The impact of GPCR structures on understanding receptor function and ligand binding

Anirudh Ranganathan
The impact of GPCR structures on understanding receptor function and ligand binding

Anirudh Ranganathan
Abstract

G protein-coupled receptors (GPCRs) form the largest superfamily of eukaryotic membrane proteins and are responsible for the action of nearly 30% of all marketed drugs. For a long period, efforts to study these receptors were limited by the paucity of atomic-resolution structural information. Numerous receptors spread across the GPCR superfamily have recently been crystallized, revealing crucial clues about receptor function and ligand recognition. The work in this thesis has primarily focused on using computational techniques to capitalize on this increasing amount of structural information. In papers I, II, and III protocols were developed to identify novel ligands for pharmaceutically important targets from in silico screens of large chemical libraries. In these papers, the fragment-based lead discovery (FBLD) approach was evaluated for GPCR targets using molecular docking screens. The high hit-rates obtained in these studies indicate promise for the use of computational approaches for fragment screening. In paper IV, molecular dynamics was used to identify a possible role for a conserved ionizable residue (Asp79^{2.50}) as a protonation switch during the activation process of the β_2 adrenergic receptor. Analyses from this paper indicated that this residue could also perform a similar function in other class A GPCRs. Papers V and VI detail the modeling strategy followed during the GPCR Dock 2013 assessment to blindly predict the structure of two serotonin receptor subtypes (5-HT_{1B} and 5-HT_{2B}) bound to ergotamine. The developed ligand-steered homology modeling protocol was largely successful resulting in the best-ranked predictions for the 5-HT_{1B} subtype. It is hoped that the work described in this thesis has highlighted the potential for structure-based computational approaches to identify novel ligands for important pharmaceutical targets and improve understanding of GPCR function.
“I am not young enough to know everything”

_Oscar Wilde_
List of papers

The following papers are included in this thesis and are referred to by their roman numerals.

PAPER I: Complementarity between in silico and biophysical screening approaches in fragment-based lead discovery against the A2A adenosine receptor.

PAPER II: Fragment-Based discovery of subtype-selective adenosine receptor ligands from homology models.

PAPER III: Ligand discovery for a peptide GPCR by structure-based screening of fragment and lead-like chemical space.

PAPER IV: Insights into the role of Asp79\textsuperscript{2.50} in β\textsubscript{2} adrenergic receptor activation from molecular dynamics simulations.


PAPER VI: Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges.
The following publications are not included in this thesis.


PAPER VIII: Discovery of GPCR Ligands by Molecular Docking Screening: Novel Opportunities Provided by Crystal Structures.

* These authors contributed equally to this work.

Reprints made with permissions from publishers.
Contents

1. Introduction ................................................................................................................. 11
  1.1. G protein-coupled receptors (GPCRs) ................................................................. 12
  1.2. GPCR Signaling and Pharmacology ................................................................. 13
  1.3. The GPCR structural revolution ...................................................................... 15
  1.4. GPCR-ligand interactions: A structural perspective ........................................ 16
  1.5. Fragment-based lead discovery: A GPCR perspective ...................................... 17
  1.6. GPCR Function: A structural perspective ......................................................... 18

2. Methods ......................................................................................................................... 20
  2.1. Molecular dynamics simulations ..................................................................... 20
  2.2. Free energy perturbation (FEP) ................................................................. 22
  2.3. Molecular docking screening ....................................................................... 23
  2.4. Homology modeling ....................................................................................... 26

3. Results and Discussion ............................................................................................... 28
  3.1. FBLD guided by GPCR crystal structures ...................................................... 28
  3.2. Discovery of subtype-selective ligands using the fragment-based approach for GPCRs of unknown structure ................................................................. 30
  3.3. Structure-based screening of fragment and lead-like chemical space against a challenging GPCR target ................................................................. 33
  3.4. GPCR activation: the role of conserved ionizable residues ......................... 35
  3.5. Computational structure prediction for GPCR-ligand complexes (The GPCR Dock 2013 assessment) ................................................................. 38

4. Conclusions and future perspectives ........................................................................... 42

5. Summary in Swedish ..................................................................................................... 45

6. Acknowledgements ....................................................................................................... 46

7. References ................................................................................................................... 48
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-HT</td>
<td>5-hydroxytryptamine, Serotonin</td>
</tr>
<tr>
<td>AC</td>
<td>Adenylyl Cyclase</td>
</tr>
<tr>
<td>AR</td>
<td>Adenosine Receptor</td>
</tr>
<tr>
<td>ADR</td>
<td>Adrenergic Receptor</td>
</tr>
<tr>
<td>CRF</td>
<td>Corticotropin Releasing Factor</td>
</tr>
<tr>
<td>EL</td>
<td>Extracellular Loop</td>
</tr>
<tr>
<td>FBLD</td>
<td>Fragment-Based Lead Discovery</td>
</tr>
<tr>
<td>FEP</td>
<td>Free Energy Perturbation</td>
</tr>
<tr>
<td>GPCR</td>
<td>G Protein-Coupled Receptor</td>
</tr>
<tr>
<td>HR</td>
<td>Histamine Receptor</td>
</tr>
<tr>
<td>HTS</td>
<td>High-Throughput Screening</td>
</tr>
<tr>
<td>ICL</td>
<td>Intracellular Loop</td>
</tr>
<tr>
<td>LE</td>
<td>Ligand Efficiency</td>
</tr>
<tr>
<td>LELP</td>
<td>Ligand Efficiency Dependent Lipophilicity</td>
</tr>
<tr>
<td>LJ</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>LLE</td>
<td>Lipophilic Ligand Efficiency</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>mGluR</td>
<td>Metabotropic Glutamate Receptor</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular Mechanics</td>
</tr>
<tr>
<td>MR</td>
<td>Muscarinic Receptor</td>
</tr>
<tr>
<td>NAM</td>
<td>Negative Allosteric Modulator</td>
</tr>
<tr>
<td>NTS</td>
<td>Neurotensin</td>
</tr>
<tr>
<td>PAINS</td>
<td>Pan-Assay Interference Compounds</td>
</tr>
<tr>
<td>PBC</td>
<td>Periodic Boundary Conditions</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-Mean-Square Deviation</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operator Characteristic</td>
</tr>
<tr>
<td>SAR</td>
<td>Structure-Activity Relationship</td>
</tr>
<tr>
<td>SBC</td>
<td>Spherical Boundary Conditions</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface Plasmon Resonance</td>
</tr>
<tr>
<td>TINS</td>
<td>Target Immobilized NMR Screening</td>
</tr>
</tbody>
</table>
1. Introduction

The need to communicate is universal and essential for organization at even the most basic level. At a cellular level, such organization is essential for the evolution and sustenance of complex life forms. We have now delved into the world of inter-cell communication where G protein-coupled receptors (GPCRs) form key links within this extensive chain. They are located in membranes that separate individual cells from the “outside world”, which in essence is the remainder of the universe. Typically these receptors are responsible for receiving a “message” from the outside. This is usually in the form of a molecule (hormones for example), which could bind to a specific GPCR producing a signal within the cell thereby achieving precise communication. This process, known as signal transduction, takes place through GPCRs in a majority of cases. Complex organisms have evolved an amazingly elaborate system of checks and balances through an array of GPCRs to keep this vital process running smoothly. However, as is easy to imagine, any major malfunction above and beyond the fail-safes, could quickly lead to drastic effects and result in various diseases. Since GPCRs play key roles in maintaining the communication network, they have naturally become important targets for drug discovery.\(^1\)

The work described in this thesis aims to capitalize on recent advances in GPCR structural biology, which have offered atomic level detail on the structure of such receptors. Such information is vital as it allows one to move from a bird’s eye view to a much higher level of resolution while studying these receptors. This increase in detail allows theoreticians to apply powerful computational tools to complement experimental methods, which aid in increasing our understanding of GPCR function.

This thesis can be divided into three distinct focus areas: (i) understanding how GPCRs signal, (ii) development of new computational protocols as tools for drug discovery, and (iii) identification of new compounds that bind to GPCRs. It is hoped that the results included in this thesis could improve understanding of GPCR function, ligand recognition, and guide future developments within this field.
1.1. G protein-coupled receptors (GPCRs)

All work reported in this thesis has been focused on GPCRs. This chapter aims to provide insights into GPCR function and the importance of these receptors for the development of therapeutics for a wide range of disorders and diseases.

GPCRs form the largest superfamily of eukaryotic membrane receptors and are essential to a large number of signal transduction pathways. They share a common seven helix topology that spans the cell membrane and these are interspersed by three extra- and intra-cellular loops. The ~800 members of the human GPCR superfamily can be classified into five main classes: Rhodopsin-like (class A), Secretin (Class B), Glutamate (Class C), Adhesion, and Frizzled/Smoother families, of which Class A is the most populated. The projects that form this thesis have focused on Class A GPCRs. For class A GPCRs, the conserved topology permits the use of the Ballesteros-Weinstein residue numbering system that allows for a direct comparison of equivalent positions across different receptors. The numbering scheme follows a X.YY syntax where X denotes the helix number and YY is a sequence-based correlative number centered on the value 50, which is assigned to the most conserved residue in each helix. Numbering of residues utilizing this scheme is indicated in superscript and is used throughout this thesis.

GPCR activation typically involves ligand recognition in an orthosteric site located in the extracellular half of the receptor. This process then triggers a conformational rearrangement allowing for the coupling of an intracellular partner (e.g. G protein), leading to downstream signaling. GPCRs have evolved to recognize a plethora of different extracellular effectors, ranging from neurotransmitters to peptides, hormones, and even light.

A fine-tuned regulation of signal transduction by GPCRs is achieved in many ways, which include the existence of multiple subtypes recognizing the same endogenous ligand, allosteric regulation, multiple signaling pathways, and tissue-specific expression levels. Malfunctioning of pathways, however, can lead to disease states, and GPCRs are implicated in serious disorders, e.g. cancer. Due to their key roles in signaling, GPCRs are valuable drug targets and nearly 30% of all marketed drugs mediate their effects via these receptors. In the following section, receptor pharmacology and the mode of action of GPCR drugs will be explored further.
1.2. GPCR Signaling and Pharmacology

GPCRs signal via G proteins and one such signaling cycle is shown in Figure 1.

Figure 1. Example of GPCR signaling via heterotrimeric G proteins\textsuperscript{10,11}. In the resting state all constituent units of the G protein heterotrimer (α, β, and γ subunits) are thought to be localized at the GPCR, with the Gα binding GDP. Upon binding of the extracellular effector (agonist) to the GPCR, conformational changes are induced in the GPCR-G protein complex, causing nucleotide exchange from GDP to GTP at the Gα. Nucleotide exchange promotes dissociation of Gα-GTP complex from the β and γ subunits. The Gα-GTP complex can then interact with other regulatory proteins such as adenylyl cyclase (AC), producing changes in second messenger levels (e.g. cAMP) inside the cell. After this process, GTP is reconverted to GDP and the Gα-GDP complex preferentially reforms the heterotrimer with the β and γ subunits restarting the signaling cycle. Image reprinted with permission\textsuperscript{11}.

GPCR ligands that are recognized at the orthosteric site can be primarily divided into three different categories based on their effect on receptor signaling (efficacy)\textsuperscript{5}:

I. **Agonists:** Endogenous compounds such as dopamine, serotonin, and adrenaline fall under this category of molecules. Binding of these ligands stimulate signaling through the receptor, and agonist activity of any ligand is measured relative to the peak effect achieved by the endogenous compound (e.g. effect of dopamine on a dopamine receptor). Ligands that achieve the same effect as the endogenous agonist are called full agonists, while those that elicit a lower response are referred to as partial agonists. Additionally, GPCRs can signal via different intracellular effectors and agonists that preferentially drive activation of a specific pathway (e.g. G protein, β-arrestin) are called biased agonists.
II. **Neutral antagonists:** GPCRs can exhibit agonist-independent or basal signaling. Neutral antagonists are ligands that do not affect the basal activity of the GPCR. However, the binding of these ligands effectively block the receptor binding site from endogenous agonists.

III. **Inverse agonists:** As the name suggests, these ligands have the opposite effect to full agonists, shifting the equilibrium towards inactive states. Hence, in addition to the blocking action of antagonists these molecules also deactivate the receptor.

![Figure 2. Receptor pharmacology](image)

**Figure 2. Receptor pharmacology.** Classification of ligands according to the biological response elicited upon binding to the GPCR.

Depending on the disease state and the specific GPCR target in question, different signaling properties of ligands are required for drug candidates. Recently, more complicated binding and potency profiles are being sought. These desired profiles include allosteric modulators that bind at sites distinct from the orthosteric site and biased ligands that preferentially activate one signaling pathway. A lot of the increased understanding of GPCR ligand recognition and signaling over the last few years has been significantly aided by the determination of crystal structures for GPCRs. These structural breakthroughs were due to major advances in GPCR crystallography, which will be described in the following section.
1.3. The GPCR structural revolution

The first crystal structure of a protein, sperm whale myoglobin\textsuperscript{30}, was solved in the late 1950’s, but it was nearly half a century later that the first structure of a GPCR in complex with a diffusible ligand was determined\textsuperscript{21,22}. The difficulty in obtaining crystal structures for GPCRs was because of the characteristics of the proteins themselves. As was detailed in the previous section, GPCRs are signaling proteins, making flexibility and the availability of multiple conformational states a functional requirement of these receptors\textsuperscript{23}. Furthermore, as membrane proteins, they become unstable outside their native environments, for example in short-chain detergents required for crystallization\textsuperscript{24}. Since they are membrane receptors, they also have large hydrophobic patches and reduced polar surface areas, which again hinders crystal formation\textsuperscript{25}. Although the above is not an exhaustive list of challenges for the determination of GPCR crystal structures, these were some major hurdles that needed to be cleared along the way\textsuperscript{23}.

The main strategies for GPCR crystallization were aimed at reducing receptor flexibility and increasing their stability outside the native membrane environment. Some successful strategies that helped to achieve these goals were increasing the polar surface area\textsuperscript{21,22}, thermostabilizing mutations\textsuperscript{24}, and the use of single-domain antibodies (nanobodies)\textsuperscript{26}. In addition to these, the discovery of specific lipidic phases that allowed for the use of suitable shorter chain detergents was also vital for crystallization\textsuperscript{21-23}. Additionally, the removal of the long, disordered intracellular loop 3 (ICL3) played an important role by decreasing flexibility in a key region where crystal contacts are established. Replacement of ICL3 with the protein lysozyme in turn increased the polar surface area, which aided in the formation of crystal contacts allowing for the achievement of well-diffracting crystals\textsuperscript{19,23}. Another orthogonal strategy for reducing the flexibility of GPCRs was the use of mutants that pushed the receptor into either inactive- or active-like conformations\textsuperscript{27,28}. Increases in thermal stability of the active or inactive states compared to the wild-type receptors permitted crystallization even with weakly binding ligands\textsuperscript{29}. The availability of structural information was immediately met with excitement for the opportunities it provided for structure-based drug design against these pharmaceutically important receptors. In the following section, the impact of crystal structures on our understanding of ligand binding will be described.
1.4. GPCR-ligand interactions: A structural perspective

With steady growth in GPCR structural information a large diversity in binding site locations and ligand recognition patterns have been revealed\(^6,16,30,31\) (Figure 3). Atomic resolution structures for GPCRs from four out of five classes in complex with orthosteric and allosteric ligands are now available\(^30\).

Figure 3. Location of binding sites from crystal structures of receptor-ligand complexes of different GPCR classes. (A, B) Main orthosteric site for class A GPCRs, exemplified by the structures of the \(\beta_2\)-ADR bound to the inverse agonist carazolol (PDB code 2RH1)\(^{21,22}\) and adrenaline (PDB code 4LDO)\(^32\), respectively. (C) Structure of the CRFR, a Class B GPCR, bound to the antagonist CP-376395 (PDB code 4K5Y)\(^33\). (D) The M\(_3\)MR (Class A GPCR) bound to the allosteric modulator LY2119620 (PDB code 4MQT)\(^34\). (E) Structure of the mGluR, a Class C GPCR, bound to the NAM mavlogurant (PDB code 4OO9)\(^35\). (F) Structure of the sodium binding site of the A\(_{2A}\)AR (PDB code 4EIY)\(^36\). Image reprinted with permission\(^30\).

A small-molecule binding class A GPCR, the \(\beta_2\) adrenergic receptor (\(\beta_2\)-ADR), is shown in panels A-B of Figure 3. In this case, there is one strong polar interaction, which is the salt-bridge to Asp113\(^33,32\). Additionally, hydrophobic residues in the binding site provide ideal desolvated environments for the more non-polar portion of the ligand resulting in increased binding affinity. This pattern is repeated across many other aminergic class A GPCRs (e.g. dopamine, histamine, and serotonin receptors)\(^35,39\) that have evolved to recognize small-molecule ligands. On the other hand, as Figure 3
also shows, new structural information about GPCRs from other classes has revealed major differences in binding site locations\(^\text{30}\). Each of these show unique interaction fingerprints that determine the nature of the ligands that are recognized by these sites\(^\text{31}\). Overall, the structures have revealed how despite sharing a common fold GPCRs have evolved binding sites capable of recognizing many different endogenous ligand chemotypes. This rapid rise in information is ushering in a new era in structure-based drug discovery for GPCRs\(^\text{30,31}\).

### 1.5. Fragment-based lead discovery: A GPCR perspective

The development of high throughput screening (HTS) has revolutionized early stage lead discovery by allowing rapid experimental screening of millions of compounds\(^\text{40}\). These screening libraries often contain many high molecular weight compounds\(^\text{41,42}\). In recent years, fragment-based lead discovery (FBLD) has steadily grown as an alternative to HTS for lead identification\(^\text{43-46}\). The philosophy of this approach is centered on identifying ligands that are approximately half the size of a typical drug molecule\(^\text{43}\). The restriction of molecular size serves a dual purpose:

- It reduces the size of chemical space such that even small fragment libraries can cover significantly larger portions of this space compared to HTS collections of drug-like compounds containing millions of molecules\(^\text{44}\).
- The smaller size of the fragment allows for it to find an optimum subpocket within the binding site with reduced steric mismatches. This is in contrast to larger molecules, which need to be ideally accommodated in the binding site\(^\text{46}\).

These factors have generally meant that fragment screens result in higher hit rates than HTS thus providing a greater number of starting points for further optimization. However, the ligands identified are often of low affinity and require further elaboration to attain high potency. For fragment screens, size-corrected metrics such as ligand efficiency (LE), lipophilic LE (LLE), or LE-dependent-lipophilicity (LELP) are used to judge ligands rather than just binding affinity\(^\text{41}\). The LE metric provides the contribution of each heavy atom to the binding free energy as shown in equation 1:

\[
\text{LE} = - \frac{RT\ln(K_d)}{N} \quad (1)
\]
where $K_d$ is the dissociation constant, $R$ the universal gas constant, $T$ the temperature and $N$ is the number of ligand heavy atoms. LE values above 0.3 kcal·mol$^{-1}$·heavy atom$^{-1}$ are considered promising for drug discovery$^{41}$.

To detect weakly binding fragment ligands sensitive screening techniques such as surface plasmon resonance (SPR)$^{47,48}$, nuclear magnetic resonance (NMR)$^{48,49}$ or crystallography$^{50}$ are usually required. Stabilization techniques that paved the way for the GPCR structural revolution also allowed for the development of such sensitive assays opening up the FBLD approach to this pharmaceutically important superfamily$^{48,51}$. The crystal structures have themselves further provided the opportunity to screen large virtual libraries of fragments using molecular docking$^{30,51}$. Many GPCRs have evolved binding sites that recognize endogenous fragment-sized ligands, which make them particularly suited for the FBLD approach. Several projects detailed in this thesis involve fragment screening against GPCRs and will be discussed in greater detail in later chapters.

1.6. GPCR Function: A structural perspective

Over the last decade, there has been a rapid increase in understanding GPCR function$^{5,6}$. The growing data has begun to paint a much more complex picture than the two state model for activation that was initially envisioned for GPCRs$^{5,52}$. As Figure 4 below shows, GPCRs can exhibit many different conformational states.

![Figure 4. Different conformational states of GPCRs revealed by NMR and MD simulations.](image)

A. An overview of the different signaling pathways. B. Conformations stabilized by agonist and G protein binding is shown for the $\beta_2$-ADR. C. The energy landscapes associated with the receptor in the basal, agonist-bound, and active (bound to agonist and G protein mimetic: nanobody Nb80) states. Image reused with permission$^{52}$. 

18
The fact that many different GPCRs recognize a wide variety of endogenous compounds but signal through relatively small numbers of G protein mediated pathways points to a relatively conserved mechanism of activation. A deeper understanding of the link between interactions of ligands in the binding site and their functional profile is just starting to be understood\(^{53-55}\). The structures are now providing atomic-level detail about the roles of conserved motifs (e.g. NPXXY, DRY, and CWXP motifs) in the function of class A GPCRs. A key development in this area was when a high-resolution crystal structure of the fully active state for the \(\beta_2\)ADR bound to an agonist and G protein was determined, highlighting the conformational changes accompanying activation\(^{11}\). This structure revealed that activation only involved subtle changes on the extracellular half of the receptor, while large changes in the intracellular portion allowed for coupling of the G protein\(^{11}\). There is also a higher sequence conservation observed for residues in the intracellular region of GPCRs compared to other parts of the receptor\(^{56}\). Together, these observations have begun to demonstrate how different GPCRs can bind a wide variety of effectors extracellularly and yet elicit responses through the same intracellular partners and pathways. NMR in conjunction with computer simulations are being used to shed further light on the intricate network of switches connecting the orthosteric site with the intracellular G protein binding site\(^{52,57}\).

There is still no clear mechanism explaining the complete process of GPCR activation\(^{58}\). However, there is now more information regarding possible microswitches that might be key players along the activation pathway\(^{59-61}\). The ability of a single receptor to activate multiple intracellular signaling pathways is also attracting focus to try to understand this from a structural perspective\(^{62,63}\). Furthermore, the discovery of ligands that are able to selectively activate one signaling pathway over others\(^{15}\) raises more questions regarding the structural drivers for such properties. Finally, despite the rapid increase in structural information, many GPCRs are yet to be crystallized. Modeling assessments have proved that computational structure prediction of GPCRs distantly related to those with an experimental structure is extremely challenging (paper VI). Hence, many aspects related to GPCR signaling and ligand binding remain largely unknown and thus provide exciting opportunities for future work.
2. Methods

In this section the computational methods used in this thesis will be covered.

2.1. Molecular dynamics simulations

Crystal structures have revealed valuable atomic level detail on the process of GPCR ligand binding and activation\(^5,6\). However, these provide a single snapshot of the receptor in complex with its binding partners. MD simulations can be used to study receptor dynamics and further understand the mechanism of GPCR activation\(^53\). MD simulations based on molecular mechanics (MM) are generally performed with the use of parameterized force fields. Descriptions of MM and force fields are provided below.

MM is based on Newtonian mechanics and is used to describe a system of interacting particles. The forces acting on such a system can be divided into bonded (up to 4 connected atoms) and non-bonded (between disconnected atoms, separated by more than 3 bonds) terms. The potential energy for a system could be described with equation 2:

\[
U_{\text{pot}} = \sum_{\text{bonds}} \frac{1}{2} K_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \\
\sum_{\text{dihedrals}} \frac{K_\phi}{2} [1 + \cos(n\varphi - \delta)] + \\
\sum_{\text{impropers}} \frac{1}{2} K_\xi (\xi - \xi_0)^2 + \sum_{\text{coulomb}} \left( \frac{q_i q_j}{r_{ij}} \right) + \sum_{\text{LJ}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \tag{2}
\]

The first four terms in equation 2 describe the bonded portion, and the latter two represent non-bonded interactions. The bonded term can be resolved into components representing bonds (harmonic oscillator, Hooke’s law), angles (between two adjacent bonds), and dihedrals (between two adjacent angles, rigid rotor). The non-bonded term can be separated into electrostatics from Coulomb’s law and the Lennard-Jones (LJ) function, which represents van der Waals interactions between two atoms. The equilibrium values for all bond lengths \(r_0\), angles \(\theta_0\), and dihedrals \(\delta\) together with the force constants \(K_b, K_\theta, K_\phi\), phases of rotation \(n\) are parameterized as part of a force field representation for a given system. Similarly, the partial charges on every atom \(q\) along with the depth and location (distance of separation)
of the potential energy well for the LJ function \((A_{ij}, B_{ij})\) are provided depending on the force field. Some popular examples of force fields for proteins are OPLS-AA\(^{64}\), AMBER\(^{65}\), and CHARMM\(^{66}\).

An MD simulation can be performed using the force field parameters. Such a simulation allows the system to evolve by calculating the force acting on every particle as a function of time after providing some initial velocity (e.g. from a Maxwell Boltzmann distribution) to the reference (static) 3D representation of the system (e.g. protein structure from PDB). Prior to simulation with MD, an initial protein structure needs to be prepared by assigning specific ionization states to titratable residues, and disulfide bridges. During the simulation, forces on an atom are calculated as a gradient of the potential energy function in equation 2. The acceleration for an atom \(i\) is then calculated using Newton’s second law of motion:

\[
F_i = -\frac{\partial U_{pot}}{\partial r_i} = ma_i \quad (3)
\]

The positions and velocities at a particular step in the simulation are calculated based on these values at the previous time or half-time step using modifications of the Taylor expansion (e.g. leap-frog Verlet scheme).

\[
r_i(t + \Delta t) = r_i(t) + v_i \left( t + \frac{\Delta t}{2} \right) \Delta t \quad (4)
\]

\[
v_i \left( t + \frac{\Delta t}{2} \right) = v_i \left( t - \frac{\Delta t}{2} \right) + a_i(t)\Delta t \quad (5)
\]

A time step \((\Delta t)\) is chosen on the basis of the fastest motion (bond vibration) and is typically set to 1-2 femtosecond. The time-averaged energies arising from simulations can then be used to measure thermodynamic and kinetic properties of the system.

The boundaries of the simulated system need to be chosen, along with proper treatment of interactions close to the edges. There are two main philosophies for assigning boundaries to the system: periodic boundary condition (PBC) and spherical boundary condition (SBC). In the former, the system is set in a solvated box and replicas of such boxes are created. The atoms close to the edge can interact with atoms in the neighboring box, and a particle crossing the boundary between these appears on the other side, thus preserving the number of particles in the system. In SBC, a sphere is used to delineate the flexible from the rigid parts of the system. All atoms within the sphere are treated as fully flexible, whereas all particles outside this bounda-
ry are rigidly held to their starting coordinates. Restraints are applied close to the sphere edge in order to preserve properties such as solvent density and polarization. There are many popular MD programs, and in this thesis GROMACS was used for PBC simulations, whereas Q was used for those involving SBC.

### 2.2. Free energy perturbation (FEP)

The free energy associated with the transformation of a state A to B can be calculated with FEP, a method based on a formulation by Zwanzig in 1954. This transformation is performed by treating the end state B as a perturbation on state A. The FEP method is used to calculate free energy changes from simulations (MD or Monte Carlo). The free energy is calculated as shown below in equation 6:

$$
\Delta G = G_B - G_A = -k_B T \ln \langle e^{-(U_B-U_A)/k_B T} \rangle_A
$$  

where $k_B$ is the Boltzmann constant, $T$ is the temperature, $G$ is the Gibbs free energy, $U_A$ and $U_B$ are the potential energies of states A and B, and $\langle .. \rangle_A$ represents the ensemble average of energies from simulations on state A.

The potential energy of state B is calculated from a simulation on state A. For such a calculation to be accurate the state B must also be adequately sampled from a simulation of state A, or in other words there must be a large overlap in the potential energy surfaces of these two states. This condition is very rarely fulfilled for two end-states A and B in a chemical transformation. However, as free energy is a state function, it is possible to divide the path along $A \rightarrow B$ into a series of smaller perturbations through “non-physical” intermediates. The free energy of transforming A to B can then be provided by the summation of each of these small perturbations, as shown in equations 7 and 8:

$$
U_m = (1 - \lambda_m)U_A + \lambda_m U_B
$$  

where $\lambda_m$ is a parameter that discretely varies between 0 and 1, whose spacing determines the number of intermediate steps in the transformation $A \rightarrow B$. Larger transformations often require many intermediate steps.

$$
\Delta G = G_B - G_A = -k_B T \sum_{m=1}^{n-1} \ln \langle e^{-(U_{m+1}-U_m)/k_B T} \rangle_m
$$  

22
Many biological processes such as GPCR activation, or ligand diffusion into the binding site, are too slow to be simulated with available computational resources. Hence, for such processes, we use a thermodynamic cycle to calculate relative free energies using FEP as shown below with an example of a calculation for obtaining relative binding free energies of two ligands L and L* (Figure 5).

![Diagram](image)

**Figure 5. The use of a thermodynamic cycle in FEP calculations.** $\Delta G_1$ and $\Delta G_3$ involve processes whose time-scales are often out-of-reach with all atom simulations (with current computational power). FEP transformations are instead performed from the ligand L to L* bound to the receptor ($\Delta G_2$) and free in water ($\Delta G_4$). From the thermodynamic cycle, the relative binding affinities ($\Delta\Delta G$) for the two ligands can be calculated.

## 2.3. Molecular docking screening

Molecular docking is a technique that aims to provide rapid structural prediction of receptor-ligand complexes and their absolute binding energies. There are two main steps within a typical docking algorithm, sampling and scoring. In the sampling step, various ligand conformations within a pre-defined binding pocket are generated, and the binding energies of these are calculated in the scoring step. To run a molecular docking calculation, a receptor structure is required, and these can be either experimentally determined (e.g. X-ray crystallography or NMR) or computationally predicted (e.g. homology modeling). A desired feature of molecular docking algorithms is computational speed. In order to achieve this, many approximations are made during these calculations. One of the primary approximations is the restriction of search space to a pre-defined binding pocket. Additionally, the receptor is often rigid and scoring functions are simplified to allow for greater computational efficiency. These approximations allow molecular docking to perform a swift ranking of large databases, which has aided the discovery of ligands in many cases$^{30,31}$. 
There are different sampling (e.g. anchor-and-grow, genetic algorithm) and scoring (e.g. force field based, knowledge-based, empirical) schemes used for molecular docking. Some examples of widely used programs are DOCK\textsuperscript{71}, GLIDE\textsuperscript{72}, GOLD\textsuperscript{73}, and AUTODOCK\textsuperscript{74}. In this thesis, DOCK was used for molecular docking. DOCK utilizes the anchor-and-grow approach for sampling and a physics-based scoring function is used to calculate the binding energy based on the united atom AMBER\textsuperscript{75} force field. The scoring function is composed of electrostatics, van der Waals (Lennard-Jones potential), and ligand desolvation energies. Desolvation is calculated as the energy for transferring the fraction of a ligand within the binding site from a dielectric of 78 to 2\textsuperscript{76}.

For sampling, the ligand is decomposed into rigid fragments called anchors, which are first sampled within the pre-defined search space (binding site) in different orientations. The remainder of the ligand is then sampled within the receptor by attaching these to the different orientations of the anchor, referred to as growing. For computational efficiency, a set of conformations are pre-generated for compounds in the library prior to docking by allowing for rotations of the flexible portions of the molecules. Of course, the binding site will strongly influence the growing step and restrict the number of possibilities at this stage. Finally, the entire ligand is minimized and scored, usually keeping the lowest energy conformation or pose. In the second step, scoring, it would be very expensive to calculate energies for every new conformation of the molecule in the binding site. Instead, since the protein is held rigid, a grid can be pre-generated containing the information regarding electrostatics and van der Waals parameters. The interaction energies of commonly occurring atoms can be pre-calculated on this grid allowing for rapid scoring.

Prior to docking, the binding site has to be prepared by assigning protonation and tautomeric states to ionizable residues and inclusion of waters or cofactors as required. Two common approaches are used to evaluate the performance of a docking setup. These are (i) redocking, where one attempts to restore the binding mode of the co-crystallized ligand in the binding site and (ii) evaluation of enrichment of known ligands over decoys, where the ability of docking to rank true ligands high compared to non-binders (decoys) is tested. The latter approach is particularly important if docking is to be used prospectively to identify novel ligands. Such evaluations of the docking setup can be done iteratively by altering sampling settings, including binding site waters, pose re-scoring with an orthogonal method, or using other receptor conformations\textsuperscript{31}. The compounds to be docked also need to be pre-prepared prior to the calculation. All relevant tautomeric and protonation states at a relevant pH need to be considered. Ligand sets are then generated by applying property filters, e.g. Lipinski’s rule of five\textsuperscript{77}, fragment rule of three\textsuperscript{78}, or project-specific properties such as blood-brain barrier penetration.
Additionally, pan-assay interfering compounds (PAINS)\textsuperscript{79} are often removed from compound libraries. Access to supercomputers and improving docking algorithms have made it possible to routinely screen databases with millions of compounds.

Compounds within the database of docked molecules are ranked based on their predicted binding energy following the calculation. Top-ranked compounds can optionally be post-processed by rescoring, filtering for contacts with key residues, clustering for chemotype diversity and evaluating novelty\textsuperscript{31}. Finally, the top-ranked molecules are generally visually inspected before selection of compounds for experimental evaluation. In this step, expert knowledge about key interactions for a receptor, specific approximations in the docking algorithm, and project-specific considerations can be added to the selection criteria. Selected molecules are tested experimentally for activity, and the success of a docking screen is quantified as a ratio of verified true ligands to the total number of compounds evaluated (hit-rate). The structure can then be used to optimize the identified ligands further for potency, selectivity or particular functional properties to generate a lead compound, which can then be progressed further along the drug discovery pipeline.

Many GPCR crystal structures have proved highly amenable to molecular docking. The success of screens as measured by the hit-rate has been remarkably high in several cases\textsuperscript{30}. Hit-rates up to 73\% against the H\textsubscript{1} histamine receptor (H\textsubscript{1}HR)\textsuperscript{80}, with ligands possessing affinities as high as 114 pM (\(\beta\text{\textsubscript{2}}\text{ADR}\))\textsuperscript{81} have been identified from such screens. Beginning in 2008, after the first structure of a pharmaceutically relevant GPCR, the \(\beta\text{\textsubscript{2}}\text{ADR}\textsuperscript{21,22},\) there has been a steady increase in in silico structure-based screens as shown in Figure 6.
Figure 6. A timeline for prospective docking screens carried out against crystal structures or homology models of GPCRs after the release of the first β₂AR crystal structure. Examples of predicted complexes for two discovered ligands from screens against the D₃DR and A₂AAR are shown to the left of the timeline. The orthosteric sites of the A₂AAR and D₃DR are shown as grey cartoons. Key residues and the predicted binding modes of the ligands are shown in sticks. (Ranganathan et al. submitted)

2.4. Homology modeling

Although the GPCR structural revolution has provided atomic-level detail for many receptors within this superfamily, structures for nearly 300 druggable members are currently unknown. At the current rate of crystal structure determination, this shortfall will continue for many years to come. Hence, computational protein structure prediction often needs to be used.

Homology modeling is a widely used method for protein structure prediction. The main principle behind this approach is that receptors that are closely related to each other in terms of sequence will share similar structures. Due to this, homology modeling can often only be reliably used to model receptors when a closely related structural template is available. Initially, the sequence of such a template with known structure is aligned to the se-
quence of the target receptor. The spatial coordinates of the receptor model are then optimized according to the template structure and sequence alignment. The generated model is typically energy minimized to resolve unfavorable residue-residue contacts. An example of a widely used program for homology modeling is MODELLER. The quality of a model can be assessed either by stereochemical properties or statistical potentials of similarity to native-like protein structures. In general, the largest uncertainties within the model are in the flexible loop regions of the receptor. Homology modeling is often carried out as an iterative process with suitable measure(s) used to pick models at each round. For example, if the model is to be used for virtual screening, then recognition of known ligands at the binding site could be used as a metric for iterative improvement. Such a combination of molecular docking and homology modeling is referred to as ligand-guided or ligand-steered homology modeling. As homology models offer little deviation from the overall template structure, these initial predictions can also be further refined using methods such as MD simulations to confer greater flexibility. Encouragingly, homology modeling of closely related receptors from the crystal structures is also beginning to contribute significantly to ligand discovery efforts (Figure 6).
3. Results and discussion

In the following sections, the main results of papers I-VI will be discussed.

3.1 FBLD guided by GPCR crystal structures

The GPCR structural revolution has opened up possibilities for structure-based drug discovery (SBDD) efforts against these important drug targets \(^{30,31}\) (see sections 1.3 and 2.3). It was hence exciting to test if computational screens using structure-based methods such as molecular docking could provide an efficient means of identifying fragment ligands. The first such screen \(^{80}\) of a fragment library against a GPCR target was performed with the H\(_1\) histamine receptor crystal structure \(^{38}\). This study demonstrated the promise of molecular docking and the fragment-based approach for GPCR drug discovery. In paper I, we took advantage of a recent crystal structure of the A\(_{2A}\) adenosine receptor (A\(_{2A}\)AR) \(^{89}\) to further test the effectiveness of in silico fragment screening using molecular docking. The A\(_{2A}\)AR is an important drug target for the development of pharmaceuticals against Parkinson’s disease and asthma \(^{90,91}\). The primary aim of our work was to evaluate the complementarity between empirical and computational approaches for fragment screening against GPCR targets.

As a first step, a small library of 500 fragments was screened in parallel using molecular docking and target immobilized NMR screening (TINS) \(^{49}\). Compounds that elicited a response with TINS (94 compounds) were shortlisted and further investigations revealed five orthosteric ligands (5% hit-rate, 14–586 µM) and 11 allosteric modulators (12% hit-rate). Three of the five orthosteric ligands identified by the TINS screen possessed LE values > 0.3 kcal·mol\(^{-1}\)·heavy atom\(^{-1}\). Docking is often judged on its ability to rank ligands high compared to non-binders and this could be done using a receiver operator characteristic (ROC) curve. ROC curves are plotted as true positives rate versus false positive rates or in this case as percentages of true actives (ligands) against non-binders. The molecular docking screen conducted in parallel to TINS placed four out of five verified orthosteric ligands in the top 5% by rank. Hence, docking could indeed rank a majority of true fragment ligands high and demonstrated that it could be used to prioritize...
molecules for screening. Conversely, allosteric ligands were identified at a rate similar to random selection, which was likely due to the fact that the docking screen only targeted the orthosteric site. However, there were also many non-actives ranked within the top 50 molecules (10% of the database), which were essentially false positives from docking. Upon inspection, many of these were suggested to be due to approximations in the scoring function (e.g. neglect of receptor desolvation). There were a few compounds for which no such issues could be identified and we selected five such top-ranked molecules for re-evaluation in radioligand assays. Three out of five compounds showed measurable affinity at the A2A AR (17.6 – 128 µM) and were in fact among the most potent from this screen. It was hence demonstrated that molecular docking could complement empirical biophysical screens by identifying false negatives.

The true power of molecular docking is the ability to screen large libraries of up to millions of compounds. Encouraged by the performance of docking in the screen of a small library we proceeded to dock the entire ZINC fragment database\(^9\) of 328,000 compounds against the A2A AR. The molecules were sampled in thousands of orientations resulting in millions of evaluated complexes. The 500 top-ranked compounds were visually inspected and 22 compounds were selected for evaluation in radioligand binding assays. Fourteen fragments were found to be true ligands of the A2A AR with affinities ranging from 2.2–240 µM. Thirteen out of fourteen ligands showed LE values greater than 0.3 kcal·mol\(^{-1}\)·heavy atom\(^{-1}\) and could be considered promising scaffolds\(^41\). The 64% hit-rate obtained was nearly 13-fold higher than that from the TINS screen and was in line with the results from the first screen against the H\(_1\) histamine receptor where a hit-rate of 73% was achieved\(^80\). Taken together these results suggest that the small size of the endogenous compounds for many class A GPCRs may allow for fragment screens to be particularly successful for these receptors. Our study reiterated the suitability of structure-based methods such as molecular docking for fragment screening against GPCR targets and highlighted the advantages of complementing biophysical screens with \textit{in silico} approaches.

Fragment ligands are often weaker in terms of binding affinity than their lead-like counterparts. For this reason, fragments need to be optimized for binding affinity, a step that involves elaboration of the scaffold towards lead-like sizes. This can be accomplished in two ways, fragment growing and linking\(^43\). In this work, we pursued fragment growing by searching for analogs for three ligand series. The initial hits with the best affinities and a large availability of analogs were selected for optimization. The analogs were screened using molecular docking and compounds for experimental evaluation were selected after visual inspection. A modest three-fold improvement in affinity was achieved for two out of three series. FEP calculations from MD simulations were used to explain the SAR observed for one of the com-
pound series. The simulations suggested that ligand desolvation may be responsible for the observed ~10-fold reduction in affinity when a phenyl group was replaced by a pyridine.

Prior to our screen, structure-based efforts had indicated that a conformational bias might be encoded in GPCR structures depending on which functional state had been captured\(^{80,81,87,93,94}\). In paper I, the crystal structure screened represented the inactive state of the A\(_2A\)AR complexed with an antagonist\(^{89}\). In agreement with previous docking screens, the functionally characterized ligands were also found to be antagonists. The small size of fragment ligands was also thought to confer greater promiscuity\(^{95}\), and no significant selectivity was found over the closely related A\(_1\) subtype.

Overall, the high hit-rate obtained in this screen demonstrated the viability of structure-based fragment screening against GPCR crystal structures. The high ranks of true ligands obtained with molecular docking in the parallel evaluation with TINS showed that such structure-based approaches could be used to prioritize compounds for biophysical screens. The discovery of false negatives from TINS using molecular docking further highlighted this complementarity. Similarly, the discovery of novel allosteric modulators using TINS further strengthened the case for using orthogonal fragment screening methods for GPCRs as such molecules were out of reach with molecular docking in this case.

### 3.2. Discovery of subtype-selective ligands using the fragment-based approach for GPCRs of unknown structure

The results from paper I showcased the promise of both structure-based approaches and fragment screening to drive GPCR ligand discovery. One of the major challenges in GPCR drug discovery is to achieve subtype selectivity, as families of receptors have evolved to recognize the same endogenous ligand. In paper I, we did not specifically target subtype-specificity and no selectivity was obtained for the A\(_2A\)AR over the closely related A\(_1\)AR. Another factor that often hinders structure-based understanding of GPCR selectivity is that, despite the revolution in structural biology, crystal structures only exist for a fraction of the entire superfamily of receptors\(^6\). However, each crystal structure opens up the possibility of using computational techniques (e.g. homology modeling) to model closely related targets by using them as templates\(^96\). For example, the crystal structure of the A\(_2A\)AR could provide the possibility to understand selectivity within the AR family by using it as a template to model the other AR subtypes (A\(_1\), A\(_3\), and A\(_{2B}\))\(^34\).
paper II, we hence wanted to build on the previous work and extend FBLD
to GPCR homology models. Fragments are compounds with lower molecular
complexity compared to their lead- and drug-sized counterparts\textsuperscript{46}. This lower
molecular complexity of fragments has often not been sufficient to achieve
receptor selectivity\textsuperscript{95}. Instead, the initial fragment ligands typically need to
be elaborated to larger sizes to achieve the desired specificity, often a chal-
genning step in the drug discovery process. For soluble targets, crystal struc-
tures of fragment ligands could be used at this stage to guide optimization\textsuperscript{50}.
However, for membrane receptors this is in most cases not possible. Hence,
in this work, we wished to extend the challenge to not only discover frag-
ment ligands for a GPCR of unknown structure, but additionally attempt to
identify molecules that were selective against a closely related subtype that
also lacked an experimental structure. In paper II, the A\textsubscript{2}AR, a potential tar-
get for asthma and cancer was the subtype for which fragment ligands were
sought\textsuperscript{97,98}. The closely related A\textsubscript{1}AR, where unintended interactions could
lead to cardiac complications\textsuperscript{99} and increased risk of seizures\textsuperscript{98}, was selected
as the antitarget. Docking screens of the ZINC fragment library against hom-
ology models of these AR subtypes were performed to identify ligands
with A\textsubscript{3}/A\textsubscript{1} selectivity.

Earlier work by Rodriguez et al. and Kolb et al. had shown that obtaining
subtype selective ligands from screens of homology models was difficult\textsuperscript{100,101}. Hence, in paper II we developed a tailored strategy to identify
subtype selective fragments from virtual screens of large libraries. As there
were no crystal structures available for the A\textsubscript{3}- and A\textsubscript{1}AR, homology models
were obtained for these receptors based on a high-resolution A\textsubscript{2A}AR crystal
structure. A ligand-steered homology modeling strategy was followed. En-
richments of known A\textsubscript{3}- and A\textsubscript{1}AR ligands from docking screens were used
for model selection. As a last step, a set of known A\textsubscript{3}/A\textsubscript{1} selective ligands
was docked to the shortlisted models to select structures for screening. The
selection was based on attaining high enrichments of the selective ligands at
the target A\textsubscript{3}AR models but poor values for this set at the antitarget A\textsubscript{1}AR. A
single structure was selected for the target (A\textsubscript{3}AR), whereas an ensemble of
ten antitarget (A\textsubscript{1}AR) conformations representing different shapes of the
orthosteric site was chosen for the docking screens. The ensemble of struc-
tures were used to account for receptor flexibility and the ability of a ligand
to be accommodated in a slightly different conformation of the A\textsubscript{1}AR bind-
ing site. The retrospective docking screens and homology models obtained
were used to identify regions within the binding site that could provide sub-
type-selectivity. Two such regions were identified above and below the
family conserved Asn\textsuperscript{6.55}, which is key for ligand recognition. Approximate-
ly half a million fragments\textsuperscript{92} were screened against the target structure and
the antitarget ensemble with molecular docking using DOCK\textsuperscript{71}. Top-ranked
compounds at the target A\textsubscript{3}AR that also had a >10,000 difference in docking
ranks to its own highest position amongst the entire antitarget ensemble were
selected for visual inspection. Among the 500 top-ranked fragments at the $A_3$AR only 27 molecules fulfilled the second criteria, highlighting the similarity between the two sites. Instead, the 4,000 highest ranked fragments at the target receptor had to be considered to obtain 500 compounds for visual inspection. 21 compounds were selected that targeted the identified regions within the $A_3$AR binding site, which were hypothesized to be responsible for $A_3$/A₁ selectivity. Additionally, the predicted poses for these compounds in the antitarget ensemble were also inspected to ensure that these could not make optimal interactions. Eight compounds were found to bind to the $A_3$AR in live-cell fluorescence assays (38% hit-rate), six of these were $A_3$/A₁ selective, and four were found to be high-affinity ligands ($pK_i > 5.6$). Remarkably, none of the compounds with significant affinity were anti-selective. This result demonstrated that it was possible to discover fragments with high hit-rates from homology models. Furthermore, it was shown that fragment screens against receptors of unknown structure could indeed identify subtype-selective ligands.

The discovery of promising selective scaffolds from the docking screen provided us with an opportunity to test if subtype-specificity could be improved during fragment optimization. This could highlight if there was an advantage in targeting selectivity in fragment screening, as previous studies had provided complex results in this regard, with this property being lost or gained during the optimization process. To verify our predictions for the receptor-ligand complexes, compounds lacking the moieties that we hypothesized to be responsible for selectivity were initially tested. These analogs lost most of their specificity. We then tested close analogs for two scaffolds that expanded further in to the subpockets believed to be responsible for selectivity. The affinities of both scaffolds were improved multifold resulting in the identification of a 40 nM, 100-fold selective $A_3$AR lead. An inactive structure of the $A_{2A}$AR used to model the target and antitarget in this study, and in line with previous results, yielded ligands with the same functional profile ($A_3$AR antagonists) as the co-crystallized ligand.

To summarize, the results of paper II demonstrated that homology modeling and fragment screening could be used to identify subtype selective scaffolds for GPCRs. Subtype specificity, if well understood in terms of its structural basis, could be a valuable property for hit prioritization and guide further optimization. It is hoped that the strategy described here could be used to aid in future drug-discovery efforts for pharmaceutically important GPCRs of unknown structure.
3.3. Structure-based screening of fragment and lead-like chemical space against a challenging GPCR target

Peptide binding GPCRs often fall into the category of promising but largely undruggable targets that resist drug discovery efforts\textsuperscript{103}. The identification of ligands to such targets has instead been based on the endogenous peptide or mimetics of these\textsuperscript{104}. Structure-based efforts against peptide GPCRs have been rare\textsuperscript{105,106}, primarily due to the paucity of structural information and difficulties associated with targeting large, solvent exposed binding sites. A high-resolution crystal structure of the neurtensin receptor 1 (NTSR1) bound to a peptide agonist NTS8\textsuperscript{-13} was recently determined, providing atomic-level detail on the recognition of the peptide in the binding site\textsuperscript{107}. Stabilization of the receptor for crystallization also provided constructs suitable for use in biophysical screening. The availability of high-resolution crystal structures and a sensitive screening assay such as SPR made the NTSR1 ideally suited for an evaluation of structure-based approaches to identify leads for challenging GPCR drug targets.

For soluble targets, the FBLD approach has yielded hits in cases where HTS has failed\textsuperscript{43}. In these cases, the weak initial hits that result from the fragment screen can often be optimized on the basis of crystal structures\textsuperscript{30}. However, for membrane receptors this is usually not the case due to the difficulties associated with obtaining crystal structures\textsuperscript{23}. In general, the optimization steps required to convert a fragment hit into a lead compound have been described as one of the most challenging aspects of FBLD\textsuperscript{44}. To evaluate the FBLD approach against peptide GPCRs, we performed what was to our knowledge the first in silico fragment screen against such a target. In order to study the effectiveness of the FBLD approach, we also screened a lead-like library against the same receptor using molecular docking. In total, 1.8 million lead-like and 0.5 million fragment compounds were evaluated against the NTSR1 crystal structure. Top-ranked compounds from the docking screen were visually inspected, resulting in the selection of 27 lead-like molecules and 25 fragments for experimental testing by SPR. The results obtained would then be used to compare FBLD and lead-like screening against a challenging drug target such as the NTSR1.

The SPR evaluation revealed that a two-fold higher hit rate (44\%) was obtained for the fragment screen compared to that from the library of lead-like compounds (19\%). However, the identified lead-sized ligands possessed affinities that were usually two orders of magnitude better than that of the fragments. Analyses using the LE measure to correct for ligand size provided
much more overlap between results of the fragment and lead-like screens, yielding median values of 0.30 and 0.33 kcal·mol⁻¹·heavy atom⁻¹, respectively. A majority of the discovered ligands were also dissimilar to known NTSR1 ligands. In order to further understand the overlap in chemotypes between ligands identified from each library, substructures of leads and superstructures of fragments were identified from commercial chemical space. This search revealed that superstructures of fragments often exist in lead-like space, but large substituents prevent optimal interactions with the receptor in many cases leading to lower docking energies. This could perhaps be one of the main reasons for lower hit-rates obtained while screening ligands of larger size. On the other hand, substructures of leads were also found within commercial fragment space, but here a combination of low heavy atom count and absence of key moieties for ligand recognition prevented the achievement of good docking ranks in many instances. However, substructures that possessed the requisite moieties and a reasonable number of heavy atoms were often high ranked by the docking screen.

As has been highlighted previously, optimization of fragment compounds into leads is in many instances one of the most challenging steps in the FBLD pipeline. A novel tetrazole ligand from the fragment screen with a $K_d$ value of 192 µM and reasonable LE of 0.3 kcal·mol⁻¹·heavy atom⁻¹ was chosen for optimization efforts. Similarly, a novel scaffold shared by two lead ligands with $K_d$ values of 1.6 and 2.1 µM was selected from the screen of this library. Close analogs of these scaffolds were identified from commercial chemical space and docked to the NTSR1. A set of analogs for the fragment and lead scaffolds selected after visual inspection were evaluated with SPR. The expansion of a cyclopropyl substituent into a linear alkyl chain that was predicted to anchor the compounds deeper in the binding site proved crucial for the tetrazole series, with many analogs displaying significant improvements in binding affinity. Overall, a 46-fold improvement was achieved resulting in the discovery of a tetrazole lead with a $K_d$ value of 4.4 µM. The series of analogs from the lead scaffold mainly explored substituents on the two phenyl rings. The additional interactions provided by the substituents yielded many submicromolar leads from this series and the best compound possessed an impressive $K_d$ value of 370 nM. LE values for the optimized leads emerging from the fragment and lead-sized hits were also highly promising (0.36 and 0.42 kcal·mol⁻¹·heavy atom⁻¹) and significantly higher than reference antagonists of NTSR1¹⁰⁸,¹⁰⁹. Hence, optimization efforts guided by molecular docking were largely successful, resulting in the discovery of high-affinity leads starting from both fragment and lead scaffolds.

An aspect of drug discovery specific to GPCR targets is that ligands with particular efficacy are often sought. For peptide GPCRs, identification of small molecule antagonists is often challenging, and finding such ligands
that further mimic the complex interactions of the much larger endogenous agonist is thought to be much more difficult\textsuperscript{103,104}. Previous docking screens against GPCR crystal structures have in many cases yielded hits whose efficacies were in accordance with the respective co-crystallized ligands\textsuperscript{93,100,110}. Upon evaluation in functional assays, compounds possessing the lead scaffold that was selected for optimization were found to be agonists of the G protein pathway. The initial tetrazole hit from the fragment screen did not possess sufficient affinity to elicit any activation of NTSR1 in functional assays. However, the optimized tetrazole lead was found to activate the \(G_q\) protein pathway. Comparing the two approaches from this perspective, the higher affinities of lead-sized hits directly allows for functional characterization, while for the fragment compounds considerable efforts often need to be blindly invested to optimize the scaffold prior to testing for efficacy.

In conclusion, the results from paper III demonstrate that the increase in atomic-level information from the revolution in GPCR structural biology could have a major impact on ligand discovery efforts against challenging targets such as peptide recognizing receptors. Furthermore, it was shown that molecular docking screens with large libraries could successfully be extended to less druggable peptide GPCRs and provide diverse, novel leads with a high hit-rate. Screening different tranches of chemical space could offer complementary advantages, and crucially, these can be performed at little extra cost with molecular docking. Fragment screens provided a larger number of weaker hits, whose affinities could be optimized using the structure, whereas the smaller numbers of lead ligands were stronger starting points for further development. Overall, the screening protocol described in paper III could be applied to identify ligands for difficult drug targets.

### 3.4. GPCR activation: the role of conserved ionizable residues

The determination of a high-resolution crystal structure of the \(\beta_2\)ADR in a ternary complex with an agonist and G protein provided an atomic-level view on the conformational changes accompanying GPCR activation\textsuperscript{11}. This structure was met with excitement for the opportunities it provided to understand the activation process, a long-standing goal in biology. In paper IV, we explored the role of conserved and buried ionizable residues on the pH dependent activation of the \(\beta_2\)ADR. Kobilka and co-workers had demonstrated that the basal activity of the \(\beta_2\)ADR was higher at a pH of 6.5 than 8\textsuperscript{111}. The dependence of the equilibrium between inactive and active states (basal activity) on pH demonstrated that there was a shift in pK\(_a\) of at least one ionizable residue in the \(\beta_2\)ADR upon activation. Unlike rhodopsin, which showed
similar pH dependence\textsuperscript{112}, the substitution of the equivalent Asp130\textsuperscript{3,49} to alanine in the β\textsubscript{2}ADR did not ablate this effect. Two other candidate ionizable residues (Glu122\textsuperscript{3,41} and Glu268\textsuperscript{6,30}) were also substituted to alanines but did not significantly alter the pH dependence. Two aspartic acid residues, Asp113\textsuperscript{3,32} and Asp79\textsuperscript{2,50}, could not be evaluated experimentally as substitutions of these to alanine resulted in non-functional receptors\textsuperscript{111}. Access to crystal structures of the inactive state of the β\textsubscript{2}ADR bound to inverse agonist carazolol\textsuperscript{21,22}, and the active state bound to G protein\textsuperscript{11} provided us with an opportunity to evaluate the roles of these residues in receptor activation. We hoped that such an evaluation would provide crucial clues about the process of activation for the β\textsubscript{2}ADR.

Asp113\textsuperscript{3,32} is located in the orthosteric site of the receptor and is important for the recognition of both agonists and antagonists. This residue is conserved among the entire family of monoamine recognizing GPCRs that include the muscarinic, serotonin, dopamine, and histamine receptors. Crystal structures of the active and inactive states highlighted that the orthosteric site itself underwent relatively small changes when compared to the large conformational differences observed intracellularly. Asp79\textsuperscript{2,50} is conserved among 94\% of all class A GPCRs\textsuperscript{56}. It is located in the middle of the transmembrane region in a smaller pocket within the β\textsubscript{2}ADR. The burial and conservation of an ionizable residue deep in the transmembrane region was indicative of an important functional role, as it is energetically unfavorable to desolvate a charge. Furthermore, from the crystal structures of the inactive and active states, it could be observed that the hydrated cavity surrounding this residue partially collapses upon activation. The more significant change in the neighborhood around Asp79\textsuperscript{2,50} was interesting as it might indicate a shift in pK\textsubscript{a} for this residue upon activation.

In order to pinpoint the residue responsible for the pH dependence, MD/FEP calculations on the β\textsubscript{2}ADR (see section 2.2) were performed. Since activation of GPCRs occurs in the order of tens of milliseconds to seconds\textsuperscript{113}, it was not possible to simulate this process using all-atom MD. However, as crystal structures of the inactive and active states were available, it was feasible to instead perform a deprotonation reaction on Asp79\textsuperscript{2,50} and Asp113\textsuperscript{3,32} in both these conformations of the β\textsubscript{2}ADR. The relative differences in the free energies of deprotonation (ΔΔG) between the two states of the β\textsubscript{2}ADR would then provide the shift in pK\textsubscript{a} upon activation for the two aspartic acid residues. The MD/FEP calculations showed that there was a significant increase in the pK\textsubscript{a} of Asp79\textsuperscript{2,50} (3.7 pK\textsubscript{a} units). In contrast, the change in pK\textsubscript{a} for Asp113\textsuperscript{3,32} was smaller (0.7 pK\textsubscript{a} units), in agreement with the subtle conformational changes observed upon activation around this residue. This highlighted the increase in Asp79\textsuperscript{2,50} pK\textsubscript{a} as possibly the main reason for the pH dependent activation of the β\textsubscript{2}ADR. The main driver for the protonation of Asp79\textsuperscript{2,50} was found to be the dehydration of the surrounding...
cavity, which destabilizes the ionized state of this residue in the active conformation of the β2-ADR. PROPKA calculations\textsuperscript{114} suggested that the absolute pK\textsubscript{a} of this residue (Asp\textsuperscript{79\textsubscript{2.50}}) was elevated to ~7.5, and hence the changes in pK\textsubscript{a} could be relevant under physiological conditions. A protonated Asp\textsuperscript{79\textsubscript{2.50}} was a likely characteristic of the active state, which was in agreement with increased basal activity at low pH\textsuperscript{111}.

In order to evaluate the effect of the protonation state of Asp\textsuperscript{79\textsubscript{2.50}} on the conformation of the β\textsubscript{2}-ADR, microsecond simulations using PBC (see section 2.1) were performed and analyzed starting from the inactive and active states of the receptor. For simulations from the inactive state, no large-scale conformational changes were observed for the receptor overall, but microsecond simulation trajectories with the ionized Asp\textsuperscript{79\textsubscript{2.50}} consistently produced larger volumes of the allosteric cavity compared to those with the neutral Asp\textsuperscript{79\textsubscript{2.50}}. This was found to be due to differences in side chain conformations for residues such as Ser\textsubscript{120\textsuperscript{3.39}} and Phe\textsubscript{282\textsuperscript{6.44}} also located in this cavity. For the active state, we analyzed >100 µs of simulation trajectories performed by Dror et al. in an earlier publication\textsuperscript{53}. In this earlier work, large conformational changes leading to deactivation of the β\textsubscript{2}-ADR were observed with both protonation states of the Asp\textsuperscript{79\textsubscript{2.50}}. Analysis of these simulations from the perspective of the cavity highlighted that, despite overall deactivation of the receptor, the volume of this cavity remained closer to that observed in the crystal structure of the active state when the Asp\textsuperscript{79\textsubscript{2.50}} was neutral. The simulations with the ionized form of the Asp\textsuperscript{79\textsubscript{2.50}} on the other hand consistently produced volumes of this cavity closer to that observed in crystal structures of the inactive state. This difference in volumes appeared to be mainly due to an undistorted conformation of transmembrane helix 7 whose intracellular end was closer to helix 2 for simulations performed with neutral Asp\textsuperscript{79\textsubscript{2.50}}. This conformation was also observed for the agonist bound structure of the A\textsubscript{2A}AR\textsuperscript{115,116}. Similarly, in one of the simulations with ionized Asp\textsuperscript{79\textsubscript{2.50}}, sodium was found to enter and reside in the allosteric cavity, which almost exactly coincided with the position of this ion in crystal structures of the inactive β\textsubscript{1}-ADR\textsuperscript{117} and inactive A\textsubscript{2A}AR\textsuperscript{56}. Although, basal activity of β\textsubscript{2}-ADR has been found to be unaffected by sodium ions\textsuperscript{118} it might point to a similar binding pocket if present in this receptor. These results indicate that the conserved Asp\textsuperscript{2.50} protonation microswitch could be present in other class A GPCRs such as the A\textsubscript{2A}AR. For rhodopsin, two serines present within this allosteric region for the β\textsubscript{2}-ADR are occupied instead by alamines. Our analysis and PROPKA calculations\textsuperscript{114} suggested that this would increase the pK\textsubscript{a} of Asp\textsuperscript{50} in both active and inactive conformations of rhodopsin/opsin due to loss of hydrogen bonding with the serines. Hence, this residue would be protonated in both states, which was also in accordance with Fourier transform infrared spectroscopy experiments\textsuperscript{119}. 
During the course of this work Nygaard et al. published results from NMR studies on the β2AR where methionine peaks were characterized under different conditions. Hence, we compared the results obtained from our simulations to the NMR data that also studied changes that occurred upon activation of the receptor. For Met82, which resides just one turn above Asp79 in the β2AR, multiple peaks were found for the apo, and inactive states of the receptor, whereas this was reduced to a single peak in the agonist-bound and fully active states of this receptor. From our results, we proposed that the multiple peaks were due to the equilibrium between the protonated and ionized forms of the Asp79 in the apo and inactive states, which from our simulations were found to be accompanied by cavities of different sizes. For the active state, MD/FEP calculations predicted a large increase in residue pKₐ that resulted in a predominantly protonated Asp79, which was observed to be consistent with a smaller, collapsed allosteric cavity. Seemingly in agreement with these predictions, NMR experiments revealed a single peak for the Met82 under these conditions. Overall, the results of paper IV suggest that Asp79 may function as a microswitch in the β2AR by changing protonation state upon activation. Our analysis further indicated that these results could hold for other, but not all, class A GPCRs. We hope that the understanding gained in paper IV regarding the role of this conserved residue for β2AR function could shed further light on the complex process of GPCR activation.

3.5. Computational structure prediction for GPCR-ligand complexes (The GPCR Dock 2013 assessment)

The GPCR Dock assessments have challenged the modeling community to blindly predict structures of GPCR-ligand complexes prior to the release of crystallographic coordinates. There have been three editions of GPCR Dock conducted in 2008, 2010, and 2013 (paper VI). We participated in the GPCR Dock 2013 challenge to predict structures of two serotonin receptor subtypes 5-HT₁B and 5-HT₂B in complex with the ligand ergotamine. The assessment also involved prediction of the smoothened receptor (a class F GPCR) crystallized with two ligands LY29400680 and SANT-1, which we did not participate in. The serotonin receptors had structural templates (crystal structures) with sequence identity up to ~40% (e.g. turkey β₁AR). Participants were provided with sequences of the receptors and given 30 days to submit up to five models each for the protein-ligand complexes. A total of 352 models for the two serotonin receptor subtypes were submitted to the assessment. Two of our models submitted for the 5-HT₁B subtype were ranked #1 and #2 overall out of a total of 181, while our best-ranked predic-
tions for the 5-HT$_{2B}$ receptor were #5 and #7 out of 171. For the 5-HT$_{1B}$ subtype, our best submissions had the most accurate predictions of the binding site and receptor-ligand contacts with the best model attaining a ligand RMSD of 1.5 Å. For the 5-HT$_{2B}$ receptor, one submission possessed the most accurate binding site in the assessment, while the lowest ergotamine RMSD achieved for this subtype by our predictions was 1.6 Å.

Our strategy for the assessment was centered around a ligand-guided homology modeling approach. The methodology was based on an iterative process aimed at improving the ability of the generated receptor models to recognize known serotonin receptor ligands at each round. The main hypothesis behind our strategy was that if a binding site capable of discriminating between true ligands and decoys would also be a good representation of the receptor structure. The iterative process involved model generation followed by docking assessments at each round with average ligand enrichment over the generated models used as the primary measure of quality. A modeling round could consist of evaluations of different structural templates, perturbation of loop conformations, or restraints on side chain rotamers for binding site residues. A total of 61 homology modeling rounds were performed for the 5-HT$_{1B}$ and 5-HT$_{2B}$ subtypes resulting in the production of more than 16,000 models. This process allowed for a thorough retrospective evaluation of the followed strategy after the crystal coordinates became available.

The turkey β$_1$ADR with an ~40% sequence identity to the two serotonin receptor subtypes and ~50% binding site residue identity was selected as the primary template for homology modeling. In previous GPCR Dock assessments, the modeling of extracellular loop 2 (EL2) has been highlighted as particularly challenging$^{120}$. The inherent flexibility of loop regions, different loop lengths, sequence diversity, and the possibility for ligand-specific conformations were some of the main reasons for the difficulties faced in modeling EL2. Analysis of topologies for the portion of EL2 involved in ligand binding (a conserved Cys$^{45.50}$ to transmembrane helix 5) from available structures of aminergic GPCRs revealed crucial clues for the modeling of this region in the serotonin receptors. When viewed from an extracellular perspective, the topologies of EL2 could be clustered into two distinct folds. The first involved a relatively straight path towards transmembrane helix 5, whereas the second followed a bell-like shape. When viewed from a perspective perpendicular to the membrane plane, there appeared to be correlation between the residue position 45.52 in EL2 (two positions after a conserved disulfide bridge involving Cys$^{45.50}$) and position 3.33 in transmembrane helix 3 (Ballesteros-Weinstein numbering). The identities for this pair of residues seemed to direct loop conformations in a way that established hydrophobic contact between them, thus anchoring the EL2 in the binding site.
Analysis of loop regions in the 5-HT$_{1B}$ and 5-HT$_{2B}$ subtypes showed that, in contrast to the large phenylalanine present as the EL2 hydrophobic anchor in the template β$_1$ADR, the target receptors had smaller residues in this position (valine and leucine respectively). This indicated that whereas overall the β$_1$ADR was a good template it would not be ideal for modeling EL2. Instead, EL2 from the D$_3$ dopamine receptor (D$_3$DR) with a valine in this position appeared to be better suited as a template for modeling the two serotonin receptor subtypes. Hence, a chimeric template between β$_1$ADR and D$_3$DR (for EL2) was used to model the 5-HT$_{1B}$ receptor. Additionally, for the 5-HT$_{2B}$ subtype, EL2 was longer by three residues compared to the D$_3$DR. It had been observed in crystal structures of peptide binding GPCRs that transmembrane helix 5 could possess an additional helical turn$^{123,124}$. We hypothesized that this could also be the case for the 5-HT$_{2B}$ receptor and modeled it accordingly. As the EL2 hydrophobic anchor for this subtype was a leucine that was larger than the valine in D$_3$DR, an ensemble of models was also generated with the β$_1$ADR as EL2 template, but with the additional turn of transmembrane helix 5 intact.

Throughout the modeling iterations side chain restraints were applied based on existing structural and mutagenesis data, e.g. on the family conserved Asp$^{332}$. The restraints that improved or maintained ligand enrichment were retained. The primary template, β$_1$ADR, also had multiple structures in the PDB with slightly varying conformations in complex with different ligands$^{125-127}$. These were also evaluated in our iterative homology modeling pipeline. Enrichment of serotonin receptor ligands from the CHEMBL database$^{128}$ over decoys was steadily improved and ergotamine was docked to models emerging from the final rounds of modeling. Ergotamine is a large ligand with many degrees of freedom. For ergotamine, it was hypothesized that the small molecule crystal structure$^{129}$, which showed the presence of an intramolecular hydrogen bond, could be representative of the conformation in the binding site. Docking predictions for the two subtypes in complex with ergotamine were filtered for contacts with key residues, energy minimized, and visually inspected. Finally, five such complexes each for the 5-HT$_{1B}$ and 5-HT$_{2B}$ receptors were submitted to the assessment.

Upon release of the crystallized coordinates for the 5-HT$_{1B}$ and 5-HT$_{2B}$ receptors we were able to retrospectively assess the effects of different steps in our modeling and docking pipeline on the accuracy of the predicted complexes. From our analysis, it was clear that the judicious use of existing experimental data in the homology modeling protocol yielded excellent results. The higher enrichment of known ligands, the main criteria for validating each modeling iteration, was correlated with a lowering of ergotamine RMSD for both subtypes. Of course, the ligand-steered modeling protocol followed here is dependent on the availability of a reasonable number of verified ligands, but this is becoming rapidly available for many receptors.
from databases such as CHEMBL\textsuperscript{128}. The crystal structures showed that ergotamine formed a hydrogen bond with Thr\textsuperscript{3.37} in both subtypes, an interaction not captured by the highest ranked models. From available mutagenesis and crystallographic data for the serotonin and other aminergic receptors we had instead predicted and filtered based on such an interaction with transmembrane helix 5, which demonstrated that mutagenesis data could provide clues regarding interactions, but needed to be used carefully. However, the use of existing crystallographic data for predicting EL2 conformations was critical to the achieved modeling accuracy. The use of chimeric templates in this region for both subtypes provided remarkably accurate predictions of this portion of the binding site, which had been highlighted as one of the major challenges in previous GPCR Dock assessments. Furthermore, the use of small molecule crystal data to predict the conformation of ergotamine in the binding site also proved to be correct. The intramolecular hydrogen bond predicted on this basis was observed in crystal structures of the ligand in complex with both subtypes. Overall, a modeling protocol encompassing a large amount of existing data provided models with ligand RMSDs < 2 Å for both subtypes, which is considered to be a cutoff for successful docking predictions\textsuperscript{136}. The results achieved in this assessment were promising for the extension of structure-based drug discovery to the many pharmaceutically important GPCR targets lacking crystal structures.

41
4. Conclusions and future perspectives

The last decade has witnessed a revolution in GPCR structural biology that has resulted in a sharp increase in atomic-level information for this important superfamily of receptors. During the same period, FBLD has increasingly become popular as a method for identifying novel pharmaceuticals against a wide range of targets. A lot of work in thesis has focused on studying structure-guided FBLD against GPCR targets. In paper I, we explored fragment screening using molecular docking against a recent crystal structure of the A2A AR, a Parkinson’s disease target. The primary goal of this work was to study the complementarity between two orthogonal approaches for fragment ligand discovery, an in silico method such as molecular docking, and empirical NMR-based screens (TINS). The remarkably high hit-rates achieved demonstrated the suitability of structure-based fragment screening for this receptor. The allosteric modulators identified by TINS were beyond the reach of docking screens in this work, which further demonstrated the complementary benefits of using orthogonal approaches for fragment screening.

The lack of specificity of the discovered A2A AR fragment ligands in paper I over the closely related A1AR served as a starting point for paper II. The achievement of subtype selectivity is often a challenging problem in GPCR drug discovery, largely stemming from the evolution of highly similar binding sites suited to recognize the same endogenous compound. Frequently, an experimental structure exists for only one member within a family, and requires the use of homology models to gain an understanding of subtype selectivity from a structural perspective. The adenosine receptor family is a prototypical example of this phenomenon, where crystal structures of only the A2A AR exist, while none were available for the A1-, A2- or A2B AR. In paper II, we selected the A3AR as our target, and wished to identify fragment ligands to this receptor that were also selective over the A1AR, the antitarget. The modeling and screening protocol used in this study was successful in identifying A3AR fragment ligands with a high hit-rate, good binding affinities and discovered compounds that displayed the desired selectivity. The structural information from the homology models was also used to progress two of the initial hits into leads, culminating in the identification of a high affinity A3-selective antagonist. Overall, papers I and II have highlighted the promise of structure-based approaches for fragment screening against GPCR targets. In future, the developed screening protocols could be extended to new targets within different branches of the GPCRome that are being opened up due to the determination of crystal structures.
Peptide recognizing GPCR targets offer exciting prospects as drug targets, but identification of small molecule ligands of these is often a major challenge\cite{103,104}. In paper III, based on previous successes in structure-based fragment screens, we evaluated this approach against the crystal structure of the NTSR1\cite{107}, a peptide-binding GPCR that is a Parkinson’s disease target\cite{131}. In parallel to the fragment screen, docking evaluations of lead-like chemical space was also performed allowing us to directly compare the advantages and drawbacks of these popular approaches for ligand discovery. Whereas the fragment screens delivered an approximately two-fold higher hit-rate, the fewer number of lead-sized ligands had better affinities, which could make them preferred starting points. This was further highlighted by the fact that the lead scaffold selected for further evaluation displayed agonism in functional assays, whereas the chosen fragment hit had to be optimized for affinity first before showing the same effect at the NTSR1. This study resulted in the remarkable discovery of potent small-molecule agonist leads for this target. Paper III clearly highlighted the advantages of structure-based screening against less druggable targets. New crystal structures of such receptors would hopefully offer opportunities to discover ligands to pharmaceutically important targets, which could aid in opening up new avenues for drug development.

The complex process of the GPCR activation is yet to be fully understood. Crystal structures at different stages of activation have provided snapshots of the conformational changes that accompany this process\cite{11,21,22}. MD simulations have been used to further study the process by which these conformational changes occur\cite{53}. These studies have suggested roles for conserved amino acid residues as microswitches that switch states as a precursor to larger conformational changes\cite{53,59}. In paper IV, we wished to study the role of highly conserved ionizable residues in the β2ADR, and the roles they played in conferring a pH-dependence for the activation of this receptor. Experiments that included mutagenesis studies had revealed that the basal activity of the β2ADR increases at lower pH, and pointed to that at least one of two aspartic acid residues had to undergo a pKₐ shift upon activation\cite{111}. One of the candidate residues was Asp₁₁₃\cite{3.32} that was present in the binding site of the β2ADR whereas the other, Asp⁷⁹\cite{2.50}, was located in a buried cavity in the middle of the transmembrane region. Access to β2ADR crystal structures in the inactive and active states allowed us to utilize FEP to calculate the pKₐ shifts for these residues upon activation of the receptor. These calculations showed that Asp⁷⁹\cite{2.50} underwent a large increase in pKₐ during the activation process in agreement with the significant changes that occurred in the allosteric cavity where this residue was located. Constant protonation state, large-scale MD simulations further showed that the ionized and neutral forms of Asp⁷⁹\cite{2.50} resulted in largely inactive- or active-like conformations of the cavity, and these results could also explain data from recent NMR experiments on this receptor. Together, the results of paper IV
pointed to the Asp79$^{2.50}$ as the residue responsible for pH dependent activation of the $\beta_2$ADR by functioning as a microswitch located in between the orthosteric and G protein coupling sites of this receptor. Its high degree of conservation (94%)$^{56}$ among class A GPCRs could also point to an important role for this residue across many receptors. We have begun to investigate the role of Asp79$^{2.50}$ protonation in other class A GPCRs such as the $\beta_1$ADR and $\alpha_2A$AR along with the effect of allosteric modulators (sodium) in studies that could shed further light on the GPCR activation process. A better comprehension of the complex process of GPCR activation could be key to understanding diseases involving these receptors and consequently be used to design more effective pharmaceuticals for these. The increase in computational power and structural information could mean that simulation data may be increasingly utilized to augment experimental data from methods such as NMR or Förster resonance energy transfer (FRET) to further understand GPCR activation.

In paper II it was shown that accurate homology models from tailored protocols could be successfully utilized in ligand discovery efforts for receptors with unknown structures. GPCR dock assessments have become important avenues for evaluating the state of receptor modeling and the development of improved protocols$^{120,121}$. Paper V provides a detailed analysis of our participation in the GPCR dock 2013 edition (paper VI) involving the prediction of structures for two serotonin receptor subtypes, 5-HT$^{1B}$ and 5-HT$^{2B}$, in complex with ergotamine. Our ligand-steered iterative homology modeling procedure with chimeric templates provided best-ranked predictions for the 5-HT$^{1B}$ subtype and among the most accurate submissions for the 5-HT$^{2B}$ receptor in the assessment. The judicious use of available experimental data, and an improved prediction for EL2 were identified as some of the key reasons behind the accuracy of our predictions. The modeling approach followed during the assessment is automatable, and there is an ongoing effort to generate accurate models of other class A GPCRs, to guide ligand discovery. A key challenge would be to modify our modeling protocols to deliver models of orphan GPCRs with sufficient accuracy to suggest mutagenesis experiments that could in turn provide valuable data to improve our understanding of these receptors.

The last two decades have provided us with a revolution in GPCR structural biology, rapid progress in fragment-based approaches, and a huge increase in computational power. The work included in this thesis has involved attempts to combine these developments to better understand GPCR function and design protocols to identify ligands for these important drug targets. As the field quickly advances, the lessons learnt in these projects would hopefully allow us to capitalize on these improvements and further progress our understanding of GPCRs.
5. Summary in Swedish

6. Acknowledgements

When some people first described the four years of a PhD as really challenging, I was sure that there was more than a hint of exaggeration at play. However, I soon realized that if one were not surrounded by wonderful people it could easily become very difficult. It is hard to express in words the gratitude I feel towards my supervisor, friends, colleagues and family for not only easing the path, but for making it one of the most enjoyable periods of my life.

First and foremost, I would like to offer thanks to my supervisor Dr. Jens Carlsson for placing his faith in me and providing me with the opportunity to work in his group. Huge thanks for being an excellent supervisor, mentor and guide throughout. So seamless was the whole process that it felt at most times like I was your tenth or twentieth PhD student and definitely not your first. However, as your first PhD student it has been wonderful to see the group continually expand with many new fantastic people and I wish this continues for the years to come. Thank you so much.

I would also like to thank my second supervisor Prof. Arne Elofsson for his valuable words of advice over the last four years. I also thank Prof. Erik Lindahl for his wonderful words of encouragement and great advice during the course of my PhD.

It is difficult to even begin thanking friends and colleagues who have all been so great at various points in time. Dr. David Rodriguez Diaz, my friend, philosopher, and guide whenever I was in need of any of these. I cannot thank you enough for answering so many of what I am sure were annoying, amateurish questions during the PhD. The long dinners and lunches with you and Tamara were so great. All the best guys and see you in India. Kostas Tsirigos, my steadfast friend and confidant through these four years, sincere thanks for all the discussions, TV shows, meals, food from home, and of course the amazing trip to Greece. I cannot wait to return the hospitality sooner rather than later on your trip. On a similar vein, Christoph Peters, thank you for all the great discussions at the coffee machine, over beers, and the many fantastic meals at your place. Again, I look forward to welcoming you, Dortje and Thade wherever I find myself. Luisa Hugerth, many thanks for all the lively discussions, and spirited debates. I will certainly miss these times. My squash guru Mirco Michel, thank you so very much for all the great early morning games. You have been a wonderful friend that I could turn to at various points during this period. Again, needless to say, you, Nina, and Lia will always be part of fond memories. Axel Rudling, the local Stockholm, thank you for educating all of us on Swedish culture. I
will miss beer club and train rides back from Uppsala very much, thanks for the many great Thursday and Friday evenings, and do travel more! Marco Salvatore the Italian masterchef, thanks a lot for many an entertaining day spent discussing everything from cuisine to champions league. I will miss you and Enrichetta very much indeed. Stephanie Heusser, thanks for being such fantastic company, answering many questions regarding procedures in the lab, and for all the cakes! Stefan Fleischmann thanks for teaching me about bourbons, and all the great movie nights. Hugo Wefer for all the great Friday after-work dinners. Pierre Matricone, thanks for being such an excellent co-worker, both in and out of office. Working with you on projects has been a real joy. Mariama Jaiteh, for being an excellent lab colleague with whom I have enjoyed many a great discussion. Barbara Männel and Ennys Gheyouche who have both been amazing people to share the office with. Dr. Alexey Zeifman, big thanks for all the fantastic inputs to many different projects. Sammy Yazdi my work neighbor for all these years, thank you very much for all the enjoyable conversation. Özge Yoluk thank you for lending a poster tube on many many occasions! Also for all the great discussions over these four years. Per Warholm, my guide for all things Swedish and non-swedish, thanks a lot for patiently answering my many questions. There are many great people I have shared the office with over the years, Marcin, Minttu, Narges, Nanjiang, Walter, Szilard, Anders, Viveca, Bjorn, Karolis, Magnus, Magnus, Harsha, Sudha, David, and many others, and I thank them all from the bottom of my heart. Please do forgive me if I have missed anyone out. Many thanks to my previous co-workers, Aitakin, Marco, Gabriele and Adolfo for great memories and excellent discussions. I would of course like to thank the many great collaborators I have had the pleasure to work with. None of these projects would have been possible without you.

Of course, I could not have done this without the love and support of my entire family. To you guys I owe an immeasurable amount of gratitude not only for my PhD, but for absolutely everything. Thanks a ton!
7. References


Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational...


