Using molecular dynamics simulations to understand receptor-complex communication and signaling

Hannah Margaret Hoag

Rowan University

Follow this and additional works at: https://rdw.rowan.edu/etd

Part of the Medicinal-Pharmaceutical Chemistry Commons

Recommended Citation
Hoag, Hannah Margaret, "Using molecular dynamics simulations to understand receptor-complex communication and signaling" (2021). Theses and Dissertations. 2874.
https://rdw.rowan.edu/etd/2874

This Thesis is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact graduateresearch@rowan.edu.
USING MOLECULAR DYNAMICS SIMULATIONS TO UNDERSTAND RECEPTOR-COMPLEX COMMUNICATION AND SIGNALING

by

Hannah Margaret Hoag

A Thesis

Submitted to the
Department of Chemistry and Biochemistry
College of Science and Mathematics
In partial fulfillment of the requirement
For the degree of
Master of Science in Pharmaceutical Sciences
at
Rowan University
November 15, 2018

Thesis Chair: Dr. Chun Wu
Dedications

I would like to dedicate this thesis to my father, Rob Hoag, and to my mother, Patricia Barry. I would not be who I am today without your endless support.

I would also like to dedicate this thesis to Parker Lee and Wilbur Lee. I would not have been able to finish this without your love and support.
Acknowledgments

I would first like to acknowledge Dr. Subash Jonnalagadda. Without your mentoring and support in my undergraduate career I would not be where I am today. Your guidance has led me to this field and I am forever appreciative of you. Thank you.

I would like to acknowledge Dr. Chun Wu. Your dedication to my success has not gone unnoticed. I have developed a passion for this field and that is because of you. Thank you for showing me what true hard work looks like.

I would also like to acknowledge Richard Norton. Thank you for spending two years of mornings with me. You helped make my experience as a TA very fun and rewarding.

Lastly, I would like to acknowledge Dr. Gregory Caputo and Dr. Thomas Keck. You are both important, contributing factors to my success in my undergraduate and graduate careers. Thanks for putting up with me for way too long.
Abstract

Hannah Margaret Hoag
USING MOLECULAR DYNAMICS SIMULATIONS TO UNDERSTAND RECEPTOR-COMPLEX COMMUNICATION AND SIGNALING
2017-2018
Dr. Chun Wu
Master of Science in Pharmaceutical Sciences

The overarching purpose of this document is to use Computer-aided drug design and Molecular dynamic simulations to better understand elusive drug-receptor interactions, as well as various types of inter-receptor signaling. Chapter One introduces the theory and importance of Computer-aided drug design and the methodology used in both Chapters Two and Three.

Chapter Two uncovers the relationship between the well-studied ABCB1 transporter and a newly identified drug known as Xanthohumol (XN). XN is compared to a commonly used drug, Doxorubicin (DOX), in this chapter. If the ABCB1 transporter can be properly inhibited, cancer-fighting drugs will be able to stay within the cancer cell and will therefore be more effective. Molecular dynamic simulations are completed and analyzed for both XN and DOX as comparison. It was determined that XN competitively blocks DOX binding and may be a stronger inhibitor than DOX.

Chapter Three uses MD simulations to study GPCR signaling when bound to an agonist or antagonist and when unbound. Through MD simulation and analysis, it was determined that the alpha subunit plays an important role in GPCR-G-protein activation. Using MM-GBSA, RMSF/D, and other various analyses, various aspects of GPCR-G-protein activation were uncovered within this chapter.
## Table of Contents

Abstract ........................................................................................................................................... v

List of Figures ................................................................................................................................... viii

List of Tables .................................................................................................................................... ix

Chapter 1: Introduction to Drug Discover, CADD and Computational Methods ....... 1

1.1. Methods Used in this Thesis ........................................................................................................ 2

1.1.1. Sequence Alignment and Homology Modeling ................................................................. 2

1.1.2. Protein-Ligand Binding ....................................................................................................... 2

1.1.3. Molecular Docking ............................................................................................................... 3

1.1.4. Binding Energy Calculations .............................................................................................. 4

1.1.5. MM-GBSA and Scoring Functions ..................................................................................... 5

1.1.6. Molecular Dynamics Simulations ....................................................................................... 6

Chapter 2: Experimental and Simulation Identification of Xanthohumol as an Inhibitor and Substrate of ABCB1 ................................................................. 8

2.1. Introduction ............................................................................................................................. 8

2.2. Methods .................................................................................................................................. 10

2.2.1. Molecule Docking ............................................................................................................. 10

2.2.2. Molecular Dynamics Simulations and Analysis ............................................................... 11

2.2.3. MM-GBSA Binding Energies ........................................................................................... 13

2.3. Results .................................................................................................................................... 13

2.4. Discussion ............................................................................................................................... 19

Chapter 3: In Silico Visioning of G-Protein Communications with GPCRs and their Ligands Using Molecular Dynamics Simulations in Explicit Membrane .......... 20

3.1. Introduction ............................................................................................................................. 20
Table of Contents (Continued)

3.2. Methods .............................................................................................................. 23

3.2.1. Homology Model of Receptors ..................................................................... 23
3.2.2. Protein Structure Preparation ....................................................................... 23
3.2.3. Ligand Preparation ......................................................................................... 23
3.2.4. Ligand Docking ............................................................................................... 24
3.2.5. MD Simulation System Setup ......................................................................... 24
3.2.6. Relaxation and Production Runs ................................................................... 24
3.2.7. SID Analysis ................................................................................................... 25
3.2.8. Converge of Simulation ............................................................................... 26
3.2.9. Trajectory Clustering Analysis ...................................................................... 26
3.2.10. Binding Energy Calculations and Decomposition Methods ....................... 26

3.3. Results .................................................................................................................. 27

3.4. Discussion ........................................................................................................... 51

3.5. Conclusions ......................................................................................................... 55

References .................................................................................................................. 57

Appendix A: Supporting Material for Chapter One .................................................... 60

Appendix B: Supporting Material for Chapter Two .................................................... 69
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1. Doxorubicin</td>
<td>9</td>
</tr>
<tr>
<td>Figure 2. Xanthohumol</td>
<td>10</td>
</tr>
<tr>
<td>Figure 3. Structural representation of the predicted ABCB1-ligands interaction</td>
<td>16</td>
</tr>
<tr>
<td>Figure 4. Protein Root Mean Square Fluctuation of DOX and XN complexes</td>
<td>17</td>
</tr>
<tr>
<td>Figure 5. Ligand Root Mean Square Fluctuation of DOX and XN complexes</td>
<td>18</td>
</tr>
<tr>
<td>Figure 6. Crystal constructs</td>
<td>29</td>
</tr>
<tr>
<td>Figure 7. Root Mean Square Deviation of the ligands in the three protein-ligand complexes against simulation time</td>
<td>31</td>
</tr>
<tr>
<td>Figure 8. Root Mean Square Deviation for seven components of the GPCR-G-protein complex</td>
<td>33</td>
</tr>
<tr>
<td>Figure 9. Trajectory images of the three protein-ligand complexes in three view points</td>
<td>35</td>
</tr>
<tr>
<td>Figure 10. Full structure images of the three protein-ligand complexes</td>
<td>37</td>
</tr>
<tr>
<td>Figure 11. Receptor only images</td>
<td>38</td>
</tr>
<tr>
<td>Figure 12. Ga protein structure images</td>
<td>39</td>
</tr>
<tr>
<td>Figure 13. 2D protein-ligand interaction diagrams and histograms for agonist vs. antagonist</td>
<td>41</td>
</tr>
<tr>
<td>Figure 14. 2D protein-ligand interaction diagrams and histograms for antagonist vs. no ligand</td>
<td>43</td>
</tr>
<tr>
<td>Figure 15. Protein secondary structures and their composition percentages</td>
<td>45</td>
</tr>
<tr>
<td>Figure 16. Snapshots of α5 helix and β6-α5 loop movement for agonist and no ligand</td>
<td>46</td>
</tr>
<tr>
<td>Figure 17. Distance measurements and RMSD for α5 and β6-α5 loop</td>
<td>47</td>
</tr>
<tr>
<td>Figure 18. Protein Cα RMSF diagrams of agonist vs. antagonist</td>
<td>49</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1. MM-GBSA Binding Energies for XN and DOX complexes</td>
<td>15</td>
</tr>
<tr>
<td>Table 2. MM-GBSA Comparisons for GPCR complexes</td>
<td>27</td>
</tr>
<tr>
<td>Table 3. Ligand-Receptor Contacts</td>
<td>40</td>
</tr>
<tr>
<td>Table 4. GDP-Gα Contacts</td>
<td>42</td>
</tr>
</tbody>
</table>
Rational drug design, or simply known as drug design, is a process of finding new therapeutic agents based on previous knowledge of a known biological target. (1) Most often a drug is designed in a way that resembles an endogenous ligand for the biological receptor, or it is designed in a way that is complementary in charge and shape to the biological target in order to induce successful binding.

Drug discovery and design has succumbed to limitations throughout the last few decades. Many clinical drugs were discovered through accidents or by using naturally occurring plants or entities for screening. (2) Only one lead compound may be identified through screening 80,000 compounds causing concern for the effectiveness of these drug development methods. Following this identification, various techniques of organic synthesis of this compound will be completed to increase efficacy, increase solubility or reduce toxicity. (2) Following this process, a structure-activity relationship (SAR) may be uncovered. This SAR would be used to further synthesize other compounds. (2) Understandably so, this process becomes lengthy, costly, and time consuming for a minimal amount of success in developing a new therapeutic agent.

Computer-aided drug design (CADD) has emerged in more recent years and helps with many stages of the drug discovery process. (3,4) CADD has significantly accelerated drug discovery in multiple aspects. These methods can predict compound affinity for a receptor before the compound is synthesized in the lab. CADD further helps to predict whether or not a particular molecule will bind to a target and ultimately
how well the molecule will bind to this target. For the medicinal chemist to better understand drug-receptor interactions, computational chemistry is now commonplace. CADD can produce quantitative predictions of drug binding affinity. A scoring function is used to provide an estimate for how well a drug will bind to its target. These methods use experimental affinities, machine learning, linear regression, and other techniques to compute an estimate of binding energies between the compound and its target. (5) Overall, computational chemistry, molecular modeling, computer graphics, and other CADD tools have proven to be cost-effective all while lessening previous limitations of designing new therapeutic targets.

1.1 Methods Used in this Thesis

  1.1.1 Sequence alignment and homology modeling. Sequence alignment and homology modeling is often used when a particular crystal structure is not available for analysis. Using various tools, in this case Prime (Schrödinger), you can align two or more protein sequences. This tool predicts the degree of homology between the two sequences. This is useful when a structure is given for one particular species, but not another. This homology analysis can be used to model the desired human structure. In this particular thesis, for example, a human GPCR/G-protein complex was modeled from a bovine crystal structure. Homology modeling and sequence alignment can also help refine crystal structures that are already available. This allows for more accurate ligand docking and further analysis.

  1.1.2 Protein-ligand binding. A ligand is an entity that forms a complex with a protein to ultimately serve a biological purpose. A ligand typically produces a signal by
binding to a site on its target protein. Various conformational changes can occur and multiple signals may be produced from this complex formation. There are three known ways that ligands have been explained to bind to their target protein. These models are known as the lock-and-key model, the induced fit model and the conformational selection model.

1.1.2.1 The lock-and-key model. This analogy was first used in 1894 by Emil Fischer. In this model the “lock” is the enzyme and the “key” is the substrate that is binding. Essentially if the compound (key) fits correctly into the target (lock) then the target will be activated. However, if an incorrect substrate, meaning incorrectly sized, shaped or positioned, tries to bind to the target, then the target will not be activated.

1.1.2.2 The induced fit model. This model assumes the enzyme or target to be somewhat flexible, as opposed to the previous model. When a compound binds to its enzyme, the enzyme assumes a final shape for activation. Not all compounds will be able to induce the final shape for the enzyme. Only the correct compound will be able to induce the correct alignment for enzyme activation.

1.1.2.3 The conformational selection model. This final model is often confused for the induced fit model but they are different. The conformational selection model involves a fluctuating ligand that binds to its target therefore changing the target conformation to a more stabilized state. An example of conformational selection model is shown in Ras proteins. This model has the least amount of “rigidity” as compared to the previous two models.

1.1.3 Molecular docking. Molecular docking is a large element of CADD. Without proper molecular docking techniques, this thesis would not have been possible. Docking is used to predict the preferred orientation of a compound when bound to its target.
Molecular docking may produce many orientations of a bound compound to its target but one orientation is deemed preferred based on the binding affinity of the two entities.

Within protein-ligand docking are two mechanisms. These mechanisms are the search algorithm and the scoring function. The search algorithm stems from the search space theory. This theory contains all possible conformations and orientations of the protein with the ligand. This would be considered impossible to examine the entire search space, so the search algorithm is used. The search algorithm is comprised of various strategies which include: systematic or random torsional searches about rotatable bonds, molecular dynamics simulations and genetic algorithms to determine new low energy conformations. The second mechanism, the scoring function, takes a binding pose and produces a number indicating the probability that this particular pose would favorably bind to the target. Scoring functions and energies will be discussed in a later section.

Docking is one of the most commonly used methods within CADD as it plays a large role in understanding biochemical processes at the molecular level before synthesizing any physical compounds. Molecular docking has quite a few applications. Two of these applications consist of hit identification and lead optimization. Docking can be used to screen databases of compounds to better identify the compounds that are most likely to bind to the target. This will help to expedite the process of finding new “hits.” This method can also help to optimize these newfound leads by using the predicted orientations to better design more selective analogs.

1.1.4 Binding energy calculations. In this field of study, a ligand forms a complex with a substrate (biomolecule). This is in order to serve a type of biological purpose. In the ligand-protein binding models, a ligand typically binds to a site on a target protein in order to induce a signal. Binding of a ligand often results in a conformational change
within the target protein. This will be seen in both Chapters Two and Three of this document. Ligands and proteins are constantly associating and dissociating in an equilibrium. This is demonstrated in the following equation:

\[ P + L \leftrightarrow P*L \]

There are many ways to calculate binding energies for ligand-protein binding. In this particular thesis binding energies are calculated using thermodynamics. This first includes starting with Gibbs free energy equation:

\[ \Delta G^\circ = -RT \ln K_B \]

Using the equation above but with enthalpy and entropic terms gives the equation:

\[ \Delta G^\circ = \Delta H^\circ - T \Delta S^\circ \]

To further understand this concept, we can change this equation around again. To look at \( \Delta E \) for our ligand-protein we can see that \( H = E + PV \). Since pressure and volume minimally change in these biological systems we can use the following equation:

\[ \Delta G^\circ = \Delta E - T \Delta S^\circ \]

Using the above equation, \( \Delta E \) can be determined for the ligand-protein system. An average \( \Delta E \) can be found from the various conformations that occur in this system. This methodology is used in both Chapters 2 and 3. The system’s energy can be determined by breaking down the individual energies. This is looks like:

\[ E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{electrostatic}} + E_{\text{VDW}} \]

The above equation is the sum of \( E_{\text{Covalent}} \) and \( E_{\text{Noncovalent}} \).

1.1.5 MM-GBSA and scoring functions.

1.1.5.1 MM-GBSA. Molecular Mechanics Generalized Born Surface Area is the force-field energy calculation used in both Chapters Two and Three. For these particular experiments, the OPLS3 force field was used in the MM-GBSA calculations. The OPLS3 (6) uses parameters to stabilize structures during MD simulations. It has an
accurate prediction of solvation free energies while minimizing errors in binding energy predictions. A VSGB 2.0 solvation model (7) was also used in the calculations for the two following chapters. This is an optimized implicit solvent model with physics-based corrections for hydrogen bonding, self-contact interactions, pi-pi interactions and hydrophobic interactions. (31) The MM-GBSA procedure consists of three steps: Receptor alone (minimization), Ligand alone (minimization) and Receptor-Ligand complex (minimization). The total binding free energy equation is:

\[ \Delta G \text{ (bind)} = E_{\text{complex (minimized)}} - (E_{\text{ligand (minimized)}} + E_{\text{receptor (minimized)}}) \]

The three components that are analyzed for each are \( E_{\text{electrostatics (H\text{bond} + E_{\text{coulomb}} + E_{\text{GB solvation}})}} \), \( E_{\text{vdW (E\text{vdW} + E_{\text{pi-pi stacking}} + E_{\text{self-contact}})}} \), and \( E_{\text{lipophilic}} \).

### 1.1.5.2 XP Glide Docking and Scoring (8)

XP Glide docking and scoring were used for both Chapters Two and Three of this document. XP stands for extra precision. Glide XP uses novel terms combined with binding free energy standard scoring approaches. (32) This has greatly enhanced experimental binding affinity accuracy. Glide XP docking uses the following equation to calculate affinities:

\[ \text{Glide XP Binding Score} = E_{\text{coulomb}} + E_{\text{vdW}} + E_{\text{bind}} + E_{\text{penalty}} \]

This can be better explained by looking at the two following equations:

\[ E_{\text{bind}} = E_{\text{hyd_enclosure}} + E_{\text{hb_nn motif}} + E_{\text{hb_cc_motif}} + E_{\text{pt}} + E_{\text{hb_pair}} + E_{\text{phobic_pair}} \]

\[ E_{\text{penalty}} = E_{\text{desolv}} + E_{\text{ligand_strain}} \]

### 1.1.6 Molecular dynamics simulations.

Molecular dynamics (MD) simulations is one of the most important methods used in this particular thesis and within the field of CADD as a whole. It is a simulation used to study the movement of atoms and molecules within a fixed period of time. Using the current position and velocity of an atom, as well as the forces acting upon it by other atoms, the future position and velocity can be projected. This is done using Newton’s law of motion and solving it numerically for
every molecule and atom within the particular timeframe. The forces between these particles and the potential energies are found using a force field. A force field is essentially a set of parameters used to calculate the potential energy of a system of atoms. Some parameters include bonding forces, bending forces, rotational forces, electrostatic forces and van der Waals forces. MD simulations can still face limitations with longer simulations but as software and computers continue to develop, the accuracy of MD simulations increase significantly. MD simulations have many applications within CADD. These simulations can show conformational changes, signaling, protein folding, ion transport, and many other important biochemical processes. In this particular thesis MD simulations are used to show interactions between various parts of the GPCR-G-protein complex.
Chapter 2

Experimental and Simulation Identification of Xanthohumol as an Inhibitor and Substrate of ABCB1 (published April 2018)

2.1 Introduction

Doxorubicin (DOX) is an anti-cancer drug used to treat various types of cancer, most notably treating breast cancer. It is often given in conjunction with other chemotherapy drugs in order to avoid multidrug resistance. (9) Multidrug resistance is a large obstacle in attacking cancer cells through various chemotherapeutic drugs. Multidrug resistance has been found to be linked to ATP-binding cassette (ABC) proteins. (10) This includes p-glycoprotein, which is also known as the ATP binding cassette subfamily B member 1 (ABCB1). ABCB1 can be found in a multitude of organs throughout the body and is specifically located within the apical membrane of the epithelia. (11) ABCB1 is crucial in the transportation, and subsequently, the elimination of many drugs. If a drug is a substrate for ABCB1 it will be pumped out of the cells and back into the blood before it is able to have a promising effect on targeting the cancer cell. (11) This transportation affects the toxicity of many drug substrates. Inhibition or blocking of these transporters could allow for better drug toxicity and overall more effective cancer treatments. Using substrates to inhibit the transportation of drugs, such as DOX, is a key focus in ABCB1 research.
The ABCB1 transporter has two transmembrane domains (TMDs) and two cytoplasmic nucleotide-binding domains (NBDs). (12) There are two main conformations of ABCB1 and the transition between these two conformations requires ATP hydrolysis. The transporter can exist as inward-facing (open to the cytoplasm) and outward-facing (open to the extracellular space). ABCB1 has two pseudo-symmetric halves that each contain six transmembrane helices as well as one of the two NBDs. (13) There is a binding pocket that is enclosed by these two halves. This binding pocket is able to bind many different ligands, all of which are structurally diverse. (13)

Xanthohumol (XN) has shown to function as a cancer chemopreventive agent within more recent years. (14) It has also shown to be anti-diabetic, anti-inflammation, anti-invasive and anti-angiogenic. (14) It is a prenylated chalcone that can be found in *Humulus lupulus* L. (the hop plant). (14) XN has been shown to sensitize previous DOX-resistant breast cancer MCF-7 cells. (15) XN has also been studied to examine the effects on the expression of efflux drug transporters, including ABCB1. (16) It was determined in this study that the presence of XN in fact decreased the mRNA expression of ABCB1.
These results are promising, but little is known about the interaction between XN and ABCB1 directly. Molecular docking, molecular dynamic simulations and various analyses have yet to be completed on XN transporters.

Using two systems, we reveal the XN binding site within ABCB1, the competitive inhibition of ABCB1 function, and the impact XN has on the transportation of DOX once bound to ABCB1.

**2.2 Methods**

**2.2.1 Molecule docking.** Model generation, ligand preparation, docking and molecular dynamics were conducted using the programs of the Schrödinger Suite 2015 (Schrödinger, LLC, New York, NY, 2015). A human ABCB1 homology model was constructed based on the crystal structure of mouse ABCB1a (PDB ID: 4Q9I) (17) using the Prime program. The ligands were prepared using Maestro Elements. The 2D structures of DOX and XN were drawn using the 2D sketcher. The ionization/tautomeric states were
generated at pH=7 using Epik. The lowest ionization/tautomeric states were selected. Minimization was used to relax the ligand structure. ADP from a maltose transporter (PDB ID: 3PUV) was modified to ATP before being transferred into the 4Q9I crystal structure. The NBD in the ABCB1 model (4Q9I) was structurally assigned with the NBD in the maltose transporter (3PUV), two ATP molecules from 3PUV were incorporated into 4Q9I. Glide XP docking followed by Induced Fit Docking (IFD) were used to dock DOX and XN into the transmembrane domain. For glide XP ligand docking, the protein grid files were generated from the prepared proteins, where the centroid of the crystal ligand (QZ-Ala) was used to specify the active site. The prepared ligands (DOX and XN) were docked separately into this generated grid using Glide XP scoring function with default procedures and parameters. The obtained complexes from Glide XP docking were further subjected to IFD for getting the final complexes (Figure S1).

2.2.2 Molecular dynamics simulations and analysis.

2.2.2.1 Molecular dynamics simulation system setup. Molecular dynamics (MD) simulation systems were constructed using the prepared protein-ligand complexes from the IFD. The pre-aligned complex was placed in a membrane of POPC lipids and solvated in an orthorhombic water box with a buffer distance of 10 Å using SPC water model. System was neutralized using Na⁺ ions, and was added with a salt concentration of 0.15 M NaCl. OPLS3 force field (6) was used to represent the receptor-ligand-lipid.

2.2.2.2 Relaxation and production runs. Using the Desmond module, the system was first relaxed using the default relaxation protocol for membrane proteins. (18) After the relaxation, a 1000.0 ns production run was conducted under the NPT ensemble for
each of the two systems using the default protocol. In details, temperature 310 K was controlled using the Nosé-Hoover chain coupling scheme (19) with a coupling constant of 1.0 ps. Pressure 1 bar was controlled using the Martyna-Tuckerman-Klein chain coupling scheme (19) with a coupling constant of 2.0 ps. M-SHAKE was applied to constrain all bonds connecting hydrogen atoms, enabling a 2.0 fs time step in the simulations. The k-space Gaussian split Ewald method was used to treat long-range electrostatic interactions under periodic boundary conditions (charge grid spacing of ~1.0 Å, and direct sum tolerance of $10^{-9}$). The cutoff distance for short-range non-bonded interactions was 9 Å, with the long-range van der Waals interactions based on a uniform density approximation. To reduce the computation, non-bonded forces were calculated using an r-RESPA integrator (20) where the short-range forces were updated every step and the long-range forces were updated every three steps. The trajectories were saved at 50.0 ps intervals for analysis.

**2.2.2.3 Simulation interaction diagrams (SID).** Desmond SID tool was used to analyze the interaction between proteins and ligands in the simulations. This included Root Mean Square Deviation (RMSD/Figure S2) and Root Mean Square Fluctuation (RMSF/Figures 7 and 8) measures, secondary structure changes (Figure S3, and protein-ligand contacts (Figure S4 and S5) and ligand torsion profiles (Figure S6).

**2.2.2.4 Convergence of simulation.** To check the convergence of MD simulations, we investigated the protein Cα and ligand RMSD plots for each trajectory (Figure S2). The relatively flat plots within the last 200ns indicate that the complex systems have reached a steady state.
2.2.3 MM-GBSA binding energies. Molecular Mechanism-General Born Surface Area (MM-GBSA) binding energies were calculated on the frames in the last 200ns of each system (Table 2). OPLS3 force field, VSGB 2.0 solvation model and the default Prime procedure was used for the MM-GBSA calculation. The default procedure consists of three steps: Receptor alone (minimization), Ligand alone (minimization), Receptor-ligand complex (minimization). The total binding free energy equation is: \( \Delta G_{\text{bind}} = E_{\text{complex}}(\text{minimized}) - (E_{\text{ligand(minimized)}} + E_{\text{receptor(minimized)}}) \). There are three components analyzed: 

- Electrostatics \( E_{\text{electrostats}}(H\text{bonds} + E_{\text{coulomb}} + E_{\text{GB-solvation}}) \),
- \( E_{\text{vdW}}(E_{\text{vdW}} + E_{\sigma-\pi \text{ stacking}} + E_{\text{self-contact}}) \), and
- Lipophilic.

2.3 Results

To probe the interactions between ligands (DOX and XN) and ABCB1, we constructed two ABCB1-Ligand-ATP complexes and ran 1000.0 ns MD simulation for each system. Our RMSD analysis indicates the convergence of the MD simulation in the last 400 ns for both systems (Figure S2). The detailed results on protein secondary structure, protein-ligand contacts, and Ligand torsion profiles are included in the supporting material (Figure S3-S6). Here we highlight some key observations. First, DOX and XN bind to a similar central TMD site. Figure 3A-C shows the superimposed structures containing the ligands from the last frames of the two MD simulations. It is visible that the ligands DOX (red) and XN (blue) are overlapping in the central TMD site. Figure 3D shows the binding of DOX to nearby residues in the central TMD site. Figure 3E shows the binding of XN to nearby residues in the central TMD site. DOX was close to residues Ile 340, Phe 983, Asn 839, Gln 990, Tyr 307, Gln 725, and Phe 343 (Fig. 3D.
and Table S1). XN is close to residues Phe 336, Phe 983, Ala 980, Phe 728, Asn 842, Tyr 953, and Phe 732 (Fig. 3E and Table S1). The good overlap between the binding sites of XN and DOX suggest that XN could competitively block the DOX binding to the transporter. Second, the MM-GBSA binding energy data (Table 1) indicates that XN (-88.0±6.3 kcal/mol) binds stronger than DOX (-86.1±8.6 kcal/mol) to the transporter at the central TMD site by -1.9 kcal/mol. The energy decomposition indicates the major contribution to the energy difference (-1.9 kcal/mol) between the two ligands is from the hydrophobic interaction (ΔΔLip=-10.7 kcal/mol), which is reduced by the VDW interaction (ΔΔVDW=6.0 kcal/mol), and the electrostatic interaction (ΔΔGBELE=2.9 kcal/mol). Third, both Protein (Figure 4) and ligand RMSF analysis (Figure 5) shows that DOX-protein system had slightly larger position fluctuation than XN-protein system. The average protein RMSF for the DOX complex was calculated to be 4.51 Å while the protein RMSF for the XN complex was calculated to be lower at 2.89 Å. Similarly, the average ligand RMSF for DOX was calculated to be 2.58 Å while the average ligand RMSF for XN was calculated to be 1.85 Å. The XN complex experienced less fluctuation overall as compared to the DOX complex. Less fluctuation could indicate a more stable ligand-protein complex, furthering the evidence that XN acts as a competitive inhibitor for DOX.
**Table 1**

*MM-GBSA Binding Energies for XN and DOX Complexes*

<table>
<thead>
<tr>
<th></th>
<th>ΔVDW</th>
<th>ΔGBELE</th>
<th>ΔLIPO</th>
<th>ΔTOT</th>
<th>ΔΔTOT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XN Complex</strong></td>
<td>-51.6±3.2</td>
<td>14.3±3.2</td>
<td>-50.7±3.1</td>
<td>-88.0±6.3</td>
<td>-1.9</td>
</tr>
<tr>
<td><strong>DOX Complex</strong></td>
<td>-57.5±3.4</td>
<td>11.4±6.5</td>
<td>-39.9±3.0</td>
<td>-86.1±8.6</td>
<td>0.0</td>
</tr>
</tbody>
</table>

ΔVDW = Change of VDW energy in gas phase upon complex formation (Units: kcal/mol)

ΔGBELE = Change of GB reaction field energy + gas phase Elec. energy upon complex formation (Units: kcal/mol)

ΔLIPO = Change of lipophilic term (lipophilic energy) upon complex formation (Units: kcal/mol)

ΔTOT = ΔVDW + ΔLIPO + ΔGBELE Change of potential energy in water upon complex formation (Units: kcal/mol)
Figure 3. Structural representation of the predicted ABCB1-ligands interaction. (A) Superimposed ABCB1 models from the last snapshot of the MD simulations. Model with DOX and ATP ligands is represented using red ligands in VDW and pink-colored ribbons. Model with XN and ATP ligands is represented using blue ligands in VDW and yellow-colored ribbons. (B) Close-up view of the superimposed complexes of DOX and XN at the central TMD of ABCB1 from the top. DOX is shown in red and XN is shown in blue. (C) Close-up view of the superimposed complexes of DOX and XN at the central TMD of ABCB1 from the bottom. (D) Protein-ligand interactions between DOX and nearby residues in the central TMD site. (E) Protein-ligand interactions between XN and nearby residues in the central TMD site.
Figure 4. Protein Root Mean Square Fluctuation (RMSF) of DOX and XN complexes. Protein RMSF is used to characterize local changes along the protein chain. Peaks indicate areas of the protein that fluctuate the most during the simulation. The average RMSF for the DOX complex is 4.51 Å. The average RMSF for the XN complex is 2.89 Å.
Figure 5. Ligand Root Mean Square Fluctuation (RMSF) of DOX and XN complexes. (A) DOX atom reference. (B) XN atom reference. (C) Ligand RMSF is used to characterize changes in the ligand atom positions. The graph shows the ligand’s fluctuations broken down by atom, corresponding to the 2D structure. Ligand RMSF is measured on the ligand heavy atoms after first being aligned on the protein backbone. The average RMSF for DOX ligand is 2.58 Å. The average RMSF for XN ligand is 1.85 Å.
2.4 Discussion

The molecular modeling and simulation results showed the detailed interactions between ligands (XN and DOX) and ABCB1, and explain the molecular basis for the competitive inhibition of DOX efflux. XN bound to ABCB1 at the central TMD site, which was overlapped with the DOX binding site; and XN binding affinity is stronger than DOX leading to smaller protein and ligand position fluctuation. These results support that XN competitively blocked the DOX binding and thus efflux by the pumper. It would be interesting to experimentally confirm the predicted XN-ABCB1 interactions, such as through mutagenesis studies.
Chapter 3

*In Silico* Visioning of G-protein Communications with GPCRs and their Ligands
Using Molecular Dynamic Simulations in Explicit Membrane

3.1 Introduction

Guanine nucleotide-binding proteins, known as G proteins, are a family of proteins that act as molecular switches inside the cell. Heterotrimeric G proteins couple to a receptor on the cell surface known as a G protein-coupled receptor (GPCR). A GPCR uses extracellular signals to transmit intracellular responses. These downstream signaling cascades within the cell are responsible for many well-known molecular pathways. One of the most common being Ras-GTP signaling. The GPCR protein family is one of the largest membrane protein families and is encoded by over 800 genes in the human genome. (21) Targeting the GPCR family can produce therapeutic agents that reduce neurological disorders, asthma, COPD, cancer, and inflammatory diseases. (21) GPCR signal transduction mechanisms have been widely researched topics, including ligand-GPCR interactions and GPCR-trimeric protein interactions. Although pathways and responses have been widely researched, coordinated communication from stimulating ligands to its effector GDP still remains elusive. GPCR conformational changes and subsequent nucleotide exchange is not well understood. Understanding GPCR conformational changes and G protein stimulation can prove to be very beneficial in developing novel drugs for various diseases.

The heterotrimeric \((\alpha\beta\gamma)\) G protein, in its inactive form, is bound to guanosine diphosphate (GDP). Once activated through the GPCR, the GDP is released and
guanosine triphosphate (GTP) spontaneously binds to the vacated binding site. (22) Upon activation of the GPCR by an agonist, the G protein undergoes a significant conformational change. (22) When GDP is released and the G protein is bound to GTP, the G protein dissociates into a Gα subunit and a Gβγ complex. The Gα subunit is evolutionarily related to the Ras family of proteins. (23) The Gα subunit contains two domains, the Ras domain (GαRas) and the alpha-helical domain (GαAH). (23) The GαRas domain contains the nucleotide binding site and the GαAH domain is responsible for domain separation by moving away from GαRas. Chung et. al. used the β2 adrenergic receptor-G protein complex to examine nucleotide exchange. They speculate that GDP release involves the β6-strand-α5 helix, β1 strand, the P-loop and the α1-helix after agonist activation. They propose the β1 strand and the P-loop disrupt interactions with the GαRas and GDP, therefore promoting nucleotide exchange. (23) GPCR’s have 7 transmembrane (TM) regions, three intracellular loops and three extracellular loops. Both agonists and antagonists bind to the GPCR binding pocket causing downstream effects. Rosenbaum et. al. used molecular dynamics (MD) simulations to better understand the conformations of β2AR. They found that an agonist-bound active conformation spontaneously relaxes to an inactive-like conformation in the absence of a G protein. (24) Their results suggest that binding energy from a G protein is required to stabilize conformational changes of the GPCR in the active state. (24)

The β2 adrenergic receptor (β2AR) is often used as a model system to study the GPCR family. (25) Dror et. al at D.E. Shaw Research addressed important questions about domain separation and GDP release using β2AR. They completed atomic-level MD simulations of heterotrimeric G proteins with and without bound GPCR’s. (22)
Using the crystal structure of the β2AR- G protein complex (nucleotide-free) 66 simulations were performed using lengths of up to 50 µs each. (22) They also completed simulations of a GDP-bound G protein. In these simulations, the GaRas and GaAH domains separated from one another. (22) This open conformation of the Ga subunit resembles the nucleotide-free β2AR- G protein complex conformation. (22) Although domain separation occurred, GDP remained bound throughout the simulations. Even removal of the GaAH did not promote GDP leaving GaRas domain in simulation. (22) Separation occurs spontaneously even when GPCR is not bound to the heterotrimeric G protein. Dror et. al.’s research suggests that domain separation is necessary but not sufficient for GDP release. A weakening of nucleotide-Ras domain contacts is also necessary. (22) An activated β2AR favors GDP detachment and nucleotide exchange through conformational changes at the GaRas domain. These changes weaken GaRas interactions with GDP. (22) In this study we aim to better understand the interaction between the β2AR and a GDP-bound Gs protein using molecular dynamic simulations. We aimed to probe the communication from β2AR with an agonist, antagonist or no ligand to a Gs protein. We used the β2AR-G protein complex crystal structure bound to agonist BI-167107 (PDB ID: 3SN6) and the heterotrimeric G protein bound to GDP (PDB ID: 1GOT) to construct three homology model systems. The antagonist, alprenolol, was docked to the 3SN6 model. 1.5 or 3.0 µS MD simulations were carried out for the three systems using the OPLS3 force field (6). Our data indicates a large conformational change within the Ga subunit when bound to either the agonist or antagonist. Our simulations show GDP leaving from the agonist system only. The ligand and protein conformational changes, simulations, molecular switches, and structure activation were
analyzed for our three systems. This data provides insight on GPCR-G-protein activation and nucleotide exchange.

3.2 Methods

3.2.1 Homology model of receptors. This predicts the protein structure of human GPCR in complex with the heterotrimeric G-protein. PDB ID 3SN6 is a crystal structure of the β2-Adrenergic receptor in complex with a Gs-protein complex. This crystal structure is part bos Taurus (domestic cattle), rattus norvegicus (brown rat) and human. Homology modeling aligns two or more sequences which is useful in this case. We were able to model our desired human structure using PDB ID 3SN6. PDB ID 3SN6 contains agonist BI-167107. To confirm the GDP binding site in our homology model we superimposed the closed Gα subunit containing GDP (PDB: 1GOT) with our human Gα subunit containing the docked GDP. To confirm the antagonist (alprenolol) binding site in our homology model we superimposed our docked antagonist with another crystal structure containing alprenolol (PDB: 2RH1).

3.2.2 Protein structure preparation. Protein structures were prepared using Maestro protein preparation wizard. First, the charge state of preprocessed protein was optimized at pH=7. Second, a restrained minimization was performed to relax the protein structure using OPLS3 force field.

3.2.3 Ligand preparation. Ligand preparation was conducted using the programs of the Schrödinger Suite 2015 (Schrödinger, LLC, New York, NY, 2015). The ligands
were prepared using Maestro Elements. The 2D structures of GDP, alprenolol and BI-167107 were drawn using the 2D sketcher. The ionization/tautomeric states were generated at pH=7 using Epik. The lowest ionization/tautomeric states were selected. Minimization was used to relax the ligand structures.

3.2.4 **Ligand docking.** The prepared receptors were used to generate grid files. The centroid of the crystal ligand (agonist BI-167107) was used to specify the active site for the antagonist (alprenolol). The prepared antagonist was docked into the grid using Glide XP scoring function with default procedures and parameters. The receptor grid was generated using Van der Waals scaling factor of 1 and partial charge cutoff 0.25. The ligand docking was performed using a ligand-centered grid using OPLS3 force field (6). GDP was docked in our homology model using site detection. This was confirmed using a pre-docked GDP structure from PDB ID 1GOT.

3.2.5 **MD Simulation system setup.** All six systems were built using the prepared receptor-ligand complexes from the IFD. The complexes were placed in POPC lipids. Each system was built using the SPC solvent model with an orthorhombic solvent box included. The buffer had a 10 Å distance. A salt concentration of 0.15 M NaCl was added and the system was neutralized using Na+ ions. The force field used for each system is OPLS3 to represent the receptor-ligand complexes.

3.2.6 **Relaxation and production runs.** Using Desmond module, the system was first relaxed using the default relaxation protocol for membrane proteins. This relaxation protocol consists of eight stages: 1) Minimization with restraints on solute heavy atoms; 2) Minimization without any restraints; 3) Simulation with heating from 0 K to 300 K,
H$_2$O barrier and gradual restraining; 4). Simulation under the NPT ensemble (constant number of particles, constant pressure of 1 bar and constant temperature of 300 K) with H$_2$O barrier and with heavy atoms restrained; 5) Simulation under the NPT ensemble with equilibration of solvent and lipids; 6). Simulation under the NPT ensemble with protein heavy atoms annealing from 10.0 kcal/mol to 2.0 kcal/mol; 7). Simulation under the NPT ensemble with Ca atoms restrained at 2 kcal/mol; and 8). Simulation for 1.5 ns under the NPT ensemble with no restraints. After the relaxation, a 3000.0 ns production run was conducted under the NPT ensemble for each of the systems using the default protocol. In details, temperature was controlled using the Nosé-Hoover chain coupling scheme with a coupling constant of 1.0 ps. Pressure was controlled using the Martyna-Tuckerman-Klein chain coupling scheme with a coupling constant of 2.0 ps. M-SHAKE was applied to constrain all bonds connecting hydrogen atoms, enabling a 2.0 fs time step in the simulations. The k-space Gaussian split Ewald method was used to treat long-range electrostatic interactions under periodic boundary conditions (charge grid spacing of ~1.0 Å, and direct sum tolerance of 10$^{-9}$). The cutoff distance for short-range non-bonded interactions was 9 Å, with the long-range van der Waals interactions based on a uniform density approximation. To reduce the computation, non-bonded forces were calculated using an r-RESPA integrator where the short-range forces were updated every step and the long-range forces were updated every three steps. The trajectories were saved at 40.0 ps intervals for analysis.

3.2.7 SID analysis. The DESMOND Simulation Interactions Diagram tool was used to analyze the three systems. This tool analyzes the interactions between proteins and ligands throughout the simulation. Analysis includes Root Mean Square Fluctuation
(RMSF), Root Mean Square Deviation (RMSD), protein-ligand contacts, secondary structure changes, and ligand properties.

3.2.8 Convergence of simulation. The RMSD plots for all three trajectories can confirm the convergence of the MD simulations. The RMSD plots for each system show a relatively flat plot within the last 150 ns. This flattening indicates that the systems have reached a steady state.

3.2.9 Trajectory clustering analysis. This analysis uses frames from the Desmond trajectory output (D.E. Shaw Research). This clusters structures from Desmond trajectories based on the RMSD matrix. This matrix is a set of atoms from Schrodinger scripts. The merging distance cutoff used was 2.0 Å. The centroid structure is considered to have the largest number of neighbors in the structural family. Therefore, this structure is used to represent the family.

3.2.10 Binding energy calculations and decomposition methods. MM-GBSA binding energies. Molecular Mechanism-General Born Surface Area (MM-GBSA) binding energies were calculated on the frames in the last 200ns of each system. OPLS3 force field, VSGB 2.0 solvation model and the default Prime procedure was used for the MM-GBSA calculation. The default procedure consists of three steps: Receptor alone (minimization), Ligand alone (minimization), Receptor-ligand complex (minimization). The total binding free energy equation is: \[ \Delta G_{\text{bind}} = E_{\text{complex (minimized)}} - (E_{\text{ligand(minimized)}} + E_{\text{receptor(minimized)}}). \] There are three components analyzed: Eelectrostatics (Hbond + Ecoulomb +EGB_solvation), EvdW (EvdW+Eπ-π stacking +Eself-contact), and Elipophilic.
3.3 Results

MM-GBSA binding energies indicate that the agonist destabilizes the GPCR-G-protein complex. We performed MM-GBSA calculations, as described in the method section, at three interfaces of the GPCR-G-protein complex. These results are found in Table 3. The agonist binds more favorable to the GPCR binding pocket than the antagonist by -60 kcal/mol. However, at the GPCR-G-protein interface, the free energy binding is much more favorable for the antagonist system at -237 kcal/mol as compared to the agonist system at -163 kcal/mol. The complex with no ligand docked to the GPCR was even more favorable at -249 kcal/mol for this interface. At the G-protein-GDP interface, the agonist has the weakest binding at -22 kcal/mol. The no ligand complex and antagonist complex have free binding energies of -48 kcal/mol and -32 kcal/mol respectively.

Table 2

MM-GBSA (kcal/mol) Comparisons for GPCR Complexes

<table>
<thead>
<tr>
<th></th>
<th>No Ligand</th>
<th>Antagonist</th>
<th>Agonist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ligand - Receptor</strong></td>
<td>-125.0 ± 5.0</td>
<td>-185 ± 17</td>
<td></td>
</tr>
<tr>
<td><strong>Receptor - G-protein</strong></td>
<td>-249.0 ± 18.0</td>
<td>-237.0 ± 15.0</td>
<td>-163.0 ± 8.0</td>
</tr>
<tr>
<td>GDP - G-protein</td>
<td>-48.0 ± 4.0</td>
<td>-32.0 ± 8.0</td>
<td>-22.0 ± 8.0</td>
</tr>
</tbody>
</table>

Human GPCR-G-protein complex has same binding site for ligands and GDP. We used PDB ID 3SN6 containing agonist BI-167107 as the template for homology modeling. The human 3SN6 crystal structure is shown in Figure 6B. To confirm the
GDP binding site in our homology model we superimposed the closed Gα subunit containing GDP (PDB: 1GOT, Figure 6A) with our human Gα subunit containing the docked GDP (Figure 6D). To confirm the antagonist (alprenolol) binding site in our homology model we superimposed our docked antagonist (Figure 6C) with another crystal structure containing alprenolol (PDB: 2RH1, Figure 6E). BI-167107, alprenolol and GDP are shown in Figure 6F-H, respectively. The final agonist complex is shown in Figure 6I, the final antagonist complex is shown in Figure 6J and the final no ligand complex is shown in Figure 6K.
Figure 6. Crystal Constructs. (A-C) Three crystal complexes. (A) PDB: 1GOT. (B) PDB: 3SN6. (C) PDB: (D): 1GOT and 3SN6 (Chain A) superimposed. (E) 2RH1 and 3NYA superimposed (receptors only). (F-H) 2D interactions for three ligands used. (F) Antagonist BI-167107 from 3SN6. (G) Agonist Alpenolol from 2RH1. (H) GDP
from 1GOT. (I-K) Three built complex models. (I) Agonist complex. (J) Antagonist complex. (K) No ligand complex.

Ligand RMSD shows similar deviations between agonist and antagonist but GDP deviation within the binding pocket for agonist complex only. The root mean square deviation (RMSD) for the agonist and the antagonist is shown in Figure 7A. Although the agonist binds more favorable to the GPCR, the RMSD values are similar for both ligands, with the antagonist having more deviation throughout the simulation comparatively. The RMSD for GDP binding is shown in Figure 7B. GDP in all three complexes start at approximately the same value. GDP in the agonist complex begins to deviate more than both the no ligand and antagonist complexes at about 800 ns into the simulation time. GDP deviation drastically changes at about 1250 ns and reaches a value of 70-75 Å. The deviation stays consistent for GDP in both the no ligand and antagonist complexes.
Figure 7. Root Mean Square Deviation (RMSD) of the ligands in the three protein-ligand complexes against simulation time. (A) Antagonist vs. agonist. (B) GDP in each protein-ligand complex.

Protein RMSD indicates the Gα subunit has the largest conformational change after agonist/antagonist binding. The RMSD for seven components of the GPCR-G-
protein structure is shown in Figure 8A-C. Figure 8A shows the RMSD for the agonist complex. There is a clear separation in RMSD values between the components containing the Gα subunit and the components that do not. The full complex, the G-protein and the Gα subunit all have a RMSD value of about 5 Å or higher, while the GPCR only, Gβ subunit, Gγ subunit and the Gβ-γ complex all have an RMSD value between 2-4 Å. The same separation is seen in the antagonist complex in Figure 3B. The no ligand complex (Figure 8C) does not have a visible separation between groups, and the Gγ subunit is comparative in RMSD value to the Gα subunit.
Figure 8. Root Mean Square Deviation (RMSD) for seven components of the GPCR-G-protein complex. (A) Agonist. (B) Antagonist. (C) No Ligand.

Trajectory images show GDP movement within agonist binding pocket before detachment. MD simulation shows GDP detachment from the agonist Gα subunit only. Trajectory images (Figure 9) show 300 locations of GDP throughout the simulation for each system. There are three viewpoints for these images. GDP movement within the antagonist and no ligand Gα subunits is minimal in these images (Figure 9B and 9C).
GDP movement within the agonist G\(\alpha\) subunit is visible. GDP appears to make conformational changes before being expelled from the G\(\alpha\) subunit. This can be seen from all three viewpoints. GDP re-attachment to the G\(\beta\) subunit is inaccurate. This is due to the limitations set by the simulation box.
Figure 9. Trajectory images of the three protein-ligand complexes in three viewpoints. (A) Agonist viewpoints. (B) Antagonist viewpoints. (C) No ligand viewpoints. 3000 frames with step 10 in VMD yielded 300 images of GDP movement.
Last snap shot images indicate changes within the full structure, GPCR and alpha subunit between systems. There are visible changes between each system. The GPCR-G-protein complex is shown for each system in Figure 10. The superimposed images show subtle changes throughout the complex. The GPCR changes are shown in Figure 11. The most noticeable changes occur in TM 6 and 7 between the superimposed GPCR’s. These GPCR images are again taken from the last snap shot of each simulation. The most noticeable changes are visible between the Gα subunits. These images are shown in Figure 12. GDP is no longer present in the agonist Gα subunit. There are slight changes at the α5 helix and the GDP binding site. There are also visible changes between the alpha helical domains for each system. The superimposed structures are shown from two different viewpoints.
Figure 10. Full structure images of the three protein-ligand complexes. (A) Agonist complex. (B) Antagonist complex. (C) No ligand complex. (D) Agonist (red) and antagonist (black) superimposed full structures. (E) Agonist (red) and no ligand (cyan) superimposed full structures.
Figure 11. Receptor only images. (A) Agonist receptor. (B) Antagonist receptor. (C) No ligand receptor. (D) Agonist (red) and antagonist (black) superimposed. (E) Agonist (red) and no ligand (cyan) superimposed.
Protein-ligand interaction analysis shows the key residual contacts in the agonist complex are the same as the antagonist complex. The ligand-GPCR contacts are listed in Table 3. The 2D interaction diagrams and contact histograms are shown in Figure 13. Asp 113, Phe 193, Phe 290, and Asn 312 are in contact with both the agonist and
antagonist for at least 30.0% of the simulation time. The agonist and antagonist involve residues mainly from TM 3, 5, 6 and 7.

Table 3

*Ligand-Receptor Contacts*

<table>
<thead>
<tr>
<th>Agonist</th>
<th>Antagonist</th>
</tr>
</thead>
<tbody>
<tr>
<td>W109&lt;sup&gt;3,28&lt;/sup&gt;</td>
<td>D113&lt;sup&gt;3,32&lt;/sup&gt;</td>
</tr>
<tr>
<td>D113&lt;sup&gt;3,32&lt;/sup&gt;</td>
<td>F193&lt;sup&gt;E2&lt;/sup&gt;</td>
</tr>
<tr>
<td>F193&lt;sup&gt;E2&lt;/sup&gt;</td>
<td>F290&lt;sup&gt;6,52&lt;/sup&gt;</td>
</tr>
<tr>
<td>S203&lt;sup&gt;5,42&lt;/sup&gt;</td>
<td>N312&lt;sup&gt;7,39&lt;/sup&gt;</td>
</tr>
<tr>
<td>S207&lt;sup&gt;5,46&lt;/sup&gt;</td>
<td>V114&lt;sup&gt;3,33*&lt;/sup&gt;</td>
</tr>
<tr>
<td>F290&lt;sup&gt;6,52&lt;/sup&gt;</td>
<td>F289&lt;sup&gt;6,51*&lt;/sup&gt;</td>
</tr>
<tr>
<td>N293&lt;sup&gt;6,55&lt;/sup&gt;</td>
<td>Y308&lt;sup&gt;7,35*&lt;/sup&gt;</td>
</tr>
<tr>
<td>N312&lt;sup&gt;7,39&lt;/sup&gt;</td>
<td>I309&lt;sup&gt;7,36*&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

*Not in contact more than 30.0% of simulation time
Figure 13. 2D protein-ligand interaction diagrams and histograms for agonist vs. antagonist.
Protein-ligand interaction analysis shows the key residual contacts for GDP in the antagonist complex are the same as the no ligand complex, while GDP in the agonist complex has little contact with the binding pocket. The GDP- Ga subunit contacts are listed in Table 4. The 2D interaction diagrams and contact histograms for GDP in the antagonist complex and no ligand complex are shown in Figure 14. Ser 50, Gly 51, Lys 52 and Ser 53 are in contact with GDP in both the antagonist and no ligand complexes for at least 30.0% of the simulation time. The no ligand complex is also in contact with Thr 54, Asn 274 and Asn 277 for at least 30.0% of the simulation time. The GDP in the agonist complex is not in contact with any residues for at least 30.0% of the simulation time.

Table 4

**GDP-Ga Contacts**

<table>
<thead>
<tr>
<th>Agonist</th>
<th>Antagonist</th>
<th>No Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>S53[^{a1}]*</td>
<td>E49[^{a1}]</td>
<td>S50[^{a1}]</td>
</tr>
<tr>
<td>K57[^{a1}]*</td>
<td>S50[^{a1}]</td>
<td>G51[^{a1}]</td>
</tr>
<tr>
<td>K274[^{aG}]*</td>
<td>G51[^{a1}]</td>
<td>K52[^{a1}]</td>
</tr>
<tr>
<td>D277[^{aG}]*</td>
<td>K52[^{a1}]</td>
<td>S53[^{a1}]</td>
</tr>
<tr>
<td>S53[^{a1}]</td>
<td>T54[^{a1}]</td>
<td>N274[^{aG}]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D277[^{aG}]</td>
</tr>
</tbody>
</table>

*Not in contact more than 30.0% of simulation time
The tyrosine toggle (NPXXY) molecular switch was present in all three systems but “switched” on in the agonist complex only. The torsion angle analysis was completed for the X1 and X2 angles within Y326 of the tyrosine toggle switch. The X1 analysis shows a relatively flat plot for the no ligand complex, with a little bit of angle change at the beginning of the trajectory. The antagonist complex shows a 100-degree change (from 300 to 200) early in the trajectory, but then stays consistently at 200 for the remainder of the trajectory. Agonist X1 shows some angle changes throughout the trajectory. X2 analysis shows much larger changes within all three systems. The no
ligand complex and antagonist complex appear to move back and forth from 0 to 360.
The agonist complex is the only system which shows a clear “switch” in conformations.
The agonist Y326 X2 visibly switches from approximately 275 degrees to 100-150 degrees halfway through the trajectory.

Protein secondary structure analysis indicates structures are the same for three complexes with subtle changes throughout. A secondary structure comparison is shown in Figure 15. TM 1, 2, 3 and 4 show little difference between the three complexes. TM 1 in the agonist complex shows subtle changes as compared to the antagonist and no ligand complexes. The largest differences between the three GPCR’s are shown in TM 6 and 7. The Gα subunit shows changes throughout for the three complexes. The Gβ subunit is conserved through each complex and the Gγ subunit has a slight loss in its second alpha helix in the no ligand complex.
Figure 15. Protein secondary structures and their composition percentages. (A) Agonist. (B) Antagonist. (C) No Ligand.

Ligand RMSF shows generally more fluctuation for the antagonist than the agonist, and much more than 10-fold fluctuation for agonist GDP than both no ligand and antagonist GDP’s. The antagonist exhibits higher RMSF values than the agonist (Figure 16). The agonist is more rigid in the GPCR. GDP in the agonist complex shows a much larger RMSF profile than GDP in both the antagonist and no ligand complexes (Figure 17). This is consistent with our MM-GBSA binding energies and our Ligand RMSD profiles.
**Figure 16.** Snapshots of α5 helix and β6-α5 loop movement for agonist and no ligand complexes. (A) First snapshot. (B) Loops separating. (C) Snapshot. (D) Snap before GDP leaves. (E) Snapshot. (F) Last snapshot.
Figure 17. Distance Measurements and RMSD for α5 and β6-α5 loop. (A) The center to center mass was calculated for α5 helix to GDP, and β6-α5 loop to GDP. (B) From top to bottom: α5 helix RMSD, β6-α5 loop RMSD, combined α5 helix and β6-α5 loop RMSD.

RMSF data shows larger fluctuation at ECL2, ICL3 for the agonist GPCR, and larger fluctuation at α4, α5 and β6 for the agonist Gα subunit. The RMSF values are
similar for the agonist as compared to the antagonist in all rigid, TM regions (Figure 18). At ECL2 there is slightly more fluctuation in the agonist GPCR. At ICL3, there is at least 2x more fluctuation in the agonist GPCR as compared to the antagonist GPCR. ICL3 is the largest ICL and the loop in closest contact with the G-protein. In Figure 18B there is visibly more fluctuation in the antagonist complex as compared to the agonist complex in the alpha helical domain of the Ga subunit. However, in the Ras domain for the agonist Ga subunit there is more fluctuation at α4, α5 and β6. The α5 helix is the last helix of the Ras domain and the only helix in direct contact with the GPCR. In Figure 18C there is similar fluctuation in the Gβ subunit for the agonist as compared to the antagonist. In Figure 18D there is more fluctuation in the first helix of the Gγ subunit for the antagonist complex, but similar fluctuation in the second helix for both subunits.
Figure 18. Protein Cα RMSF diagrams of agonist (blue) vs. antagonist (orange). (A) Receptor. (B) Gα protein. (C) Gβ protein. (D) Gγ protein.
Agonist fluctuation at the β6-α5 loop and α5 helix is much larger than both the antagonist and no ligand complexes. Simulation comparisons show a clear difference between the agonist complex and the antagonist, no ligand complexes. The agonist β6-α5 loop separates from the antagonist β6-α5 loop and shows an upward movement away from GDP. The β6-α5 loop in the no ligand complex does not show much fluctuation and movement away from GDP. The simulations also show a change in α5 position for the agonist complex as compared to the other two. The second trajectory of the agonist also shows loop fluctuation and a large α5 displacement.

Distance measurements confirm simulations and show agonist complex has the most β6-α5 loop fluctuation and α5 helix displacement. The distance was measured between the center-of-mass of the α5 helix and GDP, and between the center-of-mass of the β6-α5 loop and GDP (Figure 17). The no ligand complex remains constant for both the α5 helix and the β6-α5 loop with values of approximately 25 Å and 15 Å, respectively. The antagonist complex shows minor changes throughout the analysis for both measurements. The α5 helix distance starts around 30-31 Å and ends at approximately 27 Å. The decrease is gradual with little to no fluctuation during the simulation. The β6-α5 loop distance shows similar results for the antagonist. The distance stays between 18-22 Å for the first 600 frames, then decreases to about 13-15 Å for the remainder of the simulation. The agonist begins at approximately 11 Å then quickly begins to fluctuate between 12-22 Å distances until the simulation reaches approximately 800 µs. At 800 µs the distance increases to about 26 Å then down to 9 Å. The distance increases again to 23 Å before GDP begins to move out of the binding pocket. GDP leaves completely at about 1170 µs where the distance jumps to 108 Å.
RMSD of α5 helix and the β6-α5 loop confirms deviations of the β6-α5 loop are much larger than the α5 helix. RMSD analysis was completed for the α5 helix, the β6-α5 loop, and both the helix and loop combined. The residues used for the α5 helix were 326-335 on Chain A (Gα subunit). The residues used for the β6-α5 loop were 336-359 on Chain A. Both sets of residues were used for the combination analysis. Figure 17 shows RMSD values between 2-6 Å for the α5 helix and values between 4-12 Å for the β6-α5 loop.

3.4 Discussion

Since there have been studies regarding GPCR-G-protein complex activation and nucleotide exchange, we want to validate our results against these studies. As stated earlier, Chung et. al. completed research to better understand the molecular workings of G-protein activation through peptide amide hydrogen-deuterium exchange mass spectrometry. (23) They determined that P-loop stabilization is a key determinant of GDP binding affinity. (23) Dror et. al. used simulations to analyze the Gα subunit domains and nucleotide release. They determined that domain separation is necessary for GDP departure but not sufficient alone. GPCR’s facilitate a conformational change within the Ras domain to weaken nucleotide affinity (22). Dror used a previous mutagenesis study to confirm this conclusion about domain separation. This mutagenesis study suggested that weakening interactions between the β6-α5 loop and GDP facilitates nucleotide release to a greater extent than increasing domain separation. (26)
Our MM-GBSA binding energies (Table 2) demonstrate destabilization through agonist binding at two interfaces of the GPCR-G-protein complex. This is consistent with our initial results from the simulation trajectories. Agonist destabilization at both the GPCR-G-protein and G-protein-GDP interface may likely correlate to the downstream effects of G-protein signaling. GDP leaves the agonist complex in the simulation indicating that destabilization is necessary for nucleotide exchange. The G-protein dissociates from the GPCR after nucleotide exchange is complete, indicating that destabilization at this interface is also necessary. Both the agonist (BI-167107) and the antagonist (alprenolol) bind to the same binding pocket in the GPCR. The MM-GBSA values indicate the agonist binds more favorably to the GPCR. Ligand RMSD shows similar deviations within the GPCR binding pocket (Figure 7A) and the ligands have similar key contact residues (Table 3).

Ligand RMSD for GDP shows GDP movement at its binding pocket in the Gα subunit. While the GDP starts at relatively the same RMSD value, GDP within the agonist complex begins to show a larger deviation while still in the binding pocket. This movement within the binding pocket indicates a conformational change is necessary for GDP detachment.

The protein RMSD was determined for seven components of the complexes (Figure 8). We looked at the complex as a whole, the G-protein only, beta-gamma complex, alpha subunit, beta subunit, gamma subunit and the GPCR only. The distinction between the groups containing the alpha subunit and those that do not is very noticeable for the agonist and antagonist complexes. This separation does not occur in the complex without a ligand. The large conformational changes within the alpha subunit
prepare the G-protein for nucleotide exchange. Ligand binding at the GPCR induces a change within the alpha subunit, whether it’s an agonist or antagonist.

The trajectory images for each system show the movement of GDP every 10 frames of the simulation time. GDP movement within the GαRas binding pocket indicates GDP conformational changes are necessary for detachment. This is confirmed by the little to no movement of GDP within the binding pocket of the antagonist and no ligand complexes. The GDP conformational changes within the binding pocket are indicative of coordinated communication between the activated GPCR and the G-protein. The agonist-activated GPCR induces a change with the GαRas binding pocket that induces a change in GDP binding. This confirms previous research by Dror et. al. that opening of the Gα domains alone is not sufficient for nucleotide exchange. (22)

Protein RMSF data clearly shows differences between the three complexes. The complete protein RMSF analysis shows the rigid, TM regions for all three complexes. The agonist complex shows more fluctuation at ECL2 and much larger fluctuation at ICL3 compared to the antagonist and no ligand complexes. The large fluctuation at the ICL3 loop indicates it may play a role in GPCR-G-protein activation. It is the largest of the six loops and it is in closest contact with the G-protein. The protein RMSF for the Gα subunit is shown for both domains. In the GαAH domain, there is larger fluctuation in the antagonist and no ligand complexes as compared to the agonist complex. This data could be confirming Dror et. al’s work. (22) In the antagonist and no ligand complexes the GαAH domain should begin to close due to the fact that nucleotide exchange is not occurring. The larger fluctuation at the GαAH domain could indicate the start of this domain closing, although our simulation is not long enough to see it move back towards
the GaRas domain. From the protein RMSF we can see clear fluctuation differences at locations on the GaRas domain. These locations line up with α4, β6 and α5 on the GaRas secondary structure. The α5 is in closest contact with the GPCR, followed by the β6-α5 loop. The larger fluctuation here indicates that this region may play a role in G-protein activation by the GPCR. This would confirm results from previous research done by Dror et. al and Chung et. al. (22,23) Simulations also show a change in the β6-α5 loop for the agonist complex that is not visible in the antagonist and no ligand complexes. There is much larger fluctuation and a shift upward of this loop. The β6-α5 loop appears to move away from GDP. This is further confirming the notion of breaking GaRas domain contacts with GDP made by Dror et. al. (22) Distance analysis confirms the results from the comparative simulations (Figure 17). The center-of-mass of the α5 helix was measured from the center-of-mass of GDP over the course of the 3 µs simulation. The center-of-mass of the β6-α5 loop was also measured from the center-of-mass of GDP. For the no ligand complex, there is almost no fluctuation in distance values for both the α5 helix and the β6-α5 loop. This indicates that when activation through a ligand does not occur, there is subsequently no conformational changes within the α5 helix and the β6-α5 loop. Since GDP does not leave in the no ligand system, this also indicates that these conformational changes through activation may be necessary for nucleotide exchange. Dror et. al. showed a direct correlation between the α5 helix, β6-α5 loop and GDP leaving, therefore our results confirm this notion. (22) The antagonist complex shows slight distance fluctuation for the first 600 ns within the β6-α5 loop, but very little distance fluctuation for just the α5 helix. There does not appear to be any coordination between the α5 helix and the β6-α5 loop even during the first 600 ns. This is
an indication that fluctuation within the loop is necessary for GDP departure, and the antagonist complex did not have enough distance fluctuation in order to break GDP-GαRas bonds. GDP leaves the agonist complex at approximately 1180 ns. There is clear distance fluctuation in both the α5 helix and the β6-α5 loop throughout the 1180 ns leading up to GDP departure. There is a peak of fluctuation between 780 and 850 ns for both the α5 helix and β6-α5 loop. The fluctuation appears to level off around 975 ns for the agonist complex but it shows more fluctuation than the antagonist and no ligand complexes between 975 and 1180 ns. Only the agonist complex shows coordinated fluctuation between the α5 helix and the β6-α5 loop. The fluctuation witnessed for the β6-α5 loop in the antagonist complex is not present within the α5 helix. This could indicate that movement of both the α5 helix and the β6-α5 loop is necessary for GDP detachment and nucleotide exchange.

3.5 Conclusions

Our MD simulations demonstrate the effects of an agonist and an antagonist on G-protein activation and therefore nucleotide exchange. MM-GBSA analysis shows the destabilization of the GPCR-G-protein complex due to agonist binding. This destabilization does not occur to the same magnitude when the antagonist binds. This destabilization is likely the preparation for GDP release and G-protein dissociation. While the tyrosine toggle switch (NPXXY) appears to be activated in the agonist complex, we argue that molecular switches are not crucial for GPCR-G-protein activation. RMSD and RMSF analysis shows the importance of the conformational changes within the GPCR and G-protein. The largest changes within the receptor occur at the ECL2 and ICL3. The Gα subunit plays the largest role in G-protein activation and
GDP release. The RMSF analysis shows larger fluctuation at the α4 helix, α5 helix and β6 strand as compared to both the antagonist and no ligand complexes. Further analysis shows a clear distinction between the agonist β6-α5 loop as compared to the other two complexes. Both the first and second trajectory of the agonist show an upward movement, away from GDP, at the β6-α5 loop. This movement is predicted to break GDP-GaRas domain interactions, therefore allowing GDP to leave the binding pocket. GDP was not able to leave in the antagonist or no ligand complexes where this movement was not as noticeable. Conformational changes of GDP within the binding pocket before departure were seen in the agonist complex only. These changes may occur due to the movement of the β6-α5 loop. Coordinated movement between the α5 helix and the β6-α5 loop was also witnessed in the agonist complex using RMSD analysis of that region. This was not witnessed in the other two complexes. This is confirming the roles that the α5 helix and β6-α5 loop play in nucleotide exchange.
References


Appendix A

Supporting Material for Chapter 2

Table A1

*Residue Interactions with DOX and XN Ligands*

<table>
<thead>
<tr>
<th>DOX complex</th>
<th>XN complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILE 340&lt;sup&gt;1&lt;/sup&gt;</td>
<td>PHE 336&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>PHE 983&lt;sup&gt;1&lt;/sup&gt;</td>
<td>PHE 983&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>ASN 839&lt;sup&gt;1&lt;/sup&gt;</td>
<td>ALA 980&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>GLN 990&lt;sup&gt;1&lt;/sup&gt;</td>
<td>PHE 728&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>TYR 307&lt;sup&gt;1&lt;/sup&gt;</td>
<td>ASN 842</td>
</tr>
<tr>
<td>GLN 725&lt;sup&gt;1&lt;/sup&gt;</td>
<td>TYR 953</td>
</tr>
<tr>
<td>PHE 343&lt;sup&gt;1&lt;/sup&gt;</td>
<td>PHE 732</td>
</tr>
</tbody>
</table>

<sup>1</sup>Interactions that occur more than 30.0% of the simulation time in the selected trajectory
Figure A1. Complexes from IFD docking. Closed, original crystal structure (PDB ID: 4Q9I). (A) DOX docked. (B) XN docked.
Figure A2. Protein/Ligand Root Mean Square Deviation (RMSD) Comparison. RMSD measures the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. All protein frames are first aligned on the reference frame backbone (i.e. the initial frame), and then RMSD is calculated based on the atom selection. (A) DOX complex. (B) XN complex.
Figure A3. Protein Secondary Structure. Protein secondary structure elements (SSE) are monitored throughout simulation. This plot reports the SSE distribution by residue index.
throughout the protein structure. (A) DOX complex. (B) XN complex. Purple: alpha-helices  Blue: beta-strands White: Coils
Figure A4. Protein-Ligand Contacts. (A) DOX complex. (B) XN complex.
Figure A5. Protein Interactions Diagrams. (A) Shows the 2D interaction diagram and the 3D interaction diagram from the last snapshot structure for the DOX complex. (B) Shows the 2D interaction diagram and the 3D interaction diagram from the last snapshot structure for the XN complex.
Figure A6. Ligand Torsion Profile. The ligand torsions plot summarizes the conformational evolution of every rotatable bond in the ligand throughout the simulation trajectory. (A) DOX. (B) XN.
Appendix B

Supporting Material for Chapter 3

Figure B1. β2-Adrenoceptor. (A) Crystal structure. (B) Amino acid representation.
Figure B2. β2-Adrenoceptor FASTA sequence.
Figure B3. (A) Ga subunit. (B) WD40 Scaffold of the beta subunit within the agonist complex. (C) Beta (yellow)-gamma (blue) complex.
Figure B4. Protein Root Mean Square Deviation (RMSD) for seven individual parts of the GPCR-G-protein complexes. (A) Full complex. (B) G-protein only. (C) βγ complex. (D) Gγ. (E) Gβ. (F) Ga. (G) Receptor only.
Figure B5. Protein Cα Root Mean Square Fluctuation (RMSF) diagrams of agonist (blue) vs. antagonist (orange). (A) Receptor. (B) Gα protein. (C) Gβ protein. (D) Gγ protein.
Figure B6. Snapshot images for $\alpha_5$ helix and $\beta_6$-$\alpha_5$ loop for the agonist and antagonist complexes.
Figure B7. Ligand contacts throughout simulation for Agonist, Antagonist, No ligand.
Figure B8. Ligand properties for GDP in (A) Agonist (B) Antagonist (C) No Ligand.
Figure B9. Molecular switches. The residues are numbered according to the Ballesteros-Weinstein numbering scheme.
Figure B10. X1 Torsion angle for Y326. (A) Agonist. (B) Antagonist. (C) No ligand.
Figure B11. X2 Torsion angle for Y326. (A) Agonist. (B) Antagonist. (C) No ligand.
Second Trajectory Information

Figure B12. Root Mean Square Deviation (RMSD) of the protein and ligands in the second trajectory of the agonist complex against simulation time. (A) GDP. (B) Agonist. (C) Protein RMSD.
Figure B13. Trajectory images of the agonist second trajectory in three viewpoints.
Figure B14. Protein Cα Root Mean Square Fluctuation (RMSF) diagrams of agonist (second trajectory) compared to the antagonist. (A) Receptor. (B) Gα protein. (C) Gβ protein. (D) Gγ protein.
Figure B15. Snapshot images of GDP leaving/exchange for agonist (second T)/No Ligand
Figure B16. Snapshot images of GDP leaving/exchange for agonist (second T)/antagonist.
Figure B17. Agonist second trajectory ligand contacts and ligand properties
Figure B18. 2D protein-ligand interaction diagram and histogram for agonist in the second trajectory.
Figure B19. Ligand RMSF comparison between agonist from second trajectory and antagonist.
Figure B20. Ligand RMSF comparing GDP from second trajectory to antagonist GDP and no ligand GDP.