New Paradigms in GPCR Drug Discovery
Structure Prediction and Design of Ligands with Tailored Properties

MARIAMA JAITEH
G protein-coupled receptors (GPCRs) constitute a large superfamily of membrane proteins with key roles in cellular signaling. Upon activation by a ligand, GPCRs transduce signals from the extracellular to the intracellular environment. GPCRs are important drug targets and are associated with diseases such as central nervous system (CNS) disorders, cardiovascular diseases, cancer, and diabetes. Currently, 34% of FDA-approved drugs mediate their effects via modulation of GPCRs. Research during the past decades has resulted in a deeper understanding of GPCR structure and function. Moreover, recent breakthroughs in structural biology allowed the determination of several atomic resolution GPCR structures. New paradigms in GPCR pharmacology have also emerged that can lead to improved drugs. Together, these advances provide new avenues for structure-based drug discovery. The work in this thesis focused on how the large amount of structural data gathered over the last decades can be used to model GPCR targets for which no experimental structures are available, and the use of structure-based virtual screening (SBVS) campaigns to identify ligands with tailored pharmacological properties. In paper I, we investigated how template selection affects the virtual screening performance of homology models of the D2 dopamine receptor (D2R) and serotonin 5-HT2A receptor (5-HT2AR). In papers II and III, SBVS methods were used to identify dual inhibitors of the A2A adenosine receptor (A2AR) and an enzyme, which could be relevant for treatment of Parkinson’s Disease, and functionally selective D2R ligands from a focused library. Finally, we also investigated how structural information can complement computational and biophysical methods to model and characterize the A2AAR-D2R heterodimer (paper IV).

**Keywords:** G Protein-Coupled Receptor, Molecular Docking, Virtual Screening, Homology Modeling, Molecular Dynamics Simulation, Chemical Library, Functionally Selective Ligand, Polypharmacology, Dimerization

Mariama Jaiteh, Department of Cell and Molecular Biology, Computational Biology and Bioinformatics, Box 596, Uppsala University, SE-751 24 Uppsala, Sweden.

© Mariama Jaiteh 2020

ISSN 1651-6214
ISBN 978-91-513-0836-4
urn:nbn:se:uu:diva-399133 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-399133)
To Isatou
This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I  **Jaiteh, M.**, Rodríguez-Espigares, I., Selent, J., and Carlsson, J.  
Performance of Virtual Screening against GPCR Homology Models: Impact of Template Selection and Treatment of Binding Site Plasticity.  


The following papers were not included in this thesis.


Reprints were made with permission from the publishers.
Contents

List of Abbreviations ........................................................................................................ 9

1 Introduction .................................................................................................................... 11
  1.1 The receptor concept and birth of the G protein-coupled receptor superfamily .......... 11
  1.2 GPCR superfamily ................................................................................................. 11
  1.3 GPCR signaling ....................................................................................................... 11
    1.3.1 The canonical G protein pathway ...................................................................... 12
    1.3.2 Non-canonical pathways ................................................................................. 13
  1.4 GPCR pharmacology .............................................................................................. 13
    1.4.1 Orthosteric ligands .......................................................................................... 13
    1.4.2 Allosteric modulation and GPCR oligomerization ............................................ 14
  1.5 GPCR structural biology ....................................................................................... 15
  1.6 New paradigms in GPCR drug discovery ............................................................. 17

2 Computational structure prediction of GPCRs ................................................................ 18
  2.1 Ab initio .................................................................................................................. 18
  2.2 Threading ................................................................................................................ 19
  2.3 Homology modeling ............................................................................................... 19
    2.3.1 Template selection and sequence alignment ................................................. 19
    2.3.2 Build conserved and non-conserved regions ............................................... 20
    2.3.3 Model evaluation ............................................................................................. 20
    2.3.4 Homology modeling with MODELLER ....................................................... 22
  2.4 Protein-protein docking .......................................................................................... 22
    2.4.1 The sampling .................................................................................................... 22
    2.4.2 The scoring ....................................................................................................... 22
    2.4.3 Protein-protein docking with HADDOCK .................................................... 23

3 Molecular dynamics simulation ...................................................................................... 24
  3.1 The force field ........................................................................................................... 24
    3.1.1 The bonded interactions ............................................................................... 24
    3.1.2 The non-bonded interactions ......................................................................... 25
  3.2 The simulation ......................................................................................................... 26
    3.2.1 Initial parameters and settings ....................................................................... 26
    3.2.2 Equations of motion ...................................................................................... 27

4 Structure-based virtual screening .................................................................................. 28
  4.1 Molecular docking ................................................................................................... 28
    4.1.1 Receptor binding site preparation ................................................................... 28
4.1.2 Library preparation .......................................................... 28
4.1.3 The molecular docking process: sampling and scoring 29
4.1.4 Retrospective screening .................................................. 29
4.1.5 Molecular docking with DOCK ................................. 30

5 Paper I: Performance of Virtual Screening against GPCR Homology Models: Impact of Template Selection and Treatment of Binding Site Plasticity ........................................................................ 32

6 Paper II: Docking Screens for Dual Inhibitors of Disparate Drug Targets for Parkinson’s Disease ......................................................... 35

7 Paper III: Structure-Guided Screening for Functionally Selective D2 Dopamine Receptor Ligands from a Virtual Chemical Library .......... 39

8 Paper IV: Mapping the Interface of a GPCR Dimer: A Structural Model of A2A Adenosine and D2 Dopamine Receptor Heteromer ....... 43

9 Conclusions and perspectives ................................................ 46

10 Sammanfattning ................................................................. 48

11 Acknowledgments ............................................................. 49

References ............................................................................. 52
List of Abbreviations

3D            Three-dimensional
5-HT<sub>2A</sub>R  Serotonin 5-HT<sub>2A</sub> Receptor
7TM          Seven Transmembrane
A<sub>2A</sub>AR   A<sub>2A</sub> Adenosine Receptor
AUC         Area Under the Curve
BS           Binding Site
CAPRI     Critical Assessment of Predicted Interactions
CASP     Critical Assessment of Techniques for Protein Structure Prediction
CNS        Central Nervous System
cryo-EM     Cryogenic Electron Microscopy
D<sub>2</sub>R    D<sub>2</sub> Dopamine Receptor
DOPE       Discrete Optimized Protein Energy
ECL        Extracellular Loop
GPCR       G Protein-Coupled Receptor
HM          Homology Modeling
HTS        High-Throughput Screening
ICL        Intracellular Loop
MAO-B      Monoamine Oxidase B
MD          Molecular Dynamics
MSA        Multiple Sequence Alignment
MW          Molecular Weight
NMR        Nuclear Magnetic Resonance
PBC        Periodic Boundary Conditions
PD         Parkinson’s Disease
PDB        Protein Data Bank
RMSD       Root Mean Square Deviation
ROC        Receiver Operating Characteristic
SBVS       Structure-Based Virtual Screening
TM         Transmembrane
VS         Virtual Screening
1. Introduction

1.1 The receptor concept and birth of the G protein-coupled receptor superfamily

The idea of receptors was highly controversial when first introduced by J. N. Langley in 1905. Langley described the existence of \textit{"receptive substances which are acted upon by chemical bodies and in certain cases by nervous stimuli. The receptive substance affects or is capable of affecting the metabolism of the chief substance"} [1]. A few years later, Ehrlich proposed the \textit{"lock and key"} theory to describe the specific interaction between receptor and ligand [2]. These were the first assertions on the existence of a receptor that binds specifically to a ligand to trigger cellular responses. After eight decades of resistance, the pioneering work of Robert Lefkowitz's group and advances in biochemical techniques (e.g. radioligand binding and purification) led to the identification of the first G protein-coupled receptor (GPCR), the $\beta$-adrenergic receptor. Cloning of the $\beta_2$-adrenergic receptor [3] led to the discovery of a protein with seven transmembrane (7TM) spanning helices, which had been previously observed for rhodopsin and bacteriorhodopsin. Thereafter, the number of cloned receptors exhibiting the 7TM helix topology grew rapidly, giving birth to the GPCR superfamily [4].

1.2 GPCR superfamily

Comprising more than 800 members, the GPCRs constitute one of the largest superfamilies of membrane proteins in the human genome. Based on phylogenetic studies, they are classified into five main subfamilies [5]: the Rhodopsin-like (also known as class A, which accounts for \textasciitilde 87\% of human GPCRs), the Secretin (class B, 15 members), the Glutamate (class C, 15 members), the Adhesion (24 members), and the Frizzled/Taste2 (24 members) receptors. They are characterized by an N-terminal segment, seven transmembrane (TM) helices connected by three extracellular loops (ECLs) and three intracellular loops (ICLs), and a C-terminal segment. Because of their characteristic topology, the GPCRs are called the \textit{"7TM"} receptors (Fig. 1.1).

1.3 GPCR signaling

The role of GPCRs is to transduce signals from the extracellular to the intracellular environment. They are activated by chemically diverse endogenous
Figure 1.1. Schematic representation of GPCR topology. GPCRs are characterized by an N-terminal (N-term) segment, seven transmembrane spanning helices (TM1-7) connected by three extracellular loops (ECLs) and three intracellular loops (ICLs), and a C-terminal (C-term) segment. GPCRs have two main binding sites (BSs): the extracellular part of the receptor exposes a ligand binding pocket that recognizes chemical messengers and the intracellular part defines the BS for intracellular partners such as heterotrimeric G proteins and β-arrestins.

compounds, *e.g.* neurotransmitters, biogenic amines, nucleosides, lipids, hormones, peptides and proteins, ions, light, and also exogenous compounds [6]. When the endogenous agonist binds to the orthosteric (in the extracellular part) binding site (BS), the GPCR undergoes conformational changes in the intracellular domain, which result in the activation of signaling pathways.

1.3.1 The canonical G protein pathway

Studies of GPCR-mediated signaling have traditionally focused on the coupling to different heterotrimeric G proteins (Fig. 1.2). A heterotrimeric G protein is an assembly of a $G_\alpha$ subunit containing the active site, which interacts with a $G_\beta\gamma$ complex (formed by $\beta$ and $\gamma$ subunits). When the agonist activates the receptor and induces conformational changes, a GDP molecule is exchanged by GTP within the $G_\alpha$ subunit. The low affinity of the GTP-bound $G_\alpha$ subunit for the $G_\beta\gamma$ complex causes the dissociation of the heterotrimeric G protein, releasing the $G_\beta\gamma$ complex from the $G_\alpha$ subunit bound to GTP. Upon release, the $G_\beta\gamma$ complex and the GTP-bound $G_\alpha$ subunit can then stimulate their primary intracellular effectors [6]. There are various $G_\alpha$, $G_\beta$, and $G_\gamma$ subunit subtypes, and these interact with specific effectors. For instance, $G_\alpha_\delta$ and $G_\alpha_{i/o}$ stimulates and inhibits adenylate cyclase, respectively, the $G_\alpha_{q/11}$ subunits stimulate protein kinases C and phospholipases C, and the $G_\beta\gamma$ subunits stimulate membrane proteins such as ion channels [7]. These interactions result in downstream changes in second-messenger levels such as cAMP and inositol triphosphate (IP3) [8, 6].
1.3.2 Non-canonical pathways

Research during the past decades has shed more light on GPCR signaling and showed that receptors can also interact with a plethora of intracellular partners, including G protein-coupled receptor kinases (GRKs) and β-arrestins (Fig 1.2). To maintain biological homeostasis and quench the levels of second messengers, GPCRs are phosphorylated by GRKs on the ICLs, resulting in an increased affinity for β-arrestins. This non-canonical pathway leads to receptor desensitization via internalization, degradation or recycling. In addition to receptor desensitization, β-arrestins can also regulate trafficking, signaling after receptor internalization, and downregulation pathways [9, 10]. Although overlooked, GPCR signaling through β-arrestins affects a variety of key cellular functions, including chemotaxis, protein translation, gene transcription, and cellular apoptosis [11]. The discovery of non-canonical GPCR signaling pathways paved the way for the field of biased agonism.

Figure 1.2. GPCRs coupled to intracellular partners. Complexes of a G protein bound to the cannabinoid receptor 1 (CBR1, PDB code: 6N4B [12]) and β-arrestin 2 bound to neurotensin receptor 1 (NTSR1, PDB code: 6PWC [13]).

1.4 GPCR pharmacology

1.4.1 Orthosteric ligands

GPCRs possess an orthosteric BS where the endogenous agonist binds to activate the receptor and induces downstream signaling. In the absence of agonists, receptors can exhibit an agonist-independent activity (constitutive activity). GPCR activity can also be modulated by a wide range of xenobiotic compounds. The efficacy of a ligand is then measured relative to the maximal
efficacy of the endogenous agonist. Agonists are categorized as full agonists when they achieve the same efficacy as the endogenous agonist and as partial agonists if they reach lower maximal efficacy. In addition to agonists, GPCRs can be modulated by neutral antagonists, which prevent agonists from binding and therefore reduce activity to the basal level. Finally, inverse agonists promote the deactivation of the receptor by shifting the equilibrium towards inactive states [14].

The traditional classification of ligands assumes that GPCRs only exist in two different states: the active and inactive conformation. However, it has been shown that the two-state model cannot capture the complexity of GPCR signaling. Instead, GPCRs exist in a large spectrum of active and inactive conformations [15, 16]. Functionally selective ligands have the ability to selectively stabilize a subset of these conformations, leading to the coupling to specific intracellular signaling partners and pathways [17, 16]. The capacity of a ligand to act specifically on a signaling pathway has several therapeutic implications. For example, morphine, an agonist of opioid receptors, has been associated with analgesic effects, but also several side effects including addiction and respiratory stress. This can be explained by the fact that morphine is a balanced agonist. Whereas the activation of the G protein pathway is responsible for the analgesic effect, activation of the β-arrestin pathway mediates many of the adverse effects. An alternative to morphine is TRV130, a G protein biased ligand that exhibits better analgesic effects and reduced side effects [18] (Fig. 1.3).Selective modulation of functional pathways is expected to provide drugs with improved efficacy and safety profiles [19].

1.4.2 Allosteric modulation and GPCR oligomerization

GPCRs can be modulated from sites that are topographically and functionally distinct from the orthosteric BS [20]. The binding of allosteric modulators with intrinsic activity, e.g. allosteric agonists, triggers conformational changes that result in receptor activity modulation even in the absence of an orthosteric ligand. Positive and negative allosteric modulators (PAMs and NAMs) are able to affect GPCR signaling only in presence of an orthosteric agonist. These modulators alter the GPCR agonist-bound conformation, resulting in changes in the binding affinity of the agonist for the orthosteric site and/or its efficacy [20, 21, 22]. Several allosteric modulators have been characterized for GPCRs, including small molecules [20, 7] and proteins. G proteins are an example of protein allosteric modulators of ligand binding to GPCRs [21]. Moreover, a GPCR can also play the role of allosteric modulators of other GPCRs [23].

Receptor-receptor interactions were first characterized for two class C GPCRs: the γ-aminobutyric acid B receptor 1 and 2 (GABA$_B$R1 and
Figure 1.3. Schematic representation of functionally selective ligands influence GPCR signaling. Balanced agonist such as morphine (in orange) can activate both G protein and β-arrestin pathways. Biased ligands (green and blue) selectively activate pathways. GABA<sub>B</sub>R2). The constitutive heterodimerization of these two receptors was necessary for the effective signaling and trafficking to the cell surface [24, 25]. The constitutive dimerization among class C GPCRs has thereafter opened new perspectives for the study of GPCR pharmacology. Although GPCR pharmacology has generally assumed that GPCRs work as monomers, a large number of studies has shown that many receptors form functional dimers or higher-order oligomers in the membrane [26, 27, 28, 29]. Within the oligomers, the protomers can exert positive and negative allosteric modulation on the interacting partner [23]. However, understanding of GPCR quaternary structures is still limited [30]. Questions regarding the oligomeric interfaces, the influence of oligomers on signaling, and how these can be modulated by small molecules are driving current experimental and computational studies on GPCR oligomers. Understanding of the structure and function of GPCR oligomers could contribute to the design of more selective and tissue-specific drugs [23].

1.5 GPCR structural biology
Determination of atomic resolution GPCR structures has been challenging. However, the recent structural breakthroughs have led to the determination of ~350 GPCR structures, including 64 unique receptors [31, 32].
The two first GPCR structures, Rhodopsin [33] and the $\beta_2$ adrenergic receptor [34], were determined seven years apart. Difficulties encountered in crystallization were related to the fact that GPCRs are inherently flexible and can have low native expression levels. Since they are membrane proteins, their isolation from the membrane in a soluble, functional, and stable form was challenging [35].

Over the past years, major advances have been made to finally crystallize GPCRs. To increase receptor solubility and stability outside the membrane, different approaches such as the fusion to soluble proteins, nanobodies, and thermostabilizing mutations have been used. The replacement of the disordered ICL3 by fusion proteins (e.g. T4 lysozyme) or nanobodies increased the polar surface area of the receptor and reduced receptor flexibility [36, 34]. Increased receptor stability was also achieved by thermostabilization. This approach consists in the introduction of mutations in the receptor and the evaluation of the ability of mutant receptors to bind ligands at increasing temperatures. The reduced conformational flexibility of thermostable receptors in comparison to the wild-type facilitates the crystallization [37, 38]. This technology, as well as the use of selective nanobodies, has the ability to capture
receptors in a specific conformational state, e.g. active or inactive-like states [39, 40]. Moreover, usage of nanobodies allows the study of receptor function, e.g. activation and biased signaling. In addition to these tools, improvements of crystallization conditions were made [35]. The first β2 adrenergic receptor structure was crystallized using bicelles [34]. Lipidic cubic phase, where the receptor is embedded in a membrane-like environment, can also be used in GPCR crystallization [36].

In recent years, cryogenic electron microscopy (cryo-EM) has been applied to GPCRs [41]. Previously, cryo-EM was limited to large proteins and resulted in low-resolution structures. With the development of better microscopes, detectors, and image processing softwares, GPCR structures can now be determined at atomic resolution [41, 42]. Since the release of the first GPCR cryo-EM structure in 2017 [41], 26 structures from class A, B and F GPCRs have been solved.

1.6 New paradigms in GPCR drug discovery

GPCRs play important roles in human physiology. These receptors are involved in several diseases, e.g. CNS disorders, cardiovascular diseases, cancer, asthma, and diabetes [43]. Currently, 34% of FDA-approved drugs mediate their effects via the modulation of GPCR function [44]. However, traditional drug discovery is failing in delivering effective new therapeutics. Moreover, due to the complex nature of diseases, current therapeutic agents exhibit adverse effects and lack of efficacy. With the recent breakthroughs in structural biology and the deeper understanding of GPCR structure and function, new paradigms in GPCR pharmacology have emerged [45]. The wealth of GPCR structures has promoted the use of structure-based methods, e.g. structure-based virtual screening (SBVS) and molecular dynamics (MD) simulations, to study activation and identify ligands. The design of novel drugs will require an integration of the advances in structural biology and pharmacology, e.g. improved understanding of functionally selectivity, polypharmacology, subtype selectivity, allosteric modulation, and oligomerization. Compounds with tailored properties could lead to more efficacious drugs with less side effects [45, 44, 46, 47, 19, 48].

The work in this thesis will focus on three major areas:

• The impact of GPCR structural biology on structure prediction and virtual screening (Paper I).
• The application of structure-based methods to design ligands with tailored pharmacological properties (Papers II and III).
• Modeling of GPCR heterodimer interfaces (Paper IV).
2. Computational structure prediction of GPCRs

The past decades have seen an explosion of genome sequencing projects, e.g. the Human Genome Project [49, 50]. This resulted in the sequencing of the nearly complete human genome, yielding approximately 22 333 protein sequences [51]. As of August 2019, the number of human protein sequences available in the UniProtKB database [52] is around 20 350. The function of these proteins can often be determined using sequence analysis tools and homology. However, deeper understanding of their function and interactions with small molecules will require three-dimensional (3D) information. Experimental 3D structures can be obtained by X-ray crystallography, nuclear magnetic resonance (NMR) or cryo-EM, but the throughput of these techniques is low. As of October 2019, only 6752 unique experimentally determined protein structures for the human species are available in the Protein Data Bank (PDB). To close the gap between the number of available protein sequences and the number of proteins with 3D structures, computational structure prediction methods can be used. Experimental structures are available for only 8% of the GPCR family. However, structure-based drug design requires accurate models of the receptors [53]. Historically, various structure prediction approaches have been used to predict of GPCR structure, including *ab initio* methods, threading, and homology modeling (HM) [54, 55, 56].

2.1 *Ab initio*

*Ab initio* methods use protein amino acid sequence to predict protein fold [57], secondary and tertiary structures [58, 59]. The limited structural information available to predict GPCR structure by homology has led to the development of *ab initio* methods for structure prediction [60]. In these methods, the seven TMs of the receptor are first predicted (using *e.g.* hydropathicity analysis and multiple sequence alignment (MSA) profiles) and assembled into a 7TM bundle. Energy minimization steps and/or MD simulations are then carried out to fold and refine the protein structure [61, 60, 62].

With the popularity of machine learning applied to life sciences, *ab initio* and deep learning approaches have been combined to predict protein structures [63]. The AlphaFold system, implemented by the Google DeepMind team, has
had great outcomes in the 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13). The method ranked first in the free modeling round after the successful prediction of several targets. The Alphafold algorithm is based on co-evolutionary data analysis. More specifically, contact maps of co-evolving residues extracted from MSA of millions of sequences were used to fold proteins [64, 65]. Two residual convolutional neural networks were used. The first one was built and trained on PDB structures to predict interatomic $C\beta$ distances and angles. The second one was trained to score the geometry and structural accuracy. Protein family-specific potentials generated through pairwise distances were then minimized by gradient descent to predict the protein structure [66, 67].

2.2 Threading

Threading is based on the observation that the number of protein folds in nature is limited [68, 69]. Hence, the sequence in hand is likely to fold as a known structure regardless of sequence similarity. On these grounds, the goal of the threading algorithm is to perform fold recognition by querying the amino acid sequence against a library of folds (e.g. CATH or SCOP) and applying a scoring function to select potential fold candidates for the queried protein [70]. Similar to \textit{ab initio} methods, co-evolutionary data can be used to guide threading. Although the global fold might be well predicted using threading, the accuracy of the model might not be sufficient to study protein function and design ligands [53].

2.3 Homology modeling

Homology modelling (HM) is founded on evolutionary principles, which state that structure is more conserved than sequence [71]. Hence, proteins sharing similar sequences are likely to adopt a similar fold and have a similar function. Using a template structure, the structure of a target sequence can be predicted based on their pairwise sequence alignment. There are several steps in the process of protein HM: e.g. template selection, sequence alignment, generation of protein core, addition of non-conserved features, build full protein, minimization, and evaluation of the final models (Fig. 2.1) [72].

2.3.1 Template selection and sequence alignment

The similarity threshold for two proteins to share the same fold is debated. The empirical rules set the sequence identity cutoff to between 28 and 35%,
but this cutoff varies with the sequence alignment length [73]. When the sequence identity between two sequences (with at least 60 residues) fall under the 28% threshold, fold conservation is uncertain [73]. There are exceptions to these rules. In larger protein families such as pentameric ligand-gated ion channels and GPCRs, the fold in the family is conserved across the different kingdoms of life despite low sequence identity. Once the global fold of the protein is obtained by homology, the focus is shifted towards local sequence conservation. In protein families, the members share several motifs and signatures [74, 5]. Two proteins with similar functions should display a conservation of the residues with functional roles such as ligand binding, protein recognition, and catalysis. In the GPCR superfamily, MSA of the family has highlighted sequence motifs that are well conserved in the different subfamilies. For instance, class A GPCRs have well conserved residues in each TM helix which are used in different numbering schemes (Ballesteros-Weinstein [75], Generic GPCR numbering [76]) to find relative positions of amino acids across multiple sequences. In the Ballesteros-Weinstein [75] numbering scheme, the most conserved residue of each TM is assigned the number X.50, where X denotes the TM number. These conserved X.50 positions often have important functional roles. For example, the highly conserved Arg3.50 in the DRY motif is involved in the ionic lock and the shift of the conformational equilibrium towards the inactive state. The Pro7.50 in the NPxxY motif is also predicted to play important roles in receptor activation [77]. Therefore, both the local and global similarities must be considered in template selection (step 1 and 2, Fig. 2.1).

2.3.2 Build conserved and non-conserved regions

The conserved regions correspond to well aligned residues in the alignment between template and target. An initial model of the target, the structural core, is built by copying the coordinates of matched atoms or residues of the template. Non-conserved regions, e.g. insertions, are modeled as loops (step 5, Fig. 2.1). Finally, energy minimization (step 6, Fig. 2.1) is applied to correct geometry errors and refine the overall model [72].

2.3.3 Model evaluation

Generated models can be evaluated for their geometry and stereochemistry using knowledge-based scoring functions (step 7, Fig. 2.1). Programs such as PROCHECK [78] compare the geometry of models to that of high-quality experimental structures. The discrete optimized protein energy (DOPE) [79], an atomic distance-dependent statistical potential derived from 3D structures databases, can be used to assess model quality. In addition to structural accuracy, models can be evaluated for their ability to recognize known ligands.
This allows the selection of reliable BS conformations for ligand discovery purposes [80, 81].

Figure 2.1. Summary of the key steps of the HM process.
2.3.4 Homology modeling with MODELLER

In this thesis, HM of proteins was performed using MODELLER [82]. MODELLER builds homology models by satisfying spatial restraints [82]. The spatial restraints are encoded as probability density function built upon geometrical features, e.g. interatomic distances and torsions, derived from the alignment between the target and the template structure. The probability density function also includes stereochemical restraints (e.g. bond lengths, angles, torsion angle preferences, and non-bonded interaction distances) extracted from molecular mechanics force fields and a database of protein structures. Generated models are then minimized using conjugated gradient and MD simulations are used to reduce violations of spatial restraints. Models quality is finally evaluated by scoring functions, e.g. DOPE score [79].

2.4 Protein-protein docking

Protein-protein interactions play key roles in biological functions such as signaling and transport. The elucidation of receptor-receptor interfaces and structural details of the interactions could guide the design of oligomer-specific ligands [83, 84]. Oligomeric complex formed by multiple protein monomers can be predicted using protein-protein docking. The docking process consists of two steps: the sampling and the scoring.

2.4.1 The sampling

During the sampling step, possible binding modes for the two monomers are explored using rigid-body docking or conformational sampling. Prediction of the complex can be made using different algorithms: exhaustive global search, local shape feature-matching, or randomized search. In the exhaustive global search, a protein with three rotational and translational degrees of freedom is moved around a static protein to sample potential complexes. The extensive search is made possible by use of grid-based approaches such as fast Fourier transformations or accelerated search in a cartesian grid space. In the local shape feature-matching approach, local complementarity between molecular surfaces of the docking partners is sought. Finally, in the randomized search, a monomer is first randomly placed in the vicinity of a fixed monomer BS. Then, binding orientations are sampled using Monte Carlo simulations or genetic algorithms [85].

2.4.2 The scoring

Generally, the sampling and scoring steps are decoupled. A large number of complexes are first generated and then the scoring step is carried out. Docked
complexes are scored using empirical-based, knowledge-based, or force field-based scoring functions. Docking algorithms and their scoring functions are regularly evaluated for their performances on benchmark sets [85] during Critical Assessment of Predicted Interactions (CAPRI) competitions [86]. These benchmarking sets contain both unbound and bound experimental structures of protein-protein complexes in order to evaluate the ability of the docking algorithm to account for conformational changes upon binding.

2.4.3 Protein-protein docking with HADDOCK

The protein-protein docking program HADDOCK [87] has been extensively evaluated in the CAPRI competitions [88, 86, 89, 90, 91, 92]. HADDOCK uses a randomized search algorithm and requires a set of ambiguous interaction restraints (AIR) to guide the docking of the two monomers. A sequence of rigid-body minimization, binding conformation optimization with flexible atoms at the interface, and a short MD simulation is performed. Finally, clustering is used to provide the best docking candidates [87]. In this thesis, protein-protein interfaces were predicted using homology-based information, i.e. the 3D structure of homologous interfaces, and mutagenesis data. The docking calculations were performed using HADDOCK and the best docking candidates were refined using MD simulations.
3. Molecular dynamics simulation

Protein structures are dynamic. The interactions between a protein and ligand involve induced-fit mechanisms [93] and/or the selection of receptor conformations by ligands [94]. GPCRs have a plethora of active and inactive conformations which can be stabilized by functionally selective ligands [95]. Therefore, in order to understand protein function, flexibility must be considered. However, the majority of structure prediction methods described in previous chapters provide static protein structures. Experimental structure determination approaches such as NMR [96] or time-resolved crystallography [97] can be used to explore the motion of proteins. Besides, biomolecular motions can be modeled computationally using MD simulations [98, 99]. In MD simulations, interactions between particles are derived using molecular mechanics force fields and the time-dependent trajectory of particles can be obtained by numerical integration of Newton’s equations of motion.

3.1 The force field

Molecular mechanics force fields are used to define the potential energy $U$ of a system of $N$ interacting particles, e.g. protein atoms. Parametrization of force fields is generally performed by quantum mechanics (QM) calculations or fitting to experimental data. Several force fields for biomolecular simulations, which differ mainly in the parametrization scheme are available (e.g. CHARMM [100], OPLS-AA [101] or AMBER [102]). A typical force field is derived by the summation of bonded (bonds, angles, improper torsions, and dihedrals, see Eq. 3.2) and non-bonded (electrostatic and van der Waals interactions, see Eq. 3.3) contributions to the potential energy.

$$U = U_{bonded} + U_{non-bonded}$$

3.1.1 The bonded interactions

The intramolecular interactions between a set of bonded particles can be described by harmonic potentials and periodic functions. The contributions to the
potential energy from bond-stretching, angle-bending, improper torsion angle and dihedral angle deviation can, for example, be derived from Eq. 3.2:

\[
U_{bonded} = U_{bonds} + U_{angles} + U_{torsions} + U_{dihedrals} + U_{impropers} \\
= \sum_{bonds} \frac{k_{AB}}{2} (l_{AB} - l_{AB,e})^2 + \sum_{angles} \frac{k_{ABC}}{2} (\theta_{ABC} - \theta_{ABC,e})^2 \\
+ \sum_{dihedrals} \frac{k_{\phi}}{2} (1 + (\cos(n\phi - \gamma))) + \sum_{impropers} \frac{k_{\zeta}}{2} (\zeta - \zeta_e)^2
\]

(3.2)

where \(k_{AB}, k_{ABC}\) and \(k_{\zeta}\) are the force constants applied to bond stretching, angle bending and improper torsion in the harmonic potential when they deviate from their reference length \((l_{AB,e})\), angle \((\theta_{ABC,e})\), and improper torsion angle \((\zeta_e)\), respectively. For the periodic function of the dihedral term, \(k_{\phi}\) is the energy barrier, \(\gamma\) the phase shift, and \(n\) the periodicity of the dihedral angle \(\phi\).

3.1.2 The non-bonded interactions

The non-bonded terms describe the intermolecular interactions between particles of the system. The contributions to the potential energy accounting for these interactions are the electrostatic and the van der Waals terms. The Coulomb potential \((U_{elec})\) describes the electrostatic interactions between two particles \(i\) and \(j\) with charges \(q_i\) and \(q_j\), respectively, as a function of the distance \(l_{ij}\). The van der Waals term describes the interactions between non-bonded particles, which when close in space become transient dipoles (induced-dipoles). Generally, a Lennard-Jones potential \((U_{L-J})\) is used to model these interactions using an attractive component \((exp - 6\) potential\) for London dispersion-attraction forces and a repulsive component \((exp - 12\) potential\) to describe repulsive forces due to overlapping orbitals and the Pauli exclusion principle. The potential energy of the system for non-bonded interactions \(U_{non-bonded}\) can therefore be defined as:

\[
U_{non-bonded} = U_{elec} + U_{L-J} \\
= \sum_{charges} \frac{q_i q_j}{4\pi \epsilon_0 l_{ij}} + \sum_{L-J} 4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{l_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{l_{ij}} \right)^{6} \right)
\]

(3.3)

where \(q_i\) and \(q_j\) are the charges of particles \(i\) and \(j\), \(\epsilon_0\) is the vacuum or reference medium permittivity, \(\epsilon_{ij}\) the depth of the well, \(\sigma_{ij}\) the distance at which the potential is equal to 0 and \(l_{ij}\) the distance between the particles \(i\) and \(j\).
3.2 The simulation

3.2.1 Initial parameters and settings

As we will see later, MD simulations consist in integration Newton’s equations of motion at each time-step $\delta t$. To be able to capture all motions of the system, including high frequency motions such as bond vibrations, a small $\delta t$ (e.g. 1-2 fs) is set. To start the simulation, initial velocities (e.g. extracted from a Maxwell Boltzmann distribution) are assigned to all particles of the solvated system. The simulation can be performed with different boundary conditions. In this thesis, we simulated our solvated protein systems using periodic boundary conditions (PBC). To account for the fact that particles near the edges of the simulation box experience different forces from those near the center, a large number of identical copies of the simulation box is generated in all directions. This continuum allows particles at the edge of the box to interact with particles in the neighboring box. When particles exit a simulation box, an image enters from the opposite edge keeping the total number of particles constant. A few precautions are necessary when using PBC. The box must be large enough to prevent a particle from interacting with its periodic image and cutoff distances must be applied for long-range interactions. For instance, in ACEMD [103], cutoff distances complemented by a smooth switching function are applied for Lennard-Jones and short-range electrostatic interactions, whereas long-range electrostatic interactions are calculated using the particle-mesh Ewald algorithm [104]. A similar protocol can be used in GROMACS [105].
3.2.2 Equations of motion

For each time-step $\delta t$, a new configuration of the system is generated. By calculating the gradient of $U$ ($\nabla U$), the force applied on each particle $i$ can be derived using Newton’s law:

$$
\vec{F}_i = -\nabla_i U = m_i \frac{d^2 r_i}{dt^2} = m_i \frac{d\vec{v}_i}{dt} = m_i \vec{a}_i
$$

(3.4)

Once the force $\vec{F}_i$ is obtained using Eq. 3.4, Newton’s equation of motion can be solved using Taylor expansion with different integration schemes. For instance, the classical Verlet algorithm [106] uses the acceleration $a_i(t)$ at time $t$ and the positions $r_i$ at time $t$ and $t - \delta t$ to derive $r_i(t + \delta t)$ (Eq. 3.5):

$$
r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + a_i(t)\delta t^2
$$

(3.5)

Subsequently, the velocity $v_i(t)$ can be calculated with the following equation:

$$
v_i(t) = \frac{r_i(t + \delta t) - r_i(t - \delta t)}{2\delta t}
$$

(3.6)

The leap-frog algorithm uses the half-step velocity $v_i(t + \frac{1}{2}\delta t)$ obtained using the acceleration $a_i(t)$ (Eq. 3.7) and the current position $r_i(t)$ to obtain $r_i(t + \delta t)$ (Eq. 3.8):

$$
v_i(t + \frac{1}{2}\delta t) = v_i(t - \frac{1}{2}\delta t) + a_i(t)\delta t
$$

(3.7)

$$
r_i(t + \delta t) = r_i(t) + v_i(t + \frac{1}{2}\delta t)\delta t
$$

(3.8)

After each time-step $\delta t$, the force $\vec{F}_i$ is derived from the potential energy of the system (Eq. 3.4). Positions are saved in trajectory files and a new iteration of the algorithm is initiated.
4. Structure-based virtual screening

The increasing availability of protein structures and computational resources led to the rise of structure-based drug design [107, 108]. Traditionally, large libraries of candidate molecules have been screened in high-throughput screening (HTS) campaigns. However, structure-based virtual screening (SBVS) methods are now routinely used in drug discovery campaigns as the starting approach to identify lead candidates in large libraries of compounds [108]. Unlike experimental HTS, this approach is fast and cost efficient. Millions of molecules can rapidly be screened at supercomputer clusters and the hit rates are often greater than from HTS [109].

4.1 Molecular docking

The aim of molecular docking is to predict the binding mode of a ligand in a protein BS. Thus, the 3D coordinates of the protein, obtained either by computational modeling or from experimental methods, are required as well as knowledge about BS location. Docking of a ligand can be divided into two steps: posing (or sampling) and scoring.

4.1.1 Receptor binding site preparation

Prior to docking calculations, the receptor BS must be prepared. The protonation states of ionizable residues are assigned with careful attention on BS residues. Moreover, functional water molecules [110] and co-factors present in the BS can be kept in the docking calculations as in Paper II.

4.1.2 Library preparation

In prospective virtual screening (VS), the sets of compounds to be docked are often extracted from commercial chemical libraries with millions of compounds such as the ZINC database [111]. Prior to screening, libraries are filtered based on physicochemical properties, e.g. molecular weight (MW), lipophilicity (logP), and the number of hydrogen bond acceptors and donors. Moreover, known pan-assay interfering compounds (PAINS) [112] are filtered out. With the increasing need for ligands with tailored properties, focused libraries are generated by extracting only a subset of compounds from large
databases. For instance, in Paper III, we extracted a set of fragments to be attached to an amidopropyl or oxobutyl linker of the 2,3-dichlorophenyl piperazine moiety, a scaffold recognized by dopamine receptors [113]. The obtained focused library contained compounds with high probability of binding to the D₂ dopamine receptor (D₂R). Finally, library generation involves consideration of relevant protonation and tautomeric states, e.g. at physiological pH, for all compounds.

4.1.3 The molecular docking process: sampling and scoring

The sampling, or posing, step aims at generating a large number of orientations and conformations of the ligand in the BS. These conformations can be sampled using different algorithms, e.g. genetic algorithms, Monte Carlo simulations, or anchor-and-grow approaches [114]. For each sampled conformation, a scoring function is used to estimate the binding energy or the fitness of a binding mode. Empirical scoring functions assess how well binding energies replicate experimental data. The free energy is estimated by summation of the contributions from uncorrelated terms such as hydrogen bonding, hydrophobic interactions, ligand entropy, and lipophilic interactions [115]. Knowledge-based scoring functions assess how well binding modes and conformations replicate experimental complexes. With these functions, protein-ligand complexes are modeled using simplified pairwise potentials such as the potential mean force [116]. Finally, force field-based scoring functions predict the binding affinity of ligands for the receptor as an energy score. This energy is derived from the summation of molecular mechanics force fields terms such as van der Waals and electrostatic interactions (see Eq. 4.1) [117].

4.1.4 Retrospective screening

Molecular docking can be used to perform retrospective screening to evaluate the performance of the software, e.g. by redocking a co-crystallized ligand to its BS. Moreover, it can be employed to assess the accuracy of predicted structures by measuring the ability of models to recognize known ligands. To perform such retrospective docking screens, a set of known ligands with a biological activity for the target protein must be compiled. These ligands are often extracted from databases with bioactivity data such as ChEMBL [118]. In addition to active ligands, a set of non-active compounds is used as decoys. Decoys can be identified experimentally, e.g. compounds that were inactive in HTS, or generated computationally, e.g. using property-matched decoys [119]. The sets of actives and decoys are screened against the structure, and then the ranking results are assessed. Ligand enrichment describes how well the protein model identifies active compounds among decoys. In this thesis, we used the area under the ROC curve (AUC), which represents the percentage...
of found actives (y-axis) against the percentage of decoys screened (x-axis). Integration of the AUC allows to quantify ligand enrichment by the receptor structure. For a model to be used in prospective screening, early recognition of actives is advantageous. To favor early enrichment, a logarithmic transformation of the AUC [120] can be applied. In this work, we used the logarithm of the AUC to measure ligand enrichment (Fig. 4.1).

![Figure 4.1. Example of ligand enrichment curve. ROC curve for a database of ligands and property-matched decoys ranked by molecular docking. The percentage of ligands and decoys identified are shown on the y- and x-axis, respectively. The solid black line represents random enrichment of ligands.](image)

**Figure 4.1.** Example of ligand enrichment curve. ROC curve for a database of ligands and property-matched decoys ranked by molecular docking. The percentage of ligands and decoys identified are shown on the y- and x-axis, respectively. The solid black line represents random enrichment of ligands.

### 4.1.5 Molecular docking with DOCK

Molecular docking calculations in this thesis were performed with DOCK [117]. In DOCK, the protein is held rigid and the BS must be predefined, often using a co-crystallized ligand. The docking process consists of an anchor-and-grow approach to sample ligand conformations and a force field-based scoring function to evaluate binding modes. During the anchor-and-grow process, the rigid core of the ligand is docked in the BS, followed by a growing step where flexible branches of the ligand are added and sampled in the BS. Docking poses are then evaluated based on electrostatic and van der Waals interactions as well as their desolvation penalties:

$$E_{bind} = E_{vdw} + E_{elec} - E_{desolv}$$ (4.1)
where $E_{vdw}$ and $E_{elec}$ are the van der Waals and electrostatic energy contributions to the binding energy $E_{bind}$, and $E_{desolv}$ is the desolvation penalty of transferring the ligand from bulk solvent to the BS with dielectric constants $\epsilon_{solvent} = 78$ and $\epsilon_{protein} = 2$, respectively.

When the full library of conformers has been docked and minimized, these are scored and ranked. The conformation with the lowest energy is retrieved as the predicted binding mode of the ligand.
5. Paper I: Performance of Virtual Screening against GPCR Homology Models: Impact of Template Selection and Treatment of Binding Site Plasticity

Despite tremendous breakthroughs in crystallization and cryo-EM [121, 35, 32], 3D structures are still missing for more than 80% of GPCR family. This shortage of atomic resolution structures has led researchers to rely on computationally predicted models for their structure-based drug discovery campaigns [108, 80, 81, 109]. While HM was the most favored and successful structure prediction method in GPCR Dock competitions [54, 55, 56], no guidelines emerged for the use of HM for VS purposes. Various studies have attempted to benchmark homology models for their suitability in VS [122, 123, 124, 125]. However, questions such as how to select templates and how to use homology models in VS campaigns remained unanswered.

In Paper I, we investigated the impact of template selection on the VS performance of homology models of the D\textsubscript{2}R and serotonin 5-HT\textsubscript{2A} receptors (5-HT\textsubscript{2A}R). Both receptors belong to the aminergic subfamily of class A GPCRs and are involved in many CNS disorders, e.g. schizophrenia and Parkinson’s disease. Homology models of the two targets were generated based on 12 aminergic and four non-aminergic GPCR templates. Selected templates covered a large span of TM and BS sequence identity, ranging from 21 to 77% and from 6 to 94%, respectively. Based on each template, a set of 250 homology models was generated for each target, from which 50 were selected based on DOPE score [79] for analysis.

The evaluation of templates using the average root mean square deviation (RMSD) of homology models to the crystal structures [126, 127] suggested an improved structural accuracy of TM and BS regions when comparing templates with low (<30%) and high (>50%) sequence identity to the target. However, templates with sequence identity between 30-50% resulted in models with variable accuracy.

To assess template performance in VS, molecular docking screens of known ligands and generated decoys were carried out against homology models. For each template, we extracted the distribution of ligand enrichment yielded by
the set of 50 homology models. There was a large variation in ligand enrichment for models based on the same template, reflecting that different BS conformations were explored by the homology models. This led us to consider the median ligand enrichment as a better descriptor of template performance (Fig. 5.1). Moreover, we considered BS flexibility by deriving a ligand enrichment for an ensemble of homology models. The ensemble enrichments were often greater than the median enrichments, but did not outperform the best enriching models.

Our work identified several weaknesses of the use of homology models in VS. Some distant templates were able to provide models with good median enrichment. However, we observed that the binding modes of top-ranked compounds did not capture the charge-charge interaction with D$_3^{3.2}$ and extend towards TM5 with an aromatic moiety as typically seen for aminergic ligands. Moreover, evaluation of template bias highlighted that models preferably enriched ligands similar to the co-crystallized ligand of the template. The chemotype bias was reduced by considering an ensemble of structures. A stronger enrichment of all chemotypes and full set of ligands were observed for the ensemble of two models based on different templates.

The modeling of ECL2 using HM is challenging, in particular if the number of templates is small [54, 55, 56, 128]. We analyzed the impact of ECL2 on model performance in VS. The median ligand enrichment considerably improved if the loop was removed. These improvements were more pro-
nounced for D$_2$R, for which the crystal structure showed a loop conformation not present in the template set. Although decoupling the impact of the loop from the TM contributions to ligand enrichment is not trivial, these results suggested a more careful modeling of the ECL2 is needed.

The relationships between sequence identity, median ligand enrichment, and structural accuracy of the BS were then analyzed. We observed a moderate negative correlation between BS accuracy and ligand enrichment, suggesting that better BS representation leads to better VS performance. In addition, templates with low BS sequence identity yielded lower median ligand enrichments (poor to fair) than templates with higher sequence identity which generally produced good to excellent enrichments. However, templates sharing $>45\%$ sequence identity with the target tended to have similar ligand enrichment values. These results can be explained by many factors, including the fact that aminergic receptors are characterized by the conserved D$_3$-32, which recognizes compounds with a positively charged amine. Despite the differences in BS composition, this conserved residue might drive ligand binding.

Homology models of the two target receptors were finally refined by MD simulations. As for the homology models, sets of MD snapshots were evaluated for their structural accuracy and ligand enrichment. The MD-refined models did not lead to improved structural accuracy or better VS performance in comparison to the raw homology models.

Overall, the results of Paper I suggest that the best template is not necessarily the closest. Although, templates with higher sequence identity were expected to lead to better enriching models, the highest enrichment and most accurate BS were not always based on the closest homolog. Nonetheless, the moderate correlation between BS accuracy and ligand enrichment suggested that molecular docking could be used to identify good representations of the BS. While incorporation of BS flexibility using ensembles of models did not improve VS performance, it often resulted in enrichment greater than the median. In addition, using an ensemble of models based on different templates allowed to reduce template bias. Finally, the refinement of homology models by MD simulations did not improve VS performances.
Parkinson’s disease (PD) is a neurodegenerative disorder characterized by a progressive loss of dopaminergic neurons in the substantia nigra pars compacta. Patients suffering from PD display motor symptoms such as stiffness, bradkynesia and tremor, but also non-motor symptoms such as cognitive decline and depression [129]. They are traditionally treated with the dopamine precursor Levodopa. However, the loss of efficacy of Levodopa and the many side effects incurred after long term use [130] have led to a need for novel therapeutics [131]. Because PD is a multifactorial disease, a polypharmacology-based approach has been proposed for the design of new therapeutics [47, 132]. Polypharmacology has gained popularity for the treatment of CNS disorders after the discovery that Clozapine superiority in clinical treatment of schizophrenia was due to its complex pharmacological profile [46]. As a result, numerous antipsychotic drugs with such complex profiles have been approved by the FDA since the late 2000s [133].

Studies on PD have suggested various potential non-dopaminergic target candidates, including enzymes and other GPCRs [131, 134]. The $A_2A$ adenosine receptor ($A_2A$AR), a nucleoside GPCR, and the monoamine oxidase B (MAO-B), an enzyme, are two targets proposed as a potential dual-target candidates after the serendipitous discovery that caffeine derivatives (e.g. CSC), which are antagonists of $A_2A$AR, could also inhibit MAO-B [135]. A dual inhibitor of $A_2A$AR and MAO-B would combine the neuroprotective effects of $A_2A$AR antagonism [136] with a sustained dopaminergic signaling through inhibition of MAO-B [137].

In Paper II, we investigated if SBVS could be used to identify dual-target $A_2A$AR / MAO-B ligands. We first screened two chemical libraries containing 0.8 million fragment-like and 4.6 million lead-like compounds from the ZINC database [111] against both $A_2A$AR [38] and MAO-B [138] crystal structures. A set of 11 lead-like and 13 fragment-like compounds was then extracted from the top 500 compounds with the lowest consensus scores.

Interestingly, although the $A_2A$AR and MAO-B binding sites were distinct and no significant similarity was found between known $A_2A$AR and MAO-B
Figure 6.1. Summary of molecular docking results for the identification dual-target compounds. We experimentally evaluated the 24 predicted dual-target compounds at both A$_{2A}$AR and MAO-B. Six compounds displaced the radioligand (>60%) at 30 μM (hit rate of 25%) in A$_{2A}$AR binding assays (K$_i$ values ranging from 19 to 7100 nM). Twelve compounds showed >70% inhibition at 30 μM (hit rate of 50%) in MAO-B enzymatic assays (IC$_{50}$ values ranging from 61 to 8700 nM). Among the confirmed hits, four compounds elicited activity at both A$_{2A}$AR and MAO-B (hit rate of 17%). The most potent dual-target ligand had a K$_i$ of 19 nM at A$_{2A}$AR and an IC$_{50}$ of 100 nM for MAO-B.

ligands (T$_c$ < 0.3 for 99.9 % of the compounds), the predicted dual-ligands exhibited similarities in their shapes and binding modes. Compounds 1 and 3 in A$_{2A}$AR (Fig 6.2.a and c) and MAO-B (Fig 6.2.b and d) were anchored by hydrogen bonds to a central amino acid in both BSs (N253$^{6,55}$ in A$_{2A}$AR and Y326 in MAO-B). In addition to hydrogen bonding, the dual-target ligands were also well anchored in both binding pockets by clusters of hydrophobic residues. To summarize, the same ligand functional groups interact with similar protein residues in the two BS. [139]. Despite the structural dissimilarities,
the docking scoring function was able to capture similarities in the BS shape and polarity.

![Docking scoring function](image)

**Figure 6.2.** Predicted binding modes of two dual-target ligands (compounds 1 and 3) and structure activity relationships (SAR). The binding modes of compound 1 (a, b) and compound 3 (c, d) in the A2AAR (grey cartoon) and MAO-B (green cartoon) BSs. SAR analyses were performed by selecting and evaluating a set of analogs experimentally. The grey and green bars represent the fold changes in affinity and IC50.

Potential concerns in the development of multi-target ligands are the difficulties in optimization and the tendency of screens to identify promiscuous compounds [140, 133]. Analysis of PubChem bioassay records showed that compounds 1 and 3 did not display activity in the majority of HTS campaigns despite being assayed extensively [141]. Both dual inhibitors were selective for A2AAR and MAO-B over other drug targets and are thus not promiscuous ligands. To understand the SAR of compounds 1 and 3 at both A2AAR and MAO-B, we performed an analog search by catalog. Selected analogs were predicted to preserve the binding modes observed for compounds 1 and
3, which was confirmed for compound 3 by a crystal structure of an analog bound to the A₂AAR [142]. Out of nine analogs of compound 1, only compound 1a displayed an improvement at both A₂AAR (27-fold) and MAO-B (12-fold) (Fig 6.2). Similarly, only two analogs of compound 3 maintained activity at both targets. Replacement of the ester moiety by an amide group resulted in loss of activity at MAO-B (Fig 6.2). This result suggested that the ester moiety which was predicted to bind in a narrow channel connecting two subpockets of the MAO-B BS, could not be substituted by a more rigid group (e.g. amide). Furthermore, the most potent dual-target ligands, 1a and 3, were confirmed as selective A₂AAR antagonists and MAO-B inhibitors over other subtypes.

Finally, we tested the two rationally designed dual-target ligands on dopaminergic neuronal-like SH-SY5Y cells pre-exposed to the 6-OHDA neurotoxin for their cytoprotective effects. Both compounds 1a and 3 displayed neuroprotective effects and reduced cell death by ~14% which is similar to the effect of the known dual-inhibitor CSC.

While the design of multi-target compounds has often relied on known scaffolds, we carried out an unbiased VS to explore the availability of dual-activity compounds in commercial chemical space. The focus on the lead-like and fragment libraries allowed us to find compounds with properties compatible with oral availability and blood-brain barrier penetration. Moreover, numerous structures were available for both A₂AAR and MAO-B, allowing the exploration of different BS conformations for each target. The difficulties in optimization of dual-target activity suggested that high-level atomic details are required to develop lead candidates. In the case of A₂AAR and MAO-B, the same functional groups in the ligand interacted with similar protein residues in the two BSs. However, when the BS of the two targets are very dissimilar, e.g. in shape and polarity, different functional groups of the ligands will be involved [139]. For such scenarios, SBVS needs to be complemented with other approaches such as the design of focused libraries (Paper III).

The results of Paper II show that SBVS can be used for the identification and optimization of novel dual-target ligands. The design of compounds with polypharmacological profile could provide starting points for development of improved drugs against diseases with complex etiology.
7. Paper III: Structure-Guided Screening for Functionally Selective D₂ Dopamine Receptor Ligands from a Virtual Chemical Library

GPCRs exist in a plethora of active conformations which can be stabilized by different ligands. Although the molecular mechanism of biased signaling is not fully understood, functionally selective ligands (or biased ligands) can specifically modulate the canonical G protein-dependent or non-canonical signaling pathways by stabilizing distinct receptor conformations [143, 95, 144]. The use of morphine, an opioid ligand known for its analgesics effects, has been linked to several side effects including dependence and respiratory stress. This duality in morphine action is due to its functional promiscuity. Morphine is able to activate both the G protein pathway, which is responsible of the analgesic effect, and the β-arrestin pathway which mediates many of the adverse effects. The design of functionally selective ligands, e.g. TRV130 [18], is expected to provide safer and more effective drugs [19].

In Paper III, we investigated if SBVS could be used to identify functionally selective D₂R ligands from a tailored library. A D₂R homology model was selected based on its enrichment of known biased ligands. During this retrospective screening, a secondary binding pocket where selective compounds extended was identified. In this secondary site formed by TM1, TM2 and TM7, polar contacts between biased ligands and the Glu95².65 and Ser409⁷.36 were frequently observed.

Because of the molecular complexity of bitopic compounds, they are likely not well covered in commercial libraries. Therefore, their identification by HTS or SBVS is improbable. We thus opted for the design of a focused library of D₂R ligands that potentially bind to both orthosteric and secondary BSs. The focused virtual chemical library was built using a 2,3-dichlorophenyl piperazine moiety, a privileged structure recognized by dopamine receptors [113], as anchor connected to either an oxobutyl or amidopropyl linker. Fragments with different functional groups, sizes and shapes, but with carboxylate and hydroxyl termini were extracted from the ZINC database [111]. These fragments were attached to the amidopropyl (for the carboxylate group) or oxobutyl (for the hydroxyl group) linker using amide coupling and nucleophilic substitution, respectively. After reactions, two libraries which yielded a total of 12
985 unique compounds, were docked to the D₂R homology model (Fig. 7.1) and the 300 top-ranked compounds of each set were visually inspected. A set of 18 compounds with the piperazine moiety well anchored in the orthosteric site and forming salt-bridge interactions with D114 had the highest binding affinity. In addition, these compounds extended into a secondary binding pocket formed by TM1, TM2 and TM7. All selected compounds were successfully synthesized and tested for their signaling properties.

![Design of the virtual chemical library of potential D₂R biased ligands.](image)

**Figure 7.1.** Design of the virtual chemical library of potential D₂R biased ligands. The functionally selective ligands were designed using an anchoring moiety in the orthosteric site and a linker (cyan) that allowed extension of a fragment (pink) to a secondary binding pocket.

Canonical G protein activation was assessed in [³⁵S]GTPγS binding assays, whereas β-arrestin 2 recruitment was characterized using the DiscoveRx Path-Hunter assay. All compounds were first tested at a single-point concentration (10 μM) to estimate the Eₘₐₓ. The full agonist (Eₘₐₓ = 100%) Quinpirole, with a balanced profile, and the β-arrestin 2 recruitment partial agonist UNC0006 (Eₘₐₓ = 18%) were used as references. All 18 compounds acted as partial agonists of at least one signaling pathway (Fig. 7.2). In addition, single-point experiments demonstrated that a majority of the predicted biased compounds acted as balanced compounds and did not show clear preference for G protein or β-arrestin 2 signaling. For this reason, no relation between structural features of the moiety in the secondary pocket and the functional selectivity profile of the partial agonists could be derived.

A set of six compounds with good efficacy or a preference for the β-arrestin 2 recruitment assay were selected for full dose-response curve determination.
Similar to the reference β-arrestin 2 partial agonist UNC0006, compound 4 exhibited an EC<sub>50</sub> = 320 nM and E<sub>max</sub> = 16 % for β-arrestin 2 recruitment but did not promote [<sup>35</sup>S]GTPγS binding. Another five compounds were partial agonists (EC<sub>50</sub> from 30 to 600 nM) of both G protein binding (E<sub>max</sub> 4.5-24%) and β-arrestin 2 recruitment (E<sub>max</sub> 16-40%), but displayed preferences for the β-arrestin signaling pathway.

![Figure 7.2. Comparison between the E<sub>max</sub> values of the 18 compounds and UNC0006 in β-arrestin 2 recruitment and [<sup>35</sup>S]GTPγS binding assays. Compounds with oxybutylene linker and the reference compound, UNC0006, are shown with filled green circles. The amidopropyl containing compounds are represented with blue circles.](image)

In this proof-of-concept study, we were able to explore a specific region of chemical space to identify a set of synthetically tractable compounds. However, the lack of SAR still limits the understanding of functional selectivity. Despite that all compounds were selected based on the interactions observed for known biased ligands, a majority displayed a balanced signaling profile. A potential explanation is that compounds were identified using D<sub>2</sub>R structure modeled based on the inactive conformation of the D<sub>3</sub>R. The identified interactions might not be sufficient to stabilize the relevant subset of active conformations. Further structural information is needed. With the recent release of the D<sub>2</sub>R structure, MD simulations could be used to generate multiple
active state conformations to better understand functional bias at the molecular level.

To summarize, in Paper III we explored the design of functionally selective D₂R ligands using SBVS and a tailored virtual library. This work showcases the design of novel ligands by combining computational approaches (SBVS and virtual library design), chemical synthesis, and biological testing. However, rational design of biased ligands is still limited by the lack of structural information.
8. Paper IV: Mapping the Interface of a GPCR Dimer: A Structural Model of A$_{2A}$ Adenosine and D$_2$ Dopamine Receptor Heteromer

Dopaminergic neurons in the substantia nigra pars compacta project in the striatum, which also selectively express the A$_{2A}$AR [145]. The existence of A$_{2A}$AR-D$_2$R heterodimers in the striatum [146] was supported by the identification of a negative allosteric modulation of D$_2$R by A$_{2A}$AR activation [147]. The A$_{2A}$AR-D$_2$R heterodimer thus represents a potential drug target for diseases caused by D$_2$R signaling impairment such as PD. Blockade of the negative allosteric modulation exerted on D$_2$R by A$_{2A}$AR activation could be prevented using A$_{2A}$AR antagonist, e.g. caffeine, leading to neuroprotection [136] (see Paper II).

The structural basis of A$_{2A}$AR-D$_2$R heterodimerization is not well understood. In Paper IV, we attempted to characterize the dimer interface using biophysical experiments in combination with computational approaches. In a previous study [148], synthetic TM helices of the D$_2$R (TM-I$^{D2}$ to TM-VII$^{D2}$) were employed to competitively inhibit A$_{2A}$AR-D$_2$R dimerization. Whereas TM-I$^{D2}$-TM-III$^{D2}$ and TM-VII$^{D2}$ peptides did not affect heterodimerization, TM-IV$^{D2}$-TM-VI$^{D2}$ induced a concentration-dependent decrease of BRET$^1$ signal. These results suggested the involvement of TM-IV$^{D2}$ and TM-V$^{D2}$ in the A$_{2A}$AR-D$_2$R interface. Similarly, in Paper IV, we investigated the involvement of TM helices of the A$_{2A}$AR in the dimer interface. BRET$^1$ experiments were carried out using TM helices of the A$_{2A}$AR (TM-I$^{D2A}$ to TM-VII$^{A2A}$). TM-IV$^{A2A}$-TM-VI$^{A2A}$ displayed a concentration-dependent inhibition of the BRET$^1$ signal (Fig. 8.1). This result was also confirmed in PLA, where TM-IV$^{A2A}$ and TM-V$^{A2A}$ disrupted A$_{2A}$AR-D$_2$R dimerization in different cell systems.

Finally, the TM helices were assayed for their impact on the allosteric modulation in the A$_{2A}$AR-D$_2$R heterodimer. The A$_{2A}$AR agonist CGS-21680 reduced the affinity of the D$_2$R agonist Quinpirole in membrane preparations expressing A$_{2A}$AR-D$_2$R heterodimer. However, in the presence of TM-IV$^{A2A}$ and TM-V$^{A2A}$, the negative allosteric modulation of A$_{2A}$AR activation on D$_2$R was significantly reduced.
Figure 8.1. Effects of A$_2$AAR TM helices on A$_2$AAR-D$_2$R dimerization. The concentration-response effect of the individual A$_2$AAR TM helices on BRET$^1$ signals of A$_2$AAR-D$_2$R heterodimer.

These experimental results were subsequently used to guide protein-protein docking and modeling the A$_2$AAR-D$_2$R heterodimer. Various orientations of the A$_2$AAR and D$_2$R protomers were sampled and two hundred models of the A$_2$AAR-D$_2$R dimer were generated and clustered based on interface RMSD. One cluster contained dimer conformations involving extensive interactions between TM4 and TM5 of A$_2$AAR and D$_2$R. The predicted A$_2$AAR-D$_2$R interface had a total BSA of $\sim$1,200 Å. The modeled heterodimer structure was then refined using MD simulations. Analysis of the MD snapshots highlighted a few important interactions. The residue Y192$^{5.41}$ in D$_2$R participated in stacking interactions with a cluster of aromatic residues in TM5 of A$_2$AAR (Y179$^{5.40}$ and F183$^{5.44}$). In the intracellular part of the receptors, L207$^{5.56}$ in D$_2$R interacted with I127$^{4.48}$, V130$^{5.51}$, and L131$^{4.52}$ in A$_2$AAR. Finally, K211$^{5.60}$ in D$_2$R was in contact with I127$^{4.48}$ in A$_2$AAR (Fig 8.2).

To evaluate the importance of these interactions for dimerization, mutagenesis studies were performed for the D$_2$R. The single-mutant Y192$^{5.41}$A and double mutant L207$^{5.56}$A / K211$^{5.60}$A were evaluated in BRET$^2$, PLA, and binding assays. The BRET$^2$ signal of the A$_2$AAR-D$_2$R interaction was significantly reduced for both mutants. In addition, A$_2$AAR agonist CGS-21680 reduced the affinity of D$_2$R agonist Quinpirole for both the L207$^{5.56}$A / K211$^{5.60}$A D$_2$R and Y192$^{5.41}$A D$_2$R mutants, but not to the same extent as for the D$_2$R wild-type. This reduction in response to the negative allosteric modulation supports a role of these residues in the dimer interface and allosteric modulation.

Attempts to further characterize the dimer interface by studying its specificity are still ongoing. Comparisons between A$_2$AAR and A$_1$AR, which does not dimerize with D$_2$R, suggested potential mutation sites to understand how
the $A_2AAR-D_2R$ heterodimer is formed. Moreover, a heteromer comprising $A_2AAR$, $D_2R$ and mGluR5 has also been observed and we have also investigated the structure of this complex [149, 150].

In Paper IV we integrated biophysical experiments with computational approaches to characterize the $A_2AAR-D_2R$ heterodimer. BRET experiments allowed us to derive the TM helices involved in the interface. By combining MD simulations, BRET data and mutagenesis we were able to derive potential hotspots in the $A_2AAR-D_2R$ heterodimer interface.
9. Conclusions and perspectives

The past decades have seen several breakthroughs in GPCR signaling, pharmacology and structural biology. These discoveries resulted in a series of new paradigms in GPCR structure-based drug discovery [45]. The work in this thesis has focused on how structural information can be used to predict the structures of GPCRs, GPCR-drug complexes and GPCR heterodimer interfaces. We investigated the potential of SBVS methods for the rational design of a novel generation of GPCR therapeutics.

With the current wealth of atomic-level information, HM can be used to predict the majority of GPCR structures [151]. However, a few key results emerged from Paper I. A suitable description of template quality cannot be limited to a single model performance but should rather be evaluated based on median enrichment of an ensemble of models. Moreover, the use of a set of models to derive an ensemble enrichment can be a powerful approach to incorporate receptor flexibility, which is expected to promote the identification of a diverse set of chemotypes [152] and thus improve hit rates. For challenging cases such as deorphanization of receptors, the lack of close templates still hinders the identification of putative ligands. Based on the median enrichment and ligand binding modes obtained for remote templates, usage of distant templates is not recommended in GPCR modeling. However, the use of different templates with good local similarity for distinct parts of the receptor could contribute to better models. Finally, results in Paper I showed that MD simulations do not necessarily improve homology model performances in SBVS. Nonetheless, MD simulations in combination with VS can be an alternative means to incorporate flexibility in the docking calculations [152, 153].

In Papers II and III, we demonstrated that when experimental structures or validated models are available, SBVS can be successfully used to identify ligands with tailored properties, e.g. multi-target and functionally selective ligands. However, more structures are still needed. A rigorous study of functional selectivity will require experimental structures of different active states bound to various biased ligands and use of MD simulations to derive insightful SAR. Moreover, rational optimization of multi-target ligands will necessitate new strategies such as incorporation of BS flexibility or the use of tailored made libraries.
The combined use of biophysical and computational approaches in Paper IV allowed the elucidation of a GPCR heterodimer interface. The increasing number of GPCR oligomer structures thanks to advances in crystallization and cryo-EM is contributing to the structural and functional understanding of GPCR quaternary structure. Moreover, this structural information, e.g. common interactions observed in protein-protein complexes, could guide and contribute to the computational prediction of GPCR interactome.
10. Sammanfattning

11. Acknowledgments

To Isatou, my mom: I dedicate this thesis to you. I am deeply grateful for your unconditional support, trust and faith in me. Although it took some time to grasp what ”djon bè ni da kan” means to me, you were always there. You’ve taught me the importance and value of education and investment in oneself. You’ve also taught me one of the most important thing in life: how to read. You are my most beloved and respected teacher. Allahma i simaya la.

To my brothers and sisters Abdoula, Fatou, Zac, and Ami: merci à tous pour votre soutien, écoute et fraternité. Parler de tout et de rien, m’a permis de garder les pieds sur terre. Merci Abdoula, pour la couverture. Tu déchires !

Most of what I’ve learned during my Ph.D. did not come from literature but from fruitful discussions with my colleagues and peers. I would like to express my gratitude:

To Jens: thank you for taking a chance on me and giving me the opportunity to start this long dive into membrane proteins. Your constant enthusiasm, support, and patience was invaluable. Discussing with you on all the stimulating projects has taught me a lot and fostered my curiosity. I have learned what research means through the numerous meetings and discussions we had. For that, I am truly grateful.

To David van der Spoel: thank you very much for your steady support as second supervisor. Changing department is not easy and you helped us accommodate smoothly in this new environment.

The current and former members of the Carlsson Lab have made this Ph.D. one of the best experience in my life.

To Pierre: thank you for your friendship. We’ve joined the group roughly at the same time and yet we’ve apparently formed The French mafia. Thank you for all the fruitful discussions, for helping me with food, for sharing your experiences.

To Flavio: you still owe me a table tennis game. Thank you so much for all the talks about science, society, guitar, food, and Italy.
To **Duy**: thank you for taking me as a lab intern. You will certainly be a great chemistry teacher. Thank you so much for showing me how compounds are made and giving me the chance to learn a bit of synthetic chemistry. Can I say I can make compounds?

To **Nicolas**: another member of the French mafia. Thank you so much for your support, all the advice and discussions about living green.

To **Stefanie**: thank you for your warm support. I think our office has the best ambient temperature. I wish you a good luck with your Ph.D.

To **Andreas**: thanks for the enriching discussions about docking, teaching and all. Good luck with your Ph.D.

To **Jon**: thanks for all the discussions about coding, MD and other geeky stuffs.

To **Wilber and Marcus**: thank you so much for letting me play with cell cultures and gels. Honestly, I think I just like the colors in cell imaging! Although I cannot claim any expertise, I have learned valuable things about biological and biophysical experiments. Thank you so much.

To **David, Alexey, Anirudh and Axel**: thank you very much. Most of what I’ve learned come from the many discussions we had. I have learned from you what one needs to dive into the world of GPCRs. Thank you for sharing your knowledge and wisdom about research and Ph.D. studies. Thank you for your story telling and gossiping too.

To **the members of the GPCR Book Club**: thanks for the fruitful discussions and the knowledge sharing.

To **Tex**: thanks for sharing teas.

To **the ICM, teaching and Ph.D. studies admins**: thank you for your patience and continuous support.

To **the people at ScilifeLab, aplha6**: thank you so much for welcoming us and introducing us to the concept of ”Fika” seen through the eyes of international students and researchers.

To **Ozge**: thank you for being my Drama mate. We definitely need to visit Korea and Japan together.
To people at Orgfarm: thank you so much for welcoming us and giving me the opportunity to do teaching. It was an invaluable experience.

Moving to a new country is often scary and require some adaptation. Life is full of important meetings and I have met extraordinary people. Truly, it is the people who make one love a place.

To Samantha and Lo: thank you for being my adoptive family in Stockholm.

To Najla and Aminata: merci infiniment pour tous les délires, les voyages, votre soutien et vos encouragements.

To Ece: thank you for being my big sister. Living with you has given me a sense of peace and belonging. Thank you for all the support, the talks and everything.

To Sammy: you are the first friend I made in Sweden. I have discovered the beauty of Stockholm with you. Thank you for all the talks and experiences we shared.

To Sonia, Emmeline, Ali, and Riad, my dearest friends: thank you for your unwavering support and friendship throughout all these years.
References


[81] Katritch, V., Kufareva, I. & Abagyan, R. Structure based prediction of


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1889

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-399133

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2020