeda, S.; Zhang, Y.; Qu, Q.; Paggi, J. M.; Latorraca, N. R.; Hilger, D.; Dawson, R.; Matile, H.; et al. Structure of the μ-Opioid Receptor–Gi Protein Complex. *Nature* **2018**, *558* (7711), 547–552.

(41) Asada, H.; Horita, S.; Hirata, K.; Shiroishi, M.; Shiimura, Y.; Iwanari, H.; Hamakubo, T.; Shimamura, T.; Nomura, N.; Kusano-Arai, O.; et al. Crystal Structure of the Human Angiotensin II Type 2 Receptor Bound to an Angiotensin II Analog. *Nat. Struct. Mol. Biol.* **2018**, *25* (7), 570–576.

(42) Liu, H.; Kim, H. R.; Deepak, R. N. V. K.; Wang, L.; Chung, K. Y.; Fan, H.; Wei, Z.; Zhang, C. Orthosteric and Allosteric Action of the C5a Receptor Antagonists. *Nat. Struct. Mol. Biol.* **2018**, *25* (6), 472–481.

(43) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (32), 13118–13123.

(44) Lennard-Jones, J. E. On the Determination of Molecular Fields. —II. From the Equation of State of a Gas. *Proc. R. Soc. London. Ser. A* **1924**, *106* (738), 463–477.

(45) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N□log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(46) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8* (4), 1409–1414.

(47) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929.

(48) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F. de Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (9), 2997–3002.

(49) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566.

(50) Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem. Rev.* **2017**, *117* (1), 139–155.

(51) Dror, R. O.; Green, H. F.; Valant, C.; Borhani, D. W.; Valcourt, J. R.; Pan, A. C.; Arlow, D. H.; Canals, M.; Lane, J. R.; Rahmani, R.; et al. Structural Basis for Modulation of a G-Protein-Coupled Receptor by Allosteric Drugs. *Nature* **2013**, *503* (7475), 295–299.

(52) Saleh, N.; Ibrahim, P.; Saladino, G.; Gervasio, F. L.; Clark, T. An Efficient Metadynamics-Based Protocol To Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. *J. Chem. Inf. Model.* **2017**, *57* (5), 1210–1217.

(53) Vilardaga, J.-P.; Bünemann, M.; Krasel, C.; Castro, M.; Lohse, M. J. Measurement of the Millisecond Activation Switch of G Protein–coupled Receptors in Living Cells. *Nat. Biotechnol.* **2003**, *21*, 807–812.

(54) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. Activation Mechanism of the β_2-Adrenergic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (46), 18684–18689.

(55) Clark, T. G-Protein Coupled Receptors: Answers from Simulations. *Beilstein J. Org. Chem.* **2017**, *13*, 1071–1078.

(56) Miao, Y.; Nichols, S. E.; Gasper, P. M.; Metzger, V. T.; McCammon, J. A. Activation and Dynamic Network of the M2 Muscarinic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (27), 10982–10987.

(57) Michino, M.; Abola, E.; Participants of GPCR Dock 2008; Brooks, C. L.; Dixon, J. S.; Moult, J.; Stevens, R. C. Community-Wide Assessment of GPCR Structure Modelling and Ligand Docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* **2009**, *8* (6), 455–463.

(58) Kufareva, I.; Katritch, V.; 2013, P. of G. D.; Stevens, R. C.; Abagyan, R. Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure* **2014**, *22* (8), 1120–1139.

(59) Kufareva, I.; Rueda, M.; Katritch, V.; Participants, G. D. 2010; Stevens, R. C.; Abagyan, R. Status of GPCR Modeling and Docking as Reflected by Community-Wide GPCR Dock 2010 Assessment. *Structure* **2011**, *19* (8), 1108–1126.

(60) Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S. Protein–peptide Docking: Opportunities and Challenges. *Drug Discov. Today* **2018**, *23* (8), 1530–1537.

(61) Lensink, M. F.; Velankar, S.; Wodak, S. J. Modeling Protein–protein and Protein–peptide Complexes: CAPRI 6th Edition. *Proteins Struct. Funct. Bioinforma.* **2017**, *85* (3), 359–377.

(62) Tubert-Brohman, I.; Sherman, W.; Repasky, M.; Beuming, T. Improved Docking of Polypeptides with Glide. *J. Chem. Inf. Model.* **2013**, *53* (7), 1689–1699.

(63) London, N.; Raveh, B.; Schueler-Furman, O. Peptide Docking and Structure-Based Characterization of Peptide Binding: From Knowledge to Know-How. *Curr. Opin. Struct. Biol.* **2013**, *23* (6), 894–902.

(64) Hauser, A. S.; Windshügel, B. LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *J. Chem. Inf. Model.* **2016**, *56* (1), 188–200.

(65) Antes, I. DynaDock: A New Molecular Dynamics-Based Algorithm for Protein-Peptide Docking Including Receptor Flexibility. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (5), 1084–1104.

(66) Trellet, M.; Melquiond, A. S. J.; Bonvin, A. M. J. J. A Unified Conformational Selection and Induced Fit Approach to Protein-Peptide Docking. *PLoS One* **2013**, *8* (3), e58769.

(67) Raveh, B.; London, N.; Schueler-Furman, O. Sub-Angstrom Modeling of Complexes between Flexible Peptides and Globular Proteins. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (9), 2029–2040.

(68) Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: An FFT-Based Protein Docking Program with Pairwise Potentials. *Proteins Struct. Funct. Bioinforma. 65* (2), 392–406.

(69) Chen, R.; Li, L.; Weng, Z. ZDOCK: An Initial-Stage Protein-Docking Algorithm. *Proteins Struct. Funct. Genet.* **2003**, *52* (1), 80–87.

(70) de Lecea, L.; Kilduff, T. S.; Peyron, C.; Gao, X.-B.; Foye, P. E.; Danielson, P. E.; Fukuhara, C.; Battenberg, E. L. F.; Gautvik, V. T.; Bartlett II, F. S.; et al. The Hypocretins: Hypothalamus-Specific Peptides with Neuroexcitatory Activity. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (1), 322–327.

(71) Sakurai, T.; Amemiya, A.; Ishii, M.; Matsuzaki, I.; Chemelli, R. M.; Tanaka, H.; Williams, S. C.; Richardson, J. A.; Kozlowski, G. P.; Wilson, S.; et al. Orexins and Orexin Receptors: A Family of Hypothalamic Neuropeptides and G Protein-Coupled Receptors That Regulate Feeding Behavior. *Cell* **1998**, *92* (4), 573–585.

(72) Sakurai, T.; Amemiya, A.; Ishii, M.; Matsuzaki, I.; Chemelli, R. M.; Tanaka, H.; Williams, S. C.; Richardson, J. A.; Kozlowski, G. P.; Wilson, S.; et al. Addendum | Cell - Volume 92, Issue 5. *Cell* **1998**, *92* (5), 697.

(73) Leonard, C. S.; Kukkonen, J. P. Orexin/Hypocretin Receptor Signalling: A Functional Perspective. *Br. J. Pharmacol.* **2013**, *171* (2), 294–313.

(74) Kastin, A. J.; Akerstrom, V. Orexin A but Not Orexin B Rapidly Enters Brain from Blood by Simple Diffusion. *J. Pharmacol. Exp. Ther.* **1999**, *289* (1), 219–223.

(75) Bingham, S.; Davey, P. .; Babbs, A. .; Irving, E. .; Sammons, M. .; Wyles, M.; Jeffrey, P.; Cutler, L.; Riba, I.; Johns, A.; et al. Orexin-A, an Hypothalamic Peptide with Analgesic Properties. *Pain* **2001**, *92* (1–2), 81–90.

(76) Ehrström, M.; Näslund, E.; Levin, F.; Kaur, R.; Kirchgessner, A. L.; Theodorsson, E.; Hellström, P. M. Pharmacokinetic Profile of Orexin A and Effects on Plasma Insulin and Glucagon in the Rat. *Regul. Pept.* **2004**, *119* (3), 209–212.

(77) Kukkonen, J. P.; Leonard, C. S. Orexin/Hypocretin Receptor Signalling Cascades. *Br. J. Pharmacol.* **2014**, *171* (2), 314–331.

(78) Roecker, A. J.; Cox, C. D.; Coleman, P. J. Orexin Receptor Antagonists: New Therapeutic Agents for the Treatment of Insomnia. *J. Med. Chem.* **2016**, *59* (2), 504–530.

(79) Nishino, S.; Ripley, B.; Overeem, S.; Lammers, G. J.; Mignot, E. Hypocretin (Orexin) Deficiency in Human Narcolepsy. *Lancet* **2000**, *355*, 39–40.

(80) Scammell, T. E.; Winrow, C. J. Orexin Receptors: Pharmacology and Therapeutic Opportunities. *Annu. Rev. Pharmacol. Toxicol.* **2011**, *51*, 243–266.

(81) Rouet-Benzineb, P.; Rouyer-Fessard, C.; Jarry, A.; Avondo, V.; Pouzet, C.; Yanagisawa, M.; Laboisse, C.; Laburthe, M.; Voisin, T. Orexins Acting at Native OX_1 Receptor in Colon Cancer and Neuroblastoma Cells or at Recombinant OX_1 Receptor Suppress Cell Growth by Inducing Apoptosis. *J. Biol. Chem.* **2004**, *279* (44), 45875–45886.

(82) Laburthe, M.; Voisin, T. The Orexin Receptor OX_1R in Colon Cancer: A Promising Therapeutic Target and a New Paradigm in G Protein-Coupled Receptor Signalling through ITIMs. *Br. J. Pharmacol.* **2012**, *165* (6), 1678–1687.

(83) Yin, J.; Mobarec, J. C.; Kolb, P.; Rosenbaum, D. M. Crystal Structure of the Human OX_2 Orexin Receptor Bound to the Insomnia Drug Suvorexant. *Nature* **2015**, *519* (7542), 247–250.

(84) Suno, R.; Kimura, K. T.; Nakane, T.; Yamashita, K.; Wang, J.; Fujiwara, T.; Yamanaka, Y.; Im, D.; Horita, S.; Tsujimoto, H.; et al. Crystal Structures of Human Orexin 2 Receptor Bound to the Subtype-Selective Antagonist EMPA. *Structure* **2018**, *26* (1), 7–19.

(85) Heifetz, A.; Barker, O.; Morris, G. B.; Law, R. J.; Slack, M.; Biggin, P. C. Toward an Understanding of Agonist Binding to Human Orexin-1 and Orexin-2 Receptors with G-Protein-Coupled Receptor Modeling and Site-Directed Mutagenesis. *Biochemistry* **2013**, *52* (46), 8246–8260.

(86) Heifetz, A.; Morris, G. B.; Biggin, P. C.; Barker, O.; Fryatt, T.; Bentley, J.; Hallett, D.; Manikowski, D.; Pal, S.; Reifegerste, R.; et al. Study of Human Orexin-1 and -2 G-Protein-Coupled Receptors with Novel and Published Antagonists by Modeling, Molecular Dynamics Simulations, and Site-Directed Mutagenesis. *Biochemistry* **2012**, *51* (15), 3178–3197.

(87) Malherbe, P.; Roche, O.; Marcuz, A.; Kratzeisen, C.; Wettstein, J. G.; Bissantz, C. Mapping the Binding Pocket of Dual Antagonist Almorexant to Human Orexin 1 and Orexin 2 Receptors: Comparison with the Selective OX_1 Antagonist SB-674042 and the Selective OX_2 Antagonist N-Ethyl-2-[(6-Methoxy-Pyridin-3-Yl)-(Toluene-2-Sulfonyl)-Amino]-N-Py. *Mol. Pharmacol.* **2010**, *78* (1), 81–93.

(88) Tran, D.-T.; Bonaventure, P.; Hack, M.; Mirzadegan, T.; Dvorak, C.; Letavic, M.; Carruthers, N.; Lovenberg, T.; Sutton, S. W. Chimeric, Mutant Orexin Receptors Show Key Interactions between Orexin Receptors, Peptides and Antagonists. *Eur. J. Pharmacol.* **2011**, *667* (1–3), 120–128.

(89) Putula, J.; Kukkonen, J. P. Mapping of the Binding Sites for the OX_1 Orexin Receptor Antagonist, SB-334867, Using Orexin/Hypocretin Receptor Chimaeras. *Neurosci. Lett.* **2012**, *506* (1), 111–115.

(90) Lang, M.; Söll, R. M.; Dürrenberger, F.; Dautzenberg, F. M.; Beck-Sickinger, A. G. Structure-Activity Studies of Orexin A and Orexin B at the Human Orexin 1 and Orexin 2 Receptors Led to Orexin 2 Receptor Selective and Orexin 1 Receptor Preferring Ligands. *J. Med. Chem.* **2004**, *47* (5), 1153–1160.

(91) Ammoun, S.; Holmqvist, T.; Shariatmadari, R.; Oonk, H. B.; Detheux, M.; Parmentier, M.; Åkerman, K. E. O.; Kukkonen, J. P. Distinct Recognition of OX_1 and OX_2 Receptors by Orexin Peptides. *J. Pharmacol. Exp. Ther.* **2003**, *305* (2), 507–514.

(92) Asahi, S.; Egashira, S.-I.; Matsuda, M.; Iwaasa, H.; Kanatani, A.; Ohkubo, M.; Ihara, M.; Morishima, H. Development of an Orexin-2 Receptor Selective Agonist, [Ala(11), D-Leu(15)]Orexin-B. *Bioorg. Med. Chem. Lett.* **2003**, *13* (1), 111–113.

(93) Takai, T.; Takaya, T.; Nakano, M.; Akutsu, H.; Nakagawa, A.; Aimoto, S.; Nagai, K.; Ikegami, T. Orexin-A Is Composed of a Highly Conserved C-Terminal and a Specific, Hydrophilic N-Terminal Region, Revealing the Structural Basis of Specific Recognition by the Orexin-1 Receptor. *J. Pept. Sci.* **2006**, *12* (7), 443–454.

(94) Lee, J.-H.; Bang, E.; Chae, K.-J.; Kim, J.-Y.; Lee, D. W.; Lee, W. Solution Structure of a New Hypothalamic Neuropeptide, Human Hypocretin-2/Orexin-B. *Eur. J. Biochem.* **1999**, *266* (3), 831–839.

(95) Kim, H.-Y.; Hong, E.; Kim, J.-I.; Lee, W. Solution Structure of Human Orexin-A: Regulator of Appetite and Wakefulness. *J. Biochem. Mol. Biol.* **2004**, *37* (5), 565–573.

(96) Darker, J. G.; Porter, R. A.; Eggleston, D. S.; Smart, D.; Brough, S. J.; Sabido-David, C.; Jerman, J. C. Structure-Activity Analysis of Truncated Orexin-A Analogues at the Orexin-1 Receptor. *Bioorg. Med. Chem. Lett.* **2001**, *11* (5), 737–740.

(97) German, N. A.; Decker, A. M.; Gilmour, B. P.; Thomas, B. F.; Zhang, Y. Truncated Orexin Peptides: Structure-Activity Relationship Studies. *ACS Med. Chem. Lett.* **2013**, *4* (12), 1224–1227.

(98) Yanagisawa, M. Small-Molecule Agonists for Type-2 Orexin Receptor. U.S. Patent 8258163, 2012.

(99) Nagase, H.; Yanagisawa, M.; Saitoh, T.; Kutsumura, N.; Irukayama, Y. Sulfonamide Derivative and Pharmaceutically Acceptable Acid Addition Salt Thereof. U.S. Patent application 20180179151, 2018.

(100) Nagahara, T.; Saitoh, T.; Kutsumura, N.; Irukayama-Tomobe, Y.; Ogawa, Y.; Kuroda, D.; Gouda, H.; Kumagai, H.; Fujii, H.; Yanagisawa, M.; et al. Design and Synthesis of Non-Peptide, Selective Orexin Receptor 2 Agonists. *J. Med. Chem.* **2015**, *58* (20), 7931–7937.

(101) Ballesteros, J. A.; Weinstein, H. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Methods Neurosci.* **1995**, *25*, 366–428.

(102) Isberg, V.; de Graaf, C.; Bortolato, A.; Cherezov, V.; Katritch, V.; Marshall, F. H.; Mordalski, S.; Pin, J.-P.; Stevens, R. C.; Vriend, G.; et al. Generic GPCR Residue Numbers – Aligning Topology Maps While Minding the Gaps. *Trends Pharmacol. Sci.* **2015**, *36* (1), 22–31.

(103) Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Mathiesen, J. M.; Sunahara, R. K.; Pardo, L.; Weis, W. I.; Kobilka, B. K.; Granier, S. Crystal Structure of the μ-Opioid Receptor Bound to a Morphinan Antagonist. *Nature* **2012**, *485* (7398), 321–326.

(104) Granier, S.; Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Weis, W. I.; Kobilka, B. K. Structure of the δ-Opioid Receptor Bound to Naltrindole. *Nature* **2012**, *485* (7398), 400–404.

(105) Wu, H.; Wacker, D.; Mileni, M.; Katritch, V.; Han, G. W.; Vardy, E.; Liu, W.; Thompson, A. a; Huang, X.-P.; Carroll, F. I.; et al. Structure of the Human κ-Opioid Receptor in Complex with JDTic. *Nature* **2012**, *485* (7398), 327–332.

(106) Thompson, A.; Liu, W.; Chun, E.; Katritch, V.; Wu, H.; Vardy, E.; Huang, X.-P.; Trapella, C.; Guerrini, R.; Calo, G.; et al. Structure of the Nociceptin/Orphanin FQ Receptor in Complex with a Peptide Mimetic. *Nature* **2012**, *485* (7398), 395–399.

(107) Wu, B.; Chien, E. Y. T.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; et al. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science* **2010**, *330* (6007), 1066–1071.

(108) Wang, C.; Jiang, Y.; Ma, J.; Wu, H.; Wacker, D.; Katritch, V.; Han, G. W.; Liu, W.; Huang, X.-P.; Vardy, E.; et al. Structural Basis for Molecular Recognition at Serotonin Receptors. *Science* **2013**, *340* (6132), 610–614.

(109) Wacker, D.; Wang, C.; Katritch, V.; Han, G. W.; Huang, X.-P.; Vardy, E.; McCorvy, J. D.; Jiang, Y.; Chu, M.; Siu, F. Y.; et al. Structural Features for Functional Selectivity at Serotonin Receptors. *Science* **2013**, *340* (6132), 615–619.

(110) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G. W.; Tate, C. G.; Schertler, G. F. X. Structure of a β_1-Adrenergic G-Protein-Coupled Receptor. *Nature* **2008**, *454* (7203), 486–491.

(111) Chien, E. Y. T.; Liu, W.; Zhao, Q.; Katritch, V.; Han, G. W.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; et al. Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective AntagonistStructure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. *Science* **2010**, *330* (6007), 1091–1905.

(112) Shimamura, T.; Shiroishi, M.; Weyand, S.; Tsujimoto, H.; Winter, G.; Katritch, V.; Abagyan, R.; Cherezov, V.; Liu, W.; Han, G. W.; et al. Structure of the Human Histamine H1 Receptor Complex with Doxepin. *Nature* **2011**, *475* (7354), 65–72.

(113) Kruse, A. C.; Hu, J.; Pan, A. C.; Arlow, D. H.; Rosenbaum, D. M.; Rosemond, E.; Green, H. F.; Liu, T.; Chae, P. S.; Dror, R. O.; et al. Structure and Dynamics of the M3 Muscarinic Acetylcholine Receptor. *Nature* **2012**, *482* (7386), 552–556.

(114) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; IJzerman, A. P.; Stevens, R. C. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322* (2), 1211–1217.

(115) Zhang, C.; Srinivasan, Y.; Arlow, D. H.; Fung, J. J.; Palmer, D.; Zheng, Y.; Green, H. F.; Pandey, A.; Dror, R. O.; Shaw, D. E.; et al. High-Resolution Crystal Structure of Human Protease-Activated Receptor 1. *Nature* **2012**, *492* (7429), 387–392.

(116) Murakami, M.; Kouyama, T. Crystal Structure of Squid Rhodopsin. *Nature* **2008**, *453* (7193), 363–367.

(117) Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. The Retinal Conformation and Its Environment in Rhodopsin in Light of a New 2.2 A Crystal Structure. *J. Mol. Biol.* **2004**, *342* (2), 571–583.

(118) Hanson, M. A.; Roth, C. B.; Jo, E.; Griffith, M. T.; Scott, F. L.; Reinhart, G.; Desale, H.; Clemons, B.; Cahalan, S. M.; Schuerer, S. C.; et al. Crystal Structure of a Lipid G Protein-Coupled Receptor. *Science* **2012**, *335* (6070), 851–855.

(119) Săli, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815.

(120) Li, L.; Chen, R.; Weng, Z. RDOCK: Refinement of Rigid-Body Protein Docking Predictions. *Proteins Struct. Funct. Genet.* **2003**, *53* (3), 693–707.

(121) Discovery Studio. Biovia: San Diego, CA, USA 2015.

(122) Matlab. MathWorks, Inc: Natick, MA, USA 2017.

(123) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie Int. Ed.* **1999**, *38* (1–2), 236–240.

(124) Hubbard, S. J.; Thornton, J. M. *Naccess*; Department of Biochemistry and Molecular Biology, University College London: London, England, 1993.

(125) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(126) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. *Solving Softw. Challenges Exascale* **2015**, 3–27.

(127) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins* **2010**, *78* (8), 1950–1958.

(128) Jämbeck, J. P. M.; Lyubartsev, A. P. Another Piece of the Membrane Puzzle: Extending Slipids Further. *J. Chem. Theory Comput.* **2013**, *9* (1), 774–784.

(129) Rackers, J. A.; Laury, M. L.; Lu, C.; Wang, Z.; Lagardère, L.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. TINKER 8: A Modular Software Package for Molecular Design and Simulation. 2017.

(130) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; et al. Amber 14. 2014.

(131) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian. 2009.

(132) Brameld, K. A.; Kuhn, B.; Reuter, D. C.; Stahl, M. Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused Analysis. *J. Chem. Inf. Model.* **2008**, *48* (1), 1–24.

(133) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem. 23* (16), 1623–1641.

(134) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem. 29* (11), 1859–1865.

(135) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(136) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(137) Allen, W. J.; Lemkul, J. A.; Bevan, D. R. GridMAT-MD: A Grid-Based Membrane Analysis Tool for Use with Molecular Dynamics. *J. Comput. Chem. 30* (12), 1952–1958.

(138) Pluhackova, K.; Kirsch, S. A.; Han, J.; Sun, L.; Jiang, Z.; Unruh, T.; Böckmann, R. A. A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers. *J. Phys. Chem. B* **2016**, *120* (16), 3888–3903.

(139) Schafmeister, C. E.; Po, J.; Verdine, G. L. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *J. Am. Chem. Soc.* **2000**, *122* (24), 5891–5892.

(140) Hojo, K.; Hossain, M. A.; Tailhades, J.; Shabanpoor, F.; Wong, L. L. L.; Ong-Pålsson, E. E. K.; Kastman, H. E.; Ma, S.; Gundlach, A. L.; Rosengren, K. J.; et al. Development of a Single-Chain Peptide Agonist of the Relaxin-3 Receptor Using Hydrocarbon Stapling. *J. Med. Chem.* **2016**, *59* (16), 7445–7456.

(141) Marqusee, S.; Baldwin, R. L. Helix Stabilization by Glu-...Lys+ Salt Bridges in Short Peptides of de Novo Design. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (24), 8898–8902.

(142) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. Unusually Stable Helix Formation in Short Alanine-Based Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (14), 5286–5290.

(143) Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. A Neutral, Water-Soluble, .Alpha.-Helical Peptide: The Effect of Ionic Strength on the Helix-Coil Equilibrium. *J. Am. Chem. Soc.* **1991**, *113* (13), 5102–5104.

(144) Baldwin, R. L. α-Helix Formation by Peptides of Defined Sequence. *Biophys. Chem.* **1995**, *55* (1–2), 127–135.

Recent Publications in this Series

46/2018 Panu Luukkonen
Heterogeneity of Non-Alcoholic Fatty Liver Disease – Genetic and Nutritional Modulation of Hepatic Lipid Metabolism
47/2018 Henriikka Kentala
ORP2 – A Sterol Sensor Controlling Hepatocellular Bioenergetics and Actin Cytoskeletal Functions
48/2018 Liisa Pelttari
Genetics of Breast and Ovarian Cancer Predisposition with a Focus on *RAD51C* and *RAD51D* Genes
49/2018 Juha Gogulski
Prefrontal Control of the Tactile Sense
50/2018 Riku Turkki
Computer Vision for Tissue Characterization and Outcome Prediction in Cancer
51/2018 Khalid Saeed
Functional Testing of Urological Cancer Models by RNAi and Drug Libraries
52/2018 Johanna I. Kiiski
FANCM Mutations in Breast Cancer Risk and Survival
53/2018 Jere Weltner
Novel Approaches for Pluripotent Reprogramming
54/2018 Diego Balboa Alonso
Human Pluripotent Stem Cells and CRISPR-Cas9 Genome Editing to Model Diabetes
55/2018 Pauli Pöyhönen
Cardiovascular Magnetic Resonance Evaluation and Risk Stratification of Myocardial Diseases
56/2018 Pyry N. Sipilä
Dissecting Epidemiological Associations in Alcohol Drinking and Anorexia Nervosa
57/2018 Elisa Lahtela
Genetic Variants Predisposing to Prognosis in Pulmonary Sarcoidosis
58/2018 Ilari Sirenius
Lääkkeisiin ja lääkkeeen kaltaisiin tuotteisiin liittyvät toiveet ja illuusiot – psykodynaaminen näkökulma
59/2018 Nuno Nobre
Quality of Life of People Living with HIV/AIDS in Finland
60/2018 Pedro Miguel Barroso Inácio
The Value of Patient Reporting of Adverse Drug Reactions to Pharmacovigilance Systems
61/2018 Taru A. Muranen
Genetic Modifiers of *CHEK2*-Associated and Familial Breast Cancer
62/2018 Leena Seppä-Lassila
Acute Phase Proteins in Healthy and Sick Dairy and Beef Calves and Their Association with Growth
63/2018 Pekka Vartiainen
Health-Related Quality of Life in Patients with Chronic Pain
64/2018 Emilia Galli
Development of Analytical Tools for the Quantification of MANF and CDNF in Disease and Therapy
65/2018 Tommi Anttonen
Responses of Auditory Supporting Cells to Hair Cell Damage and Death: Cellular Stress Signalling and Epithelial Repair
66/2018 Muntasir Mamun Majumder
Improving Precision in Therapies for Hematological Malignancies

Helsinki 2018 ISSN 2342-3161 ISBN 978-951-51-4576-5