5-2017

Parameterization of AR231453: A Potent Agonist of the GPR119 Receptor, a Target for Diabetes Treatment

John E. Hamilton
The University of Texas Rio Grande Valley

PARAMETERIZATION OF AR231453: A POTENT AGONIST OF THE GPR119 RECEPTOR, A TARGET FOR DIABETES TREATMENT

A Thesis

by

JOHN E. HAMILTON

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

May 2017

Major Subject: Chemistry

PARAMETERIZATION OF AR231453: A POTENT AGONIST OF THE GPR119 RECEPTOR, A TARGET FOR DIABETES TREATMENT

A Thesis
by
JOHN E. HAMILTON

COMMITTEE MEMBERS

Dr. Evangelia Kotsikorou
Chair of Committee

Dr. Frank Dean
Committee Member

Dr. Jason Parsons
Committee Member

Dr. Jose Gutierrez
Committee Member

May 2017

ABSTRACT

Hamilton, John E., Parameterization of AR231453: A Potent Agonist for the GPR119 Receptor, a Target for Diabetes Treatment. Master of Science (MS), May, 2017, 36 pp., 4 tables, 10 figures, references, 36 titles.

The first reported potent agonist for the GPR119 receptor, a potential target for the treatment of diabetes, is 2-fluoro-4-methanesulfonyl-phenyl-{6-[4-(3-isopropyl-[1,2,4]oxadiazol-5-yl)-piperidin-1-yl]-5-nitro-pyrimidin-4-yl}-amine (AR231453). The dynamic interactions of AR231453 with GPR119 using molecular dynamics simulation are of great interest. However, parameters for AR231453 that describe the behavior of the molecule in the CHARMM force field have not been determined. The following study produces parameters by creating model fragments and compares the produced parameters to quantum calculations. The produced parameters are then further refined within the Visual Molecular Dynamics (VMD) program's plugin program, the Force Field Toolkit as well as Gaussian 09 and CHARMM. Molecular dynamics simulations are done to study the orientation of AR231453 in a lipid bilayer.

DEDICATION

I would like to dedicate this thesis to my dogs, Dude and Emma.

ACKNOWLEDGMENTS

I would like to acknowledge Robert Jackson for technical support and the Texas Advanced Computing Center (TACC) for use of computational resources on the Stampede supercomputer.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

GPR119 is a G-protein-coupled receptor (GPCR) which has become a potential target of antidiabetic drugs. The first potent and orally efficacious agonist of GPR119 was identified by Semple et al. known as AR231453. Recent difficulties including other GPR119 agonists losing efficacy due to tachyphylaxis and several of the candidate agonists having low efficacies have led to shifting studies to other targets and antidiabetic drugs. One hurdle in the understanding of GPR119 is the lack of a crystal structure for the protein. Homology modeling of the protein may be able to overcome this missing information. With a working model, molecular dynamics studies of the protein could shed light on how the protein interacts with a variety of different agonists. The interaction of GPR119 and AR231453 could reveal important insights into how the compound activates the protein and lead to renewed interest in developing better antidiabetic drugs for this target. In order to run molecular dynamics simulation studies, molecular mechanics parameters for AR231453 must be developed and validated. The aim of this thesis is to obtain parameters to be used for further computational studies.

G-Protein-Coupled Receptors and the GPR119 Receptor

GPCRs are transmembrane proteins which contain seven α helices. These proteins are very diverse in mammals and the superfamily is one of the largest [1,2]. GPCRs have a wide

range of functions in the physiology of organ systems [3]. Located at the cellular surface, GPCRs transmit a signal which is converted into an intracellular response through heterotrimeric G-proteins [3]. Due to their importance, GPCRs have become targets for drug development with an estimated 40-45% of all modern drugs targeting these receptors [4].

Within the superfamily of GPCRs, the rhodopsin family (class A) is the largest with approximately 90% of all GPCRs belonging to this subfamily [5]. There are many ligands which interact with these proteins including biogenic amines, peptides, lipids, nucleosides and nucleotides, and other large proteins to name a few [3]. Rhodopsin family GPCRs have become the target of pharmaceutical studies because these proteins can bind small molecule amines which could be exploited to produce better drugs and many of these GPCRs are orphans with no known ligands [3]. The potential for discovery of new ligands and understanding of interactions between protein and ligand promises to be fertile ground for future works.

GPR119 is a class A rhodopsin-like GPCR which has received attention because of studies showing that modulation of the receptor influence glucose homeostasis and weight gain. Fredriksson et al. described human GPR119 during a charting of GPCRs in the human genome [3]. Many studies have shown human and rat GPR119 expression in the pancreas, fetal liver, and gastrointestinal tract [6,7,8,9,10]. In cells transfected with GPR119, intracellular cAMP levels were shown to increase hinting that the receptor couples to the $G_{\alpha s}$ subunit which then leads to the stimulation of adenylate cyclase [7,11,12]. In rodent pancreatic islets, Chu et al. deduced that β-cells were the primary site of GPR119 expression; however, other sites could not be ruled out for expression [12]. Sakamoto et al. provided contrary evidence showing that pancreatic polypeptide-secreting cells as the only site of GPR119 expression in mouse and rat islets [13].

The hypothesis that GPR119 could be involved in modulation of insulin secretion is due to evidence that it is expressed in pancreatic β-cells and that modulation of glucose stimulated insulin secretion could proceed by similar mechanisms such as in the incretin hormones glucagon like peptide-1 (GLP-1) and gastric inhibitory polypeptide/glucose dependent insulinotropic peptide (GIP) which also act through a $G_{\alpha s}$ subunit in β-cells [6,14]. Enhancement of glucose stimulated insulin secretion in β-cell lines was seen using a GPR119 endogenous agonist, oleoyl lysophosphatidylcholine [9]. The glucose stimulated insulin secretion using oleoyl lysophosphatidylcholine was dampened in the presence of adenylate cyclase inhibitor or GPR119-selective siRNA [9]. In another study, Lan et al. showed that there was no glucose stimulated insulin secretion in GPR119 deficient mice [15]. Using the GPR119 agonist AR231453, Chu et al. showed enhancement of glucose stimulated insulin secretion in hamster insulinoma cells while GPR119 deficient mice showed no effect using the compound [12].

AR231453: a potent and orally efficacious agonist for GPR119

2-fluoro-4-methanesulfonyl-phenyl-{6-[4-(3-isopropyl-[1,2,4]oxadiazol-5-yl)-piperidin-1-yl]-5-nitro-pyrimidin-4-yl}-amine, also known as AR231453, was the first reported potent agonist of GPR119 [16]. Semple et al. identified a compound from an inverse agonist screening hit that served as the starting point for the development of AR231453. Using the knowledge that the nonpeptide antagonists of GPR119 have a similar structure to agonists, the nitro pyrimidine core served as the starting block for further synthesis [16]. Also, the piperidine group was deemed important due to the lack of activity of similar compounds without this group in the screening hit [16]. Structure-activity relationship studies were then conducted which identified the remaining important functional groups including adding a phenol group with a methyl

sulfone in the 4-position of the phenol ring leading to an agonist and replacing an ester group with a more metabolically stable isopropyl oxadiazole group which is capable of forming hydrogen bonding and donating as the ester [16]. The resulting compound, AR231453, showed in vivo activity for enhanced glucose dependent insulin release in mice after oral administration but not in GP119 deficient mice [16].

Since the discovery of AR231453, several studies of the interaction of AR231453 and GPR119 have been conducted. Chu et al. showed that AR231453 significantly increased cAMP accumulation and insulin release in both HIT-T15 cells and rodent islets [12]. However in GPR119 deficient systems, AR231453 produced no activity [12]. AR231453 also improved glycemic control in both normal and diabetic mice [12]. In the same study, the actions of AR231453 were observed to increase blood insulin levels in a glucose dependent manner in isolated rat islets and in vivo as opposed to other glucose-independent agents such as sulfonylureas [12]. AR231453 was also shown to increase cAMP and GLP-1 release in mice [17]. Gao et al. demonstrated that AR231453 can stimulate β-cell replication and improve islet graft function in diabetic mice transplanted with mouse islets although the direct or indirect effect of acting of GPR119 is unclear [18].

The ligand binding pocket of GPR119 was studied using mutational mapping on residues required for GPR119 activation using AR231453 and oleoylethanolamide [19]. The mutational mapping for AR231453 and OEA were similar and that residues in extracellular loop-2b were important for activity [19]. Using a hybrid multiple template, a GPR119 model was built to be used for docking studies with AR231453 [19]. Docking of AR231453 into the GPR119 model showed that binding occurs in a vertical pocket [19]. The study revealed the major AR231453 binding mode to be SO_2 up with the sulfonyl group of AR231453 pointing to the extracellular

region [19]. However, the SO_2 down pose was still observed and is energetically allowed [19]. The mutational mapping and docking study will help inform and potentially corroborate future molecular dynamics studies such as could be done with AR231453 parameters.

Molecular Mechanics and Force Field Parameter Development

Since quantum mechanical and semi-empirical calculations are computationally expensive for large scale system like proteins, molecular mechanics methods such as dynamics simulations are the standard computational choice for studying biomolecular problems. Given that protein flexibility is crucial for understanding drug binding, molecular mechanics is central to the study of computational structure-based drug discovery [20]. Molecular mechanics (MM) force fields satisfactorily represent a quantum mechanical surface by using a classical model [20]. A force field is defined by a potential energy function and a set of parameters that are used by the potential energy function. The potential energy function represents the bonded (bonds, angles, and dihedrals) and nonbonded (electrostatics and Van der Waals) energy terms as additive terms [20]. The parameters used in the equation represent the second important part of a force field. Unrealistic parameters can give unrealistic results when performing molecular dynamics. The potential energy function represents the heart of molecular mechanics and has been developed in detail in numerous reviews [20,21]. Biomolecular force fields with similar energy terms in the potential energy function include CHARMM, AMBER, GROMOS, OPLS, and CGenFF [21-26].

The CGenFF force field represents a force field aimed at simulating drug-like molecules in a biological environment represented by CHARMM force fields [26]. The CGenFF force field

uses the same potential energy function and conventions as CHARMM [26-28]. In an effort to streamline parameterization, Mayne et al. developed the force field toolkit to aid in parameter development [29].

CHAPTER II

METHODOLOGY

The CHARMM force field is a potential energy function described in equation (1)

$$E = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{U-B} K_{U-B}(s - s_0)^2 + \sum_{dihedrals} K_\phi[1 + \cos(n\phi - \delta)] + \sum_{improper} K_\psi(\psi - \psi_0)^2 + \sum_{nonbonded} \left\{ \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}} \right\} \quad (1)$$

where, K_b, K_θ, $K_{U\text{-}B}$, K_φ, K_ψ represent the force constants, b represents bond length, θ represents bond angle, s is the Urey-Bradley distance, ϕ is the dihedral angle, and ψ is the improper dihedral angle. The equilibrium values of each term is indicated with a subscript of zero. The dihedral term includes n and δ which represent the periodicity and the phase of the function, respectively. Nonbonded interactions are contained in the Lennard Jones potential and electrostatic interactions are included in the Coulombic potential. In the Lennard-Jones potential, R_{min} is the distance between atoms at the Lennard-Jones minimum and ϵ is the well depth. In the Coulombic potential, q represent the partial atomic charges and ϵ_1 represent the effective dielectric constant. The distance between atoms i and j is represented by r_{ij}. These values constitute the parameters needed to describe AR231453 in the CHARMM force field. The bond, angle Urey-Bradley, and improper terms utilize Hooke's law. The dihedral term is represented by a sinusoidal function.

Molecular mechanics parameters for AR231453 for the CHARMM force field were developed by comparison to quantum mechanical (QM) data using the Schrödinger molecular modeling suite, the VMD plugin force field toolkit (FFTK), Gaussian 09 program, and CHARMM program [30,29,31,32,22]. The FFTK uses an objective function along with target data to optimize parameters. The objective function provides for an optimization algorithm in which to compare QM and MM data and optimize until convergence to an objective value [29]. Objective functions for charges, bonds and angles which are done simultaneously, and dihedrals are included in the optimization workflow. The structure of AR231453 is displayed in Figure 1 along with the atomic naming convention used in this work. CGenFF database atom types were assigned to AR231453 and are listed in Table 1 for each atom label in Figure 1.

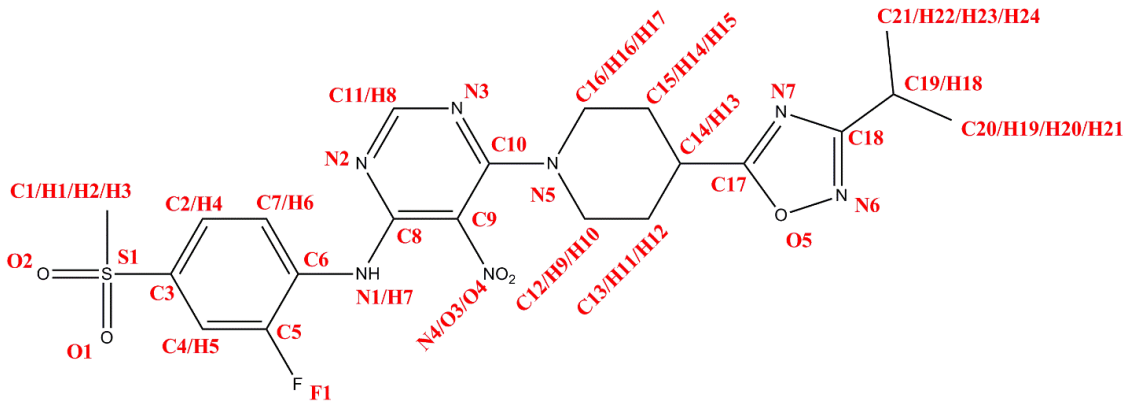


Figure 1: Structure of AR231453

Atom	Type	Partial Charge
C6	CG2R61	0.05
C2/C7	CG2R61	-0.115
C3	CG2R61	0.29
C4	CG2R61	-0.1
C5	CG2R66	0.11
F1	FGR1	-0.21
S1	SG3O2	0.14
O1/O2	OG2P1	-0.36
C1	CG331	0.02
N1	NG311	-0.219
N3	NG2R62	-0.73
C10	CG2R64	0.683
C9	CG2R61	0.32
C8	CG2R64	0.117
N2	NG2R62	-0.73
C11	CG2R64	0.5
N4	NG2O1	0.4
O3/O4	OG2N1	-0.34
C13/C15	CG321	-0.18
C12/C16	CG321	0
N5	NG301	-0.5
C14/C19	CG311	-0.09
C17	CG2R53	0.62
N7	NG2R50	-0.64
C18	CG2R53	0.62
N6	NG2R50	-0.42
O5	OG2R50	-0.18
C20/C21	CG331	-0.27
H4/H6	HGR61	0.115
H5	HGR62	0.15
H1/H2/H3	HGA3	0.09
H7	HGPAM1	0.339
H8	HGR62	0.14
H9/H10/H11/H12	HGA2	0.09
H14/H15/H16/H17	HGA2	0.09
H13/H18	HGA1	0.09
H19/H20/H21/H22/H23/H24	HGA3	0.09

Table 1: Listing of Atom Types, Names, and Partial Charges for AR231453

Model Compounds

The AR231453 compound was strategically divided into model compounds so that parameter determination becomes computationally tractable. The models were chosen to preserve the atom types present in AR231453 and to represent the torsion parameters realistically. Additionally, each fragment contains dihedrals that have missing parameters in the CGenFF to be determined. The Figures 2A-2E show the model compounds used for parameter development.

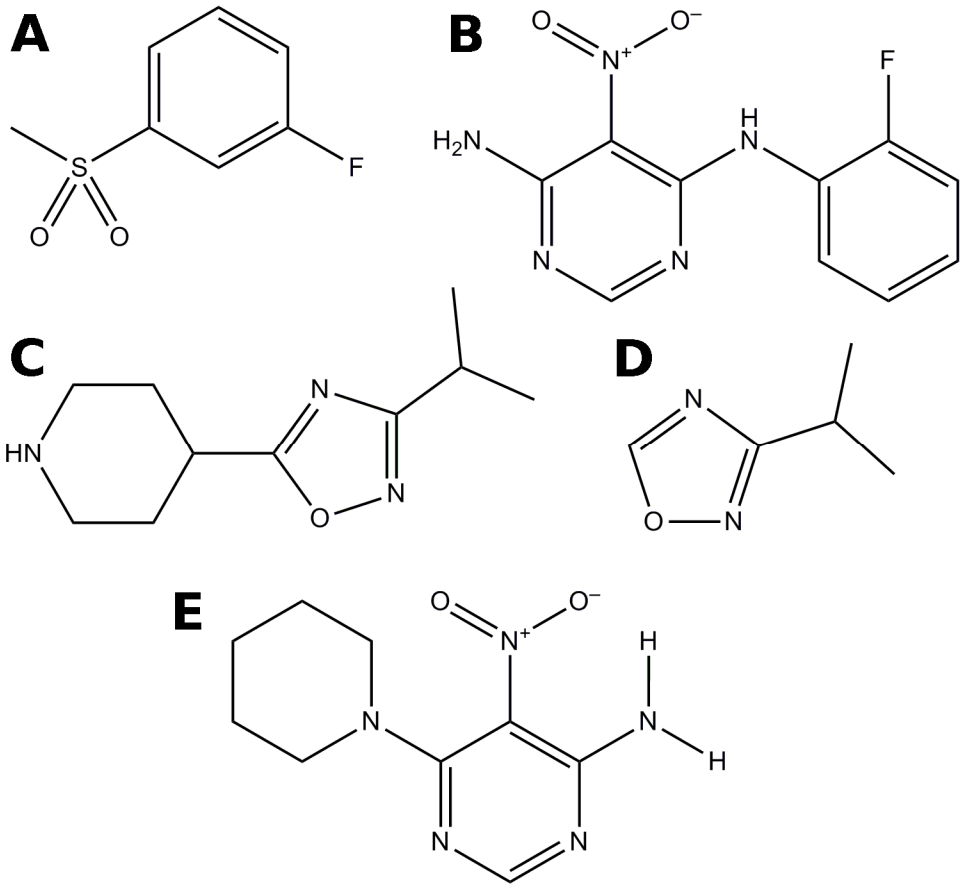


Figure 2: Model Compound Fragments

System Preparation

System preparation began with building the model compounds in Maestro (Schrödinger program), initially optimizing the geometry using the Hartree-Fock (HF) method with a 6-31G* basis set, and writing out a protein data bank file (PDB) which contains the Cartesian coordinates for each atom of the fragment [30]. The PDB file was then uploaded to the ParaChem website which assigns a set of parameters by analogy from CGenFF [27,28]. These parameters are the initial guesses to start the calculations. The PDB file is imported into the visualizing program VMD and initial partial charges from the ParaChem guess are used to generate an initial protein structure file (PSF). Using the PDB and PSF files, an initial parameter file with missing parameters is generated in FFTK. Each fragment is geometry optimized again using HF 6-31-G* level of theory in the program Gaussian 09 generated from FFTK [31,32].

Nonbonded Parameters

Using Gaussian 09, QM water interaction data for each water-accessible atom of the molecule were obtained. The information was used in FFTK to optimize the partial charges of each atom. Bounds were placed on partial charges if unrealistic deviations from the CGenFF are observed. All model compounds except model E were assigned partial charges by analogy to CGenFF. Model compound E had no close analogue in the CGenFF database and therefore the partial charges had to be determined using FFTK.

Internal Parameters

Through a Gaussian 09 QM Hessian calculation, bonds and angles can be optimized. The Hessian, a matrix of the second derivatives of the potential energy with respect to input coordinates, can be used to determine the change of energy when a distortion is applied to the bonds and angles [29]. The bonded objective function then measures the difference in QM and MM energies of distortion to a defined threshold objective value [29]. Dihedral parameters are generated by doing a QM potential energy scan (PES) for each missing dihedral at the Moller-Plesset (MP2)/6-31g* level of theory in Gaussian 09. Molecular mechanics (MM) potential energy scans are calculated in FFTK and were compared to QM PES scans produced by Gaussian 09. A simulated annealing method was used to fit the MM to the QM PES resulting in an initial set of parameters. The simulated annealing method is more effective at exploring parameter space in order to find a global minimum. Simulated annealing methods employ a metropolis sampling criteria and a temperature cooling scheme. Those parameters were then used in NAMD to perform a 1000 step conjugate gradient minimization [33]. The conjugate gradient method was used because a line search like steepest descent can take a long time to converge near the minimum where conjugate gradient methods use information from the previous step to converge faster. The resulting minimum geometry structure was used as input for a second round of optimization.

Infrared Spectrum of AR231453

N-(2-fluoro-4-methanesulfonylphenyl)-(6-[4-(3-isopropyl-[1,2,4]oxadiazol-5-yl)-piperidin-1-yl]-5-nitropyrimidin-4-yl)amine (AR231453) was obtained from Enzo Life Sciences

in solid phase. The infrared spectrum was recorded on a Bruker FT-IR with ATR attachment. A vacuum molecular dynamics simulation with the developed parameters was conducted in order to compute an IR spectrum. The model is first minimized 100 steps using steepest descent followed by 1000 steps of adopted basis Newton-Raphson minimization. The molecule was heated from 100K to 300K. 200 picoseconds of equilibration at 300K was performed on the molecule followed by 400 picoseconds of production run at 300K. The calculated infrared spectrum was obtained by importing the last 200 picoseconds of the production run into IR Spectral Density Calculator Plugin for VMD Version 1.9.2a42. The Fourier correction option was chosen. To produce agreement with the experimental spectrum, force constants were adjusted when necessary.

The normal modes are the independent vibrations of groups of atoms which are independent in the sense that they do not affect the other normal modes upon their excitation. To calculate the normal modes, a consideration of the quantum mechanical harmonic oscillator equation is necessary. The eigenvalues of the equation can be written in terms of the force constants in the Hessian matrix. Therefore, the normal modes are determined from diagonalization of the Hessian matrix. In order to test and validate the parameters, a QM normal mode analysis using HF (6-31G*) was conducted using Jaguar. A MM normal mode analysis using the developed parameters was obtained using the VIBRAN command in CHARMM. The QM and MM normal mode analyses were compared.

Molecular Dynamics Simulations of AR231453 in a Lipid Bilayer

A 60 x 60 square patch of lipid POPC was generated with the membrane plugin in VMD. AR231453 was oriented with the sulfone group pointed towards the charged head of the lipids and placed into the bilayer as shown in Figure 3. One molecule of A231453 were inserted into each leaflet of the bilayer and two lipids were removed out of the patch to make room for the compounds. The patch was then solvated with more water in case the compound could escape the bilayer by adding two slabs of water above and below the patch.

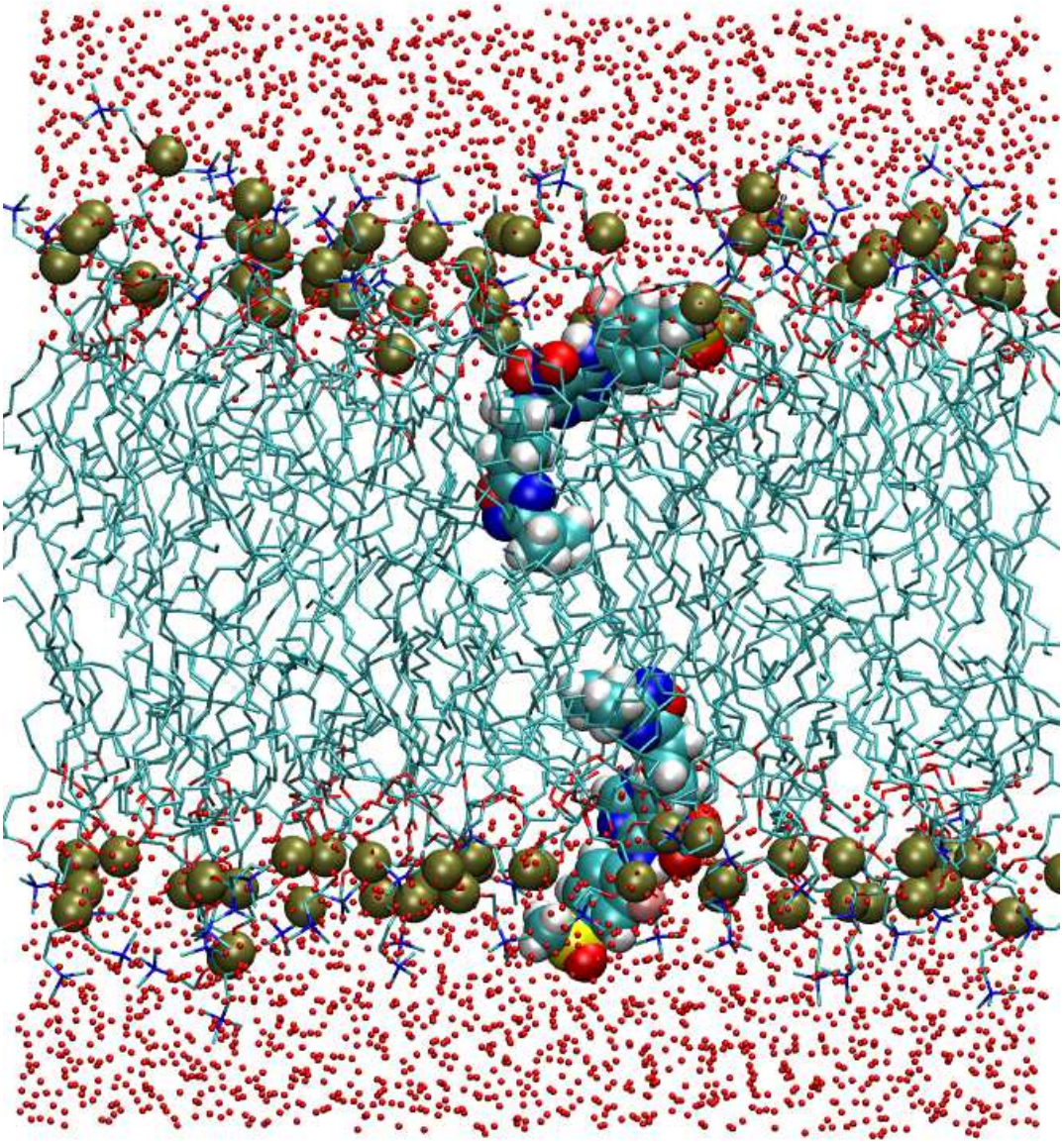


Figure 3: AR231453 in the lipid bilayer

A series of three constrained minimizations of 500 steps were conducted to let AR231453 settle into place. A final minimization of 20000 was then done with no constraints. The simulation cell needed to be warmed up to 310 K which is the physiological temperature. A

warmup was done by increasing the temperature from 0 K to 310 K in 10K increments of 20 ps each while holding everything except waters fixed allowed the waters to mix appropriately. The warmup was then repeated on the system without holding lipids and AR231453 fixed. A second warmup of 0.25 ns of NVT (constant particle number/constant volume/constant temperature) at 310 K was then ran.

Equilibration of the system for 0.25 ns was done with NPAT (constant particle number/constant pressure/constant area/constant temperature) instead of NVT to keep the area per lipid within a certain value. A second equilibration for 0.25 ns was conducted with the langevin piston decay changed from 100 to 500 and the dampening changed from 10 to 1. Production runs were then done for a total of 75 ns. A trajectory of 1500 frames (7.5 ns) was selected from the total production runs using a stride of 10 frames to avoid correlation between frames. The number of hydrogen bonds were calculated from the trajectory. The tilt angle of AR231453 was measured along the trajectory with the VMD plugin MEMBPLUGIN [34].

CHAPTER III

RESULTS AND DISCUSSION

Charge Parameters

Since the fragments chosen for parameter determination had close analogues in the CGenFF database, most of the charges for the atoms in the molecule were assigned by analogy. However, the fragment in Figure 2E required charge calculation because there was not a suitable analogue. The rules of assigning charges to aliphatic carbons and hydrogens not adjacent to heteroatoms for the CGenFF database have been followed. Aliphatic hydrogens in CHARMM are assigned a charge of 0.09 with carbon being adjusted to give a unit charge [35]. Aromatic carbon and hydrogens not adjacent to a heteroatom are assigned a charge of -0.115 and 0.115 [35]. Therefore, C13, C14, C15, and H11-H15 in the piperidine ring of the fragment in Figure 2E have been excluded from charge optimization and assigned charges according to these rules since they are aliphatic carbons and hydrogens. The remaining atoms in the fragment were allowed to vary but C12 and C16 in the piperidine ring and N2 and N3 in the nitro ring were constrained to match due to symmetry. The charge optimization required bounds to be placed to enforce realistic charges on certain atoms. The use of bounds is due to the charge-fitting problem being overdetermined from a large amount of QM data in FFTK [29].

Internal Parameters

Since many small molecules with heterocycle scaffolds are included in the CGenFF database, many of the bonds and angles in AR231453 are already determined. However, torsions generated by linking these groups need to be determined as shown in Figure 4. Known bond, angle, and torsion parameters were assigned by analogy to the CGenFF database. Unknown bond and angle parameters were determined using FFTK method of computing and comparing QM and MM potential energy surfaces through a hessian matrix and internal coordinates. Unknown torsion parameters were determined using FFTK method of bidirectional torsion scanning. The bidirectional torsion scanning overcomes the problem of high energy conformations such as torsions in a ring which cause the scan to terminate. A QM potential energy surface is generated for a set dihedral while the rest of the molecule is allowed to relax [29]. The net energy for the set dihedral contribution is then determined. The MM potential energy surface is then calculated with a geometry minimization done after to ensure that distortions from QM bonds and angles do not contaminate the fitting [29]. The torsions are then simultaneously fitted using simulated annealing. The unknown torsions were determined from the smaller model compound fragments (Figure 2A, 2D, 2E). Once those torsions were determined, the larger fragments (Figure 2B and 2C) were used to determine missing parameters at the connections between smaller fragments. The Figure 5 A-E show the fitting of the torsions for each fragment.

The fittings shown in Figure 5 A-E illustrate the improvement in the optimization from the initial MM energy to the final MM energy in comparison to the QM energy. A successful optimization is considered to have been achieved if the shape of the MM barriers and location of the minima coincide with the QM energy. Since high energy barriers are not likely to be important for most biological situations, the height of the MM barriers is generally not

considered to be important. An energy cutoff of 2 kcal/mol was used for most situations. In some cases, a higher energy cutoff gave a better line shape. Optimization was conducted by setting values of the periodicity and phase (0° and 180°) and adjusting them as necessary in relation to the force constant. In a few cases, adding a second or third dihedral function with a different periodicity gave a better line shape to the initial dihedral function. The reason this gives a better shape is that the dihedral can be thought of as a linear combination of sinusoidal functions. The final dihedral MM energy was determined to be complete when no further improvement to the line shape could be seen.

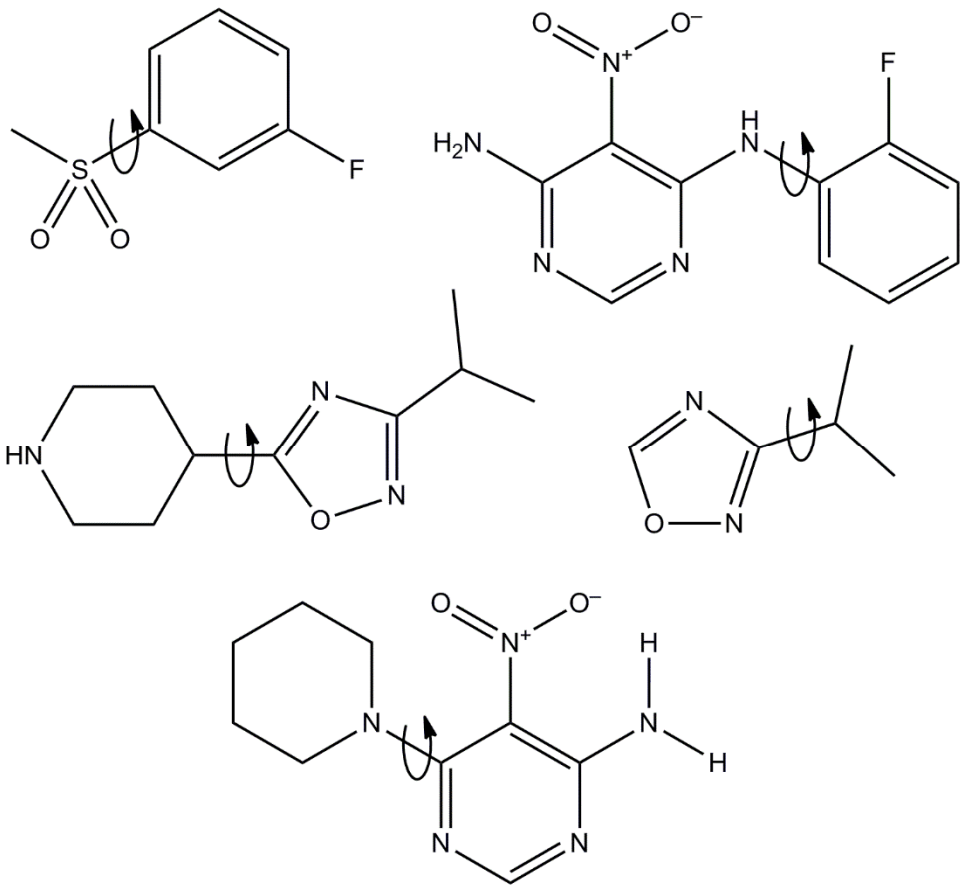
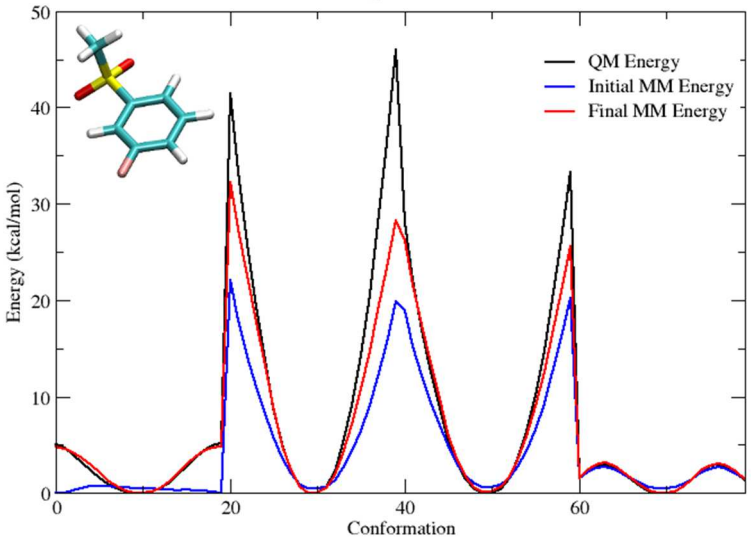
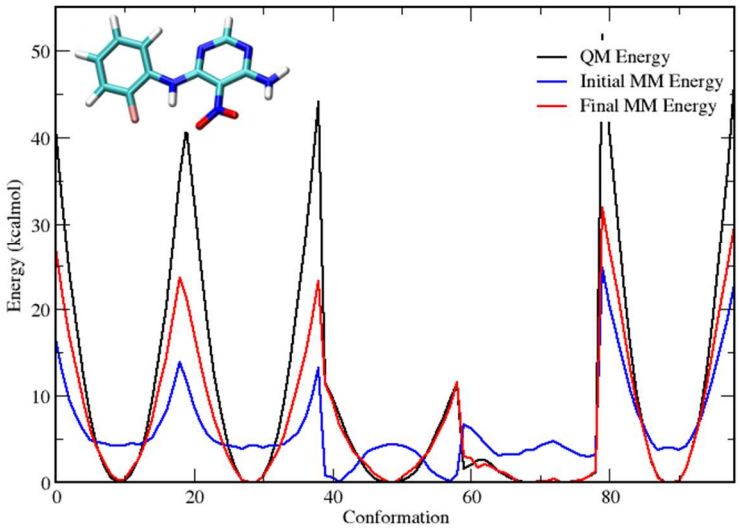


Figure 4: Linkage torsions that needed to be determined

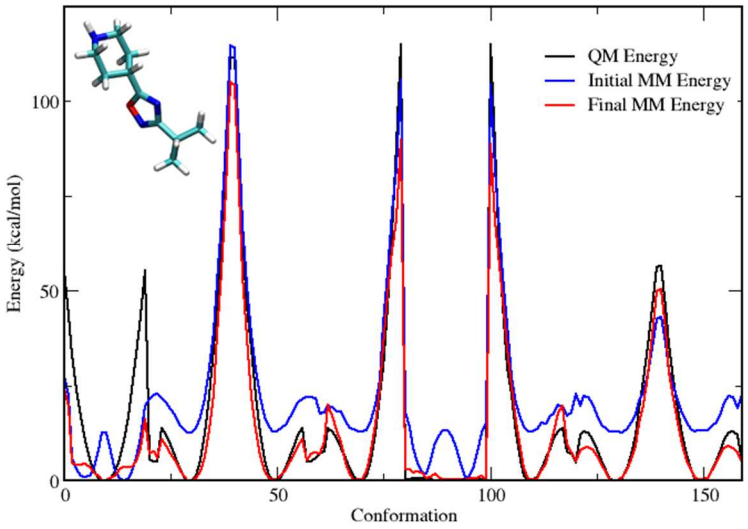
Fragment A



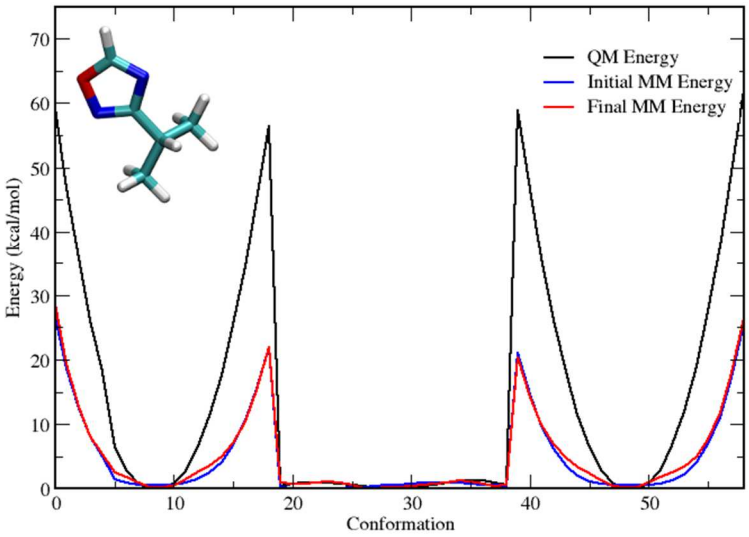
Fragment B



Fragment C



Fragment D



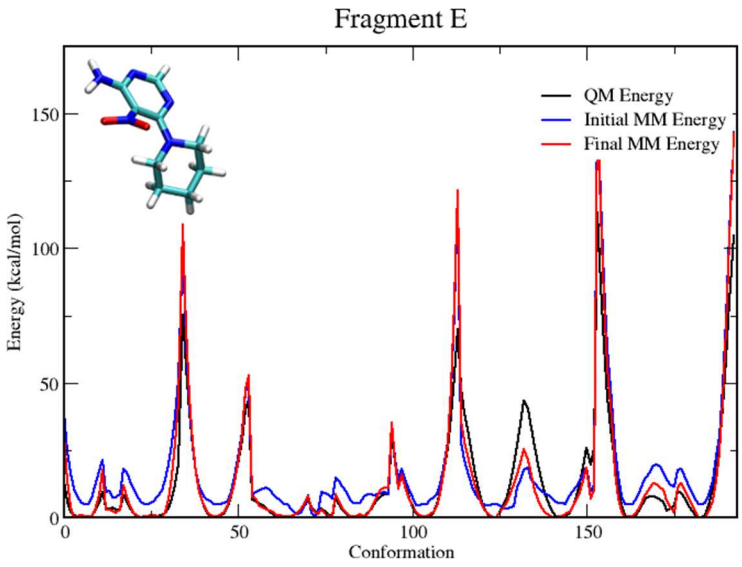


Figure 5: Torsions

Refinement

In order to test how the parameters for the fragments preform when integrated into the full molecule, lowest energy conformers were used as starting structures for a CHARMM minimization to ensure the reproduction of the QM structure of AR231453. The two lowest energy QM conformations were obtained using a conformational search. The lowest energy conformation was MM minimized and the structure was compared to the QM minimized structure. Differences in the structure were identified and adjustments to the dihedral parameters were conducted until another MM minimization produced a better overlay. The change of one parameter can alter the minimizations of both low energy conformations. Therefore, a MM minimization on the other second lowest conformer with the changed parameter was run to check that the change in dihedral parameter from the first run did not adversely affect the second structure. If the changed dihedral parameter did not adversely affect the second structure, then

the changed parameter was kept in the parameter file. The next dihedral parameter that needed adjustment was then identified and the same iterative approach by going back and forth between each conformer to ensure that one change does not negatively impact the other minimized structure was conducted until the entire MM structure had a decent overlay. Optimal structures were obtained by manually adjusting several dihedral force constants and phase values as shown in Table 2. The adjusted dihedrals are highlighted in red and K_ϕ and δ that were changed are bolded. Figure 6 shows the overlay of the QM global energy minimum conformer (green) and the CHARMM minimized structure (orange) after all improved dihedral changes had been conducted.

Type	Type	Type	Type	Kϕ	n	delta
CG2R61	NG311	CG2R64	CG2R61	3	2	180
CG2R61	NG311	CG2R64	NG2R62	2.454	2	180
CG2R61	CG2R61	NG311	CG2R64	1.976	2	180
CG2R61	CG2R61	SG3O2	CG331	1.303	2	0
CG2R61	SG3O2	CG331	HGA3	0.148	3	0
NG311	CG2R61	CG2R66	CG2R61	2.871	2	180
SG3O2	CG2R61	CG2R61	CG2R66	2.999	2	180
CG2R66	CG2R61	NG311	HGPAM1	0.366	2	**180**
CG2R66	CG2R61	NG311	CG2R64	1.288	2	**180**
NG311	CG2R61	CG2R66	FGR1	0.581	2	180
SG3O2	CG2R61	CG2R61	HGR62	1.29	2	180
CG2R64	CG2R61	CG2R64	NG311	2.237	2	180
NG2O1	CG2R61	CG2R64	NG311	0.979	2	180
NG311	CG2R64	NG2R62	CG2R64	2.994	2	180
NG2R62	CG2R64	CG2R61	NG2O1	**3**	2	**180**
NG2R62	CG2R64	CG2R61	CG2R64	**3**	2	180
NG2R62	CG2R64	NG301	CG321	**2.5**	2	180
CG321	CG321	NG301	CG2R64	1.753	3	**180**
HGA2	CG321	NG301	CG2R64	0.657	3	0
CG2R64	CG2R61	NG2O1	OG2N1	1.33	2	180
HGPAM1	NG311	CG2R64	CG2R61	2.998	2	180
CG2R61	CG2R64	NG301	CG321	2.744	2	180
NG301	CG2R64	CG2R61	CG2R64	2.997	2	180
HGPAM1	NG311	CG2R64	NG2R62	2.999	2	180

CG2R64	NG2R62	CG2R64	NG301	**3**	2	180
NG301	CG2R64	CG2R61	NG2O1	**3**	2	180
CG321	CG311	CG2R53	NG2R50	2.759	3	180
CG321	CG311	CG2R53	NG2R50	2.521	2	180
CG321	CG311	CG2R53	NG2R50	2.973	4	**180**
CG321	CG311	CG2R53	OG2R50	0.401	2	**180**
CG321	CG311	CG2R53	OG2R50	1.511	4	**180**
CG321	CG321	NG301	CG321	2.866	3	0
CG321	NG301	CG321	HGA2	0.06	3	0
CG321	CG321	CG311	CG2R53	**1.5**	3	**0**
NG301	CG321	CG321	HGA2	1.009	3	**0**
NG301	CG321	CG321	CG311	1.08	3	0
CG311	CG2R53	NG2R50	CG2R53	0.886	2	**180**
CG311	CG2R53	OG2R50	NG2R50	1.282	2	**180**
CG311	CG2R53	OG2R50	NG2R50	2.933	4	180
HGA2	CG321	CG311	CG2R53	0.045	3	180
NG2R50	CG2R53	CG311	CG331	8.979	4	**180**
NG2R50	CG2R53	CG311	CG331	0.39	6	**180**
NG2R50	CG2R53	CG311	HGA1	3	4	0
NG2R50	CG2R53	CG311	HGA1	0.485	6	180
CG2R53	CG311	CG331	HGA3	0.645	3	0
CG311	CG2R53	NG2R50	OG2R50	0.844	2	180
HGA1	CG311	CG2R53	OG2R50	1.583	2	**180**

Table 2: Dihedral Parameters

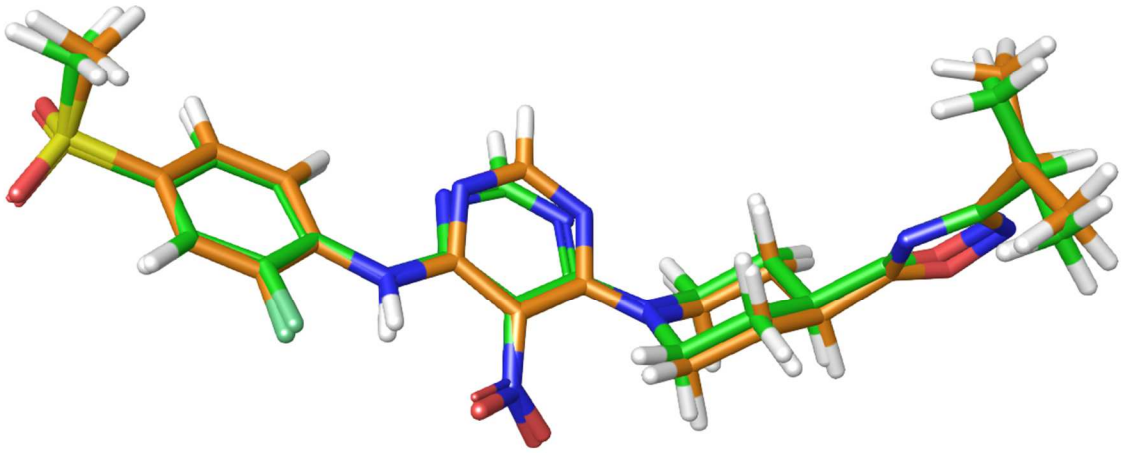


Figure 6: Overlay of QM global energy minimum conformer and CHARMM minimum structure

Force Constant Refinements

Comparison of the experimental IR spectrum and the calculated is shown in Figure 7. Bond force constants are shown in Table 3. The angle force constants were adjusted to reproduce the experimental spectrum and are shown in Table 4. The only force constants adjusted were of the newly developed parameters which are a subset of the total parameters required to describe AR231453. Table 4 has the adjusted angles highlighted in yellow and the adjusted K_θ are bolded. Only adjusting the subset of the angle parameters may result in slight differences between the calculated and experimental spectra. When comparing the experimental and IR spectra, the intensity of the peaks can be hard to match because of limitations in the fixed charge model to represent IR intensities [36]. Additionally, the experimental spectra is solid phase and the

calculated spectra is conducted *in vacuo* which could lead to some differences in peak positions. However, the shape of the spectra and the general position of the characteristic peaks are in place. Taking these factors into account, the spectra are in relatively good agreement.

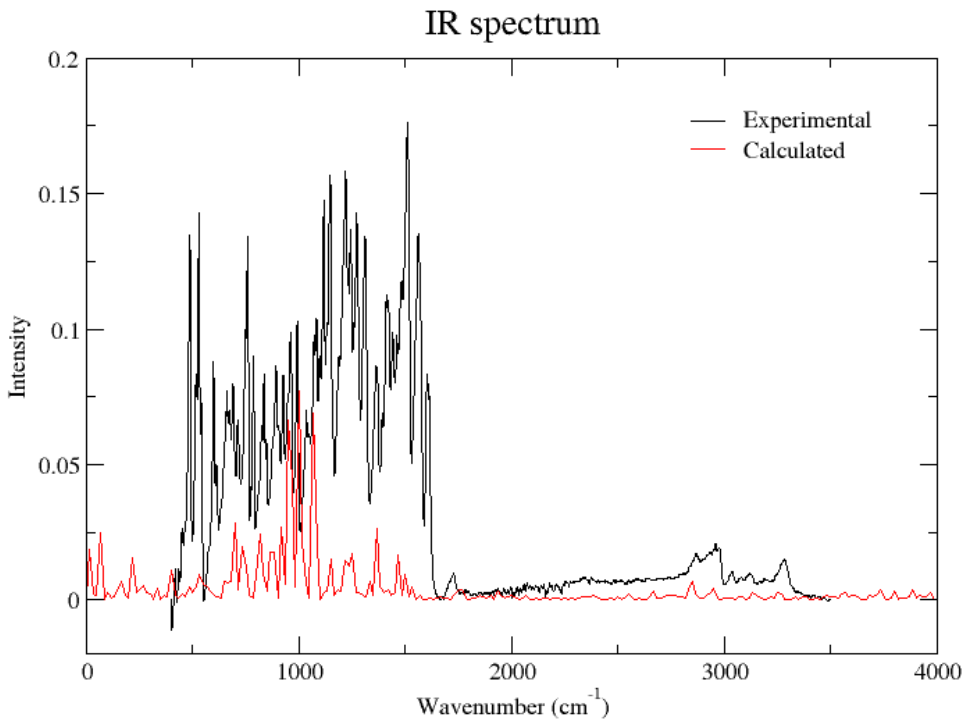


Figure 7: Infrared Spectrum of AR231453 overlaid with calculated spectrum

Type	Type	Kb	b0
NG311	CG2R64	509.84	1.349
CG2R64	NG301	411.53	1.3
CG321	NG301	214.547	1.375
CG311	CG2R53	279.114	1.462

Table 3: Bonded Parameters

Type	Type	Type	K_θ	θ_0
CG2R61	NG311	CG2R64	130.667	127.194
CG2R61	SG3O2	CG331	52.295	94.256
CG2R66	CG2R61	NG311	**40**	112.094
NG311	CG2R64	NG2R62	**100**	118.135
NG311	CG2R64	CG2R61	76.882	118.52
NG2R62	CG2R64	NG301	33.683	126.954
CG2R64	NG301	CG321	46.86	139.209
CG2R64	CG2R61	CG2R64	**60**	120.428
CG2R64	CG2R61	NG2O1	56.35	122.474
CG2R61	CG2R64	NG301	31.246	122.444
CG321	CG311	CG2R53	31.924	103.358
CG321	CG321	NG301	**85**	110.386
CG321	NG301	CG321	**350**	114.6
NG301	CG321	HGA2	**40**	105.574
CG311	CG2R53	OG2R50	26.346	117.395
CG311	CG2R53	NG2R50	38.523	122.983
CG2R53	CG311	HGA1	7.958	110.534
CG2R53	CG311	CG331	65.078	107.783
CG2R64	NG311	HGPAM1	**20**	117.925

Table 4: Angle Parameters

A Jaguar calculation of the normal mode frequencies at HF 6-31G* and a CHARMM calculation of the normal modes were compared to validate the parameters. The lowest energy conformer was used to conduct the analysis. The comparison is shown in Figure 8 and shows a direct correspondence between the QM and MM normal modes. The only significant QM and

MM frequency differences occurred near 1836 and 1847 which is due to N-H bending and the oxadiazole carbon-nitrogen stretching. The N-H bending mode was probably effected by differences in the MM and QM geometry of the nitro group. Adjusting the CG2R64 NG311 HGPAM1 K_θ in Table 4 brought the CHARMM normal mode closer to the QM normal mode as seen in Figure 8.The R^2 value was 0.9994 showing that the frequencies fit the regression model.

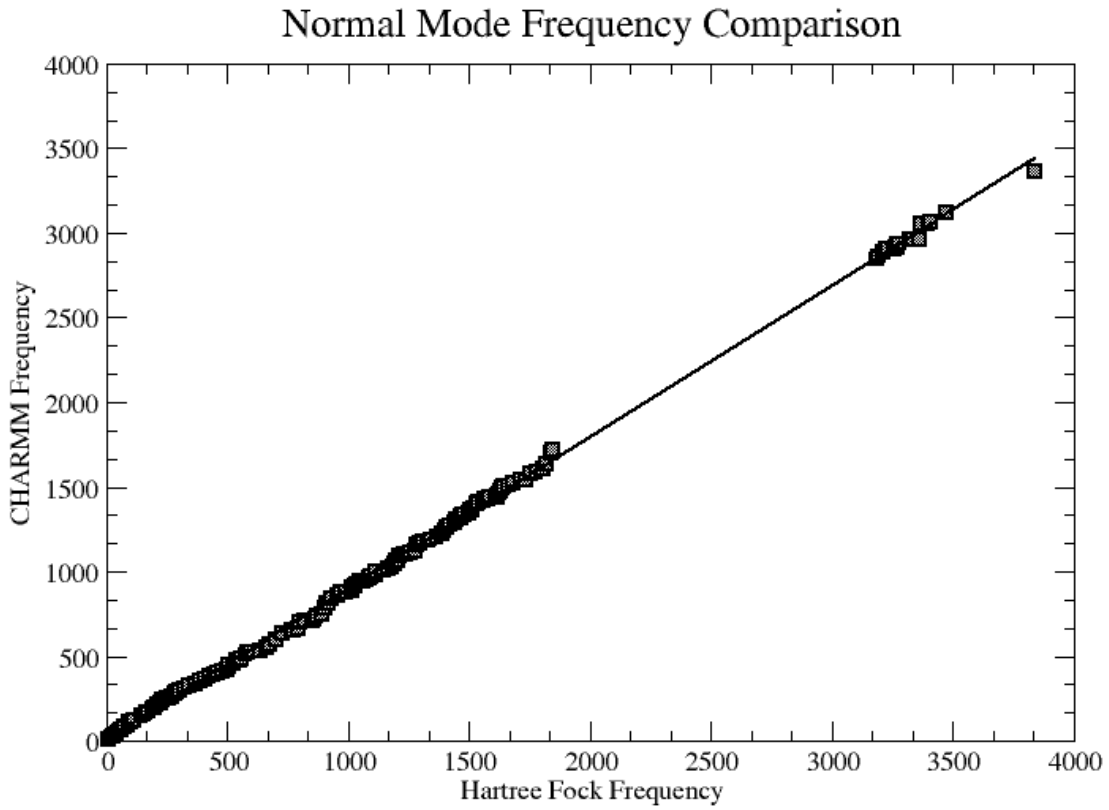


Figure 8: Normal Mode Analysis

Hydrogen Bonding and Tilt Angle from MD Trajectory

From the 75 ns of molecular dynamics trajectory, a frame was selected every ten frames to ensure that no biasing occurs. The number of hydrogen bonds was counted to see if the position of the AR231453 compound was buried in the bilayer or in the water. However, no correlation could be determined since the sulfone group of the compound was located around the hydrophilic ends of the lipids where waters could hydrogen bond to the compound. The compound was found to have 5 or less hydrogen bonds 88% of the time from the trajectory. The maximum number of hydrogen bonds counted in the trajectory was 11. The tilt angle of AR231453 was measured with respect to the z axis of the lipid bilayer to understand the orientation of AR231453. The orientation of the compound therefore ranges from vertical to parallel in relation to the lipid head groups. As can be seen from Figure 9 and 10, the tilt angle spends most of the time at a 50° to 60° angle with respect to the z axis.

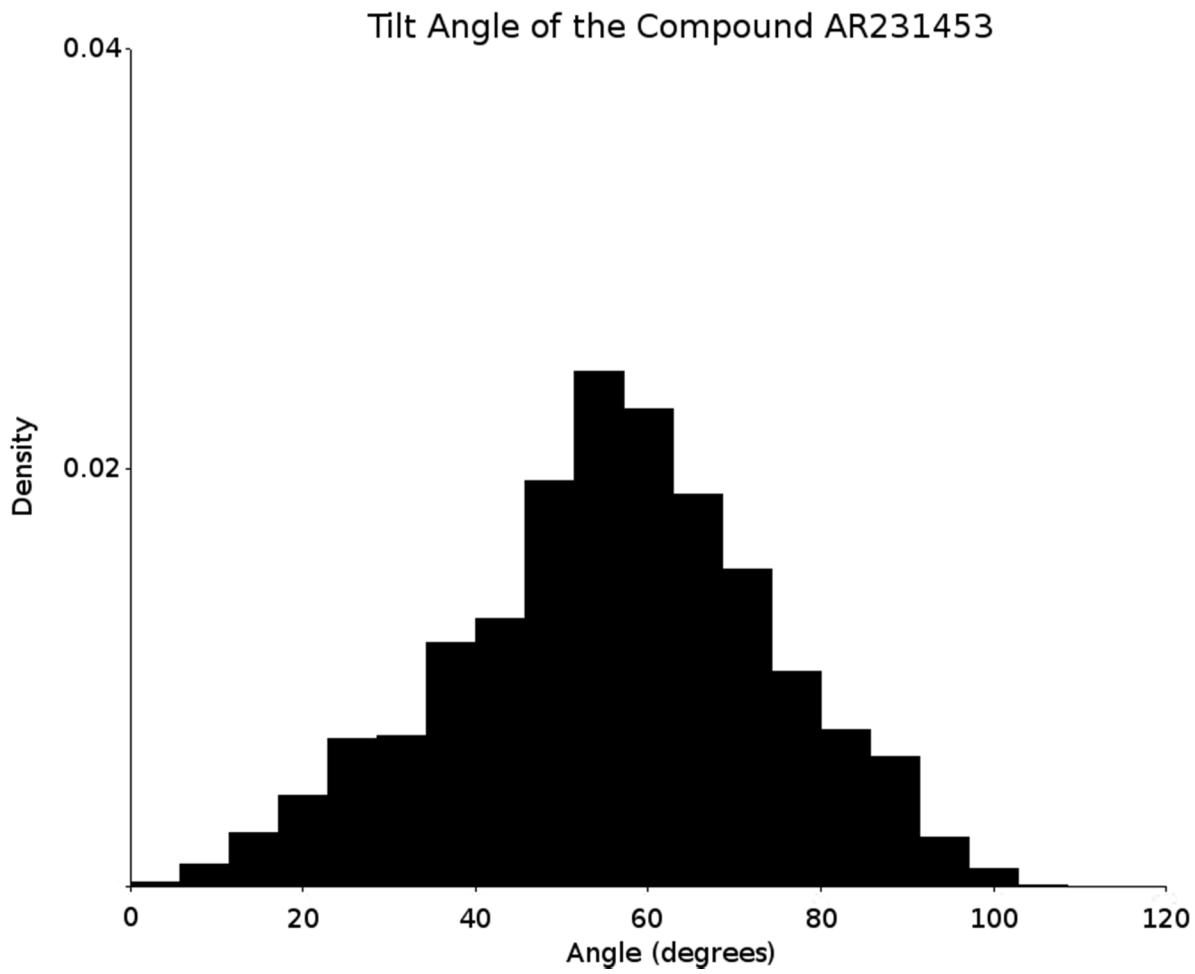


Figure 9: Histogram of Tilt Angles of the Compound AR231453

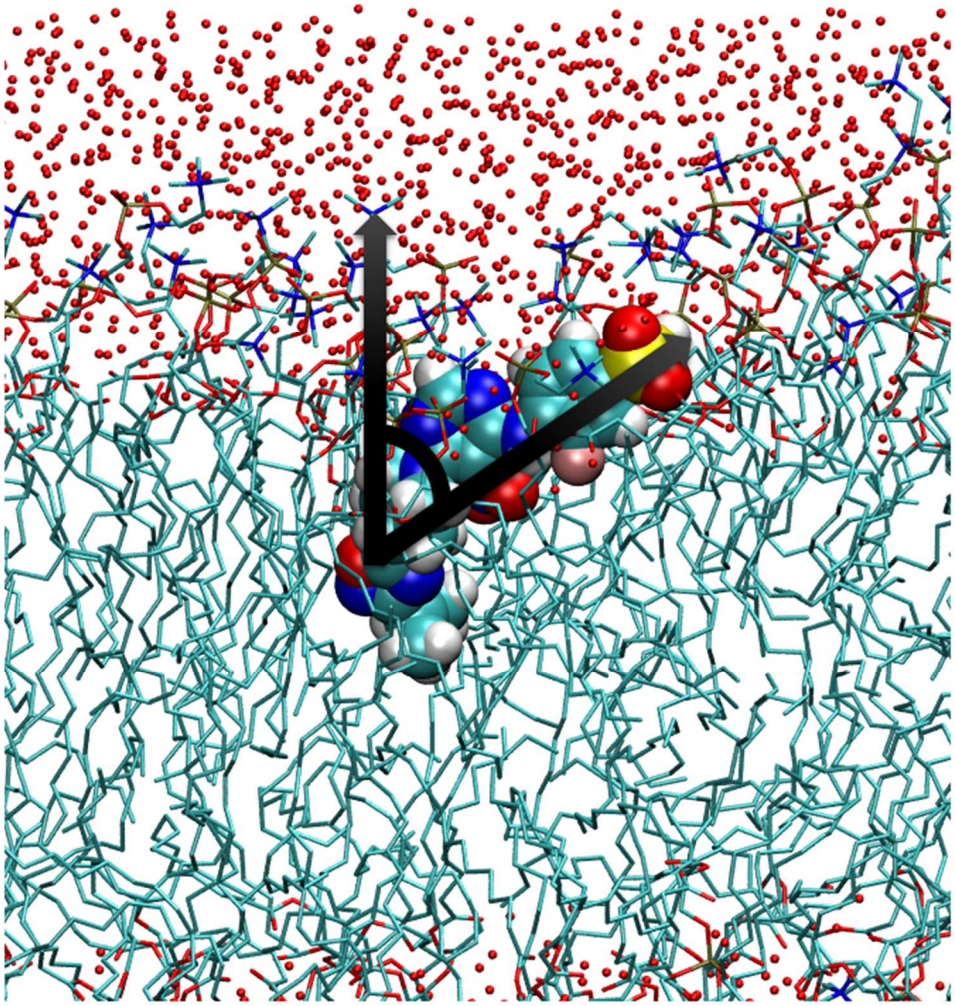


Figure 10: Tilt angle of AR231453 in lipid bilayer

CHAPTER IV

CONCLUSION

To summarize, parameters for AR231453 were developed using the FFTK and CGenFF database. The use of five model compounds facilitated the calculation of missing parameters. Refinement of the parameters after combining the fragments into the whole molecule using lowest energy conformers and proceeding to calculate minimum energy structures using CHARMM was accomplished. The QM and MM minimized structures using the refined parameters overlay well. After adjusting force constants, the calculated and experimental IR spectra are in relatively good agreement. The QM and MM normal mode analysis comparison shows a good correspondence. Along with the similarity of the calculated and experimental IR spectra, the normal mode analysis comparison give satisfactory evidence that the final parameters are able to give a molecular mechanics reproduction of the structure and dynamics of AR231453. Molecular dynamics simulations with the newly developed parameters were ran and gave insight into the behavior of the AR231453 compound in a lipid bilayer.

REFERENCES

1. Lander, E.S. et al. Nature 2001, 409, 860-921.

2. Venter, J.C. et al. Science 2001, 291, 1304-1351.

3. Fredricksson, R.; Hoglund, P.J.; Gloriam, D.E.I.; Lagerstrom, M.C.; Schioth, H.B. FEBS Lett. 2003, 554, 381-388.

4. Flower, D.R. Biochim. Biophys. Acta 1999, 1422, 207-234.

5. Palczewski, K. et al. Science 2000, 289, 739-745.

6. Overton, H.A.; Fyfe, M.C.T.; Reynet C. Br. J. Pharmacol. 2008, 153, S76-S81.

7. Bonini, J.A.; Borowsky, B.E.; Adham, N.; Boyle, N.; Thompson, T.O. DNA encoding SNORF25 receptor US patent 6,221,660-B1 2001.

8. Jones, R.M. et al. 1,2,3-trisubstituted aryl and heteroaryl derivatives as modulators of metabolism and the prophylaxis and treatment of disorders related thereto such as diabetes and hyperglycaemia International Patent Publication WO 065380 2004.

9. Soga, T. et al. Biochem. Biophys. Res. Commun. 2005, 326, 744-751.

10. Griffin, G. Methods for identification of modulators of OSGPR116 activity US patent 7,083,933-B1 2006.

11. Ohishi, T. et al. Method of screening remedy for diabetes Eur Patent Application EP

1338651-A1 2003.

12. Chu, Z. et al. Endocrinology 2007a, 148, 2601-2609.

13. Sakamoto, Y. et al. Biochem. Biophys. Res. Commun. 2006, 351, 474-480.

14. Furman, B.; Pyne, N.; Flatt, P.; O'Harte, F. J. Pham Pharmacol. 2004, 56, 1477-1492.

15. Lan, H. et al. Mice lacking GPR119 maintain metabolic homeostasis. Proceedings of Keystone Symposium, Keystone, Colorado, USA., January 14-19 2007, abstract 253.

16. Semple, G. et al. J. Med. Chem. 2008, 51, 5172-5175.

17. Chu, Z. et al. Endocrinology 2008, 149, 2038-2047.

18. Gao, J.; Tian, L.; Weng, G.; O'Brien, T.D.; Luo, J.; Guo, Z. Transplant. Proc. 2011, 43, 3217-3220.

19. Englestoft, M.S.; Norn, C.; Hauge, M.; Holliday, N.D.; Elster, L.; Lehmann, J.; Jones, R.M.; Frimurer, T.M.; Schwartz, T.W. Br. J. Pharmacol. 2014, 171, 5774-5789.

20. Vanommeslaeghe, K.; Guvench, O.; MacKerell, A.D. Curr. Pharm. Des. 2014, 20, 3281-3292.

21. MacKerell, A.D. J. Comput. Chem. 2004, 25, 1584-1604.

22. Brooks, B. R. et al. J. Comp. Chem. 2009, 30, 1545-1615.

23. Case, D.A. et al. 2016, AMBER 2016, University of California, San Francisco.

24. Christen, M. et al. J. Comput. Chem. 2005, 26, 1719-1751.

25. Jorgensen, W.L.; Tirado-Rives, J. J. Am. Chem. Soc. 1988, 110(6), 1657-1666.

26. Vanommeslaeghe, K. et al. J. Comput. Chem. 2010, 31, 671-690.

27. Vanommeslaeghe, K.; MacKerell, A.D. J. Chem. Inf. Model. 2012, 52, 3144-3154.

28. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A.D. J. Chem. Inf. Model. 2012, 52, 3155-3168.

29. Mayne, C.G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J.C. J. Comput. Chem. 2013, 34, 2757-2770.

30. Schrödinger Release 2016-3: MS Jaguar, Schrödinger, LLC, New York, NY, 2016.

31. Humphrey, W.; Dalke, A.; Schulten, K. J. Molec. Graphics 1996, 14, 33-38.

32. Gaussian 09, Revision E.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

33. Phillips, J.C. et al. J. Comput. Chem. 2005, 26, 1781-1802.

34. Guixà-González, R.; Rodriguez-Espigares, I.; Ramírez-Anguita, J.M.; Carrió-Gaspar, P.; Martinez-Seara, H.; Giorgino, T.; Selent, J. Bioinformatics 2014, 30, 1478-80.

35. MacKerell, A.D. Atomistic Models and Force Fields. In Computational Biochemistry and Biophysics, Becker, O.; MacKerrel, A. D.; Roux, B.; Watanabe, M., Eds.; Marcel Dekker Inc, New York, 2001.

36. Iliff, H.A.; Lynch, D. L.; Kotsikorou, E.; Reggio, P. H. J. Comput. Chem. 2011, 32, 2119-2126.

BIOGRAPHICAL SKETCH

John Eli Hamilton attended the University of Texas- Rio Grande Valley (formerly Pan American) where he earned degrees in chemistry and mathematics. He earned a MS in chemistry in May 2017. He can be contacted by email at john.hamilton01@utrgv.edu.