Theoretical Studies of G-Protein-Coupled Receptors

Xianqiang Sun
Abstract

The family of G-protein-coupled receptors (GPCRs) contains the largest number of drug targets in the human body, with more than a quarter of the clinically used drugs targeting them. Because of the important roles GPCRs play in the human body, the mechanisms of activation of GPCRs or ligands binding to GPCRs have captivated much research interest since the discovery of GPCRs. A number of GPCR crystal structures determined in recent years have provided us with unprecedented opportunities in investigating how GPCRs function through the conformational changes regulated by their ligands. This has motivated me to perform molecular dynamics (MD) simulations in combination with a variety of other modeling methods to study the activation of some GPCRs and their ligand selectivity.

This thesis consists of six chapters. In the first chapter, a brief introduction of GPCRs and MD simulation techniques is given. Detailed MD simulation techniques, including pressure controlling methods and temperature coupling approaches, are described in chapter 2. The metadynamics simulation technique, used to enhance conformational sampling, is described in chapter 3. In chapter 4, I outline the inhomogeneous fluid theory used to calculate the thermodynamics properties of interfacial water molecules. Using the methods described in chapters 2-4, I carried out theoretical investigations on some GPCRs with the results summarized in chapter 5. In chapter 6, I provide a summary of the thesis with future work outlined in an outlook.
The studies presented in this thesis were carried out at the Division of Theoretical Chemistry and Biology, School of Biotechnology, KTH-Royal Institute of Technology, Sweden.

**List of papers included in this thesis**

**Paper I.** Functional Water Molecules in Rhodopsin Activation.

**Xianqiang Sun, Hans Ågren, Yaoquan Tu**

*The journal of physical chemistry B. 08/2014;118(37):10863-10873.*

**Paper II.** Function of the sodium ion in the activation of the δ-Opioid receptor

**Xianqiang Sun, Hans Ågren, Yaoquan Tu**

*In manuscript*

**Paper III.** Microsecond molecular dynamics simulations provide insight into the allosteric mechanism of the Gs protein uncoupling from the β2 adrenergic receptor.

**Xianqiang Sun, Hans Ågren, Yaoquan Tu**

*The journal of physical chemistry. B. 12/2014;118 (51):14737–14744*

**Paper IV.** Residues remote from the binding pocket control the antagonist selectivity towards the corticotropin-releasing factor receptor-1.

**Xianqiang Sun, Jianxin Cheng, Xu Wang, Hans Ågren, Yun Tang, Yaoquan Tu**

*Scientific Reports 01/2015; 5:8066*
Comments on my contributions to the papers included

As the first author, I was responsible for all the calculations and the writing of the first draft of the papers.

List of publications not included in this thesis


Xianqiang Sun, Lei Chen, Yaozong Li, Weihua Li, Guixia Liu, Yaoquan Tu, Yun Tang:


Paper II. Modification of the anticancer drug tamoxifen to avoid CYP2D6 polymorphism.

Li Gao, Xianqiang Sun, Yaoquan Tu, Hans Ågren, Leif A. Eriksson

*Canadian Journal of Chemistry* 08/2013; 91(9):916-924.
Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisor Dr. Yaoquan Tu, without whom this thesis would not have been possible. I am deeply grateful for your guidance, support, and insightful suggestions. Your advice on my research as well as on my career is invaluable.

I am also deeply grateful to my co-supervisor, Prof. Hans Ågren. You have been a tremendous mentor for me. Your support and advice contribute immensely to my career.

I wish to thank Prof. Yun Tang, at East China University of Science and technology. In his active group, I initiated my research in computational chemistry.

I am deeply grateful to Prof. Yi Luo, Prof. Faris Gel’mukhanov, Prof. Olav Vahtras, Prof. Boris Minaev, Dr. Zivinas Rinkevicius, Dr Arul Murugan, Dr. Mårten Ahquist, and Dr. Stefan Knippenberg for many helpful suggestions.

I would also like to acknowledge my colleagues Dr. Cui Li, Guanglin Kuang, Yan Wang, Dr. Xin Li, Dr. Qiong Zhang, Dr. Li Gao, Dr. Lijun Liang, Dr. Lu Sun, Xu Wang, Zhengzhong Kang, Wei Hu, Dr. Xinrui Cao, Dr. Sai Duan, Dr. Quan Miao, Dr. Guangjun Tian, Dr. Yongfei Ji, Keyan Lian, Dr. Guangping Zhang, Dr. Liqin Xue, Ignat Harczuk, Kayathri Rajarathinam, Asghar Jamshidi Zavaraki, Dr. Zhexia Zhang, Dr. Chunjie Yuan, Jing Huang Junfeng Li, Susanna Monti, Rafael Carvalho Couto, Dr. Peng Cui, Dr. Xiuneng Song, Dr. Hongbao Li, who have been a source of friendships as well as good advice and collaboration. I appreciate Yaozong Li, Junfu Bu, Dr. Hongtao Zhao, Dr. Ximing Xu, Dr. Hao Zhou, Xin Cai and Dr. Jiao Chen for many helpful discussions. I also thank Nina Bauer for providing the enjoyable environment for us.
I must also express my thanks to the China Scholarship Council for the financial support for my study in Sweden and the Swedish National Infrastructure for providing computational recourses.

Lastly, I would like to thank my family for all their love and encouragement.
CONTENTS

CHAPTER 1. INTRODUCTION ................................................................. 1

CHAPTER 2 MOLECULAR DYNAMICS SIMULATIONS .......... 9
  2.1 Classical mechanics ................................................................. 9
  2.2 Hamiltonian of a classical system ............................................. 9
  2.3 Temperature control ............................................................... 11
  2.4 Pressure control ..................................................................... 15
  2.5 Force fields ............................................................................. 18
  2.6 Periodic boundary conditions ................................................... 21
  2.7 Cut-off ..................................................................................... 22
  2.8 Ewald summation .................................................................... 24
  2.9 The Verlet algorithm: ............................................................... 27

CHAPTER 3 METADYNAMICS .......................................................... 31

CHAPTER 4 INHOMOGENEOUS FLUID THEORY .............. 35

CHAPTER 5 SUMMARY OF THE INCLUDED PAPERS......... 39
  Paper I, Functional water molecules in rhodopsin activation ................. 39
  Paper II. Function of the sodium ion in the activation of the δ-Opioid receptor.41
  Paper III. Microsecond Molecular Dynamics Simulations Provide Insight into
            the Allosteric Mechanism of the Gs Protein Uncoupling from the β2 Adrenergic
            Receptor .............................................................................. 42
  Paper IV. Residues remote from the binding pocket control the antagonist
            selectivity towards the corticotropin-releasing factor receptor-1 ............ 45
CHAPTER 6 SUMMARY AND OUTLOOK ........................................49

6.1 Summary ..................................................................................49

6.2 Outlook ....................................................................................50

REFERENCES ..................................................................................53
Chapter 1. Introduction

G-protein-coupled receptors (GPCRs) belong to one of the most important protein families in the human body.\textsuperscript{1-4} They can selectively bind to a wide variety of ligands, including light-sensitive compounds, pheromones, hormones, and neurotransmitters, and transmit the signals from the extracellular side to the intracellular side of the cell. Through regulating the coupling and decoupling of their effector proteins, the heterotrimeric G-protein or arrestin, GPCRs transmit a signal to the effector proteins. This signal is further amplified to modulate the physiological behavior of the cell. Because of the important functions of GPCRs, it is not surprising that GPCRs are correlated with many disorders in the human body and are the targets for more than a quarter of the clinically used drugs.\textsuperscript{5}

GPCRs have captivated researchers for more than 110 years.\textsuperscript{6} The first study of GPCRs can be traced back to 1904, in which year the British physiologist John Newport Langley coined the term ‘receptive substance’. He documented that the ‘chief substance’, which is the chemical that the cell secreted, should act on the ‘receptive substance’ to regulate the cell behavior, rather than directly act on the body tissues.\textsuperscript{7} The concept of the ‘receptive substance’ was further developed by his student Sir Henry Dale, who attributed the variation of the binding affinities of adrenaline to the differences of ‘adrenaline receptors’.\textsuperscript{8} The American pharmacologist Raymond Perry Ahlquist further proved the assumption addressed by Sir Henry Dale and suggested the existence of two types of the adrenaline receptor (a prototype of GPCR), the $\alpha$ type and the $\beta$ type.\textsuperscript{9}

All the pioneering studies mentioned above provided alluring perspectives to differentiate the GPCRs. Up to now, genes encoding about 838 GPCRs have been identified. These GPCRs can be classified into four classes: Class A, B, C, and F, according to the database GPCRDB.\textsuperscript{2} Class A is the largest group among these classes with 714 members, including hormone-, neurotransmitter-, and light receptors. Class B consists of 47 receptors,
including those of vasoactive intestinal peptides, secretin receptors, calcitonin, and parathyroid hormone/parathyroid hormone-related peptides. The receptors in Class C are glutamate receptors that are activated through an indirect metabotropic process. Class F includes 10 different types of Frizzled and one smoothened receptor, which play pivotal roles in the Wnt signaling pathway and in the Hedgehog signaling pathway, respectively.\(^2\)\(^10\)

Despite the existence of so many different classes of GPCRs, all of them share a common seven-transmembrane helix architecture for the transmembrane part.\(^10\) In addition, the GPCRs in the same class always contain conserved sequential motifs in the transmembrane part of the receptors. A highly conserved ‘DRY’ motif can be found in transmembrane helix 3 (TM3) for receptors in Class A. The ‘CWxP’ motif and ‘NPxxY’ motif are two other important motifs in Class A receptors. In Class B, a conserved ‘GWGxP’ motif is present in TM4.\(^11\)\(^12\) The conserved motifs ‘F/Y/HxPKxY’ on TM7 and ‘FxxCWxP’ on TM6 have been found for receptors in Class C. The receptors in Class F have the conserved ‘KTxxxW’ (‘ATxxW’ in smoothened) motif.\(^13\) These key conserved motifs undergo substantial conformational changes upon activation of the GPCRs.
Figure 1.1 Structure of Rhodopsin (PDB ID 1F88).

Structures of GPCRs with high resolution are essential for investigating the conformational changes of these motifs upon receptor activation. The crystallographic structures of GPCRs with high resolution were not available until 2000, when Palczewski and his coworkers made a breakthrough in determining the crystal structure of rhodopsin, a prototype Class A GPCR, at the resolution of 2.8 Å (Figure 1.1).14 The architecture of the seven transmembrane helices linked by three intracellular loops and three extracellular loops has been verified by this structure. This provides pivotal information in understanding the activation mechanism of rhodopsin. Using the crystallographic structure of rhodopsin as template, great success has been achieved in homology modeling studies of the other GPCRs in the same class.15, 16 Nevertheless, the technique used to crystallize rhodopsin is not successful for crystallizing the other GPCRs, because of the intrinsically unstable properties of these membrane proteins.

To improve the stability of membrane proteins, several innovative techniques have been developed and applied to the crystallization of GPCRs. One of the
most important innovations is to substitute the flexible region of the protein that hampers the crystallization with a structurally rigid domain to increase the stability of the GPCR\textsuperscript{17-19}. This leads to high success rates for crystallizing GPCRs. Another innovation is to increase the thermodynamic stability of a GPCR by selected mutations to facilitate the GPCR crystallization.\textsuperscript{20} In addition, the use of the lipid cubic phase\textsuperscript{21} and new detergents\textsuperscript{22} also plays a crucial role in obtaining the crystals.\textsuperscript{1}

Through application of the innovative techniques, Brian Kobika and his coworkers solved the dilemma in crystallizing GPCRs in 2007 and determined the crystal structure of the $\beta_2$ adrenergic receptor (\(\beta_2\)AR).\textsuperscript{17} The techniques applied to crystallize the $\beta_2$AR have also been successfully applied to the other GPCRs and have led to a flurry of GPCR structures. The $\beta_2$AR complex with its heterotrimeric Gs-protein was first crystallized in 2011 (Figure 1.2).\textsuperscript{23} The structure confirmed the outward tilting motion of TM6 upon the coupling of the Gs-protein to the $\beta_2$AR. The structure can certainly be used as a template to investigate the coupling between the other GPCRs and G-proteins.

By now a total of 119 structures for 22 GPCRs belonging to four classes have been determined.\textsuperscript{24} All these structures have greatly advanced the studies of GPCRs. However, a crystal structure is a static picture corresponding to the conformation in an energetically favorable state, whereas a GPCR is a dynamic entity that can exist in various states.\textsuperscript{25,26} Such dynamic behavior of GPCRs is not provided by the crystallographic structures. As a result, little is known about how the conformations of GPCRs can be regulated by different ligands and how these conformational changes are associated with the signaling of the GPCRs.
Molecular dynamics (MD) simulations provide us with a fundamental tool to explore different conformational states of a protein and to study the dynamic behavior of the protein. In an MD simulation, the classical Newton’s equation of motion is applied to govern the temporal evolution of a many-body system. The temporal average of the system is often assumed as the statistical ensemble average of the system for a sufficiently long simulation. Thus, the microscopic properties of the system can be calculated from a MD simulation.

The first MD simulation, which described the motion of atoms in a deterministic way, was performed on hard spheres by Alder and Wainwright in 1957, which was one year before the first x-ray crystallography structure of a protein was determined. Since then, scientists have successfully carried out MD simulations on various systems, such as those involving water molecules, solutes and peptides.
The first MD simulation of proteins was carried out by McCammon, Gelin, and Karplus in 1977. In this seminal work, the protein BPTI was simulated in vacuum with a crude interatomic potential and lasted for only 9.2 ps. The simulation provided the atomistic dynamic behavior of the protein close to its native structure. Compared to the MD simulations performed today, the simulation time was evidently much shorter. However, the result provided a novel concept that proteins are dynamically equilibrated at the energetically global minimum state, and that the X-ray crystal structure of a protein provides the averaged structure of the protein. The concept that proteins are dynamical soft molecules rather than rigid bodies formed the basis for MD simulations of proteins later on.

To obtain more reliable simulation results for proteins, more sophisticated MD simulation techniques and potentials have been developed. The two most important techniques developed for MD simulations are associated with temperature and pressure control. MD simulations are based on numerically solving the Newton’s equation of motion where the Hamiltonian of a system should be conserved. To keep a system around a desired temperature, several thermostats have been developed, such as the V-scale thermostat and the Nosé-Hoove thermostat. Moreover, most experiments are carried out under a constant pressure. To be in line with the experimental condition, barostats to keep simulations at a constant pressure, such as the Berendsen barostat and the Parrinello-Rahman barostat, have been developed. As a result, the simulations can be carried out corresponding to a desired NVT (N: constant particle number, V: constant volume, T: constant temperature) ensemble by introducing the thermostat to the system, or to a desired NPT (P: constant pressure) ensemble by introducing both a thermostat and a barostat to the system.

Besides the development of simulation techniques, more accurate force fields have also been developed to describe the interatomic interactions used in MD simulations. The most widely used force fields for bimolecular simulations are CHARMM (Chemistry at HARvard Molecular Mechanics), OPLS (Optimized Potentials for Liquid Simulation), and Amber (Assisted Model building with Energy Refinement). These force fields can be easily
embedded into the commonly used MD simulation packages to calculate the desired properties of the system.

With sophisticated MD simulation programs run on parallel supercomputers, it is of great interest to simulate protein activation processes or processes of ligand binding to proteins. However, these processes take milliseconds, the time scale of which is not reachable with the computational resources available for most researchers. To alleviate the limitation of computational resources, several advanced sampling techniques have been developed. Replica-Exchange Molecular Dynamics (REMD) is a technique developed to improve the sampling of unbiased MD simulations by simulating multiple replicas of a system at different temperatures. The energy barriers between different local minima can be overcome using this method as more conformational spaces are explored. The string method utilizing carefully selected collective variables provides us with an efficient sampling technique in determining the minimum free energy transition path. Parallel MD simulations, together with the Markov state model or accelerated MD simulations, can also be applied to improve the sampling to provide more insight into the dynamic modulation of protein functions. Metadynamics, developed by Parrinello and his coworker in 2002, is another accelerated sampling method, developed by Parrinello and his coworker in 2002. In a metadynamics simulation, biased potentials are added during the simulation to disfavor the system to revisit the historically reached conformations. This method has been widely used in the study of biomolecules, materials, and chemical reactions.

MD simulations have also been successfully carried out to study the dynamics of GPCRs. Grossfield and his coworkers performed MD simulations to elucidate the activation mechanism of rhodopsin. As presented in his work, a counter-ion switch from Glu181 to Glu113 was observed in a long-time unbiased MD simulation upon manually switching the protonation from Glu181 to Glu131. An increase of hydration was observed in the simulation. The hydration of the GPCR was further verified by the follow-up MD simulations and X-ray crystallography structures. The interaction between GPCRs and their ligands have also been studied using MD simulations and novel ligands with high potency have been discovered to modulate the
coupling of the G-protein or β-arrestin to the GPCRs to regulate the downstream proteins.\textsuperscript{54,55} Computational investigations of the interactions between a GPCR and its G-protein have been carried out based on the homology model.\textsuperscript{56} High flexibility of the helix domain of the Gα-protein was observed in the simulation of the heterotrimeric G-protein, which is in line with the explanation provided by the structure of the GPCR-Gs protein complex.\textsuperscript{57,58} Moreover, long MD simulations have been successfully applied to study the activation mechanism of a GPCR,\textsuperscript{59} a ligand binding to the GPCR,\textsuperscript{60,61} allosteric modulation of the GPCR\textsuperscript{62} and ligand dependent activation of the GPCR\textsuperscript{60} by the research group of D. E. Shaw.

Albeit the great success in the simulations of GPCRs, there are still a lot of issues that need to be resolved and addressed with MD simulations in combination with other modeling techniques. Two of the most important issues are, i) the signaling in GPCRs and, ii) the selectivity of a ligand towards different GPCRs with a high rate of sequence identity. In this thesis, MD simulations were performed to study these two issues. The work on the signaling in GPCRs is shown in Papers I, III and IV. In Paper I, I carried out MD simulations in combination with inhomogeneous fluid theory to elucidate the function of water molecules in the activation of rhodopsin.\textsuperscript{63} In Paper II, I combined MD simulations and metadynamic simulations to investigate the function of the sodium ion in the allosteric modulation of the δ-opioid receptor. I found that the sodium ion can stabilize the key residue Trp274 at different conformations in the presence or absence of the sodium ion.\textsuperscript{64} In Paper III, contributions of different conformations of the Gs domain in a GPCR-Gs protein complex were investigated to gain insight into the signaling from the GPCR to the Gs-protein. The selectivity of an antagonist towards CRF\textsubscript{1}R and CRF\textsubscript{2}R was investigated, as reported in paper IV.\textsuperscript{65} I found that one hydrogen bond significantly contributes to the high selectivity of the antagonist towards CRF\textsubscript{1}R.