UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Applications and Development of the MMPBSA Method for Rational Drug Design

Permalink
https://escholarship.org/uc/item/9rk3b66d

Author
Greene, D'Artagnan Gene

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Applications and Development of the MMPBSA Method for Rational Drug Design

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

D'Artagnan Greene

Dissertation Committee:
Professor Ray Luo, Chair
Professor Celia Goulding
Professor Sheryl Tsai

2019

Chapters 1, 2, & 3 © American Chemical Society
All Other Materials © 2019 D'Artagnan Greene

DEDICATION

For my parents.

TABLE OF CONTENTS

Page

LIST OF FIGURES vi

LIST OF TABLES ix

ACKNOWLEDGMENTS xi

CURRICULUM VITAE xii

ABSTRACT OF THE DISSERTATION xiv

INTRODUCTION 1

REFERENCES 8

CHAPTER 1: Computational Analysis for the Rational Design of Anti-Amyloid Beta (Aβ) Antibodies 17

1.1 INTRODUCTION 18

1.2 METHODS 22
- 1.2.1 Computational docking of amino acid residues to anti-Aβ antibodies 22
- 1.2.2 Molecular dynamics simulations 24
- 1.2.3 MMPBSA binding free energy calculations 25

1.3 RESULTS 26
- 1.3.1 Computational docking of amino acid residues to anti-Aβ antibodies 26
- 1.3.2 MD method validation 30
- 1.3.3 The importance of phenylalanine to the stable binding of Aβ_{2-7} to PFA1 33
- 1.3.4 Analysis of gantenerumab and crenezumab binding to multiple Aβ epitopes 37
- 1.3.5 Predicting the binding affinity between pE3-Aβ_{3-8} and PFA1 mutants 45

1.4 DISCUSSION 53

1.5 CONCLUSIONS 58

1.6 REFERENCES 59

CHAPTER 2: Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor 69

2.1 INTRODUCTION 70

2.2 METHODS 75
2.2.1 Preparation of the $P2Y_{12}R$ complex structures 75
2.2.2 Preparation of the lipid membrane model 76
2.2.3 MD simulation protocol 76
2.2.4 Binding free energy calculations 78
2.2.5 Estimation of the free energy penalty upon the loss of Mg^{2+} 80
2.2.6 Additional computational details 81

2.3 RESULTS AND DISCUSSION 82
2.3.1 Impact of the non-polar solvation model 82
2.3.2 Impact of the protein dielectric constant 84
2.3.3 Impact of the membrane dielectric constant 86
2.3.4 Effect of the magnesium correction 88

2.4 CONCLUDING REMARKS AND FUTURE DIRECTIONS 92

2.5 REFERENCES 94

CHAPTER 3: Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies 112

3.1 INTRODUCTION 113

3.2 METHODS 119
3.2.1 Preparation of amino acid sidechains 119
3.2.2 Z-restraint, de-charging MD simulations in explicit solvent 120
3.2.3 BAR free energy calculations in explicit solvent 123
3.2.4 PBSA-BAR free energy calculations in implicit solvent 124
3.2.5 RMSD calculations 125
3.2.6 Implementation of the dielectric profile into PBSA 126
3.2.7 The natural abundance of each amino acid in different membrane regions 127
3.2.8 MMPBSA calculations for the $P2Y_{12}R$ and M2R/M3R test systems 129

3.3 RESULTS AND DISCUSSION 131
3.3.1 BAR results for de-charging amino acid side chains in an explicit membrane 131

3.3.2 PBSA results for de-charging amino acid side chains in an implicit membrane 135
3.3.3 Implementation of the heterogeneous dielectric membrane into Amber 139
3.3.4 MMPBSA results for the $P2Y_{12}R$ and M2R/M3R test systems 141

3.4 CONCLUSION 150

3.5 REFERENCES 152

CHAPTER 4: Calibration of the Non-Polar Terms for the Heterogeneous Dielectric Implicit Membrane Model 166

4.1 INTRODUCTION 166

4.2 METHODS 170

4.3 RESULTS AND DISCUSSION 172

4.4 CONCLUSION 176

4.5 REFERENCES 177

SUMMARY AND CONCLUSIONS 183

APPENDIX A: Supporting Information for Computational Analysis for the Rational Design of Anti-Amyloid Beta (Aβ) Antibodies 189

APPENDIX B: Supporting Information for Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor 199

APPENDIX C: Supporting Information for Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies 202

LIST OF FIGURES

Page

FIGURE 1.1 Method validation for MMPBSA free energy calculations of various Aβ peptides bound to the antibodies PFA1 and PFA2 31

FIGURE 1.2 Comparison of the binding pose for Aβ_{2-7} peptide variants bound to PFA1 35

FIGURE 1.3 Electrostatic contacts for Aβ_{1-8} bound to PFA1 37

FIGURE 1.4 Gantenerumab bound to both N-terminal and central Aβ peptides 40

FIGURE 1.5 Crenezumab bound to both central and N-terminal Aβ peptides 43

FIGURE 1.6 Snapshots from the MD trajectories of pE3-Aβ_{3-8} bound to wild type and mutant forms of PFA1 49

FIGURE 1.7 Snapshots from the MD trajectories of Aβ_{1-8} and Aβ_{2-7} bound to wild type and mutant forms of PFA1 51

FIGURE 2.1 Dielectric constant regions in the P2Y$_{12}$R system 73

FIGURE 2.2 Thermodynamic cycle of the MMPBSA method for a membrane protein-ligand system 78

FIGURE 2.3 Parameter optimization for the non-polar solvation model, inp (epsin=20, epsmem=4) 82

FIGURE 2.4 Parameter optimization for the protein dielectric constant, epsin (inp=2, epsmem=4) 84

FIGURE 2.5 Parameter optimization for the membrane dielectric constant, epsmem (inp=2, epsin=20) 86

FIGURE 2.6 Absolute binding free energy (ΔG) correlation plots (inp=2, epsin=20, and epsmem=4) 88

FIGURE 2.7 Relative binding free energy ($\Delta\Delta G$) correlation plots (inp=2, epsin=20, and epsmem=4) 89

FIGURE 2.8A Cross-sectional comparison of the ATP binding site for $P2Y_{12}R$ 91

FIGURE 2.8B Cross-sectional comparison of the ATP binding site for GRK1 91

FIGURE 2.8C Cross-sectional comparison of the ATP binding site for FlaH 91

FIGURE 3.1 The evolution of a heterogeneous dielectric implicit membrane model 115

FIGURE 3.2A Model of the antagonist-bound $P2Y_{12}R$ system used in our MD simulations (side view) 117

FIGURE 3.2B Model of the antagonist-bound $P2Y_{12}R$ system used in our MD simulations (top down view) 117

FIGURE 3.3A Model of the antagonist-bound M2R and M3R systems used in our MD simulations (M2R side view) 118

FIGURE 3.3B Model of the antagonist-bound M2R and M3R systems used in our MD simulations (M2R top down view) 118

FIGURE 3.3C Model of the antagonist-bound M2R and M3R systems used in our MD simulations (M3R side view) 118

FIGURE 3.3D Model of the antagonist-bound M2R and M3R systems used in our MD simulations (M3R top down view) 118

FIGURE 3.4 The difference in ΔG_{total} values for de-charging an amino acid side chain at height z 132

FIGURE 3.5A RMSD between ΔG_{total} values calculated using PBSA-BAR and Explicit BAR for various dielectric constants at different z-values (z = 0.0 Å) 135

FIGURE 3.5B RMSD between ΔG_{total} values calculated using PBSA-BAR and Explicit BAR for various dielectric constants at different z-values (z = 5.0 Å) 135

FIGURE 3.5C RMSD between ΔG_{total} values calculated using PBSA-BAR and Explicit BAR for various dielectric constants at different z-values (z = 10.0 Å) 135

FIGURE 3.5D RMSD between ΔG_{total} values calculated using

PBSA-BAR and Explicit BAR for various dielectric constants at different z-values (z = 15.0 Å) 135

FIGURE 3.5E RMSD between ΔG_{total} values calculated using PBSA-BAR and Explicit BAR for various dielectric constants at different z-values (z = 20.0 Å) 135

FIGURE 3.6 PCHIP and Spline fittings of a z-depth dependent membrane dielectric profile 140

FIGURE 3.7 Binding free energy correlation plots for the $P2Y_{12}R$ test system 143

FIGURE 3.8 Binding free energy correlation plots for the M2R/M3R test system 145

FIGURE 3.9A Snapshots of compound 6B bound to the M2R and M3R receptors from our MD simulations (M2R, just before the heating step) 147

FIGURE 3.9B Snapshots of compound 6B bound to the M2R and M3R receptors from our MD simulations (M2R, last frame of the production run) 147

FIGURE 3.9C Snapshots of compound 6B bound to the M2R and M3R receptors from our MD simulations (M3R, just before the heating step) 147

FIGURE 3.9D Snapshots of compound 6B bound to the M2R and M3R receptors from our MD simulations (M3R, last frame of the production run) 147

FIGURE 4.1 Optimal α scaling parameters as a function of z 175

LIST OF TABLES

Page

TABLE 1.1 Top binding affinities for single amino acid residues bound to the antigen-combining site of anti-Aβ antibodies 27

TABLE 1.2 Number of docked structures found at the antigen-combing site of anti-Aβ antibodies 29

TABLE 1.3 Method validation for MMPBSA calculations of Aβ peptides bound to PFA1 and PFA2 32

TABLE 1.4 The importance of phenylalanine in the binding of Aβ_{2-7} to PFA1 34

TABLE 1.5 MMPBSA binding free energy results for N-terminal and central Aβ peptides bound to gantenerumab and crenezumab 38

TABLE 1.6 MMPBSA binding free energy results for Aβ_{1-8}, Aβ_{2-7}, and pE3-Aβ_{3-8} bound to the wildtype and several mutant forms of PFA1 47

TABLE 2.1 Effect of the non-polar solvation model on absolute (ΔG) and relative ($\Delta\Delta G$) binding affinities 83

TABLE 2.2 Effect of the protein dielectric constant on absolute (ΔG) and relative ($\Delta\Delta G$) binding affinities 85

TABLE 2.3 Effect of the membrane dielectric constant on absolute (ΔG) and relative ($\Delta\Delta G$) binding affinities 87

TABLE 3.1 Optimum values of the dielectric constant within the implicit membrane 136

TABLE 3.2 RMSD results for PBSA-BAR calculations using the heterogeneous dielectric membrane model 138

TABLE 3.3 Results of MMPBSA calculations for antagonist ligands bound to P2Y$_{12}$R 142

TABLE 3.4 Results of MMPBSA calculations for antagonist ligands bound to M2R and M3R 144

TABLE 4.1 Convergence data for $\Delta G_{explicit,solv}$ 172

TABLE 4.2 RMSD values for a given value of the scaling parameter α using the classical non-polar implicit membrane model 173

TABLE 4.3 RMSD values for a given value of the scaling parameter α using the modern non-polar implicit membrane model 174

ACKNOWLEDGMENTS

I would like to thank my committee chair Ray Luo for giving me an opportunity to do research here at the University of California, Irvine and for providing me with guidance and support on all my projects. I would also like to thank my committee members, Professors Celia Goulding and Sheryl Tsai for their additional guidance and support.

For our study on anti-Aβ antibodies, I would like to thank Drs. Hartmut Luecke, Charles Glabe, and Hiromi Arai for both inspiring this project and for providing us with helpful advice that led to the improvement of our computational docking protocol. I would also like to thank Theodora Po, who assisted in the model building and analysis of gantenerumab and crenezumab and also helped produce the majority of the figures that appear throughout the study. Additionally, I want to thank both Jennifer Pan, who worked on modeling and analyzing the PFA1 mutations, and Tanya Tabibian, who built and analyzed the initial PFA1 models that were used in our method validation, for their contributions to this project.

For our study on the analysis of $P2Y_{12}R$ using the single dielectric implicit membrane model, I would like to thank Wesley Botello-Smith and Alec Follmer for their initial work that helped build the foundation for this study, Li Xiao for his concurrent work on membrane proteins that had an influence on this project, and Eleftherios Lambros for his help parameterizing both nucleotide agonists and for producing the striking images seen in Figure 2.8.

For our study on the development and application of the heterogeneous dielectric implicit membrane model, I would like to thank Ruxi Qui for implementing the PCHIP and spline fittings into Amber 18, Remy Nguyen for her constant assistance with data analysis and also for producing many beautiful figures that appear throughout the study, and Tianyin Qiu for his help with building the M2R and M3R models and for producing the many convergence plots for these two systems.

I would like to thank the American Chemical Society for permission to include Chapters 1, 2, and 3 of my dissertation, which were originally published in the Journal of Physical Chemistry B (Chapters 1 and 2) and in the Journal of Chemical Information and Modeling (Chapter 3). Financial support was provided for this work by the National Institute of Health/NIGMS (GM093040 & GM079383).

CURRICULUM VITAE

Educational Background:

University of California, Irvine PhD, Biological Sciences	September 2019
California State University, Northridge Master's degree, Physics	August 2014
California State University, Northridge Bachelor of Science, Biochemistry	May 2009

Publications:

Greene, D.; Qi, R.; Nguyen R.; Qiu, T.; Luo R. Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies. *J. Chem. Inf. Model.* **2019**, *59*, 3041-3056.

Greene, D.; Po, T.; Pan J.; Tabibian T.; Luo R. Computational Analysis for the Rational Design of Anti-Amyloid Beta (Aβ) Antibodies. *J. Phys. Chem. B*. **2018**, *122*, 4521-4536.

Wang, C.; **Greene, D.**; Xiao, L.; Qi, R.; Luo R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci*. **2018**, *4*, 87.

Xiao, L.; Diao, J.; **Greene, D.**; Wang, J.; Luo R. A Continuum Poisson-Boltzmann Model for Membrane Channel Proteins. *J. Chem. Theory Comput*. **2017**, *13*, 3398-3412.

Greene, D.; Botello-Smith, W. M.; Follmer, A.; Xiao, L.; Lambros, E.; Luo R. Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor. *J. Phys. Chem. B*. **2016**, *120*, 12293-12304.

Greene, D.; Shiferaw, Y. Approximate analytical solutions for excitation and propagation in cardiac tissue. *Phys. Rev. E*. **2015**, *91*, 042719.

Research Experience:

University of California, Irvine
Graduate Research 2016 - 2019

Worked on molecular dynamics simulations and other computational projects in Dr. Ray Luo's Computational Biology Research Laboratory.

California State University, Northridge
Graduate Research 2012 - 2014

Carried out computational research on the nature of cardiac electrical excitation in Dr. Yohannes Shiferaw's Cardiac Research Laboratory.

Teaching Experience:

University of California, Irvine
Teaching Assistant 2015 - 2019
Courses: Bio Sci 98, Bio Sci 99, Mol Bio 204, Bio Sci M114, Bio Sci M116L, Bio Sci M118L, and Bio Sci M126

California State University, Northridge
Physics Learning Assistant 2012 - 2013
Courses: Physics 100A

California State University, Northridge
Teaching Assistant 2010 - 2012
Courses: Physics 100AL, Physics 220AL, and Physics 220BL

California State University, Northridge
Supplemental Instruction (SI) Leader 2007 - 2008
Course: Biology 360 SI

California State University, Northridge
Tutor, Subject Area Tutor Lab in the Learning Resource Center 2006 - 2013
Courses: All undergraduate biology, chemistry, math, and physics courses

Awards:

Molecular Biology & Biochemistry 2018 Departmental Retreat: Oral Presentation Award Graduate Student, 1st place	2017-2018
Donald E. Bianchi Graduate Student Award for Outstanding Achievement in Research	2013-2014
C.Y. Liang Outstanding Graduate Student Award	2013-2014
The Outstanding Physics Learning Assistant Award	2012-2013
The Michael G. Dickson Memorial Award	2011-2012
Summer Research Grant	2011-2012
The Best Physics Graduate Teaching Assistant Award	2010-2011
CRC PRESS Chemistry Achievement Award: Outstanding Freshman Chemistry Award	2005-2006

ABSTRACT OF THE DISSERTATION

Applications and Development of the MMPBSA Method for Rational Drug Design

By

D'Artagnan Greene

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2019

Professor Ray Luo Irvine, Chair

The growing cost of new drugs has become a concern for the biopharmaceutical industry, which depends on innovation to sustain itself. As a result, computational methods have been increasingly implemented into the drug design workflow in an effort to reduce the cost of finding new lead candidates. Here, we focus on several applications and the development of the Molecular Mechanics Poisson-Boltzmann (MMPBSA) method for its use in rational drug design efforts. Chapter 1 demonstrates the use of computational methods in the analysis and design of anti-Aβ antibodies for their prospective use in the treatment of Alzheimer's Disease. Chapter 2 applies our single dielectric implicit membrane model to MMPBSA calculations of the membrane-bound human purinergic platelet receptor, a prominent target for treating myocardial infarction and stroke. Chapter 3 documents the development, implementation, and application of a new heterogeneous dielectric implicit membrane model for MMPBSA calculations. Chapter 4 shows the validity of our method to parameterize the non-polar terms in a depth dependent manner within our implicit membrane model. This work as a whole demonstrates both the present utility and ongoing improvement of the MMPBSA method for its use in the rational design of new drugs.

INTRODUCTION

The biopharmaceutical industry has become a leading source of economic activity in the United States, accounting for more than 1.3 trillion in total economic output in 2015 while being a generator of high quality jobs, with its 800,000 workers receiving an average compensation of $129,527 in comparison to an average of $58,603 for workers in all industries[1]. What sustains this success is innovation; in particular, a primary focus is the ongoing development of new drugs to treat various ailments that do not currently have an effective treatment strategy.

The growing cost of new drug discovery has become a concern as of late. It was recently estimated that bringing a single new drug to approval costs approximately $2.8 billion dollars, and the total capitalized costs for new drug development have increased above general price inflation at an 8.5% annual rate[2]. Furthermore, rising drug costs and patent expirations have led to an increase in prescriptions of generic versions of popular old drugs. For example, prescriptions of generic drugs increased from 51% in 2002 to 67% in 2007, resulting in $113 billion lost in sales to generic substitution[3]. If these trends remain, it has been estimated that large pharmaceutical companies will only be able to recoup 26 cents with revenue generated from new products for every dollar lost in declining product revenue due to generic substitution[4-5].

In light of these issues, new approaches are being taken that are aimed at improving the efficiency of discovering new drugs[6]. Discovering new drugs has traditionally been an empirical process that depended on either luck, knowledge of a well characterized target, or carrying out expensive, brute-force high throughput screening searches for lead compounds[7-9]. Recently, computational methods have been implemented into the drug design workflow with the aim to remove some of the guesswork associated with discovering promising new lead compounds. In this approach, rational insights gained from computational analysis direct drug development towards more promising targets, thereby reducing the cost of both lead generation and lead optimization[10-11]. Several of the computational methods employed to accomplish this are collectively referred to as structure-based drug design (SBDD) methods. SBDD methods make use of the knowledge acquired from computationally analyzing the interactions between a potential drug candidate (referred to more generally as a "ligand" molecule) and the three-dimensional structure of a protein macromolecule, which is obtained from such experimental methods as x-ray crystallography[12], nuclear magnetic resonance (NMR) spectroscopy[13], or cryo-electron microscopy (cryo-EM)[14-15].

A central application in SBDD is carrying out a virtual screen of potential drug compounds to assist in lead generation[10-11, 16-19]. In a virtual screening method, a library of potential compounds, or fragments of compounds, are computationally docked to the known active site of a protein in an attempt to identify a new lead compound with a specific bioactivity against a certain target. In most common docking protocols, the ligand is allowed to be conformationally flexible to sample various orientations while the active site of the receptor is a static structure. This limitation was originally due to the increased computational expense of sampling conformations for the larger protein receptor. More recently, dynamical docking using molecular dynamics (MD) simulations is being employed to allow both the ligand and the receptor to sample conformational space simultaneously[20-24]. Virtual Screening can be performed to narrow the initial chemical search space from a large library of potential compounds down to a few promising lead compounds that can be tested for their effectiveness experimentally. Promising compounds can undergo further refinement through a process called lead optimization, where small changes are made to the compound of interest with the usual goal of optimizing its binding affinity or specificity for a selected target[10-11].

A method to analyze and rank the binding prospects of compounds is central to computational screening and lead optimization methods. For lead generation, large-scale virtual screening protocols typically use fast, empirical scoring methods to rank compounds[25-27]. These scoring methods can be relied upon to screen out compounds with obvious structural incompatibilities to the active site of interest. However, one drawback is they typically do not reliably reproduce the correct rank order of similar compounds. For lead optimization, more rigorous physics-based binding free energy calculations are typically used in place of the empirical scoring functions that are used in standard docking studies. The most prominent of these are thermodynamic integration (TI) and free energy perturbation (FEP) methods[28-29]. These methods are based on molecular dynamics (MD) simulations. In TI, the binding free energy is obtained by calculating the difference in energy brought about by slowly varying a suitably chosen parameter. A series of sequential MD simulations are carried out to sample each small change in the simulation due to the change in the parameter, and at the end of the process, the energy can be recovered by integrating over the full set of data. While such methods are considered to be the most accurate methods available for computational free energy calculations, they are also the most time-consuming, making them somewhat impractical to apply to any large set of ligands or to a large set of proposed structural modifications in a typical drug design endeavor[28-29].

In between these two extremes are several approximation methods which speed up the calculation time by making simplifications to the calculation of binding free energies at the expense of accuracy. In the following, we focus on one common binding free energy calculation of this type, which is the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method[28-36]. The overall objective of the MMPBSA method is to calculate the free energy difference between the bound and unbound states of two molecules, $[A]$ and $[B]$, in a solvated environment:

$$[A]_{solv} + [B]_{solv} \rightleftarrows [AB]_{solv}. \tag{1}$$

Under ideal circumstances, we would like to calculate the binding free energy, $\Delta G_{bind,solv}$, by the direct application of eq. (1). However, in such an approach most of the total energy would be calculated for solvent-solvent interactions that do not contribute directly to the binding free energy calculation. In practice, a thermodynamic cycle may be used to divide the calculation of the binding free energy into separate steps, and the surrounding solvent is modeled implicitly to increase the efficiency of the calculation.

In the MMPBSA method, the binding free energy for eq. (1) can be calculated using:

$$\Delta G_{bind,solv} = \Delta G_{bind,vac} + \Delta G_{solv,complex} - \Delta G_{solv,ligand} - \Delta G_{solv,receptor}. \tag{2}$$

The first term on the right-hand side of eq. (2) directly measures the binding interaction between molecules $[A]$ and $[B]$ in a vacuum environment:

$$\Delta G_{vac} = \Delta E_{molecular\ mechanics} - T\Delta S_{normal\ mode\ analysis} \tag{3}$$

where the average interaction energy between $[A]$ and $[B]$ accounts for the first term on the right in eq. (3). An optional normal mode analysis can be carried out to estimate the entropy contribution for the second term on the right in eq. (3)[37-38].

The last three terms on the right in eq. (2) correspond to the solvation free energy that is required to remove the solvent from $[A]$ and $[B]$ and to add the solvent back to the $[AB]$ complex while carrying out the thermodynamic cycle. The solvation free energy is calculated by solving the linearized Poisson-Boltzmann equation to account for the electrostatic contribution[30, 39-42], and there is also an additional, empirical non-polar interaction term that is added to account for the hydrophobic contribution[43]:

$$\Delta G_{solv} = \Delta G_{electrostatic} + \Delta G_{non-polar}. \tag{4}$$

Here, we will focus on the application and development of the MMPBSA method for use in rational drug design efforts. In Chapter 1, we will consider an application of the MMPBSA method towards the study and improvement of anti-amyloid beta (Aβ) antibodies[44]. Anti-Aβ antibodies are being used as a prospective drug candidate to treat Alzheimer's Disease (AD), a crippling neurological disease with no known treatment options[45-47]. Here we show how a combination of fragment-based docking and MMPBSA binding free energy calculations can reproduce key experimental Aβ binding epitopes and how they can help illuminate mysterious binding trends. We also show how this approach can be used to generate new hypotheses in the field, such as predicting epitope cross-binding trends or using MMPBSA as a tool to generate prospective mutations for the rational drug design of new anti-Aβ antibody drug candidates.

The remaining portion of this work is dedicated to the application and development of the MMPBSA method towards the analysis of membrane protein-ligand systems, with a particular focus on the study of G-protein coupled receptors (GPCRs). GPCRs are a primary drug target, with nearly 40% of all drug targets being classified as either GPCRs or nuclear receptors[48-49]. GPCRs are openly accessible to binding on the cell surface, making them a very attractive drug target. Despite this, until very recently SBDD efforts for GPCRs have been minimal. This had to do with the difficulty in obtaining high quality structural data for membrane bound proteins in comparison to globular, water soluble proteins[50]. At present, these issues have been ameliorated to some extent, and crystal structures of all major classes of GPCRs are now available for SBDD studies[51-54].

The potential for computational methods to assist in the development of new GPCR drugs has been recently demonstrated by several groups[55-57]. As an example, Heifetz et al. incorporated analysis of MD simulations to improve a drug for OX_1 vs. OX_2 orexin receptor selectivity and used computational docking to assist in the discovery of novel MCH-1R antagonists for use in the treatment of diabetes[58-59]. Very recently, the potential of using MMPBSA for the study of GPCRs was evaluated using 20 GPCR structures with 934 known ligands[60]. While the authors concluded that MMPBSA remained a viable tool for SBDD, they found that the performance for GPCRs varied greatly depending on the system under study, and a low average correlation of 0.183 was obtained across all 20 structures[60].

One possible reason for the poor overall performance in this study was that the membrane proteins were simulated in a non-native, purely aqueous solvent, making use of the older Amber 99 force field. In recent years, the issue with using a non-native solvent for MD simulations of membrane proteins has been addressed with the introduction of an explicit membrane model into Amber 14[61-63]. This model allows users to place a membrane protein in its natural, hydrophobic membrane environment while carrying out MD simulations. In addition, recent work from our lab has focused on the introduction of an implicit membrane model into the PBSA module in Amber 18[64-68]. The introduction of an implicit membrane model into the Amber software suite creates a more physically relevant implicit solvent environment to assist in the calculation of MMPBSA binding free energies for membrane protein-ligand systems.

In Chapter 2, we examine an application for the MMPBSA method towards the calculation of binding free energies of membrane protein-ligand systems using a single dielectric implicit membrane model[67]. Here we use MMPBSA calculations to calculate the binding affinities of several agonist and antagonist ligands bound to the human purinergic platelet receptor ($P2Y_{12}R$), an important drug target for the treatment of myocardial infarction or stroke[69-72]. We show that our single dielectric implicit membrane model and the MMPBSA method can be used to successfully rank order the binding affinities of several different compounds that are relevant to the drug design process.

In Chapter 3, we document the development of an improved heterogeneous dielectric implicit membrane model for use in MMPBSA calculations of membrane protein-ligand systems[68]. Our new model supersedes our previous uniform, single dielectric implicit membrane model by allowing the dielectric constant to vary with depth within the membrane. As an application of our method to various drug design systems, we calculated MMPBSA binding free energies for the human purinergic platelet receptor ($P2Y_{12}R$) and two of the muscarinic acetylcholine receptors (M2R and M3R) bound to various antagonist ligands[73]. We found that the new heterogeneous dielectric membrane model has a stronger correlation with experimental binding affinities compared to the older single dielectric membrane model under otherwise identical simulation conditions.

In Chapter 4, we briefly explore the future development of our implicit membrane model by further improving its accuracy via the parameterization of the non-polar van der Waals terms[43, 74].

The application of MMPBSA methods towards rational drug design has been increasing as of late. This trend should only continue as advances in cryo-EM electron microscopy will allow access to more membrane protein systems for SBDD in the near future[75]. Membrane protein systems are large and computationally expensive to analyze, making the application of rigorous TI or FEP binding free energy calculations problematic. Our implicit membrane models make these large systems more accessible for lead generation and lead optimization-based binding free energy calculations, which can assist in the rational design of future drug candidates.

REFERENCES

1. The Economic Impact of the U.S. Biopharmaceutical Industry: 2015 National and State Estimates. TEConomy Partners, LLC: **2017**.

2. DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *J. Health Econ.* **2016,** *47*, 20-33.

3. Aitken, M.; Berndt, E. R.; Cutler, D. M. Prescription Drug Spending Trends in the United States: Looking Beyond the Turning Point. *Health Aff. (Millwood)* **2009,** *28*, w151-w160.

4. Goodman, M. Market Watch : Pharma Industry Strategic Performance: 2007-2012E. *Nat. Rev. Drug Discov.* **2008,** *7*, 967.

5. Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge. *Nat. Rev. Drug Discov.* **2010,** *9*, 203-214.

6. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011,** *162*, 1239-1249.

7. Ban, T. A. The Role of Serendipity in Drug Discovery. *Dialogues Clin. Neurosci.* **2006,** *8*, 335-344.

8. Broach, J. R.; Thorner, J. High-Throughput Screening for Drug Discovery. *Nature* **1996,** *384*, 14-16.

9. Engels, M. F.; Venkatarangan, P. Smart Screening: Approaches to Efficient HTS. *Curr. Opin. Drug Discov. Devel.* **2001,** *4*, 275-283.

10. Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009,** *42*, 724-733.

11. Macalino, S. J.; Gosu, V.; Hong, S.; Choi, S. Role of Computer-Aided Drug Design in Modern Drug Discovery. *Arch. Pharm. Res.* **2015,** *38*, 1686-1701.

12. van Montfort, R. L. M.; Workman, P. Structure-Based Drug Design: Aiming for a Perfect Fit. *Essays Biochem.* **2017,** *61*, 431-437.

13. Carneiro, M. G.; Ab, E.; Theisgen, S.; Siegal, G. NMR in Structure-Based Drug Design. *Essays Biochem.* **2017,** *61*, 485-493.

14. Venien-Bryan, C.; Li, Z.; Vuillard, L.; Boutin, J. A. Cryo-Electron Microscopy and X-Ray Crystallography: Complementary Approaches to Structural Biology and Drug Discovery. *Acta Crystallogr. F Struct. Biol. Commun.* **2017,** *73*, 174-183.

15. Boland, A.; Chang, L.; Barford, D. The Potential of Cryo-Electron Microscopy for Structure-Based Drug Design. *Essays Biochem.* **2017,** *61*, 543-560.

16. Xia, X. Bioinformatics and Drug Discovery. *Curr. Top. Med. Chem.* **2017,** *17*, 1709-1726.

17. Kontoyianni, M. Docking and Virtual Screening in Drug Discovery. *Methods Mol. Biol.* **2017,** *1647*, 255-266.

18. Shukla, R.; Singh, T. R. Virtual Screening, Pharmacokinetics, Molecular Dynamics and Binding Free Energy Analysis for Small Natural Molecules Against Cyclin-Dependent Kinase 5 for Alzheimer's Disease. *J. Biomol. Struct. Dyn.* **2019**, 1-15.

19. Gupta, S.; Lynn, A. M.; Gupta, V. Standardization of Virtual-Screening and Post-Processing Protocols Relevant to In-Silico Drug Discovery. *3 Biotech* **2018,** *8*, 504.

20. Grinter, S. Z.; Zou, X. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules* **2014,** *19*, 10150-10176.

21. Ganesan, A.; Coote, M. L.; Barakat, K. Molecular Dynamics-Driven Drug Discovery: Leaping Forward with Confidence. *Drug Discov. Today* **2017,** *22*, 249-269.

22. Gioia, D.; Bertazzo, M.; Recanatini, M.; Masetti, M.; Cavalli, A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. *Molecules* **2017,** *22*.

23. Tuffery, P.; Derreumaux, P. Flexibility and Binding Affinity in Protein-Ligand, Protein-Protein and Multi-Component Protein Interactions: Limitations of Current Computational Approaches. *J. R. Soc. Interface* **2012,** *9*, 20-33.

24. Lee, Y.; Lazim, R.; Macalino, S. J. Y.; Choi, S. Importance of Protein Dynamics in the Structure-Based Drug Discovery of Class A G Protein-Coupled Receptors (GPCRs). *Curr. Opin. Struct. Biol.* **2019,** *55*, 147-153.

25. Li, J.; Fu, A.; Zhang, L. An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. *Interdiscip. Sci.* **2019**.

26. Guedes, I. A.; Pereira, F. S. S.; Dardenne, L. E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Front. Pharmacol.* **2018,** *9*, 1089.

27. Huang, S. Y.; Grinter, S. Z.; Zou, X. Scoring Functions and Their Evaluation Methods for Protein-Ligand Docking: Recent Advances and Future Directions. *Phys. Chem. Chem. Phys.* **2010,** *12*, 12899-12908.

28. Reddy, M. R.; Reddy, C. R.; Rathore, R. S.; Erion, M. D.; Aparoy, P.; Reddy, R. N.; Reddanna, P. Free Energy Calculations to Estimate Ligand-Binding Affinities in Structure-Based Drug Design. *Curr. Pharm. Des.* **2014,** *20*, 3323-3337.

29. Shirts, M. R., Mobley, D., Brown, S.P. *Free Energy Calculations in Structure-Based Drug Design*. Cambridge University Press: **2010**.

30. Gilson, M. K.; Honig, B. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis. *Proteins* **1988,** *4*, 7-18.

31. Bashford, D.; Karplus, M. pKa's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model. *Biochemistry* **1990,** *29*, 10219-10225.

32. Srinivasan, J.; Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate - DNA Helices. *J. Am. Chem. Soc.* **1998,** *120*, 9401-9409.

33. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889-897.

34. Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005,** *48*, 4040-4048.

35. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015,** *10*, 449-461.

36. Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004,** *25*, 238-250.

37. Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual Screening Using Molecular Simulations. *Proteins* **2011,** *79*, 1940-1951.

38. Sun, H.; Duan, L.; Chen, F.; Liu, H.; Wang, Z.; Pan, P.; Zhu, F.; Zhang, J. Z. H.; Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 7. Entropy Effects on the Performance of End-Point Binding Free Energy Calculation Approaches. *Phys. Chem. Chem. Phys.* **2018,** *20*, 14450-14460.

39. Wang, J.; Cai, Q.; Li, Z. L.; Zhao, H. K.; Luo, R. Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms. *Chem. Phys. Lett.* **2009,** *468*, 112-118.

40. Wang, J.; Luo, R. Assessment of Linear Finite-Difference Poisson-Boltzmann Solvers. *J. Comput. Chem.* **2010,** *31*, 1689-1698.

41. Jeancharles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. Electrostatic Contributions to Solvation Energies - Comparison of Free-Energy Perturbation and Continuum Calculations. *J. Am. Chem. Soc.* **1991,** *113*, 1454-1455.

42. Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann Calculations for Static and Dynamic Systems. *J Comput Chem* **2002,** *23*, 1244-1253.

43. Tan, C.; Tan, Y. H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007,** *111*, 12263-12274.

44. Greene, D.; Po, T.; Pan, J.; Tabibian, T.; Luo, R. Computational Analysis for the Rational Design of Anti-Amyloid Beta (Abeta) Antibodies. *J. Phys. Chem. B* **2018,** *122*, 4521-4536.

45. Masters, C. L.; Simms, G.; Weinman, N. A.; Multhaup, G.; McDonald, B. L.; Beyreuther, K. Amyloid Plaque Core Protein in Alzheimer Disease and Down Syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **1985,** *82*, 4245-4249.

46. Iqbal, K.; Alonso Adel, C.; Chen, S.; Chohan, M. O.; El-Akkad, E.; Gong, C. X.; Khatoon, S.; Li, B.; Liu, F.; Rahman, A.; Tanimukai, H.; Grundke-Iqbal, I. Tau Pathology in Alzheimer Disease and Other Tauopathies. *Biochim. Biophys. Acta* **2005,** *1739*, 198-210.

47. Selkoe, D. J.; Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease at 25 Years. *EMBO Mol. Med.* **2016,** *8*, 595-608.

48. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets are There? *Nat. Rev. Drug Discov.* **2006,** *5*, 993-996.

49. Drews, J. Drug Discovery: A Historical Perspective. *Science* **2000,** *287*, 1960-1964.

50. Seddon, A. M.; Curnow, P.; Booth, P. J. Membrane Proteins, Lipids and Detergents: Not Just a Soap Opera. *Biochim. Biophys. Acta* **2004,** *1666*, 105-117.

51. Tautermann, C. S. GPCR Structures in Drug Design, Emerging Opportunities with New Structures. *Bioorg. Med. Chem. Lett.* **2014,** *24*, 4073-4079.

52. Jazayeri, A.; Dias, J. M.; Marshall, F. H. From G-Protein Coupled Receptor Structure Resolution to Rational Drug Design. *J. Biol. Chem.* **2015,** *290*, 19489-19495.

53. Cong, X.; Topin, J.; Golebiowski, J. Class A GPCRs: Structure, Function, Modeling and Structure-Based Ligand Design. *Curr. Pharm. Des.* **2017,** *23*, 4390-4409.

54. Gentry, P. R.; Sexton, P. M.; Christopoulos, A. Novel Allosteric Modulators of G Protein-Coupled Receptors. *J. Biol. Chem.* **2015,** *290*, 19478-19488.

55. Lee, Y.; Basith, S.; Choi, S. Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. *J. Med. Chem.* **2018,** *61*, 1-46.

56. Miao, Y.; McCammon, J. A. G-Protein Coupled Receptors: Advances in Simulation and Drug Discovery. *Curr. Opin. Struct. Biol.* **2016,** *41*, 83-89.

57. Yuan, X.; Xu, Y. Recent Trends and Applications of Molecular Modeling in GPCR(-)Ligand Recognition and Structure-Based Drug Design. *Int. J. Mol. Sci.* **2018,** *19*.

58. Heifetz, A.; Morris, G. B.; Biggin, P. C.; Barker, O.; Fryatt, T.; Bentley, J.; Hallett, D.; Manikowski, D.; Pal, S.; Reifegerste, R.; Slack, M.; Law, R. Study of Human Orexin-1 and -2 G-Protein-Coupled Receptors with Novel and Published Antagonists by Modeling, Molecular Dynamics Simulations, and Site-directed Mutagenesis. *Biochemistry* **2012,** *51*, 3178-3197.

59. Heifetz, A.; Barker, O.; Verquin, G.; Wimmer, N.; Meutermans, W.; Pal, S.; Law, R. J.; Whittaker, M. Fighting Obesity with a Sugar-Based Library: Discovery of Novel MCH-1R Antagonists by a new Computational-VAST Approach for Exploration of GPCR Binding Sites. *J. Chem. Inf. Model.* **2013,** *53*, 1084-1099.

60. Yau, M. Q.; Emtage, A. L.; Chan, N. J. Y.; Doughty, S. W.; Loo, J. S. E. Evaluating the Performance of MM/PBSA for Binding Affinity Prediction Using Class A GPCR Crystal Structures. *J. Comput. Aided Mol. Des.* **2019,** *33*, 487-496.

61. Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014,** *10*, 865-879.

62. Madej, B. D.; Gould, I. R.; Walker, R. C. A Parameterization of Cholesterol for Mixed Lipid Bilayer Simulation within the Amber Lipid14 Force Field. *J. Phys. Chem. B* **2015,** *119*, 12424-12435.

63. Madej, B. D., Walker, R. C. An Amber Lipid Force Field Tutorial: Lipid14 Edition. http://ambermd.org/tutorials/advanced/tutorial16/ (accessed Jul 26, **2019**)

64. Botello-Smith, W. M.; Liu, X.; Cai, Q.; Li, Z.; Zhao, H.; Luo, R. Numerical Poisson-Boltzmann Model for Continuum Membrane Systems. *Chem. Phys. Lett.* **2013,** *555*, 274-281.

65. Botello-Smith, W. M.; Luo, R. Applications of MMPBSA to Membrane Proteins I: Efficient Numerical Solutions of Periodic Poisson-Boltzmann Equation. *J. Chem. Inf. Model* **2015,** *55*, 2187-2199.

66. Xiao, L.; Diao, J.; Greene, D.; Wang, J.; Luo, R. A Continuum Poisson-Boltzmann Model for Membrane Channel Proteins. *J. Chem. Theory Comput.* **2017,** *13*, 3398-3412.

67. Greene, D.; Botello-Smith, W. M.; Follmer, A.; Xiao, L.; Lambros, E.; Luo, R. Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor. *J. Phys. Chem. B* **2016,** *120*, 12293-12304.

68. Greene, D.; Qi, R.; Nguyen, R.; Qiu, T.; Luo, R. Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies. *J. Chem. Inf. Model* **2019,** *59*, 3041-3056.

69. Zhang, K.; Zhang, J.; Gao, Z. G.; Zhang, D.; Zhu, L.; Han, G. W.; Moss, S. M.; Paoletta, S.; Kiselev, E.; Lu, W.; Fenalti, G.; Zhang, W.; Muller, C. E.; Yang, H.; Jiang, H.; Cherezov, V.; Katritch, V.; Jacobson, K. A.; Stevens, R. C.; Wu, B.; Zhao, Q. Structure of the Human P2Y12 Receptor in Complex with an Antithrombotic Drug. *Nature* **2014,** *509*, 115-118.

70. Zhang, J.; Zhang, K.; Gao, Z. G.; Paoletta, S.; Zhang, D.; Han, G. W.; Li, T.; Ma, L.; Zhang, W.; Muller, C. E.; Yang, H.; Jiang, H.; Cherezov, V.; Katritch, V.; Jacobson, K. A.; Stevens, R. C.; Wu, B.; Zhao, Q. Agonist-Bound Structure of the Human P2Y12 Receptor. *Nature* **2014,** *509*, 119-122.

71. Savi, P.; Zachayus, J. L.; Delesque-Touchard, N.; Labouret, C.; Herve, C.; Uzabiaga, M. F.; Pereillo, J. M.; Culouscou, J. M.; Bono, F.; Ferrara, P.; Herbert, J. M. The Active Metabolite of Clopidogrel Disrupts P2Y12 Receptor Oligomers and Partitions them out of Lipid Rafts. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 11069-11074.

72. Algaier, I.; Jakubowski, J. A.; Asai, F.; von Kugelgen, I. Interaction of the Active Metabolite of Prasugrel, R-138727, with Cysteine 97 and Cysteine 175 of the Human P2Y12 Receptor. *J. Thromb. Haemost.* **2008,** *6*, 1908-1914.

73. Liu, H.; Hofmann, J.; Fish, I.; Schaake, B.; Eitel, K.; Bartuschat, A.; Kaindl, J.; Rampp, H.; Banerjee, A.; Hubner, H.; Clark, M. J.; Vincent, S. G.; Fisher, J. T.; Heinrich, M. R.; Hirata, K.; Liu, X.; Sunahara, R. K.; Shoichet, B. K.; Kobilka, B. K.; Gmeiner, P. Structure-guided Development of Selective M3 Muscarinic Acetylcholine Receptor Antagonists. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115*, 12046-12050.

74. Dutagaci, B.; Sayadi, M.; Feig, M. Heterogeneous Dielectric Generalized Born Model with a van der Waals Term Provides Improved Association Energetics of Membrane-Embedded Transmembrane Helices. *J. Comput. Chem.* **2017,** *38*, 1308-1320.

75. Rawson, S.; Davies, S.; Lippiat, J. D.; Muench, S. P. The Changing Landscape of Membrane Protein Structural Biology Through Developments in Electron Microscopy. *Mol. Membr. Biol.* **2016,** *33*, 12-22.

CHAPTER 1

Computational Analysis for the Rational Design of Anti-Amyloid Beta (Aβ) Antibodies

Reprinted (adapted) with permission from Greene, D.; Po, T.; Pan, J.; Tabibian, T.; Luo, R. Computational Analysis for the Rational Design of Anti-Amyloid Beta (Aβ) Antibodies. *J. Phys. Chem. B* **2018,** *122*, 4521-4536. Copyright 2018 American Chemical Society.

ABSTRACT

Alzheimer's Disease (AD) is a neurodegenerative disorder that lacks effective treatment options. Anti-amyloid beta (Aβ) antibodies are the leading drug candidates to treat AD, but the results of clinical trials have been disappointing. Introducing rational mutations into anti-Aβ antibodies to increase their effectiveness is a way forward, but the path to take is unclear. In this study, we demonstrate the use of computational fragment-based docking and MMPBSA binding free energy calculations in the analysis of anti-Aβ antibodies for rational drug design efforts. Our fragment-based docking method successfully predicts the emergence of the common EFRH epitope. MD simulations coupled with MMPBSA binding free energy calculations are used to analyze scenarios described in prior studies, and we computationally introduce rational mutations into PFA1 to predict mutations that can improve its binding affinity towards the pE3-Aβ_{3-8} form of Aβ. Two out of our four proposed mutations are predicted to stabilize binding. Our study demonstrates that a computational approach may lead to an improved drug candidate for AD in the future.

1.1 INTRODUCTION

Alzheimer's Disease (AD) is an incurable neurodegenerative disorder that leads to steady memory and cognitive function loss, culminating in death. At present there is no cure for AD, and there is a notable absence of treatment options that can reverse or effectively slow progression of the disease. At the level of brain tissue, AD is characterized by both the appearance of extracellular, fibrous plaques that are built up from the polymerization of amyloid beta (Aβ) peptides[1] and the appearance of intracellular neurofibrillary tangles that consist of hyperphosphorylated tau proteins[2]. Amyloid fibril deposits are hallmarks of several neurodegenerative diseases[3], and the amyloid hypothesis states that an excessive buildup of Aβ plaques in the brain is responsible for the cognitive decline observed in AD patients. It suggests that clearing Aβ plaques from the brain would help inhibit or reverse progression of the disease.

The amyloid hypothesis has been the leading theory driving therapeutic approaches for the treatment of AD for over two decades[4]. The most common therapeutic approach to AD treatment is immunotherapy[5-7]. Several active and passive anti-Aβ immunotherapies that target Aβ species in the brain have advanced to clinical trials, but the results thus far have been disappointing. The vaccine AN1792, which targeted a full-size Aβ 1-42 peptide, advanced to human clinical trials in 2001 but was terminated in the phase II trial after 6% of the treated patients developed meningoencephalitis[8]. The result of the AN1792 trial led to the development of several passive immunization approaches over the next several years. Shorter regions of the Aβ sequence were used to develop monoclonal antibodies that target different cell types in the immune system. The N-terminal Aβ_{1-15} sequence was used for Aβ specific B cell epitopes while Aβ_{16-42} was used for Aβ specific T cell epitopes[9]. The choice of the epitope has a crucial effect on the ability of the antibody to bind to its amyloid target species. The N-terminal epitope is accessible to antibody binding in aggregated forms of Aβ while the central and C-terminal epitopes are only able to bind to the antibody in monomeric, or perhaps small oligomeric, forms of Aβ due to the central and C-terminal epitopes being inaccessible in mature fibril structures.

Initially, two anti-Aβ monoclonal antibodies advanced to clinical trials targeting distinct epitopes/species of Aβ. Bapineuzumab primarily targets insoluble amyloid plaques via the hydrophilic N-terminal epitope of Aβ_{1-5}. Initial results looked promising, but during phase II of the clinical trial a serious side effect appeared as 10% of patients developed vasogenic edema[10]. Solanezumab targets soluble monomeric Aβ peptides via the hydrophobic central Aβ_{16-24} epitope. In clinical trials, solanezumab had a much-improved safety profile as adverse side effects such as meningoencephalitis, microhemorrhage, and vasogenic edema were not obserseved[11-12]. On the other hand, questions about solanezumab's efficacy in reducing neuritic plaque burden arose, and recently it was announced that the phase III trials for solanezumab had failed to show a significant benefit in slowing cognitive decline for mild-to-moderate AD patients[13-16].

The question of how to improve a monoclonal antibody to treat AD is not straightforward. Removing harmful side effects is one issue, but to complicate matters, Aβ species exhibit a high degree of structural polymorphism, and several other Aβ species have emerged as potential disease-causing agents that would presumably need to be removed in an effective AD treatment[17]. For example, normally rare, N-terminal truncated variants of Aβ have been found in much higher concentrations in the stable, neurotoxic Aβ aggregates that are found in severe AD cases. For example, plaques can be enriched by as much as 50% with the pE3-Aβ form of Aβ[18-19], and pE3-Aβ has become a target in antibody development[20-22].

The existence of such polymorphic amyloid targets implies that targeting a single epitope associated with a single Aβ species might not be enough for an antibody to treat AD effectively. At the moment, next generation monoclonal antibodies that can bind to multiple Aβ epitopes and species are currently in clinical trials. Gantenerumab binds at nanomolar affinity to several Aβ species (with a K_D of 0.6 nM for Aβ fibrils, 1.2 nM for Aβ oligomers, and 17 nM for Aβ monomers), and it recognizes two epitopes within Aβ: the N-terminal EFRHDSGYEV sequence and a central region from the sequence VFFAEDVGSN[23]. Similarly, crenezumab, despite being generated by immunization with the N-terminal $A\beta_{1-16}$ epitope[24-25], has been shown to bind to monomeric and oligomeric forms of Aβ via the central Aβ epitope[7, 25-26]. Due to similarities it shares with solanezumab, the ability to bind to the central epitope of Aβ has been emphasized in studies of crenezumab[27], and a co-crystal structure of crenezumab (more specifically, CreneFab) was recently obtained bound to an Aβ peptide containing the central epitope[26]. Nevertheless, crenezumab has also been shown to bind to amyloid fibril species[25-26], which is puzzling since the central epitope is not readily accessible for binding in mature fibril structures. Recently, aducanumab was heralded as possibly the first “successful” anti-Aβ antibody as it was able to clear Aβ plaques thoroughly at the highest dosage and was shown to reduce cognitive decline in an early Phase III clinical trial that took place over the course of a year[28]. Aducanumab has been reported to bind to both N-terminal and central epitopes of Aβ, accounting for binding to both fibril and oligomeric forms of Aβ[29].

In this study, we have taken a computational approach to explore how an anti-Aβ antibody may bind to one or more Aβ epitopes in its antigen-combining site and to demonstrate how to computationally predict rational mutations aimed at modifying the antigen-combining site in rational drug design efforts. Using computational techniques, we explored several open questions from the literature including: 1) which residues in Aβ_{1-42} are most important in the initial binding event that anchors Aβ to a given antibody structure?, 2) how might antibodies like gantenerumab, crenezumab, and aducanumab bind to both the hydrophillic N-terminal and hydrophobic central epitopes?, and 3) can we predict a useful mutation that improves the binding affinity of a polymorphic form of Aβ towards an anti-Aβ antibody?

To address the first question, we employed an unbiased, fragment-based docking method to probe the antigen-combining site of various anti-Aβ antibodies using single amino acid residues. To address the second question, we ran molecular dynamics (MD) simulations using the available crystal structures of gantenerumab and crenezumab bound to short Aβ peptides[23, 26] (a crystal structure of aducanumab was not available for us to analyze at this time). Using Amber 16[30], we carried out Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) binding free energy calculations to help us explore the possibility that both N-terminal and central Aβ epitopes are recognized by each antibody. For the third question, we built on the study of PFA1 bound to Aβ_{2-7} and pE3-Aβ_{3-8} that was carried out previously by Gardberg et al.[31] Their work gained inspiration from a study by Clark et al. where it was demonstrated that a combination of computational and experimental mutagenesis techniques could be used to improve the binding affinities of therapeutic antibodies[32]. The study by Gardberg et al., where the goal was to improve the binding affinity between pE3-Aβ_{3-8} and PFA1, was an accessible test case given that the sequence, charge characteristics, and binding pose for pE3-Aβ_{3-8} are similar to the wildtype. The authors offered several suggestions for point mutations that may produce a mutant form of PFA1 with an enhanced affinity for pE3-Aβ_{3-8}.[31] To aid in the selection of prospective mutant PFA1 antibodies, a computational approach can provide a way to probe the interactions between an Aβ peptide and an anti-Aβ antibody. It can also act as a means to test the efficacy of point mutations made to the antigen-combining site before committing to testing them in the lab. This cost effective, rational approach to predicting mutant antibodies might be a key to producing the most effective drug candidate for this disease in the long term.

1.2 METHODS

1.2.1 Computational docking of amino acid residues to anti-Aβ antibodies

We carried out an unbiased, fragment-based computational docking study to examine the initial binding characteristics of the antigen-combining site for various anti-Aβ antibodies. 16 amino acids were individually docked to each antibody, comprising the full Aβ 1-42 monomer sequence. Each single amino acid residue was generated using the sequence command in xleap from Amber 16[30]. Three types of amino acid residue fragments were initially tested using our docking protocol: 1) the default amino acid residue that contained charged N- and C-terminal groups on the backbone, 2) a neutralized version where methyl groups were attached to both the N- and C-terminal groups to remove the backbone charges, and 3) a non-physiological fragment where two hydrogens on the N-terminus and an oxygen on the C-terminus were omitted from the structure. We compared the results of the docking for a few test residues using each method above to the actual binding sites observed in the holo PFA1 and PFA2 crystal structures and also to the leading ligand-free hotspots found by submitting the PFA1 and PFA2 apo structures to the FTMap server[33-34]. Of the three options, we found that method 3 worked the best. For method 1, the zwitterionic terminal backbone charges were capable of binding to antibody hot spots in place of the side chain functional groups, and while method 2 neutralized the terminal backbone charges, we found that some fragments were unable to bind to certain hot spots due to steric issues brought about by the attached methyl groups. Method 3 removed both the terminal charges and the steric issue of adding a methyl group to the backbone. Therefore, residues were generated using method 3 for our full analysis.

We should note that there are other possible ways to generate single amino acid fragments that were implemented in other computational studies, such as removing the backbone entirely and replacing the alpha carbon with a proton[35]. In order to obtain a single amino acid fragment that behaves more like an amino acid in the middle of a longer peptide sequence as opposed to a terminal amino acid, the side chain for each amino acid should be unaltered in the fragment while the effect of the backbone should be minimized in some way to prevent it from interfering or competing with the binding of the side chain.

After obtaining an amino acid residue, Open Babel[36] was used to convert pdb files to pdbqt files for both the antibody and the residue, and then the residue was docked to the antibody using Autodock Vina/SMINA[37]. The residue was allowed to search the entire antibody surface within a 100 Å^3 box that was centered on the antigen-combining site with a search exhaustiveness of 128. Using the default settings in Autodock Vina/SMINA, the top nine docked results for each residue, ranked by their most stable binding free energy values, were subsequently used for our analysis.

To carry out a comparison study with a variety of anti-amyloid antibodies, this method was performed on the previously published crystal structures of bapineuzumab, solanezumab, gantenerumab, crenezumab, ponezumab, PFA1, and PFA2 in both the holo and apo forms (if available) of each structure (PDB IDs: 4OJF, 4XXD, 5CSZ, 5VZX, 5VZY, 3UOT, 2IPU, 2IPT, 2R0W, and 2IQ9). The PDB file for each antibody complex was edited prior to docking to remove everything except for the residues of the isolated antibody structure.

1.2.2 Molecular dynamics simulations

To prepare the various complexes that were used in this study for MD simulations, the program Modeller[38] was used to accomplish two purposes: 1) to model in any missing residues that were not present in the original PDB file, and 2) if applicable, to generate a homology model containing a new Aβ peptide that was not present in the original PDB file. Prior to using Modeller, the PDB file was processed to remove everything except for the Aβ peptide-antibody complex. After using Modeller, the structure was further processed using the program xleap in Amber 16 to add in hydrogens, water, counter ions, and disulfide bonds. A TIP3P water box was constructed using the solvatebox command to ensure that all atoms in the starting structure were no less than 12 Å from the edge of the water box. Counter ions (i.e. chloride ions) were added to the solvated system to give a total net charge of zero. Disulfide bonds were added between pairs of residues that were indicated as having disulfide bonds in the accompanying PDB files. For the pE3-Aβ_{3-8} model (PDB ID: 3EYS), antechamber was used to parameterize the pyroglutamate (PCA) moiety.

For each MD run, a 1000 step minimization was carried out with 500 steps of steepest descent followed by 500 steps of conjugate gradient using a non-bonded cutoff of

8.0 Å. The system was then heated up to a constant 300 K over a period of 50 ps under NVT conditions employing the Langevin thermostat. The density was equilibrated over an additional 50 ps under NTP conditions, and an equilibration under NVT conditions was undertaken for approximately 250 ns. Finally, a 50 ns production run was carried out to bring the total simulation time to 300 ns. Due to the size and high conformational flexibility of both the amyloid peptide and the antibody, and the extensive use of homology modeling, a long equilibration/production run was needed in order to achieve acceptable convergence for our MMPBSA binding free energy calculation (see **MD method validation** in Results and Discussion). Afterwards, the MD trajectory of our production run was visualized using UCSF Chimera[39]. In addition, to characterize alterations in binding patterns in the antigen-combining sites of our peptide-antibody structures, residue-to-residue percent occupancy calculations were carried out between select residues in the Aβ peptide and on the antibody surface. These calculations used a cutoff distance of 10 Å over 5000 frames collected at equal intervals from the 50 ns production run. This relatively high cutoff distance was chosen due to the high conformational flexibility of certain ligands in the antigen-combing site which led to a difficulty in identifying specific binding interactions; as a result, we do not use this measure to indicate such binding interactions per say, but rather, a low percent occupancy indicates a major decrease in the likelihood that the two residues in question can participate in such a binding interaction at a given site.

1.2.3 MMPBSA binding free energy calculations

5000 frames, taken at equal intervals over the 50 ns production run, were used to calculate the MMPBSA binding free energy[40-45] for each Aβ peptide-antibody complex. All

PBSA calculations were carried out using the PBSA program[46-54] in the AMBER 16 package.[30, 55] For this calculation, inp = 2 was used for the non-polar solvation model,[56-57] radiopt = 0 was used for the intrinsic atomic radii, and the ionic strength was set to 100 mM. All other settings were kept at the default values used in Amber 16.[30, 55] Due to the rather high uncertainties in normal mode analysis, the entropy contribution was neglected in our binding affinity calculations. Experimental binding affinities were compared to our calculated binding affinities by converting K_D values into binding free energies using:

$$\Delta G = \mathrm{RTln}(\mathrm{K_D})$$

where R = $1.987 * 10^{-3}$ kcal mol^{-1} K^{-1}, T = 300 K, and K_D is the dissociation constant in units of M.

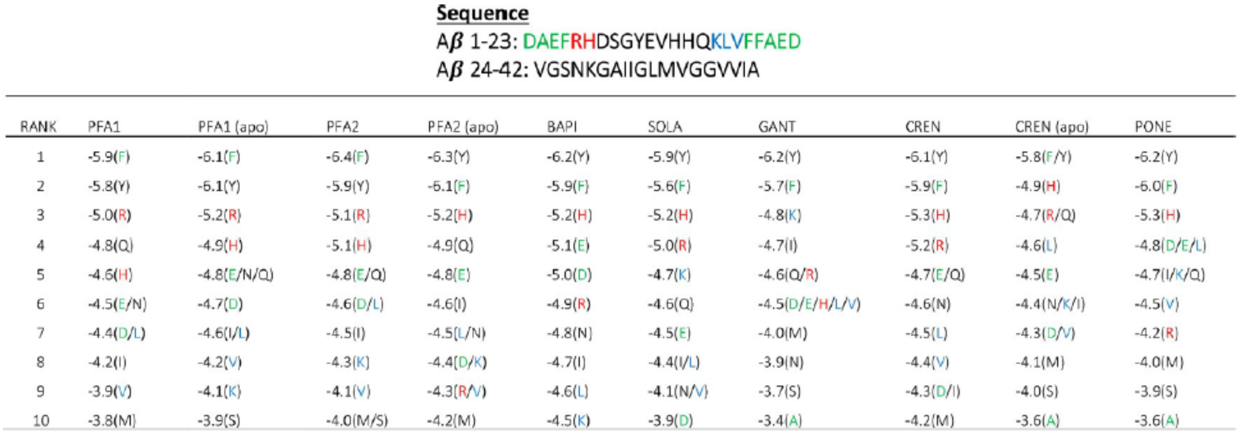
1.3 RESULTS

1.3.1 Computational docking of amino acid residues to anti-Aβ antibodies

To address the question of which amino acids are key to the initial binding and anchoring of an Aβ peptide to an anti-amyloid antibody, we carried out an unbiased fragment-based computational docking search using all 16 unique amino acid residues that appear in the Aβ_{1-42} sequence. To minimize conformational effects from the polypeptide ligand, we chose to dock single amino acid residue fragments to the antibody surface. The goal in carrying out this step is to determine which single amino acid(s) are most important to the initial binding event. Previous studies have indicated that antibodies that exhibit promiscuous binding with extended peptide epitopes have a tendency to rely on a small number of key semi-conserved anchor residues on the peptide when binding[58-59]. The ability to identify key anchor residue(s) provides an important clue towards understanding how certain anti-Aβ antibodies can initially recognize both the N-terminal and central Aβ

epitopes despite the low sequence similarity between the two regions. Two criteria from the output were considered to be relevant in determining the likelihood that an amino acid will bind with high affinity to the antigen-combining site of the antibody: 1) which residues have the strongest binding interaction at the antigen-combining site?, and 2) which residues find the antigen-combining site most consistently?

TABLE 1.1 Top binding affinities for single amino acid residues bound to the antigen-combining site of anti-Aβ antibodies



Sequence
Aβ 1-23: DAEFRHDSGYEVHHQKLVFFAED
Aβ 24-42: VGSNKGAIIGLMVGGVVIA

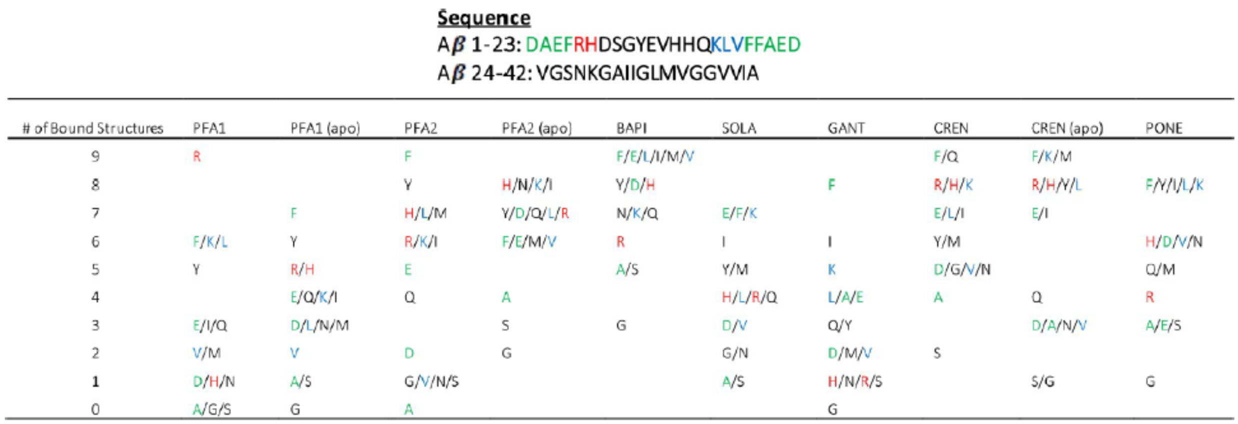
RANK	PFA1	PFA1 (apo)	PFA2	PFA2 (apo)	BAPI	SOLA	GANT	CREN	CREN (apo)	PONE
1	-5.9(F)	-6.1(F)	-6.4(F)	-6.3(Y)	-6.2(Y)	-5.9(Y)	-6.2(Y)	-6.1(Y)	-5.8(F/Y)	-6.2(Y)
2	-5.8(Y)	-6.1(Y)	-5.9(Y)	-6.1(F)	-5.9(F)	-5.6(F)	-5.7(F)	-5.9(F)	-4.9(H)	-6.0(F)
3	-5.0(R)	-5.2(R)	-5.1(R)	-5.2(H)	-5.2(H)	-5.2(H)	-4.8(K)	-5.3(H)	-4.7(R/Q)	-5.3(H)
4	-4.8(Q)	-4.9(H)	-5.1(H)	-4.9(Q)	-5.1(E)	-5.0(R)	-4.7(I)	-5.2(R)	-4.6(L)	-4.8(D/E/L)
5	-4.6(H)	-4.8(E/N/Q)	-4.8(E/Q)	-4.8(E)	-5.0(D)	-4.7(K)	-4.6(Q/R)	-4.7(E/Q)	-4.5(E)	-4.7(I/K/Q)
6	-4.5(E/N)	-4.7(D)	-4.6(D/L)	-4.6(I)	-4.9(R)	-4.6(Q)	-4.5(D/E/H/L/V)	-4.6(N)	-4.4(N/K/I)	-4.5(V)
7	-4.4(D/L)	-4.6(I/L)	-4.5(I)	-4.5(L/N)	-4.8(N)	-4.5(E)	-4.0(M)	-4.5(L)	-4.3(D/V)	-4.2(R)
8	-4.2(I)	-4.2(V)	-4.3(K)	-4.4(D/K)	-4.7(I)	-4.4(I/L)	-3.9(N)	-4.4(V)	-4.1(M)	-4.0(M)
9	-3.9(V)	-4.1(K)	-4.1(V)	-4.3(R/V)	-4.6(L)	-4.1(N/V)	-3.7(S)	-4.3(D/I)	-4.0(S)	-3.9(S)
10	-3.8(M)	-3.9(S)	-4.0(M/S)	-4.2(M)	-4.5(K)	-3.9(D)	-3.4(A)	-4.2(M)	-3.6(A)	-3.6(A)

Binding free energies are given in units of kcal/mol while the corresponding amino acid is indicated by using the standard single letter amino acid code. All of the structures given are based on the holo crystal structure form of the antibody except for those marked "(apo)" which are based on the apo crystal structure. Antibody abbreviations: PFA1 = protofibril antibody 1, PFA2 = protofibril antibody 2, BAPI = bapineuzumab, SOLA = solanezumab, GANT = gantenerumab, CREN = crenezumab, PONE = ponezumab. The Aβ 1-42 amino acid sequence is also provided where residues in the N-terminal epitope are highlighted in red, residues in the central epitope are highlighted in blue, and residues that appear in both epitopes are highlighted in green.

TABLE 1.1 lists the most stable binding affinities for the top 10 ranked amino acid residues bound to the antigen-combining sites of ten anti-Aβ antibody structures. From **TABLE 1.1**, we observe that approximately 95% of the amino acid residues that are ranked in the top 5 appear in the Aβ 1-23 region while 13% of the amino acid residues that are ranked in the top 5 appear in the Aβ 24-42 region. Furthermore, 57% of the amino acid residues that are ranked in the top 5 are found in the N-terminal DAEFRH epitope while 39% appear in the central KLVFFAED epitope. In particular, nine out of the ten antibodies have at least three out of four residues in the EFRH sequence appear in their top 5 ranking; the only exception is gantenerumab where E and H are tied for sixth place. These results are consistent with the noted tendency of the N-terminal DAEFRH epitope to bind to most anti-Aβ antibodies[60].

In **TABLE 1.2**, we examine the number of docked amino acids found at the antigen-combining site for each residue across the 10 different anti-Aβ antibodies. If we consider only the amino acids that were able to dock at the antigen-combining site more than half of the time, 79% appear in the Aβ 1-23 region while 47% appear in the Aβ 24-42 region. Of the residues that found the correct binding site more than half the time, 39% are found in the N-terminal DAEFRH epitope while 46% appear in the central KLVFFAED epitope. In contrast to **TABLE 1.1**, which displayed similar results in the residue ranking for each antibody, **TABLE 1.2** displayed clear differences between the different highly ranked residues docked to different antibodies and even showed some noticeably large differences for the holo and apo forms of the same antibody. These latter differences are presumably due to differences in the specific conformation of the antibody in the holo and apo crystal structures. It is worth pointing out that while charged and aromatic residues appear to dominate the top ranks in **TABLE 1.1**, the ability of fragments to find the active site, as observed in **TABLE 1.2**, does not correlate as strongly with polarity as many polar and non-polar residues tend to locate the antigen-combining site with similar ease.

TABLE 1.2. Number of docked structures found at the antigen-combing site of anti-Aβ antibodies



Sequence
Aβ 1-23: DAEFRHDSGYEVHHQKLVFFAED
Aβ 24-42: VGSNKGAIIGLMVGGVVIA

# of Bound Structures	PFA1	PFA1 (apo)	PFA2	PFA2 (apo)	BAPI	SOLA	GANT	CREN	CREN (apo)	PONE
9	R		F		F/E/L/I/M/V			F/Q	F/K/M	
8			Y	H/N/K/I	Y/D/H		F	R/H/K	R/H/Y/L	F/Y/I/L/K
7		F	H/L/M	Y/D/Q/L/R	N/K/Q	E/F/K		E/L/I	E/I	
6	F/K/L	Y	R/K/I	F/E/M/V	R	I	I	Y/M		H/D/V/N
5	Y	R/H	E		A/S	Y/M	K	D/G/V/N		Q/M
4		E/Q/K/I	Q	A		H/L/R/Q	L/A/E	A	Q	R
3	E/I/Q	D/L/N/M		S	G	D/V	Q/Y		D/A/N/V	A/E/S
2	V/M	V	D	G		G/N	D/M/V	S		
1	D/H/N	A/S	G/V/N/S			A/S	H/N/R/S		S/G	G
0	A/G/S	G	A				G			

The number of docked structures for each amino acid residue that were found in the antigen-combining site of the antibody are reported in the table above. Each amino acid is indicated by using the standard single-letter amino acid code. Antibody abbreviations: PFA1 = protofibril antibody 1, PFA2 = protofibril antibody 2, BAPI = bapineuzumab, SOLA = solanezumab, GANT = gantenerumab, CREN = crenezumab, PONE = ponezumab. The Aβ 1-42 amino acid sequence is also provided where residues in the N-terminal epitope are highlighted in red, residues in the central epitope are highlighted in blue, and residues that appear in both epitopes are highlighted in green.

An examination of **TABLES 1.1 and 1.2** reveals an important trend in the binding pattern that appears to be consistent across the various antibodies that we have studied; the top two residues on Aβ 1-42 that have the most stable binding free energy are either phenylalanine or tyrosine in all cases, and both residues are also near the top of the residues that are consistently docked to the antigen-combining site. Since phenylalanine and tyrosine are structurally identical except for a hydroxyl group, the strong binding free energy points to the presence of an aromatic binding pocket on the antibody that is important for binding Aβ. The existence of such a binding pocket has been pointed out before as a key binding site for phenylalanine by several anti-Aβ antibody crystallographers[23, 26, 60-61]. In addition, the two prominent Aβ epitopes, the N-terminal DAEFRH epitope and the central KLVFFAED epitope, correspond to the only two locations in the Aβ 1-42 sequence where phenylalanine appears. Our docking results, taken together with these observations, point towards phenylalanine as a leading candidate for the most important anchor residue in the initial binding of Aβ epitopes to an anti-amyloid antibody.

1.3.2 MD method validation

Although the computational docking of single amino acid residues may give us some insight as to which amino acids might bind first to the antigen-combining site of an anti-Aβ antibody, it does not necessarily help us understand how binding affinity emerges for an extended polypeptide. After the first residue in a polypeptide binds to the antigen-combining site, a previously accessible, high affinity binding site now becomes unavailable to other residues, and the binding of the first residue also restricts the search space where other residues can bind. In addition, our docking protocol lacks many key factors, such as the presence of water and the conformational motion of the full polypeptide and antibody, which have both been identified as being important factors in determining the selectivity of Aβ peptide-antibody binding[60].

Therefore, to study the binding affinity of extended Aβ peptides to amyloid antibodies, we turned to MD simulations and MMPBSA binding free energy calculations using the Amber 16 software suite[30]. First, we needed to validate our computational approach using previously published experimental data. To do this, we compared our MMPBSA binding free energy calculations to the full set of experimental binding affinities reported in the study carried out by Gardberg et al. on PFA1 and PFA2 bound to various Aβ peptides[61]. **FIGURE 1.1** shows the correlation of our MMPBSA calculations with the experimental data (the numerical data are given in **TABLE 1.3)**.

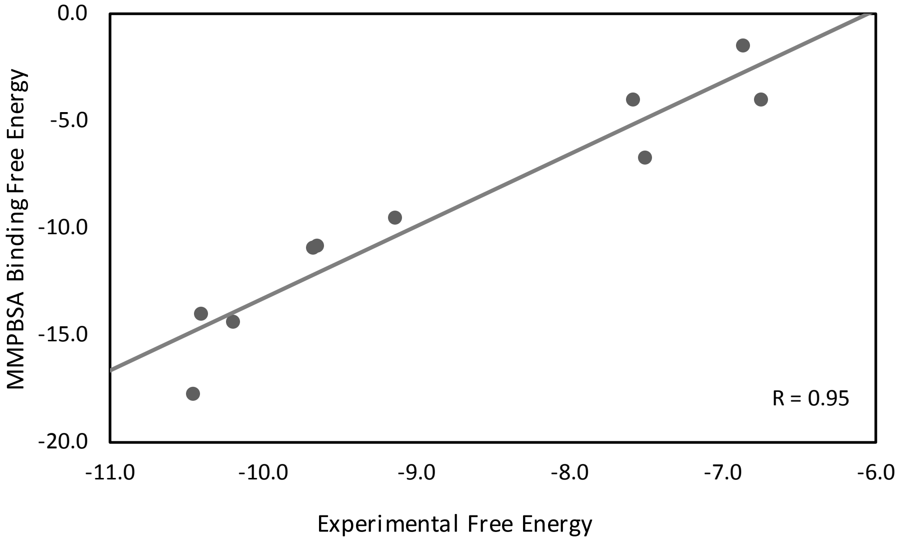


FIGURE 1.1. Method validation for MMPBSA free energy calculations of various Aβ peptides bound to the antibodies PFA1 and PFA2. The original PFA1, PFA2, and pE3-Aβ_{3-8} crystal structures (PDB IDs: 2IPU, 2R0W, and 3EYS) served as the basis for constructing homology models for all the other bound peptide structures. Our calculated MMPBSA binding free energy values were compared to the experimental binding affinity values reported by Gardberg et al. for their entire data set.[61] All free energy values are given in units of kcal/mol.

With a Pearson's R value of 0.95, our MMPBSA calculations show very good agreement with the trend seen in the experimental binding affinity data. **FIGURE A.1** shows the convergence of our MMPBSA free energy values taken over the entire 50 ns production run. It is seen that our data set shows reasonable convergence over this time frame. One data point, for the Aβ_{1-8} peptide bound to PFA1, converged very slowly and underwent a substantial change in its binding affinity over the 50 ns MMPBSA calculation.

TABLE 1.3. Method validation for MMPBSA calculations of Aβ peptides bound to PFA1 and PFA2

Complex Structure	MMPBSA	Experiment
PFA1 (Aβ_{1-8})	-14.3	-10.2**
PFA1 (Aβ_{2-7})	-10.9	-9.7*
PFA1 (pE3-Aβ_{3-8})	-3.9	-7.6
PFA1 (Grip1)	-6.7	-7.5
PFA1 (Ror2)	-17.7	-10.5
PFA2 (Aβ_{1-8})	-14.0	-10.4**
PFA2 (Aβ_{2-7})	-9.5	-9.1*
PFA2 (pE3-Aβ_{3-8})	-4.0	-6.7
PFA2 (Grip1)	-1.4	-6.9
PFA2 (Ror2)	-10.8	-9.7

The results of MMPBSA binding free energy calculations and experimental values are given in units of kcal/mol. Experimental values were taken from Gardberg et al.[61] and were converted to free energy values as described in **METHODS**. * denotes cases where the experimental binding affinity values were reported as a range, and we used the average value of that range as the experimental binding affinity. ** denotes the experimental value given for Aβ_{1-40} binding to both PFA1 and PFA2. The standard error of the mean was 0.1 kcal/mol for each structure.

To verify that the MMPBSA result for Aβ_{1-8} bound to PFA1 had converged properly, we collected 10 ns of additional MD simulation data and ran a 60 ns MMPBSA calculation using the simulation data from 250 ns to 310 ns. The 60 ns MMPBSA result showed that the 50 ns MMPBSA result had indeed already converged (see **FIGURE A.2**). In general, we found that running MD simulations for a total simulation time of 300 ns, and using 5000 frames taken from the last 50 ns for MMPBSA calculations, was sufficient to produce acceptable convergence for our data. This became the standard protocol that we used for any subsequent analysis.

1.3.3 The importance of phenylalanine to the stable binding of A$\boldsymbol{\beta}_{2-7}$ to PFA1

To test the importance of phenylalanine to Aβ peptide-antibody binding, we studied an experimental scenario discussed by Gardberg et al.[61]. The authors carried out a binding assay for Aβ_{2-7} and several other Aβ_{2-7} sequence variants bound to PFA1. They demonstrated that the binding affinity was lowered (from 60 nM to 3400 nM), but not completely abolished, when the wild type sequence, AEFRHD, was changed to the human glutamate receptor interacting protein 1 (or Grip1) sequence, AKFRHD. This was surprising since the charge characteristics completely changed from the negative glutamate residue to the positive lysine residue in the Grip1 peptide[60]. On the other hand, no binding to PFA1 was observed at all when AEFRHD was mutated to AEIRHD (the Position 4 or Pos4 mutant) despite the swapping of two non-polar hydrophobic residues[60-61]. We constructed homology models for the Grip1 and Pos4 mutants, carried out MD simulations, calculated the MMPBSA binding free energies for each, and compared the results to the Aβ_{2-7} MMPBSA binding free energy from our method validation. **TABLE 1.4** shows that the MMPBSA binding affinities qualitatively reproduce the experimental results from the Gardberg study quite well.

TABLE 1.4. The importance of phenylalanine in the binding of A$\boldsymbol{\beta}_{2-7}$ to PFA1

Aβ Peptide	MMPBSA	Experiment
Aβ_{2-7} (AEFRHD)	-10.9	-9.7
Grip1 (AKFRHD)	-6.7	-7.5
Pos4 (AEIRHD)	6.8	No Binding

The results of MMPBSA binding free energy calculations, in units of kcal/mol, are given to demonstrate the severity of the phenylalanine Pos4 mutation (AEIRHD) in the binding of Aβ_{2-7} to the antibody PFA1. The less severe Grip1 mutation (AKFRHD) was also included for comparison. The crystal structure of Aβ_{2-8} bound to PFA1 (PDB ID: 2IPU) was used both for constructing the wildtype Aβ_{2-7} peptide and for generating the homology models of the two other Aβ peptide mutants. Experimental values were taken from Gardberg et al.[31, 61] and were converted to free energy values as described in **METHODS**. The standard error of the mean was 0.1 kcal/mol for each structure.

To examine why the Pos4 (AEIRHD) mutation is much more severe than the Grip1 (AKFRHD) mutation, we first constructed RMSD plots of the 50 ns production runs for each of the three relevant structures. The average RMSD values for the two mutants, Grip1 and Pos4, are clearly higher than the average RMSD for the Aβ_{2-7} structure indicating a larger degree of structural change from the initial structure for the two mutants (see **FIGURE**

A.3). However, the RMSD plots provide no indication that the Pos4 mutant has a significantly lower binding affinity than the Grip1 mutant.

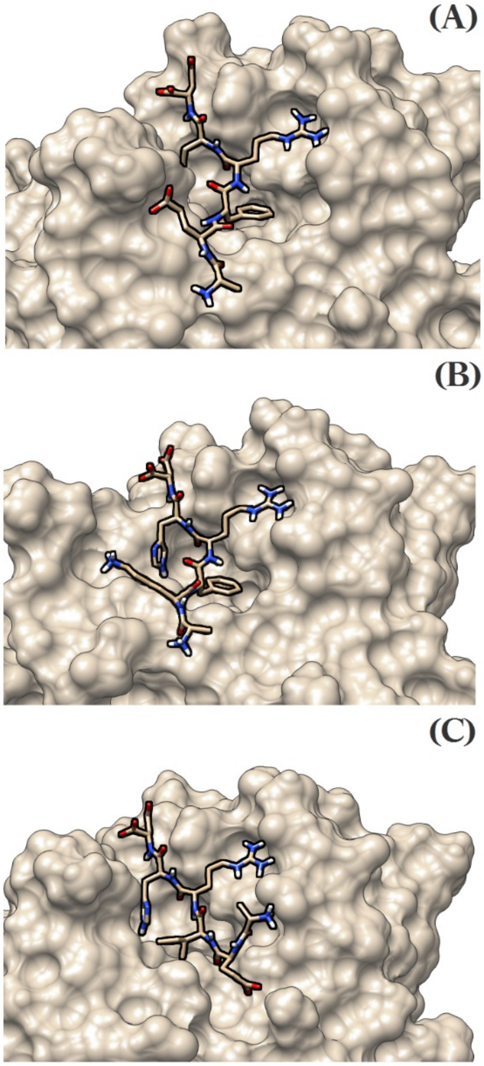


FIGURE 1.2. Comparison of the binding pose for Aβ_{2-7} peptide variants bound to PFA1. Three Aβ peptides are shown bound to PFA1: (A) Aβ_{2-7}, (B) Grip1, and (C) the Pos4 mutant. All three images were taken at the halfway point of the production portion of the MD simulation.

Next, the MD trajectories of the three complexes were visualized and compared with one another. It was observed that both mutant residues were no longer able to bind to their original binding pockets in comparison to the wildtype $A\beta_{2-7}$ peptide. **FIGURE 1.2** shows a representative frame, taken from the halfway point of the MD trajectory, to illustrate the situation for each structure. In the Grip1 structure, the binding pocket and the lysine residue have separated from each other indicating that the binding contact between the two has been disrupted (see **FIGURE 1.2B**). However, the rest of the residues in the $A\beta_{2-7}$ sequence remain bound in their proper orientations (in agreement with structural observations made by Gardberg at al.[61]), which suggests that the loss or alteration of the single glutamate binding interaction is the major cause of the decreased binding affinity of PFA1 for Grip1.

The Pos4 mutation is more complicated. We note first that the isoleucine residue is displaced outside of the deep phenylalanine binding pocket (**FIGURE 1.2C**). Another noteworthy difference is that, when the isoleucine residue was pushed out of the pocket, the front half of the $A\beta$ peptide chain rotated. The glutamate residue disassociated from its normal binding pocket and was found instead binding to a nearby binding pocket that normally binds to either aspartate or alanine, which appear at positions 1 and 2 of the full $A\beta_{1-8}$ peptide respectively (**FIGURE 1.3**). Our unbiased computational docking data predicted that the Pos4 mutant would have a larger destabilizing effect on the binding affinity compared to the Grip1 mutant. From **TABLE 1.1**, phenylalanine had a top binding affinity of -5.9 kcal/mol and -6.1 kcal/mol for the holo and apo forms of PFA1 respectively, while glutamate had a top binding affinity of -4.5 kcal/mol and -4.8 kcal/mol for the holo and apo forms. However, the change in orientation of the binding pose in the Pos4 mutant for the full peptide clearly could not be captured by calculating the binding affinity for a single amino acid residue docked to the antibody surface.

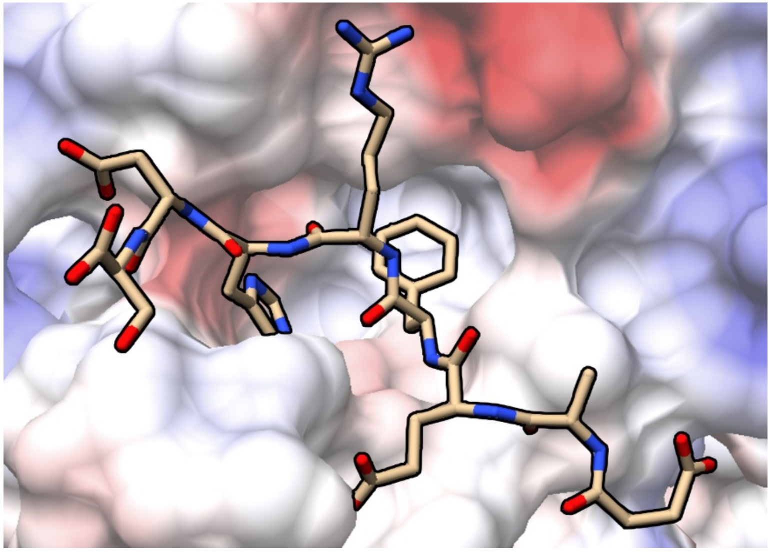


FIGURE 1.3. Electrostatic contacts for Aβ_{1-8} bound to PFA1. A surface map of PFA1 is provided which shows the key electrostatic contacts made between the PFA1 antibody and charged residues in Aβ_{1-8}. The N-terminal aspartate residue (D1) on the Aβ peptide appears in the lower right corner of the figure. Negatively charged regions are depicted in red while positively charged regions are shown in blue. Coulombic Surface Coloring was used to depict electrostatic contacts on the antibody surface where the scale ranges from -10 kcal/mol*e (red, negative region) to 0 kcal/mol*e (white, neutral region) to 10 kcal/mol*e (blue, positive region).

1.3.4 Analysis of gantenerumab and crenezumab binding to multiple Aβ epitopes

Using our prior observation that phenylalanine is a very important residue for the binding of Aβ peptides to anti-amyloid antibodies, we also examined the antigen-combining sites of gantenerumab and crenezumab with the aim of discovering how these two antibodies might bind to both N-terminal and central epitopes of Aβ peptides. We first ran MD simulations and calculated MMPBSA binding free energies for both gantenerumab and crenezumab bound to the Aβ peptides observed in their crystal structures (PDB IDs: 5CSZ and 5VZY). Afterwards, we generated homology models containing the other prominent Aβ epitope that was not present in the original crystal structure for both gantenerumab and crenezumab.

TABLE 1.5. MMPBSA binding free energy results for N-terminal and central Aβ peptides bound to gantenerumab and crenezumab

Complex Structure	MMPBSA	Experiment
gantenerumab (DAEFRHDSGYE)	-33.8	-10.7
gantenerumab forward (HHQKLVFFAEDV)	1.8	-
crenezumab (EVHHQKLVFFAEDVG)	-15.2	-11.7
crenezumab reverse 2 (SDHRFEAD)	-3.1	-

The crystal structures of N-terminal Aβ (DAEFRHDSGYE) bound to gantenerumab (PDB ID: 5CSZ) and central Aβ (EVHHQKLVFFAEDVG) bound to crenezumab (PDB ID: 5VZY) were used both to calculate the MMPBSA binding free energy for the crystal structure-based complexes and for generating the four homology models of the transposed epitopes for each. Only the most stable of the four homology models of the transposed epitopes for each antibody were used for analysis. These are listed as gantenerumab forward and crenezumab reverse 2. The other structures were less stable and were omitted from our analysis (data not shown). The colored type for phenylalanine reveals the residue that was used to line up the phenylalanine residue in each homology model to the phenylalanine residues(s) in the original crystal structure (residues that are colored the same indicate the phenylalanine residue that was used for the alignment). Experimental binding free energies are given for the antibody structures bound to the monomer form of Aβ[23, 26]. The results of MMPBSA binding free energy calculations are given in units of kcal/mol. The standard error of the mean was 0.1 kcal/mol for each structure.

Four homology models that featured the new peptide were generated where: 1) each new Aβ peptide was modeled in both forward and reverse orientations across the antigen-combining site, and 2) the location of phenylalanine residues in the Aβ epitopes were used for the initial alignment in the antigen-combining site. Calculated MMPBSA binding free energies are given in **TABLE 1.5** while the convergence for our MMPBSA values is demonstrated in **FIGURE A.4**. Only the most stable homology model of the transposed Aβ peptide bound to each antibody was used for analysis; we have labeled them as “gantenerumab forward” and “crenezumab reverse 2” to reflect their initial alignment. The other transposed epitope structures were less stable and were deemed unsuitable for further analysis (data not shown).

The N-terminal peptide, DAEFRHDSGYE, bound to gantenerumab yielded a calculated binding free energy of -33.8 kcal/mol. **FIGURE 1.4** compares the initial pose of the wildtype N-terminal peptide bound to gantenerumab (**FIGURE 1.4A** and 1.**4B**) to snapshots taken at the first, middle, and last frames of the MD production run (**FIGURE 1.4C**, **1.4D**, **1.4E**, and **1.4F**) for our most stable model of gantenerumab bound to the central Aβ peptide. The source of the strong binding affinity for the N-terminal Aβ peptide seems apparent from an examination of **FIGURE 1.4B** as the antigen-combining site of gantenerumab exhibits several positive (blue) electrostatic contact sites on its surface that correlate with the many negatively charged aspartate and glutamate (red) residues that appear on the N-terminal Aβ peptide. Calculated percent occupancy values for various sites can be found in **TABLE A.1**.

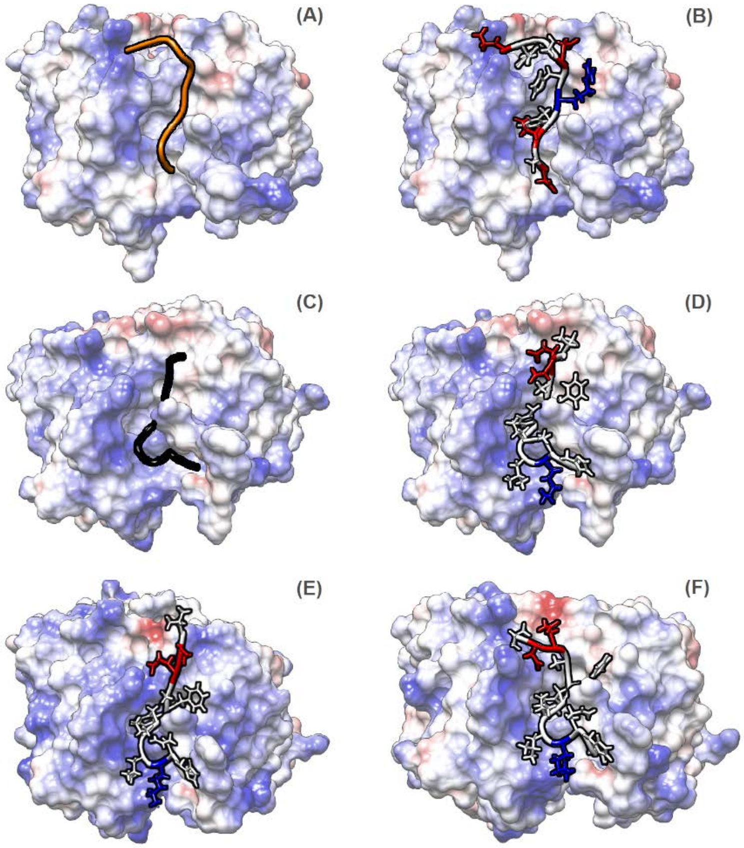


FIGURE 1.4. Gantenerumab bound to both N-terminal and central Aβ peptides. Gantenerumab is shown bound to the Aβ peptide containing the N-terminal epitope (PDB ID: 5CSZ) in the first frame of the MD simulation in (A) and (B). Structures (C), (D), (E), and (F) show the most stable central Aβ peptide bound to gantenerumab in the forward sequence (HHQKLVFFAEDV) across the gantenerumab antigen-combining site taken from the first frame (C and D), the middle frame (E), and the last frame (F) of the MD production run. In all structures, the N-terminus end of the peptide appears towards the bottom of the

antigen-combining site while the C-terminus appears near the top. Coulombic Surface Coloring was used to depict electrostatic contacts on the antibody surface where the scale ranges from -10 kcal/mol*e (red, negative) to 0 kcal/mol*e (white, neutral) to 10 kcal/mol*e (blue, positive). The residues on the peptide were colored as acidic (red), basic (blue), or neutral (white).

Comparing the original N-terminal Aβ peptide binding pose in **FIGURE 1.4B** to the binding pose of the transposed central Aβ peptide in **FIGURE 1.4D**, **1.4E**, and **1.4F**, we see that most of these electrostatic contacts have been lost when the central Aβ peptide is bound as it does not contain as many charged residues as the N-terminal peptide. However, there appear to be at least two potential binding interactions when the central Aβ peptide is bound to gantenerumab: E22 to R57 and F19 to F119 (both from the VFFAED portion of the epitope) appear to be close enough to interact in both structures. The percent occupancy for E3 to R57 of 100% when the N-terminal Aβ peptide is bound only drops to 85.58% for E22 to R57 when the central Aβ peptide is bound, and the percent occupancy for F4 to F119 of 96.12% when the N-terminal Aβ peptide is bound actually increased to 99.92% for F19 to F119 when the central Aβ peptide is bound. However, it is worth noting that D23 appears to be competing to some extent with E22 for occupancy near the R57 binding site as indicated by a noticeable 45.54% occupancy for D23 to R57 when the central Aβ peptide is bound. In our figures, this competition can even be seen, as E22 is clearly bound to the pocket in **FIGURE 1.4E** whereas D23 is bound to that same site in **FIGURE 1.4F**. Outside of these two interactions, a likely interaction for the N-terminal peptide between R5 and Y93, with a percent occupancy of 100% (see **FIGURE 1.4B)**, is clearly lost for the central Aβ peptide in **FIGURE 1.4D**, **1.4E**, and **1.4F** as the possible corresponding interaction between K16 and Y93 has a 0% occupancy.

The above observations suggest that the FAED region of the alternate Aβ epitope still may form at least a few binding contacts with the gantenerumab antigen-combining site. This is in qualitative agreement with experimental observations for gantenerumab, which display an epitope in the VFFAEDVGSN region, but do not display an epitope emerging from neighboring regions of the central epitope sequence that include the sequence HHQKL for instance[23]. The much less stable binding free energy of 1.8 kcal/mol that we obtained for the central Aβ peptide bound to gantenerumab is also consistent with the observation that gantenerumab is unable to bind and alter soluble Aβ levels in contrast to what was observed for solanezumab, which preferentially recognizes the central Aβ epitope[23, 62].

The central Aβ peptide bound to crenezumab yielded a calculated binding free energy of -15.2 kcal/mol. The transposed N-terminal peptide had a weaker binding free energy of -3.1 kcal/mol. While the calculated free energy was smaller for the N-terminal peptide, the gap in free energy between the bound N-terminal and central Aβ peptides was much smaller for crenezumab than it was for gantenerumab. It is noteworthy to mention that the most stable transposed N-terminal peptide was obtained when the phenylalanine residue in the reverse N-terminal epitope, SDHRFEAD, was aligned with F20 in the original central Aβ epitope, HHQKLVFFAEDV, prior to model building. Binding in this reverse sense in the antigen-combining site of an anti-Aβ antibody has been noted before as solanezumab binds to the central Aβ peptide in this reverse sense compared to the orientation adopted by the N-terminal peptide in gantenerumab.[60] The original binding pose of the central Aβ

peptide bound to crenezumab is shown in **FIGURE 1.5A** and **1.5B**. This may be compared to the binding pose observed for the N-terminal peptide bound to crenezumab during the production portion of our MD simulation in **FIGURE 1.5C**, **1.5D**, **1.5E**, and **1.5F**. Compared to the initial binding pose of the central Aβ peptide bound to crenezumab, it is seen that the binding site of crenezumab has had to distort considerably to bind to the N-terminal Aβ peptide in an alternate conformation.

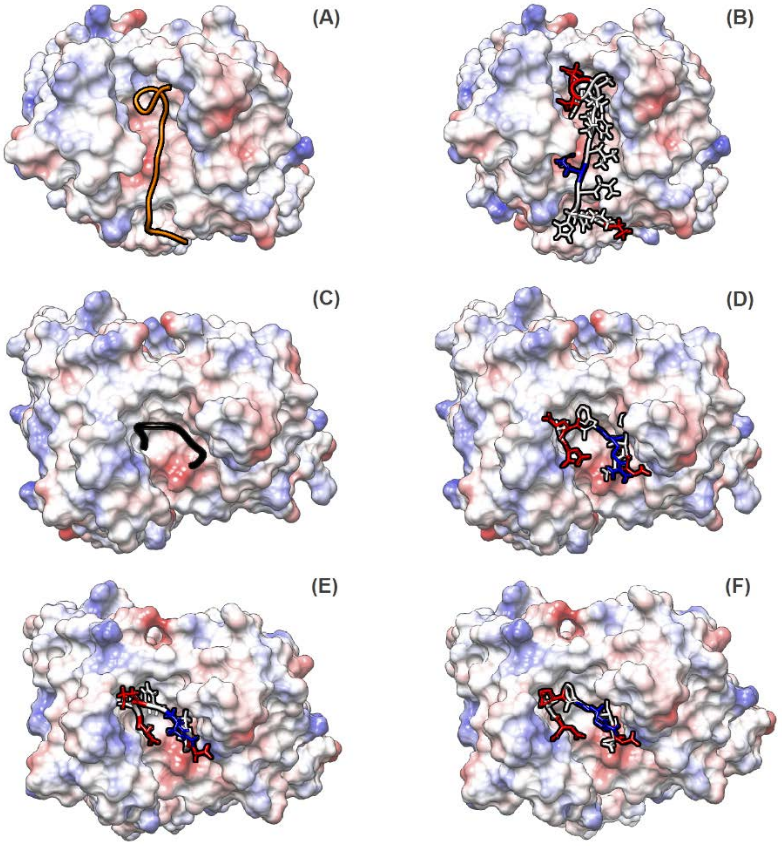


FIGURE 1.5. Crenezumab bound to both central and N-terminal Aβ peptides. Crenezumab is shown bound to the Aβ peptide containing the central epitope (PDB ID: 5VZY) in the first frame of the MD simulation in (A) and (B). Structures (C), (D), (E), and (F) show the N-terminal Aβ peptide bound to crenezumab in the reverse sense (SDHRFEAD) across the crenezumab antigen-combining site as observed in the first frame (C and D), the

middle frame (E), and the last frame (F) of the MD production run. In structures (A) and (B), the N-terminus end of the peptide appears towards the bottom of the antigen-combining site while the C-terminus appears near the top. For (C), (D), (E), and (F), the N-terminus end appears towards the upper left of the antigen-combining site while the C-terminus end appears towards the lower right of the antigen-combining site. Coulombic Surface Coloring was used to depict electrostatic contacts on the antibody surface where the scale ranges from -10 kcal/mol*e (red, negative region) to 0 kcal/mol*e (white, neutral region) to 10 kcal/mol*e (blue, positive region). The residues on the peptide were colored as acidic (red), basic (blue), or neutral (white).

The pattern in potential binding interactions for crenezumab is qualitatively similar to what we saw before with gantenerumab, but in general the percent occupancy for the various sites that we examined are usually higher for the transposed Aβ peptide bound to crenezumab. For comparison with **TABLE A.1**, the results of percent occupancy calculations for crenezumab are available in **TABLE A.2**. The percent occupancy is 99.0% for F19 to V94 when the central Aβ peptide is bound and is 100% for F4 to V94 when the N-terminal Aβ peptide is bound. The negatively charged E3 residue also appears to be attracted to a nearby binding pocket as seen in **FIGURE 1.5E** and **1.5F**; the percent occupancy for E22 to N52 is 100% when the central Aβ peptide is bound and only drops to 92.28% for E3 to N52 when the N-terminal peptide is bound. In addition, the D23 to N53 percent occupancy of 36.94% when the central Aβ peptide is bound has actually increased to 86.3% for D1 to N53 when the N-terminal Aβ peptide is bound. Also, similar to the situation for gantenerumab, the positively charged residue in the alternate peptide does not stably bind to crenezumab. The percent occupancy of 100% for K16 to D101 when the central Aβ peptide is bound drops to 0% for R5 to D101 when the N-terminal Aβ peptide is bound.

In an attempt to understand why we were able to obtain a slightly stable binding free energy for the reverse epitope bound to crenezumab, it is noteworthy to point out the electrostatic similarities between the N-terminal DAEFRHD epitope and the reverse sequence of the central epitope KLVFFAED, which when written out in reverse becomes DEAFFVLK. When both sequences are compared in this way, it becomes apparent that the first four amino acids in the N-terminal sequence and the first four amino acids in the reverse central epitope sequence are very similar to each other, with DAEF and DEAF differing only by switching the internal positions of the alanine and glutamate residues. However, despite this similarity, the inability to interchange the binding of key arginine and lysine residues elsewhere in the sequence may help explain how crenezumab achieves its preference for the central Aβ peptide. The inability for lysine and arginine residues to cross bind for both gantenerumab and crenezumab is likely due to a difference in the positioning of these residues in both the sequence and three-dimensional space; the arginine residue appears next to the DAEF sequence in the N-terminal epitope (DAEFRHDS) whereas the lysine residue appears three residues away from the corresponding FAED sequence in the central epitope (HHQKLVFFAEDV).

1.3.5 Predicting the binding affinity between pE3-Aβ_{3-8} and PFA1 mutants

With an eye towards the future, we wanted to see if we could use a visual inspection of our MD simulations to rationally plan out single amino acid mutations that improve the calculated MMPBSA binding free energy between an anti-Aβ antibody and a polymorphic

Aβ species. Such a computational approach has been successfully demonstrated before[32], and it is a cost-effective way to probe for antibody mutations that could potentially improve the binding strength and specificity of an antibody for an additional Aβ target. The most promising mutants that are identified may then be produced and tested in a laboratory to confirm the predicted improvement in binding affinity at a later point.

In order to illustrate this approach, we examined a question posed by Gardberg et al. in their study of pE3-Aβ_{3-8} bound to the anti-protofibril antibody, PFA1. In that study, the pE3-Aβ_{3-8} amyloid peptide was shown to bind to PFA1 with less affinity ($K_D = 3000$ nM) than the wild type Aβ_{2-7} peptide ($K_D = 60$ nM)[31]. pE3-Aβ_{3-8} is still considered to be dangerous due to its prominent presence in Alzheimer's plaques, and a mutant antibody that can bind to both Aβ species with high affinity would be more desirable as a potential drug candidate. To probe for such a mutant, we first ran a preliminary MD simulation of PFA1 bound to pE3-Aβ_{3-8} and compared it to a MD simulation of PFA1 bound to Aβ_{1-8} that had been used previously in our method validation. Of particular interest in the MD simulation for Aβ_{1-8} was that the glutamate residue was found to be localized fairly well inside of its binding pocket as would be expected from analyzing the initial structure (see **FIGURE 1.3**). In contrast, the MD simulation for pE3-Aβ_{3-8} revealed that the terminal pyroglutamate (PCA3) residue was engaged in a tug-of-war of sorts between the glutamate binding pocket and another nearby binding pocket that normally belongs to either alanine or aspartate.

Based on the MD simulations, we proposed that the difference in the observed binding affinity was caused by either: 1) the PCA3 residue in pE3-Aβ_{3-8} lacking the full negative charge of the glutamate residue, weakening its attraction to the glutamate binding pocket, or 2) the loss of alanine in pE3-Aβ_{3-8}, opening up the possibility for PCA3 to be attracted away from the glutamate binding pocket towards the other nearby unoccupied binding pocket. On these grounds, we computationally introduced two sets of single amino acid substitution mutations into the binding pocket of PFA1, carried out MD simulations, and calculated the MMPBSA binding free energies between the mutant structures and the pE3-Aβ_{3-8} peptide. The first two mutants, Y59A and N60A on the H chain, were designed to weaken the attraction of PCA3 towards the other nearby binding pocket. The next two mutants, S92K and H93K on the L chain, were designed to strengthen the positive electrostatic character in the glutamate binding pocket.

TABLE 1.6. MMPBSA binding free energy results for A$\boldsymbol{\beta}_{1-8}$, A$\boldsymbol{\beta}_{2-7}$, and pE3-A$\boldsymbol{\beta}_{3-8}$ bound to the wildtype and several mutant forms of PFA1

Complex Structure	MMPBSA
Aβ_{1-8}-PFA1 wildtype	-14.3
Aβ_{2-7}-PFA1 wildtype	-10.9
pE3-Aβ_{3-8}-PFA1 wildtype	-3.9
pE3-Aβ_{3-8}-PFA1 mutant (Y59A (H chain))	-8.4
pE3-Aβ_{3-8}-PFA1 mutant (N60A (H chain))	-10.2
pE3-Aβ_{3-8}-PFA1 mutant (S92K (L chain))	5.3
pE3-Aβ_{3-8}-PFA1 mutant (H93K (L chain))	-2.7
Aβ_{1-8}-PFA1 mutant (N60A (H chain))	-18.0
Aβ_{2-7}-PFA1 mutant (N60A (H chain))	-8.3
Aβ_{1-8}-PFA1 mutant (Y59A (H chain))	-16.7
Aβ_{2-7}-PFA1 mutant (Y59A (H chain))	-9.7

The crystal structure of pE3-Aβ_{3-8} bound to PFA1 (PDB ID: 3EYS) was used both for the wildtype pE3-Aβ_{3-8}-PFA1 complex and for generating the homology models of the four pE3-Aβ_{3-8}-PFA1 mutant complexes. The crystal structure of Aβ_{2-8} (PDB ID: 2IPU) was used to generate the homology models of the Aβ_{1-8} and Aβ_{2-7} peptides bound to the N60A and Y59A mutant forms of PFA1. The results of MMPBSA binding free energy calculations are given in units of kcal/mol. The standard error of the mean was 0.1 kcal/mol for each structure.

The MMPBSA results are given in **TABLE 1.6**, convergence plots for the calculated MMPBSA binding free energies are shown in **FIGURE A.5**, and the results of percent occupancy calculations are available in **TABLE A.3**. Two out of the four proposed mutants were able to lower the calculated binding free energy by an appreciable amount, predicting that these mutations will stabilize the bound structure. The calculated binding free energy for pE3-Aβ_{3-8} bound to the N60A PFA1 mutant was -10.2 kcal/mol, and the calculated binding free energy for the Y59A PFA1 mutant was -8.4 kcal/mol. These were both more favorable than the binding free energy of -3.9 kcal/mol calculated for pE3-Aβ_{3-8} bound to the PFA1 wildtype structure. The more favorable binding free energies were also comparable to values we obtained of -14.3 kcal/mol and -10.4 kcal/mol for the original wild type Aβ_{1-8} and Aβ_{2-7} peptides bound to PFA1 respectively.

Examining the MD trajectories (**FIGURE 1.6**) revealed that our N60A mutation worked out more or less as we expected. The PCA3 residue was now localized closer to the two histidine residues near the glutamate binding pocket for the majority of the simulation time. For the N60A mutant, the percent occupancy for PCA3 to H27D of 97.9% and for PCA3 to H93 of 99.9% were large increases over the corresponding wildtype values of 19.7% and 44.0% respectively. In addition, the percent occupancy for PCA3 to S58, a residue near the N60A mutation site, was reduced from 50.9% in the wildtype to 8.6% in the N60A mutant.

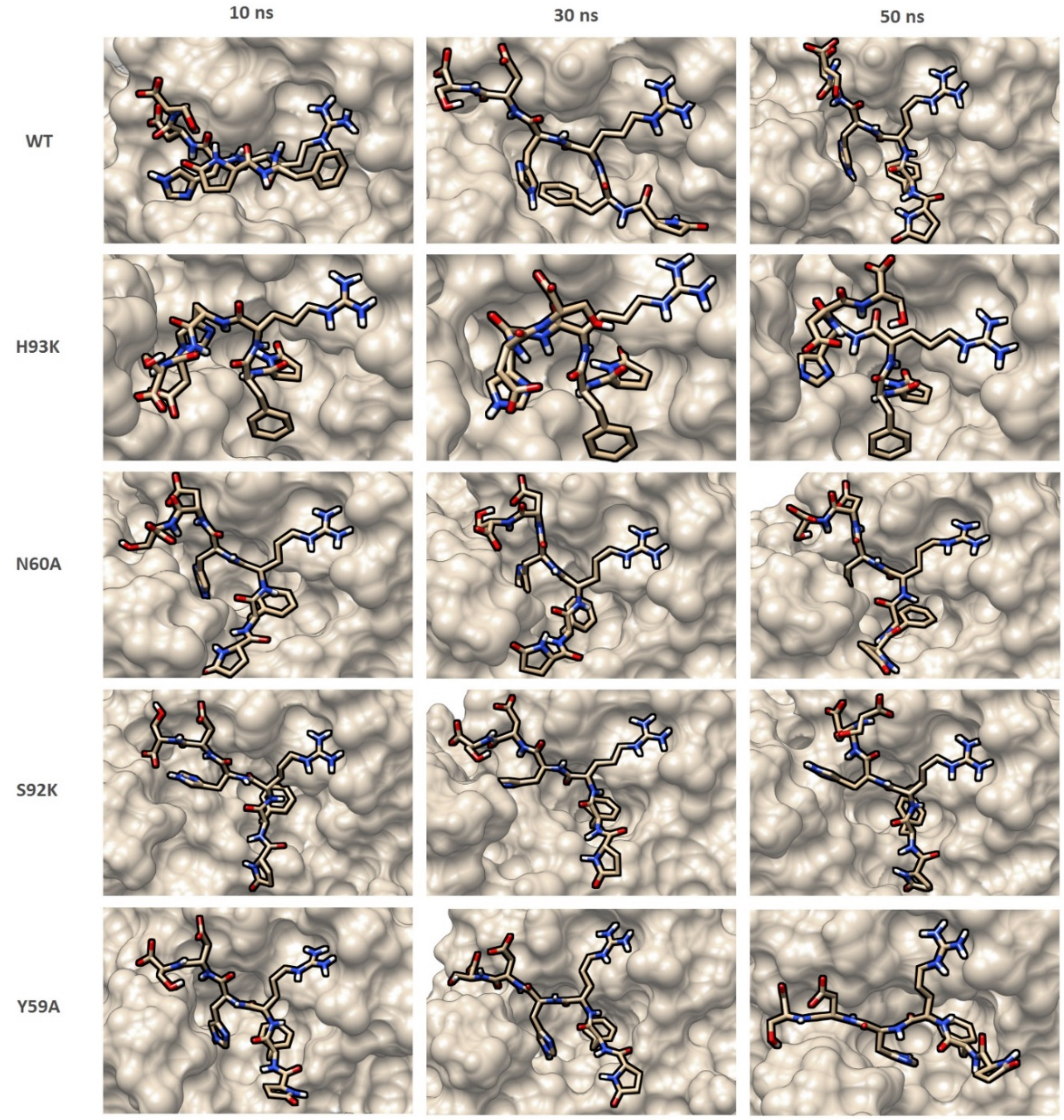


FIGURE 1.6. Snapshots from the MD trajectories of pE3-Aβ_{3-8} bound to wild type and mutant forms of PFA1. The structures here are visualized as snapshots taken at 10 ns, 30 ns, and 50 ns during the production run of the MD simulation.

In a somewhat similar way, our Y59A mutant appeared to stabilize PCA3 by reducing its movement, but this time the PCA3 residue localized more towards the other nearby binding pocket, on the central right side, as indicated by an increase in percent occupancy from 50.9% to 81.0% for PCA3 to S58. Unlike in the N60A mutant, the PCA3 residue showed only a modest increase in its localization towards the H27D and H93 residues with percent occupancy values of 25.8% and 62.4% compared to 19.7% and 44.0% in the wildtype respectively.

The other two mutations are not predicted to improve the binding affinity of pE3-Aβ_{3-8} towards the mutant antibody. The S92K mutant had a calculated binding free energy of 5.3 kcal/mol that indicated a strong destabilization of the bound structure. For the S92K mutant, we note that the PCA3 to H27D percent occupancy decreased considerably from 19.7% to 1.0%, and also the PCA3 to L96 percent occupancy increased from 17.3% to 76.1%, indicating that perhaps an increase in percent occupancy at this particular site, which happens to be the phenylalanine binding site, is disruptive to binding.

The H93K mutant was also destabilized as its calculated binding free energy of -2.7 kcal/mol was a bit lower than that of the wild type pE3-Aβ_{3-8} peptide bound to PFA1. It failed to pull PCA3 residues towards the H27D and H93 locations as we intended; the PCA3 to H27D percent occupancy was lowered from 19.7% in the wildtype to 0% in the H93K mutant while the PCA3 to K93 (the position which was previously H93) percent occupancy was lowered from 44.0% in the wildtype to 9.8% in the H93K mutant. Analysis of the MD trajectory also revealed that the pE3-Aβ_{3-8} peptide had undergone drastic changes in its binding pose for the H93K mutant. The entire backbone of pE3-Aβ_{3-8} was flipped outwards toward the solvent, which allowed the PCA3 residue to move into the deep central binding pocket, displacing phenylalanine. The PCA3 to L96 percent occupancy increased to 100% for the H93K mutant compared to 17.3% in the wildtype.

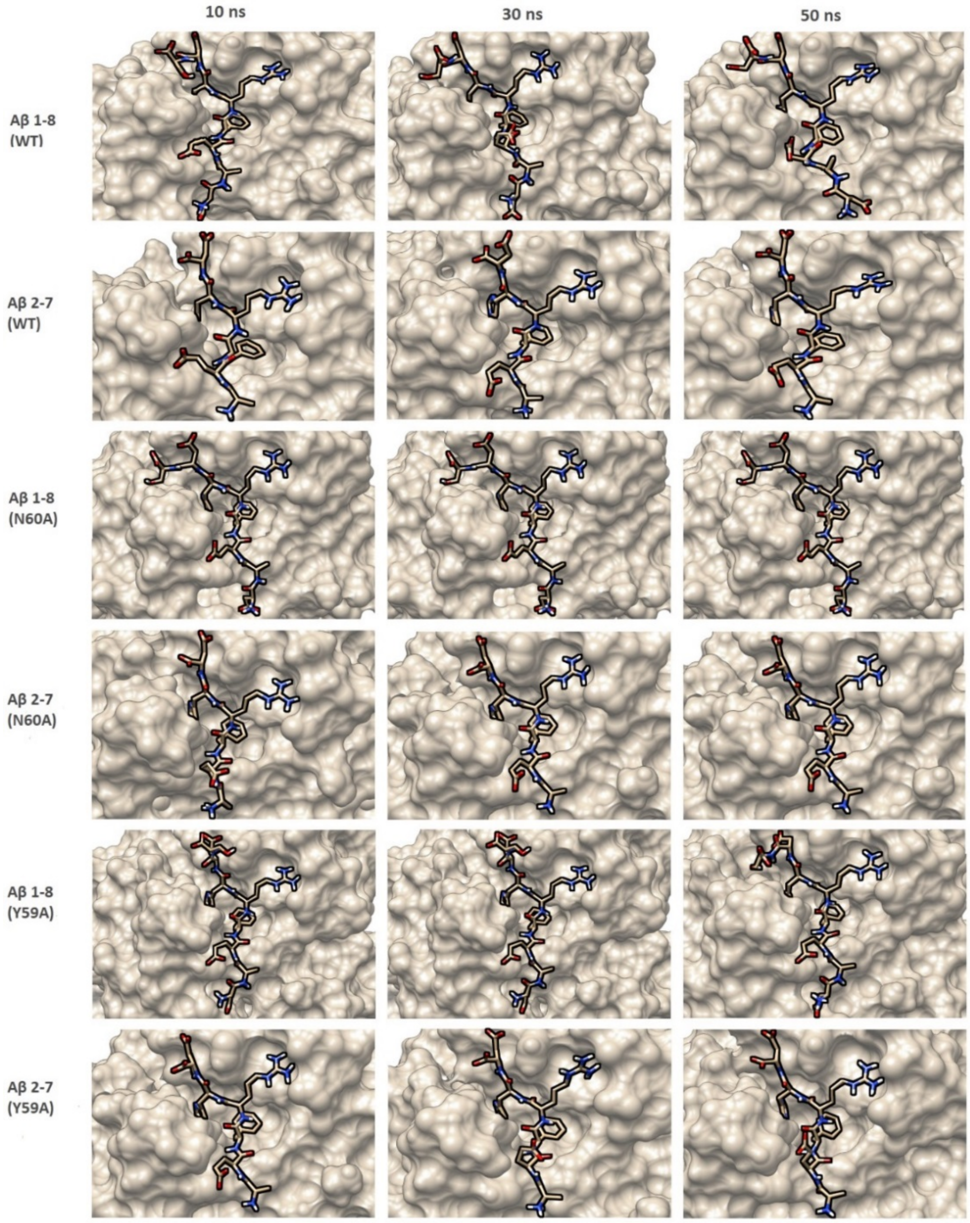


FIGURE 1.7. Snapshots from the MD trajectories of A$\boldsymbol{\beta}_{1-8}$ and A$\boldsymbol{\beta}_{2-7}$ bound to wild type and mutant forms of PFA1. The structures here are visualized as snapshots taken at 10 ns, 30 ns, and 50 ns during the production run of the MD simulation.

Given the rather large changes in binding pose and affinity for the pE3-Aβ_{3-8} peptide towards our PFA1 mutants, it is also reasonable to ask what changes might take place for the binding of the original wild type Aβ_{1-8} and Aβ_{2-7} peptides to our two successful mutant antibodies. This is an important issue if we wish to find a single antibody that is capable of binding with high affinity to multiple Aβ species. We therefore also examined MD simulations and calculated MMPBSA binding free energies for Aβ_{1-8} and Aβ_{2-7} bound to our N60A and Y59A mutant forms of PFA1. The binding free energies are given in **TABLE 1.6**, representative snapshots from the MD trajectories are shown in **FIGURE 1.7**, and the results of percent occupancy calculations for both wild type and mutant forms are available in **TABLE A.4** and **A.5** for Aβ_{1-8} and Aβ_{2-7} respectively.

For both mutants, the binding affinity for Aβ_{1-8} was actually predicted to be improved over the wildtype PFA1 antibody. The calculated binding free energy for Aβ_{1-8} bound to the N60A PFA1 mutant was -18.0 kcal/mol, and the calculated binding free energy for the Y59A PFA1 mutant was -16.7 kcal/mol. One possible explanation for such an increase in binding affinity is that the percent occupancy of D1 to N27 increased to 99.9% in the N60A mutant and to 83.8% in the Y59A mutant compared to 79.1% in the wildtype. In contrast, both mutations are predicted to slightly destabilize the binding of the Aβ_{2-7} peptide as the calculated binding free energy for Aβ_{2-7} bound to the N60A PFA1 mutant was -8.3 kcal/mol, and the calculated binding free energy for the Y59A PFA1 mutant was -9.7 kcal/mol. A plausible explanation is that the mutation interfered with the binding of alanine, which is present in Aβ_{2-7}, and also that a compensating D1 to N27 interaction is completely missing for this case. Nevertheless, if we compare the computed binding free energies of the wild type PFA1 for Aβ_{1-8}, Aβ_{2-7}, and pE3-Aβ_{3-8} species to that of our N60A and Y59A mutants, we predict a general increase in the binding affinity for our various Aβ species to the mutant antibodies albeit with a modest sacrifice in affinity for Aβ_{2-7}.

1.4 DISCUSSION

In the preceding sections, we have shown how fragment-based docking and MD simulations, accompanied by MMPBSA binding free energy calculations, can assist in the study of anti-Aβ antibodies for rational drug design efforts. Employing a fragment-based docking method provided us with a means to dock single amino acid residues in an unbiased fashion to the surface of anti-Aβ antibodies to probe for key anchoring residues that are involved in the initial binding interaction. Our docking approach borrows heavily from current ideas being used in computational fragment-based drug design. In these methods, potential binding sites are located by probing the surface of a large protein receptor using small fragments of a ligand as opposed to attempting to dock the entire ligand[33-34]. A potential disadvantage to using small fragments to probe the receptor is a decrease in the binding selectivity for the ligand; additional binding sites other than the primary site of interest are often identified as potential binding sites[63]. On the other hand, conformational possibilities for a small fragment are much lower than for a larger ligand, and it has been shown that the leading hot spots identified using computational fragment-based methods correlate well with actual ligand binding sites[34]. In the present case, we were fortunate to be examining antibodies where the location of the ligand binding site was already known beforehand.

The primary shortcoming of using single amino acid fragment-based docking to study the binding of peptide ligands is that spatial and conformational information that is important to the binding of the full peptide is lost. For example, our MD simulation of Aβ_{2-7}

peptide variants bound to PFA1 suggested that the large decrease in binding affinity of the Pos4 mutant was not solely due to a disruption of the single phenylalanine binding interaction as several neighboring amino acid binding interactions were also disrupted. Attempting to dock dipeptide (or larger) fragments from the full amino acid sequence to test for spatial effects introduces additional complications to the docking protocol. For example, multiple possibilities now exist when ranking dipeptide fragments, such as the first amino acid binding to a pocket, but not the second, or both amino acids binding to separate pockets simultaneously. An additional conformational search step of the dipeptide ligand might be carried out in advance of the binding affinity calculation to optimize the binding pose, but this would come at the cost of greatly increasing both the difficulty and computational expense of applying this approach[64-65].

For situations where only an apo crystal structure of the antibody is available to study, docking single amino acids might yield some useful information as to which amino acids are crucial to the binding of an extended epitope. We showed that this approach was able to predict the emergence of the EFRH epitope observed in many anti-Aβ antibodies. In particular, phenylalanine emerged as a dominant interaction, displaying the most stable binding free energy and very consistently docking into the antigen-combining sites of all ten antibodies. This observation that phenylalanine is a central anchor in the binding of Aβ peptides to anti-Aβ antibodies is corroborated by experimental observations from various crystallographers[23, 26, 60-61]. However, we believe that using MD simulations and MMPBSA calculations to probe the ligand binding properties in holo structures is highly preferable, if this option is available, since the binding pose of the full peptide ligand is already established in this case.

The most noteworthy observation from our MD exploration of Aβ epitope cross binding was that our data indicated that crenezumab may bind to both N-terminal and central Aβ epitopes, although the binding to the N-terminal region is predicted to be much weaker. This topic is of interest because cross binding between these two epitopes may be related to aducanumab's reported ability to bind to both oligomeric and fibril Aβ species, and this may be a factor in its more optimistic outlook as a drug candidate. Additionally, it has been observed experimentally that crenezumab has a puzzling ability to bind to both soluble monomers and oligomers, using the central epitope of Aβ, and to insoluble amyloid plaques where it is believed that only the N-terminal epitope is readily accessible[25-26]. This behavior contrasts sharply with solanezumab which can only bind to soluble monomers using the central epitope of Aβ and not to fibril structures via the N-terminal epitope.

To explain the cross binding to both fibril and oligomeric species observed for crenezumab, Ultsch et al. suggested that the Aβ fibril species may have defects that expose the central Aβ epitope along the fiber axis to allow crenezumab to bind to it in a few locations[26]. While this is certainly possible, the slightly stable MMPBSA binding affinity that we observed for binding the N-terminal peptide to crenezumab presents an alternative explanation; a weak binding affinity for the N-terminal epitope would give crenezumab a chance to bind to fibril structures to some extent. Although crenezumab's binding affinity for the fiber form would be lower than for the monomer or oligomer forms that bind via the central epitope, the high effective concentration of potential N-terminal binding sites along a fiber axis could still account for the sporadic fiber binding pattern observed by Ultsch et al.[26]. We should note that other possibilities for cross binding to fibers exist than the two proposed above. Ma et al. has recently pointed out that crenezumab can recognize the Aβ 13-16 epitope which may also be exposed in fibers to allow the antibody to bind[66]. It is important to note that our analysis assumed that a conserved interaction between the antibody and phenylalanine was maintained if cross binding occurred, but this assumption may not necessarily hold if Aβ 13-16 is responsible for cross binding. An interesting feature of the Aβ 13-16 epitope is that it contains two side by side histidine residues, implying that a strong pH dependence for fiber binding may help decide experimentally between this possibility and the others mentioned above.

Concerning our MD methodology, our analysis of gantenerumab and crenezumab is necessarily more limited and qualitative than our analysis of the PFA1 system. In contrast to our study of PFA1, we lacked a full set of experimental binding affinities that would have allowed us to validate our gantenerumab and crenezumab data and draw more quantitative conclusions for this portion of the study. In addition, while we have identified a few short sequences in the N-terminal and central Aβ epitopes that may initiate a cross binding event, other important aspects that can affect binding selectivity, in particular the role of entropy in conformational selection for the full extended polypeptide sequence[60, 66-68], still need to be quantified and studied in more detail. In our MD simulations, we observed a large amount of conformational flexibility in the both the bound Aβ peptides and the amyloid antigen-combining sites of our various model systems. Such conformational movement made it difficult to obtain converged free energy values for our system, which required long 300 ns MD simulations and extensive 50 ns MMPBSA calculations to obtain acceptable convergence. We also used residue-to-residue percent occupancy values in our analysis, as opposed to more precise atom-to-atom calculations, due to the large amount of conformational flexibility that we observed. However, available methods that attempt to quantify these effects for protein-ligand interactions in the form of entropy calculations are unsatisfactory[43], and much effort is currently aimed at improving the accuracy and efficiency of such calculations[69-72]. In particular, the single trajectory MD/MMPBSA method that we have employed does not take into account any significant conformational changes that may occur between the bound and unbound states, and the multi-trajectory method that attempts to address this issue is known to have major issues with convergence[71]. Therefore, omitting entropy calculations is a standard practice when we apply the single trajectory MD/MMPBSA approach as we have done here and in past work[43, 73]. A truly deep understanding of binding selectivity will need to take into account entropic effects in a more quantitative manner, and finding a way to accurately quantify the entropic effects in a highly flexible system like this would be an interesting avenue for a future study.

Finally, we computationally introduced rational mutations into the PFA1 antibody in an attempt to predict mutants that will improve the binding affinity of PFA1 towards the pE3-Aβ species of Aβ. Given the high conformational flexibility of both the Aβ peptide and the antibody itself, predicting useful mutations from static crystal structures would be difficult, if not impossible, to do. On the other hand, visual inspection of the MD simulations of the amyloid-antibody complex allowed us to rationally identify potential mutation hot spots on the antibody surface with relative ease. Two out of the four mutants we tested are predicted to stabilize the binding of pE3-Aβ_{3-8} to PFA1. However, this came at a slight cost as the computed binding affinity for Aβ_{2-7} is predicted to be slightly lower in the mutant antibodies. In practice, this may not be much of an issue, as a compromise between the binding of Aβ species for a given antibody can be circumvented entirely by going beyond the single antibody approach. Instead, a cocktail of similar antibodies can be used to target various key Aβ species in a treatment regime. In this case for instance, both wildtype and N60A mutant forms could be used together in a proposed treatment option to maximize effectiveness.

Our results here only serve to illustrate the potential for applying a computational method to help design prospective mutant structures; many other possible mutations may exist that can impact the binding of pE3-Aβ_{3-8} in a similar or even better way. If a computational pre-screening approach such as this is carried out, it should produce a list of potential hot spot mutations that can then be tested in the lab to confirm the predicted effectiveness. If the results from the lab correlate with the predicted computational results, it may lead to an improved drug candidate with a greater potential to treat AD.

1.5 CONCLUSIONS

Anti-Aβ antibodies are currently the most advanced treatment option on the horizon for patients suffering from AD. The recent positive clinical results reported for adacanumab indicate the potential for these drugs to be effective at clearing plaque burden and reducing cognitive decline. The main issues at present are the serious autoimmune side effects associated with certain anti-Aβ antibodies and the presence of important polymorphic forms of Aβ that may not bind to an antibody drug candidate with the same high affinity as its primary Aβ target, decreasing its effectiveness.

Although it remains a possibility to obtain new antibody drug candidates from standard drug screening procedures, these are expensive and time-consuming endeavors to undertake, and there is no way of knowing what impact the drug will have on a human population until a very late stage of drug development. Alternatively, there exists the possibility to rationally modify and improve current anti-Aβ antibody drugs that have already undergone clinical trials and whose strengths and weaknesses as a drug candidate are at least somewhat understood.

For the purpose of rationally improving anti-Aβ antibody drugs, we have outlined a computational approach for studying the antigen-combining site of anti-Aβ antibodies using fragment-based docking and full molecular dynamics simulations accompanied by MMPBSA binding free energy calculations. Our fragment-based docking method successfully predicted the emergence of the common EFRH epitope, and its application identified phenylalanine in particular as a key anchoring residue. In our study of Aβ_{2-7} peptide variants bound to PFA1, we showed that MD simulations were capable of reproducing experimental observations as well as providing new insights into previously observed experimental results. We also demonstrated that MD simulations can be used as a tool to assist in exploratory research aimed at unravelling the cross-binding of different Aβ epitope sequences to certain anti-Aβ antibodies and to predict useful mutations to engineer into anti-Aβ antibody drug candidates. Computational methods are a cost-effective way to study the binding properties of anti-Aβ antibodies whose crystal structures have previously been made available for analysis, and they may be an important tool when it comes to optimizing prospective drug candidates in the future.

1.6 REFERENCES

1. Masters, C. L.; Simms, G.; Weinman, N. A.; Multhaup, G.; McDonald, B. L.; Beyreuther, K. Amyloid Plaque Core Protein in Alzheimer Disease and Down Syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **1985,** *82*, 4245-4249.

2. Iqbal, K.; Alonso Adel, C.; Chen, S.; Chohan, M. O.; El-Akkad, E.; Gong, C. X.; Khatoon, S.; Li, B.; Liu, F.; Rahman, A., et al. Tau Pathology in Alzheimer Disease and Other Tauopathies. *Biochim. Biophys. Acta* **2005,** *1739*, 198-210.

3. Dobson, C. M. Getting Out of Shape. *Nature* **2002,** *418*, 729-730.

4. Selkoe, D. J.; Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease at 25 Years. *EMBO Mol. Med.* **2016,** *8*, 595-608.

5. Weiner, H. L.; Frenkel, D. Immunology and Immunotherapy of Alzheimer's Disease. *Nat. Rev. Immunol.* **2006,** *6*, 404-416.

6. Pul, R.; Dodel, R.; Stangel, M. Antibody-based Therapy in Alzheimer's Disease. *Expert Opin. Biol. Ther.* **2011,** *11*, 343-357.

7. Mavoungou, C.; Schindowski, K.; Atta-ur-Rahman, Ed.; *Immunotherapy with Anti-Aβ Monoclonal Antibodies in Alzhiemer's Disease: A Critical Review on the Molecules in the Pipelines with Regulatory Considerations*. Bentham Books: **2013**; Vol. 1, p 3-85.

8. Orgogozo, J. M.; Gilman, S.; Dartigues, J. F.; Laurent, B.; Puel, M.; Kirby, L. C.; Jouanny, P.; Dubois, B.; Eisner, L.; Flitman, S., et al. Subacute Meningoencephalitis in a Subset of Patients with AD After Abeta42 Immunization. *Neurology* **2003,** *61*, 46-54.

9. Lemere, C. A.; Maier, M.; Peng, Y.; Jiang, L.; Seabrook, T. J. Novel Abeta Immunogens: Is Shorter Better? *Curr. Alzheimer Res.* **2007,** *4*, 427-436.

10. Salloway, S.; Sperling, R.; Gilman, S.; Fox, N. C.; Blennow, K.; Raskind, M.; Sabbagh, M.; Honig, L. S.; Doody, R.; van Dyck, C. H., et al. Bapineuzumab 201 Clinical Trial, I., A Phase 2 Multiple Ascending Dose Trial of Bapineuzumab in Mild to Moderate Alzheimer Disease. *Neurology* **2009,** *73*, 2061-2070.

11. Siemers, E. R.; Friedrich, S.; Dean, R. A.; Gonzales, C. R.; Farlow, M. R.; Paul, S. M.; Demattos, R. B. Safety and Changes in Plasma and Cerebrospinal Fluid Amyloid Beta After a Single Administration of an Amyloid Beta Monoclonal Antibody in Subjects with Alzheimer Disease. *Clin. Neuropharmacol.* **2010,** *33*, 67-73.

12. Farlow, M.; Arnold, S. E.; van Dyck, C. H.; Aisen, P. S.; Snider, B. J.; Porsteinsson, A. P.; Friedrich, S.; Dean, R. A.; Gonzales, C.; Sethuraman, G., et al. Safety and Biomarker Effects of Solanezumab in Patients with Alzheimer's Disease. *Alzheimers Dement.* **2012,** *8*, 261-271.

13. Abbott, A.; Dolgin, E. Failed Alzheimer's Trial Does Not Kill Leading Theory of Disease. *Nature* **2016,** *540*, 15-16.

14. The Lancet, N., Eds.; Solanezumab: Too Late in Mild Alzheimer's Disease? *Lancet Neurol.* **2017,** *16*, 97.

15. Hawkes, N. Promise of New Alzheimer's Drug is Dashed After Lack of Evidence. *BMJ* **2016,** *355*, i6362.

16. Imbimbo, B. P.; Ottonello, S.; Frisardi, V.; Solfrizzi, V.; Greco, A.; Seripa, D.; Pilotto, A.; Panza, F. Solanezumab for the Treatment of Mild-To-Moderate Alzheimer's Disease. *Expert Rev. Clin. Immunol.* **2012,** *8*, 135-149.

17. Finder, V. H.; Glockshuber, R. Amyloid-Beta Aggregation. *Neurodegener. Dis.* **2007,** *4*, 13-27.

18. Russo, C.; Violani, E.; Salis, S.; Venezia, V.; Dolcini, V.; Damonte, G.; Benatti, U.; D'Arrigo, C.; Patrone, E.; Carlo, P., et al. Pyroglutamate-modified Amyloid Beta-Peptides--AbetaN3(pE)--Strongly Affect Cultured Neuron and Astrocyte Survival. *J. Neurochem.* **2002,** *82*, 1480-1489.

19. Saido, T. C.; Iwatsubo, T.; Mann, D. M.; Shimada, H.; Ihara, Y.; Kawashima, S. Dominant and Differential Deposition of Distinct Beta-Amyloid Peptide Species, A Beta N3(pE), in Senile Plaques. *Neuron* **1995,** *14*, 457-466.

20. Cynis, H.; Frost, J. L.; Crehan, H.; Lemere, C. A. Immunotherapy Targeting Pyroglutamate-3 Abeta: Prospects and Challenges. *Mol. Neurodegener.* **2016,** *11*, 48.

21. Demattos, R. B.; Lu, J.; Tang, Y.; Racke, M. M.; Delong, C. A.; Tzaferis, J. A.; Hole, J. T.; Forster, B. M.; McDonnell, P. C.; Liu, F., et al. A Plaque-Specific Antibody Clears Existing Beta-Amyloid Plaques in Alzheimer's Disease Mice. *Neuron* **2012,** *76*, 908-920.

22. Frost, J. L.; Liu, B.; Rahfeld, J. U.; Kleinschmidt, M.; O'Nuallain, B.; Le, K. X.; Lues, I.; Caldarone, B. J.; Schilling, S., et al. An Anti-Pyroglutamate-3 Abeta Vaccine Reduces Plaques and Improves Cognition in APPswe/PS1DeltaE9 Mice. *Neurobiol. Aging* **2015,** *36*, 3187-3199.

23. Bohrmann, B.; Baumann, K.; Benz, J.; Gerber, F.; Huber, W.; Knoflach, F.; Messer, J.; Oroszlan, K.; Rauchenberger, R.; Richter, W. F., et al. Gantenerumab: A Novel Human Anti-Abeta Antibody Demonstrates Sustained Cerebral Amyloid-Beta Binding and Elicits Cell-Mediated Removal of Human Amyloid-Beta. *J. Alzheimers Dis.* **2012,** *28*, 49-69.

24. Muhs, A.; Hickman, D. T.; Pihlgren, M.; Chuard, N.; Giriens, V.; Meerschman, C.; van der Auwera, I.; van Leuven, F.; Sugawara, M.; Weingertner, M. C., et al. Liposomal Vaccines with Conformation-Specific Amyloid Peptide Antigens Define Immune Response and Efficacy in APP Transgenic Mice. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 9810-9815.

25. Adolfsson, O.; Pihlgren, M.; Toni, N.; Varisco, Y.; Buccarello, A. L.; Antoniello, K.; Lohmann, S.; Piorkowska, K.; Gafner, V.; Atwal, J. K., et al. An Effector-Reduced Anti-Beta-

Amyloid (Abeta) Antibody with Unique Abeta Binding Properties Promotes Neuroprotection and Glial Engulfment of Abeta. *J. Neurosci.* **2012,** *32*, 9677-9689.

26. Ultsch, M.; Li, B.; Maurer, T.; Mathieu, M.; Adolfsson, O.; Muhs, A.; Pfeifer, A.; Pihlgren, M.; Bainbridge, T. W.; Reichelt, M., et al. Structure of Crenezumab Complex with Abeta Shows Loss of Beta-Hairpin. *Sci. Rep.* **2016,** *6*, 39374.

27. Crespi, G. A.; Hermans, S. J.; Parker, M. W.; Miles, L. A. Molecular Basis for Mid-Region Amyloid-Beta Capture by Leading Alzheimer's Disease Immunotherapies. *Sci Rep* **2015,** *5*, 9649.

28. Sevigny, J.; Chiao, P.; Bussière, T.; Weinreb, P. H.; Williams, L.; Maier, M.; Dunstan, R.; Salloway, S.; Chen, T.; Ling, Y., et al. The Antibody Aducanumab Reduces Aβ Plaques in Alzheimer's Disease. *Nature* **2016,** *537*, 50-56.

29. Kastanenka, K. V.; Bussiere, T.; Shakerdge, N.; Qian, F.; Weinreb, P. H.; Rhodes, K.; Bacskai, B. J. Immunotherapy with Aducanumab Restores Calcium Homeostasis in Tg2576 Mice. *J. Neurosci.* **2016,** *36*, 12549-12558.

30. Case, D.A.; Betz, R.M.; Botello-Smith, W.; Cerutti, D.S.; Cheatham, T.E., 3rd; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W., et al. AMBER 2016, University of California, San Francisco, **2016**.

31. Gardberg, A.; Dice, L.; Pridgen, K.; Ko, J.; Patterson, P.; Ou, S.; Wetzel, R.; Dealwis, C. Structures of Abeta-Related Peptide-Monoclonal Antibody Complexes. *Biochemistry* **2009,** *48*, 5210-5217.

32. Clark, L. A.; Boriack-Sjodin, P. A.; Eldredge, J.; Fitch, C.; Friedman, B.; Hanf, K. J.; Jarpe, M.; Liparoto, S. F.; Li, Y.; Lugovskoy, A., et al. Affinity Enhancement of an In Vivo Matured

Therapeutic Antibody Using Structure-Based Computational Design. *Protein Sci.* **2006,** *15*, 949-960.

33. Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-Based Identification of Druggable 'Hot Spots' of Proteins Using Fourier Domain Correlation Techniques. *Bioinformatics* **2009,** *25*, 621-627.

34. Kozakov, D.; Hall, D. R.; Jehle, S.; Luo, L.; Ochiana, S. O.; Jones, E. V.; Pollastri, M.; Allen, K. N.; Whitty, A.; Vajda, S. Ligand Deconstruction: Why Some Fragment Binding Positions are Conserved and Others are Not. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112*, E2585-E2594.

35. MacCallum, J. L.; Bennett, W. F.; Tieleman, D. P. Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **2008,** *94*, 3393-3404.

36. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011,** *3*, 33.

37. Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013,** *53*, 1893-1904.

38. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **2016,** *54*, 5.6.1-5.6.37.

39. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera-A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.

40. Srinivasan, J.; Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998,** *120*, 9401-9409.

41. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W., et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889-897.

42. Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004,** *25*, 238-250.

43. Yang, T. Y.; Wu, J. C.; Yan, C. L.; Wang, Y. F.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Y. Virtual Screening Using Molecular Simulations. *Proteins: Struct., Funct., Bioinf.* **2011,** *79*, 1940-1951.

44. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314-3321.

45. Wang, C. H.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B. N.; Wang, H. L.; Ren, P. Y.; Luo, R. Calculating Protein-Ligand Binding Affinities with MMPBSA: Method and Error Analysis. *J. Comput. Chem.* **2016,** *37*, 2436-2446.

46. Luo, R.; Moult, J.; Gilson, M. K. Dielectric Screening Treatment of Electrostatic Solvation. *J. Phys. Chem. B* **1997,** *101*, 11226-11236.

47. Wang, J.; Cai, Q.; Li, Z.-L.; Zhao, H.-K.; Luo, R. Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms. *Chem. Phys. Lett.* **2009,** *468*, 112-118.

48. Wang, J.; Luo, R. Assessment of Linear Finite-Difference Poisson-Boltzmann Solvers. *J. Comput. Chem.* **2010,** *31*, 1689-1698.

49. Cai, Q.; Hsieh, M.-J.; Wang, J.; Luo, R. Performance of Nonlinear Finite-Difference Poisson-Boltzmann Solvers. *J. Chem. Theory Comput.* **2010,** *6*, 203-211.

50. Wang, J.; Cai, Q.; Xiang, Y.; Luo, R. Reducing Grid Dependence in Finite-Difference Poisson-Boltzmann Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 2741-2751.

51. Botello-Smith, W. M.; Luo, R. Applications of MMPBSA to Membrane Proteins I: Efficient Numerical Solutions of Periodic Poisson-Boltzmann Equation. *J. Chem. Inf. Model.* **2015,** *55*, 2187-2199.

52. Cai, Q.; Wang, J.; Zhao, H.-K.; Luo, R. On Removal of Charge Singularity in Poisson-Boltzmann Equation. *J. Chem. Phys.* **2009,** *130*.

53. Lwin, T. Z.; Luo, R. Overcoming Entropic Barrier with Coupled Sampling at Dual Resolutions. *J. Chem. Phys.* **2005,** *123*.

54. Lwin, T. Z.; Zhou, R. H.; Luo, R. Is Poisson-Boltzmann Theory Insufficient for Protein Folding Simulations? *J. Chem. Phys.* **2006,** *124*.

55. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.

56. Tan, C.; Tan, Y.-H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007,** *111*, 12263-12274.

57. Cai, Q.; Ye, X.; Wang, J.; Luo, R. On-the-fly Numerical Surface Integration for Finite-Difference Poisson-Boltzmann Methods. *J. Chem. Theory Comput.* **2011,** *7*, 3608-3619.

58. Olsson, N.; Wallin, S.; James, P.; Borrebaeck, C. A.; Wingren, C. Epitope-Specificity of Recombinant Antibodies Reveals Promiscuous Peptide-Binding Properties. *Protein Sci.* **2012,** *21*, 1897-910.

59. DiBrino, M.; Parker, K. C.; Margulies, D. H.; Shiloach, J.; Turner, R. V.; Biddison, W. E.; Coligan, J. E. The HLA-B14 Peptide Binding Site Can Accommodate Peptides with Different Combinations of Anchor Residues. *J. Biol. Chem.* **1994,** *269*, 32426-34.

60. Ma, B.; Zhao, J.; Nussinov, R. Conformational Selection in Amyloid-Based Immunotherapy: Survey of Crystal Structures of Antibody-Amyloid Complexes. *Biochim. Biophys. Acta* **2016,** *1860*, 2672-2681.

61. Gardberg, A. S.; Dice, L. T.; Ou, S.; Rich, R. L.; Helmbrecht, E.; Ko, J.; Wetzel, R.; Myszka, D. G.; Patterson, P. H.; Dealwis, C. Molecular Basis for Passive Immunotherapy of Alzheimer's Disease. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 15659-15664.

62. DeMattos, R. B.; Bales, K. R.; Cummins, D. J.; Dodart, J. C.; Paul, S. M.; Holtzman, D. M. Peripheral Anti-Abeta Antibody Alters CNS and Plasma Abeta Clearance and Decreases Brain Abeta Burden in a Mouse Model of Alzheimer's Disease. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 8850-8855.

63. Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of Secondary Binding Sites in Proteins Using Fragment Screening. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112*, 15910-15915.

64. Seeliger, D.; de Groot, B. L. Conformational Transitions upon Ligand Binding: Holo-Structure Prediction from Apo Conformations. *PLoS Comput. Biol.* **2010,** *6*, e1000634.

65. Ding, F.; Dokholyan, N. V. Incorporating Backbone Flexibility in MedusaDock Improves Ligand-Binding Pose Prediction in the CSAR2011 Docking Benchmark. *J. Chem. Inf. Model* **2013,** *53*, 1871-1879.

66. Zhao, J.; Nussinov, R.; Ma, B. Mechanisms of Recognition of Amyloid-Beta (Abeta) Monomer, Oligomer, and Fibril by Homologous Antibodies. *J. Biol. Chem.* **2017,** *292*, 18325-18343.

67. Nussinov, R.; Ma, B.; Tsai, C. J. Multiple Conformational Selection and Induced Fit Events Take Place in Allosteric Propagation. *Biophys. Chem.* **2014,** *186*, 22-30.

68. Weikl, T. R.; Paul, F. Conformational Selection in Protein Binding and Function. *Protein Sci.* **2014,** *23*, 1508-1518.

69. Meirovitch, H.; Cheluvaraja, S.; White, R. P. Methods for Calculating the Entropy and Free Energy and Their Application to Problems Involving Protein Flexibility and Ligand Binding. *Curr. Protein Pept. Sci.* **2009,** *10*, 229-243.

70. de Ruiter, A.; Oostenbrink, C. Free Energy Calculations of Protein-Ligand Interactions. *Curr. Opin. Chem. Biol.* **2011,** *15*, 547-552.

71. Steinbrecher, T.; Labahn, A. Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies. *Curr. Med. Chem.* **2010,** *17*, 767-785.

72. Lin, J. H. Accommodating Protein Flexibility for Structure-Based Drug Design. *Curr. Top. Med. Chem.* **2011,** *11*, 171-178.

73. Greene, D.; Botello-Smith, W. M.; Follmer, A.; Xiao, L.; Lambros, E.; Luo, R. Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor. *J. Phys. Chem. B* **2016,** *120*, 12293-12304.

CHAPTER 2

Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor

Reprinted (adapted) with permission from Greene, D.; Botello-Smith, W. M.; Follmer, A.; Xiao, L.; Lambros, E.; Luo, R. Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor. *J. Phys. Chem. B* **2016,** *120*, 12293-12304. Copyright 2016 American Chemical Society.

ABSTRACT

Membrane proteins, due to their roles as cell receptors and signaling mediators, make prime candidates for drug targets. The computational analysis of protein-ligand binding affinities has been widely employed as a tool in rational drug design efforts. Although efficient implicit solvent-based methods for modeling globular protein-ligand binding have been around for many years, the extension of such methods to membrane protein-ligand binding is still in its infancy. In this study, we extended the widely used Amber/MMPBSA method to model membrane protein-ligand systems, and we used it to analyze protein-ligand binding for the human purinergic platelet receptor ($P2Y_{12}R$), a prominent drug target in the inhibition of platelet aggregation for the prevention of myocardial infarction and stroke. The binding affinities, computed by the Amber/MMPBSA method using standard parameters, correlate well with experiment. A detailed investigation of these parameters was conducted to assess their impact on the accuracy of the method. These analyses show the importance of properly treating the non-polar solvation interactions and the electrostatic polarization in the binding of nucleotide agonists and non-nucleotide antagonists to $P2Y_{12}R$. Based on the crystal structures and the

experimental conditions in the binding assay, we further hypothesized that the nucleotide agonists lose their bound magnesium ion upon binding to $P2Y_{12}R$, and our computational study supports this hypothesis. Ultimately, this work illustrates the value of computational analysis in the interpretation of experimental binding reactions.

2.1 INTRODUCTION

Membrane proteins provide a range of important functions as cell receptors, signaling proteins, transmembrane channels, and more. Their roles as receptors and signaling proteins make them particularly relevant as candidates for drug targets. However, the study of membrane proteins is more complicated than the study of globular proteins. Particularly, the presence of the membrane complicates structural studies, both experimentally and computationally. The presence of the membrane makes it more difficult to employ experimental techniques such as NMR and X-ray crystallography. For instance, the signal from the membrane must be disentangled from that of the protein when using NMR, and membrane proteins are notoriously difficult to crystallize. For computational studies, modeling of the membrane becomes an important consideration.

An active area of computational studies of proteins is the prediction of protein-ligand binding affinities. The Amber 16[1] and AmberTools 16[2] suites currently provide the capability of performing such calculations for globular proteins via the widely used MMPBSA module[3-8]. The consideration of solvation effects in these computational approaches is quite important. It is relatively common knowledge that solvent-solute interactions provide the primary driving force for producing and maintaining the properly folded structures of proteins[9-11].

Inclusion of the solvent into a computational model or simulation can generally be classified into one of two different categories: explicit and implicit solvation. In explicit solvation, each atom or molecule of the solvent is modeled individually. While this is generally agreed to be the most accurate method, one is often not interested in the properties of the solvent itself, but rather in the behavior it induces upon the solute. Unfortunately, accurately capturing statistically meaningful characteristics requires sampling either from ensembles of trajectories or from a single very long trajectory. Implicit solvents provide an attractive alternative wherein the effects of the solvent are modeled as a continuum[10-26]. While the fine-grained details of individual solvent-solute particle interactions are lost, the relevant statistically averaged effects may still be captured when a properly parameterized and transferrable model is used. In addition, since the individual solvent molecules are no longer modeled directly, there are far fewer particles to simulate which reduces the sampling challenges in molecular simulations.

In the case of membrane proteins, the membrane must also be included when modeling solvation effects[27-33]. In general, the molecules that make up a lipid membrane are much more complex than water or other small organic solvents, and this increases the computational expense of their inclusion. Thus, there has been much effort put into the development and testing of implicit membrane solvent models[27-33]. Implicit membranes have appeared in several recent computational studies as they can assist in finding the proper native fold of a membrane protein for structural studies and calculations[34-36]. Implementation of an implicit membrane is currently available in packages such as APBS[32], Delphi[33, 37], and both the Amber 16[1] and AmberTools 16[2] suites. With the implementation of an implicit membrane model into the Amber/PBSA program[38-42], the implicit membrane model can be more readily interfaced with the existing MMPBSA framework[3-8].

One of the key features to consider in implicit solvent models is the modeling of electrostatic interactions. This is most readily accomplished by employing the Poisson-Boltzmann equation (PBE)[43-60]. In cases where the ion concentration is relatively low (a few hundred millimolar or less), this equation may be approximated as the linear PBE:

$$\nabla \cdot \varepsilon \nabla \phi = -4\pi\rho_0 + \varepsilon_v \kappa^2 \phi \tag{1}$$

where $\kappa^2 = \frac{8\pi e^2 I}{\varepsilon_v k_B T}$. Here v denotes the solvent, and $I = z^2 c$ represents the ionic strength of the solution. The PBE-based solvent models have many biological applications. For example, they have been applied to the prediction of pKa values for ionizable groups in biomolecules[61-65], solvation free energies[66, 67], binding free energies[68-73], and protein folding and design[74-83]. Even in its simplified linear form, solving the PBE is a non-trivial endeavor. Due to its complexity, there is no general closed form solution; a numerical solution must be sought with the exception of very simplified geometries[27, 38-40, 42, 48, 59, 60, 84-115]. A semi-analytical Generalized Born (GB) equation was also developed to approximate the PBE solution and is quite popular in biomolecular applications.

In order to apply the PBE or GB frameworks to implicit membrane solvent models, an additional dielectric region must be added (see **FIGURE 2.1**). The appropriate dielectric constant of the membrane region is generally thought to be quite low relative to the bulk solvent dielectric constant that is typically set to be between 60 (mimicking SPC water models) and 80 (typical for TIP3P models)[116-119]. Various dielectric constant profiles were explored during the development of GB implicit membrane models[116, 118, 120], and it was demonstrated that a simple two dielectric constant model can reproduce electrostatic free energies relatively well by modeling the membrane as a slab like region with a uniform dielectric constant of about 2. Models with 3 or more layers have been shown to improve the accuracy of the results; however, beyond three layers, the improvement was shown to be marginal[118].

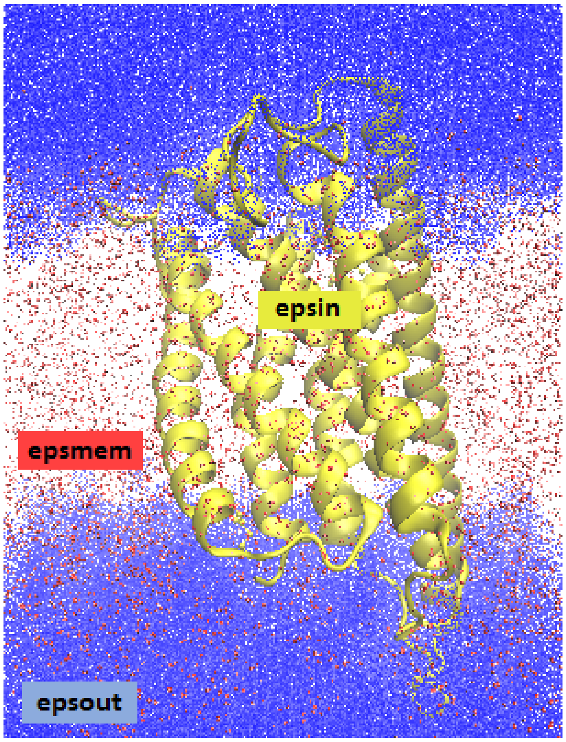


FIGURE 2.1. Dielectric constant regions in the P2Y$_{12}$R system. The protein/ligand is shown in yellow (dielectric: epsin), the implicit membrane is shown in red (dielectric: epsmem), and the surrounding implicit water solvent is shown in blue (dielectric: epsout).

A complete implementation of an implicit membrane under the Amber/PBSA program[38-42] requires implementation of appropriate membrane to protein non-polar solvation free energy terms[121]. While development of these terms is still underway, it is not expected to impact binding free energy calculations for protein-ligand systems in which the binding pocket is sequestered away from the membrane in the protein interior. Thus, such systems make good candidates for testing the electrostatic free energy calculations provided by the current implementation of the implicit membrane model within the Amber/PBSA program[38-42].

While three-dimensional structures for globular proteins are quite abundant, such data is less prevalent for membrane proteins. To validate the Amber/PBSA program, and the MMPBSA framework for membrane protein applications, it is necessary to find a protein for which experimental binding affinities and structures of the associated protein-ligand complexes are both available. Recently, a study on the human purinergic platelet receptor ($P2Y_{12}R$) was reported[122, 123], providing crystal structures of the receptor bound to three different ligands and experimental measurements of dissociation constants for the wild type, several select mutants, and two additional ligands. Several antithrombotic drugs target $P2Y_{12}R$ to inhibit platelet aggregation for the purpose of preventing myocardial infarction and stroke. However, limitations of these drugs have motivated the development of a new generation of $P2Y_{12}R$ inhibitors[124, 125]. Thus, computational modeling, based on the latest structural and functional data, will further facilitate developmental efforts aimed in this direction.

This work documents both the development of an MMPBSA algorithm for membrane protein-ligand binding applications and the validation of the algorithm using $P2Y_{12}R$ complexes with nucleotide agonists and non-nucleotide antagonists. We analyze the sensitivity and the quality of predicted binding affinities with respect to several key polar and non-polar components in the MMPBSA framework. In addition, we also analyze the role of magnesium ions in the binding of nucleotide agonists to $P2Y_{12}R$. Our results show the application of protein-ligand binding affinity prediction methods towards interpreting experimental binding affinities.

2.2 METHODS

2.2.1 Preparation of the $P2Y_{12}R$ complex structures

Three separate crystal structures of $P2Y_{12}R$[122, 123], two complexed with agonist ligands: 2-methylthio-adenosine-5'-triphosphate and 2-methylthio-adenosine-5’-diphosphate (2MeSATP and 2MeSADP respectively) and one with an antagonist ligand: AZD-1283 (AZD), were downloaded from the protein databank. As was noted in the corresponding literature[122, 123], each of the crystal structures contained several sequence gaps for which no structure could be resolved. The program Modeller[126] was used to generate initial structures for these gaps. These homology models were then merged into the crystal structures using the Multi-SEQ[127] module in the program VMD[128]. The single point mutation D294N was modeled in the Amber/LEAP program[1, 2, 129].

In addition to the ligands obtained from the three $P2Y_{12}R$ crystal structures, two additional antagonist ligands, with reported binding affinity data[122], were docked to the receptor from the $P2Y_{12}R$-AZD crystal structure. The structure for Ticagrelor (TIQ) was extracted from a previously published crystal structure[130] (PDB ID: 5ALB). The structure for PSB-0739 (PSB) was generated using MarvinSketch[131] to produce a 2D structure file which was subsequently converted into a 3D structure file using the OpenEye toolkits[132]. The structures were independently docked to the receptor using AutoDock Vina/SMINA[133].

The ligands in the three $P2Y_{12}R$ complex structures (2MeSATP, 2MeSADP, and AZD) were extracted to individual structure files for parameterization. Two additional ligands, PSB and TIQ, were obtained from the docking analysis documented above. The two agonists are simple derivatives of ATP and ADP, so their parameters were obtained from the literature[134] except for the atomic charges for the 2-methylthio-adenine group. For the rest of the ligands, the Amber/ANTECHAMBER[135] program was used to generate force field parameters.

2.2.2 Preparation of the lipid membrane model

$P2Y_{12}R$ is found embedded within platelet outer membranes. The membrane environment for $P2Y_{12}R$ was modeled with an explicit all-atom model for the molecular dynamics (MD) simulation and with an implicit continuum model for the post-processing binding affinity calculation. Construction of the explicit all-atom membrane model was accomplished using the CHARMM-GUI membrane builder web server[136]. The membrane was constructed with a POPC, POPS, and POPE ratio of 3:2:3 and a cholesterol to lipid ratio of 2:5[137]. Sphingomyelin lipids were not included since their force field parameters were not yet available.

The aqueous phase of the $P2Y_{12}R$ membrane protein-ligand system was modeled using an explicit all-atom approach with the TIP3P model along with sufficient potassium and chloride ions to mimic a roughly 150 millimolar KCl concentration. The constructed membrane protein-ligand system was loaded into the LEAP[1, 2, 129] program for the generation of simulation force field topology and coordinate files.

2.2.3 MD simulation protocol

Each system was first minimized using 500 steps of steepest descent followed by 500 steps of conjugate gradient optimization. All residues taken directly from the crystal structures were held fixed. Residues generated from homology modeling were left unrestrained, along with all solvent molecules including membrane lipids, water molecules, and ions. After minimization, heating was performed in two phases. In the first phase, systems were brought up to 100 K over 2500 time steps (5 ps) under the NTV condition using a Langevin thermostat with a collision frequency of 1.0 per ps. This was followed by heating from 100 K to 303 K over 100 ps under the NPT condition with anisotropic pressure scaling using the Berendsen barostat with a pressure relaxation constant of 2.0 ps and a target pressure of 1.0 atm. In both cases, a cutoff radius of 10 Å was used when computing non-bonded interactions. All these preparation simulations were performed using the MPI parallelized SANDER program from the Amber 16 suite[1, 2, 129].

After the initial heating was completed, it was necessary to equilibrate the membrane density prior to the production run. The production run utilized Amber's GPU accelerated PMEMD program that does not allow frequent updating of the box size[1, 2, 129]. Density equilibration was performed over 10 identical 500 ps NTP simulations at a constant temperature of 303 K. As with heating and minimization, portions of the protein substructure containing coordinates from the crystal structure were held fixed. Afterwards, all restraints were removed, and a 20 ns equilibration run using the GPU accelerated PMEMD program was performed to fully relax the system. Finally, a 10 ns production run was used for the MMPBSA calculation, which was found to be sufficient to achieve the averaging required for free energy calculations (**FIGURE B.1**). In order to take full advantage of the GPU accelerated code, a Monte-Carlo thermostat was employed instead of the Berendsen thermostat used during the previous simulation phases.

2.2.4 Binding free energy calculations

Binding free energies were computed using the SANDER/PBSA module in the Amber 2016 release[1, 2, 129]. The production run trajectory was post processed with CPPTRAJ[138] in order to remove the solvent, membrane, and counter ions from the receptor-ligand complex. 1000 frames, taken at equal intervals over the 10 ns production run, were processed using SANDER/PBSA to compute molecular mechanics potential energies and solvation free energies.

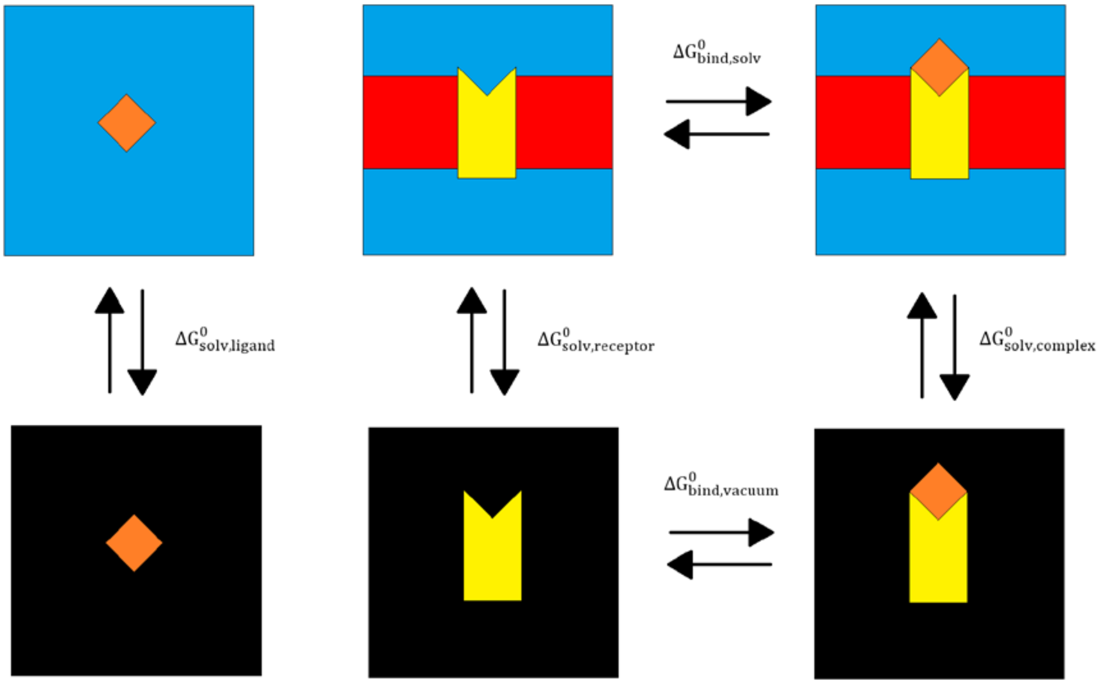


FIGURE 2.2. Thermodynamic cycle of the MMPBSA method for a membrane protein-ligand system. The membrane protein is depicted in yellow, the ligand is depicted in orange, the implicit membrane is shown in red, the water solvent is shown with blue, and the vacuum is shown with black. ΔG^0 values are labelled for the various transitions from one state to another.

The binding free energy for the protein-ligand complex was computed as the difference between the complex free energy and the sum of the receptor and ligand free energies as shown schematically in **FIGURE 2.2** for the membrane protein-ligand system. The SANDER/PBSA calculations were conducted either inside the implicit membrane/water solvent or the pure implicit water solvent according to the following thermodynamic cycle:

$$\Delta G^0_{bind,solv} = \Delta G^0_{bind,vacuum} + \Delta G^0_{solv,complex} - \Delta G^0_{solv,ligand} - \Delta G^0_{solv,receptor}. \quad (2)$$

The solvation free energies were calculated using

$$\Delta G^0_{solv} = \Delta G^0_{elec} + \Delta G^0_{np} \quad (3)$$

where the electrostatic terms, ΔG^0_{elec},were calculated using the linearized PBE solver as implemented in PBSA[38-42]. The non-polar solvation terms, ΔG^0_{np}, were calculated using either the classical model or the modern model as documented previously[139]. It is worth pointing out that **FIGURE 2.2** shows that the ligand-binding site of the protein is in the aqueous phase. This implies that the standard non-polar solvent models, optimized for solvation in water, are reasonable approximations. However, it is not unusual to observe binding cavities buried deep within the lipid bilayer. These binding reactions will require a recalibration of the non-polar solvent model, and this scenario will be explored in a future study.

The calculated binding free energies were then compared against the experimental results[122]. The experimental dissociation constants, K_D, were converted to appropriate binding free energies using the formula:

$$\Delta G = \mathrm{RTln}(\mathrm{K_D}) \quad (4)$$

where R is the gas constant, and T is the temperature.

2.2.5 Estimation of the free energy penalty upon the loss of Mg^{2+}

It is well known that both ATP and ADP have a tendency to form a complex with a magnesium ion in solution[140-142]. However, in the crystal structures for 2MeSATP and 2MeSADP bound to $P2Y_{12}R$, magnesium ions were notably absent. Nevertheless, the experimental binding affinity data for 2MeSATP and 2MeSADP bound to $P2Y_{12}R$ was obtained in the presence of 10mM $MgCl_2$[122]. When the experimental binding affinity measurements were carried out, the presence of magnesium ions in solution may have imposed an energetic penalty on the ligands 2MeSATP and 2MeSADP; if either compound was bound to a magnesium ion in solution, the bound magnesium ion would have to be removed before the ligand could bind to the $P2Y_{12}R$ active site in the same manner that was observed in the crystal structure.

In order to estimate the magnitude of this penalty, we modeled the removal of the bound magnesium ion from both 2MeSATP and 2MeSADP in solution as a binding affinity calculation. The overall reaction is given as:

$$\mathrm{Mg(2MeSAXP)} + \mathrm{P2Y_{12}R} \rightarrow \mathrm{(2MeSAXP)P2Y_{12}R} + \mathrm{Mg^{2+}} \qquad (5)$$

where AXP = ATP or ADP depending on the ligand in question, and ΔG_5 is the overall free energy change for this reaction. We can separate the removal of Mg^{2+} and the binding of the ligand to $P2Y_{12}R$ into two steps:

$$\mathrm{Mg(2MeSAXP)} \rightarrow \mathrm{2MeSAXP} + \mathrm{Mg^{2+}} \qquad (6)$$

$$\mathrm{2MeSAXP} + \mathrm{P2Y_{12}R} \rightarrow \mathrm{(2MeSAXP)P2Y_{12}R} \qquad (7)$$

with the free energy change for equation (6) given as ΔG_6, and the free energy change for equation (7) given as ΔG_7. Following Hess's Law, we can sum equations (6) and (7) to recover equation (5), and similarly, we can obtain the overall binding free energy as:

$$\Delta G_5 = \Delta G_6 + \Delta G_7. \tag{8}$$

To carry out the free energy calculation for equation (6), the 2MeSATP and 2MeSADP ligands were isolated from their respective complex structures. The bound magnesium ion was placed above and in between neighboring phosphate oxygen atoms on the ligand using UCSF Chimera[143]. The MD simulation, and MMPBSA binding free energy calculation, was carried out as described previously.

2.2.6 Additional computational details

In each PBSA calculation, a grid spacing of 0.5 Å was used with a grid to solute dimension ratio (fillratio) of 1.5. The geometric multigrid solver option was employed with a convergence threshold of 1.0×10^{-3}, and electrostatic focusing was turned off. All PB calculations were conducted using the periodic boundary condition in the PBSA program[38-42] in the Amber 2016 suite[1, 2, 129].

The solvation system physical constants were set up as follows. The membrane was modeled as a solid slab of 40 Å. The water relative dielectric constant (epsout) was fixed at 80.0. The water phase ionic strength (istrng) was set to be 150 mM. The lower dielectric region within the molecular solutes was defined using the classical solvent excluded surface model with a water solvent probe of 1.4 Å and a membrane solvent probe of 2.7 Å, which is used to account for the larger effective size of a lipid molecule when compared to a water molecule. Further details for extending the classical solvent excluded surface model to membrane systems will be discussed in a separate publication. The default weighted harmonic averaging was employed to assign dielectric constants for boundary grid edges to reduce grid dependence. Charges and radii were assigned using the same parameters as the simulation topology files.

The accuracy of the computed binding free energies was assessed by comparing the RMSD of the calculated versus the experimental values. The Pearson correlation coefficient (R), the slope, and the associated p-value of the linear regression were also analyzed. These were computed for both the absolute binding free energies (ΔG) and the relative binding free energies ($\Delta\Delta G$) for two different P2Y$_{12}$ receptors (wild type and mutant D294N) and five different ligands (2MeSATP, 2MeSADP, PSB, TIQ, and AZD).

2.3 RESULTS AND DISCUSSION

2.3.1 Impact of the non-polar solvation model

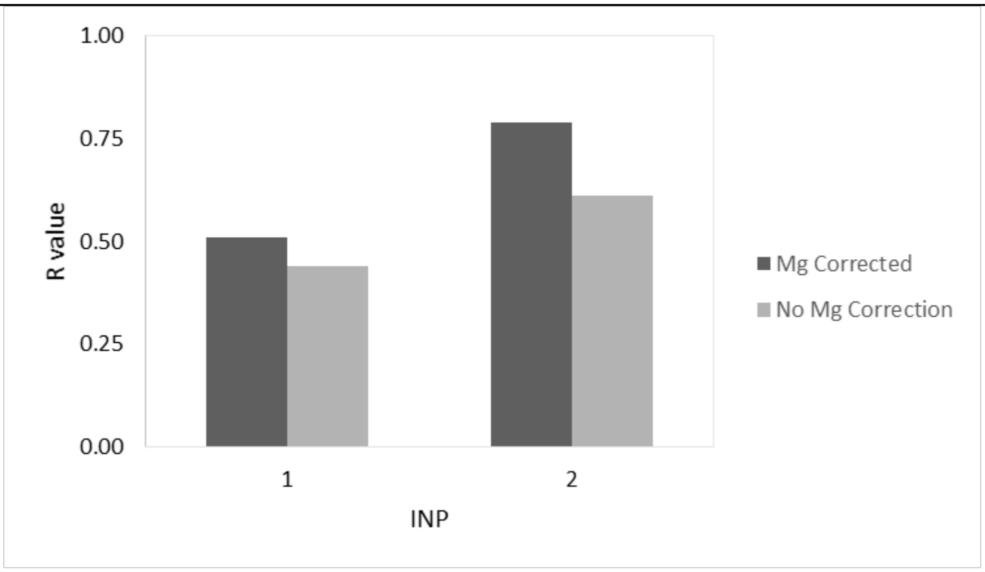
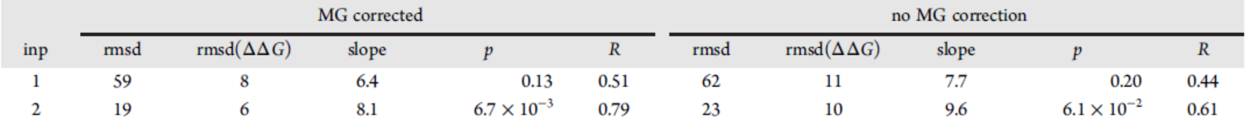


FIGURE 2.3. Parameter optimization for the non-polar solvation model, inp (epsin = 20, epsmem = 4). The R value indicates the correlation for the data set at the given inp value.

While electrostatic interactions play a major role in PBE-based implicit solvent models, various non-polar interactions, such as cavity surface tension and dispersion must also be accounted for. The PBSA module of Amber currently provides two options for computing non-polar solvation energy terms. The first method (inp = 1) uses a linear function of the solvent accessible surface area/volume[144]. The second, more sophisticated, method (inp = 2) decomposes the non-polar contribution into separate cavity and dispersion terms for better transferrable modeling of the non-polar solvation effects[121]. To test the relative effectiveness of these two methods, two sets of computations were run using inp = 1 or inp = 2 with all other parameters set at their optimal values.

FIGURE 2.3 shows the values of the Pearson correlation coefficient when inp = 1 or inp = 2 are used in the analysis. Additional metrics to assess the agreement are provided in **TABLE 2.1**. Using the classical method, inp = 1, the RMSDs are 59 and 8 kcal/mol (for ΔG and $\Delta\Delta G$ respectively) with the magnesium correction, and 62 and 11 kcal/mol without the magnesium correction. The modern method, inp = 2, yields RMSDs of 19 and 6 kcal/mol with the magnesium correction, and 23 and 10 kcal/mol without. The lower RMSD values indicate a better agreement using inp = 2.

TABLE 2.1. Effect of the non-polar solvation model on absolute (ΔG) and relative ($\Delta\Delta G$) binding affinities



	MG corrected					no MG correction				
inp	rmsd	rmsd($\Delta\Delta G$)	slope	p	R	rmsd	rmsd($\Delta\Delta G$)	slope	p	R
1	59	8	6.4	0.13	0.51	62	11	7.7	0.20	0.44
2	19	6	8.1	6.7×10^{-3}	0.79	23	10	9.6	6.1×10^{-2}	0.61

Results that are corrected for the removal of the magnesium ion appear in columns 2-6 while the results that do not take into account the magnesium correction appear in columns 7-11. RMSD values are given in units of kcal/mol. The raw data used to generate these results is available in Appendix B (**TABLE B.1**).

The same trend holds for the p-value of the correlation; inp = 1 gives p-values of 0.13 with the magnesium correction and 0.20 without the correction while inp2 yields p-values of 0.0067 and 0.061 respectively. In addition, the R values for inp = 1 are 0.51 and 0.44 with and without the magnesium correction respectively while the corresponding R values for inp2 are 0.79 and 0.61. In this case, the higher R values for inp = 2 indicate a stronger correlation. Taken together, the modern inp = 2 method clearly yields improved results over the classical inp = 1 method for our system.

2.3.2 Impact of the protein dielectric constant

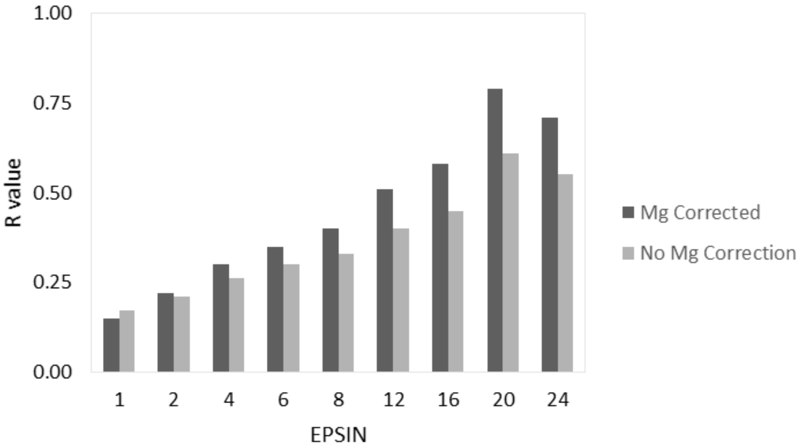


FIGURE 2.4. Parameter optimization for the protein dielectric constant, epsin (inp = 2, epsmem = 4). The R value indicates the correlation for the data set at the given epsin value.

At neutral pH, $P2Y_{12}R$ and all associated ligands except AZD and TIQ are non-neutral in solution. For neutral receptor and ligand systems, the protein dielectric constant is typically assigned to a relatively low value, such as 1 or 2. Charged systems often require a higher dielectric constant to be assigned to the protein in order to compensate for the lack of polarization treatment in typical MMPBSA calculations. To test the effect of the protein dielectric constant upon binding prediction efficacy, a series of calculations was performed with protein dielectric constants of 1, 2, 4, 6, 8, 12, 16, 20, and 24 while all other parameters were held at their optimal values.

TABLE 2.2. Effect of the protein dielectric constant on absolute (ΔG) and relative ($\Delta\Delta G$) binding affinities

	MG corrected					no MG correction				
epsin	rmsd	rmsd($\Delta\Delta G$)	slope	p	R	rmsd	rmsd($\Delta\Delta G$)	slope	p	R
1	87	110	25	0.67	0.15	120	160	39	0.64	0.17
2	53	60	20	0.53	0.22	78	89	27	0.56	0.21
4	37	34	15	0.41	0.30	50	50	19	0.47	0.26
6	31	24	13	0.32	0.35	40	36	16	0.40	0.30
8	27	19	11	0.25	0.40	35	28	14	0.35	0.33
12	23	12	9.6	0.13	0.51	29	19	11	0.25	0.40
16	22	10	8.5	7.9×10^{-2}	0.58	26	15	10	0.19	0.45
20	19	6	8.1	6.7×10^{-3}	0.79	23	10	9.6	6.1×10^{-2}	0.61
24	20	7	7.4	2.1×10^{-2}	0.71	23	11	8.7	0.10	0.55

Results that are corrected for the removal of the magnesium ion appear in columns 2-6 while the results that do not take into account the magnesium correction appear in columns 7-11. RMSD values are given in units of kcal/mol. The raw data used to generate these results is available in Appendix B (**TABLE B.2**).

FIGURE 2.4 and **TABLE 2.2** show that a general improvement in the agreement with experiment can be obtained by increasing the protein/solute dielectric constant to about epsin = 20. This is indicated by the lowering of the RMSD and p-values and an increase in the value of the correlation coefficient in comparison to these same metrics when measured at nearby epsin values. It is noted that raising the protein dielectric constant has a particularly profound effect on the agreement with heavily charged ligands such as 2MeSATP and 2MeSADP in our system as often observed when modeling binding reactions involving charged ligands/active sites in globular proteins[6, 8, 145, 146]. The agreement is also expected given that the binding pocket of $P2Y_{12}R$ is still in the water-soluble region of the protein. However, it is worth noting that this finding may not hold for membrane proteins with membrane-accessible binding pockets. High quality structure and affinity data would be necessary to establish a standard practice for such cases.

2.3.3 Impact of the membrane dielectric constant

Currently, the implicit membrane model implemented in PBSA allows for only a single membrane region. While this may be extended with relative ease in the future, this study focuses on the single membrane dielectric constant protocol.

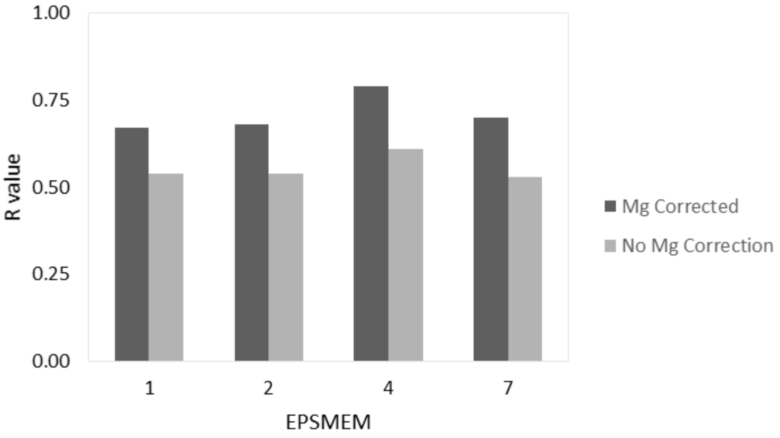
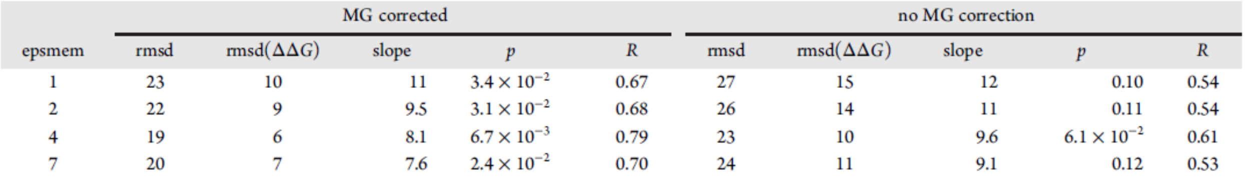


FIGURE 2.5. Parameter optimization for the membrane dielectric constant, epsmem (inp = 2, epsin = 20). The R value indicates the correlation for the data set at the given epsmem value.

This protocol follows the classical solvent excluded surface definition of molecular surface, in analogy to globular proteins. In doing so, it was reasonable to examine membrane dielectric constants of 1, 2, 4, and 7 with all other parameters set to their optimal values.

From examination of **FIGURE 2.5** and **TABLE 2.3**, it is evident that the accuracy of the binding affinity calculations is less sensitive to changes in the membrane dielectric constant in comparison to the differences observed when changing the non-polar solvation model and the protein dielectric constant. Although there is a detectable improvement in the RMSD, p-value, and the correlation coefficient for epsmem = 4, other epsmem values are comparable to each other. Again, the finding here may be attributed to the water-accessible nature of the binding pocket in P2Y$_{12}$R. It is expected that membrane-accessible binding may behave quite sensitively to the membrane dielectric constant.

TABLE 2.3. Effect of the membrane dielectric constant on absolute (ΔG) and relative ($\Delta\Delta G$) binding affinities



	MG corrected					no MG correction				
epsmem	rmsd	rmsd($\Delta\Delta G$)	slope	p	R	rmsd	rmsd($\Delta\Delta G$)	slope	p	R
1	23	10	11	3.4×10^{-2}	0.67	27	15	12	0.10	0.54
2	22	9	9.5	3.1×10^{-2}	0.68	26	14	11	0.11	0.54
4	19	6	8.1	6.7×10^{-3}	0.79	23	10	9.6	6.1×10^{-2}	0.61
7	20	7	7.6	2.4×10^{-2}	0.70	24	11	9.1	0.12	0.53

Results that are corrected for the removal of the magnesium ion appear in columns 2-6 while the results that do not take into account the magnesium correction appear in columns 7-11. RMSD values are given in units of kcal/mol. The raw data used to generate these results is available in Appendix B (**TABLE B.1**).

2.3.4 Effect of the magnesium correction

Finally, we address the effect of modeling explicit binding of the two nucleotide agonists to the magnesium ion. **FIGURES 2.3-2.5** and **TABLES 2.1-2.3** show that modeling the removal of a magnesium ion, as described in the Methods section, leads to much higher R values, lower p values, and lower RMSD values compared to the uncorrected data across the board. The effect of the magnesium correction can be seen most clearly in **FIGURES 2.6** and **7** as the data points for 2MeSATP and 2MeSADP rise in free energy to improve the agreement with the linear trend established by the other less charged or neutral ligands.

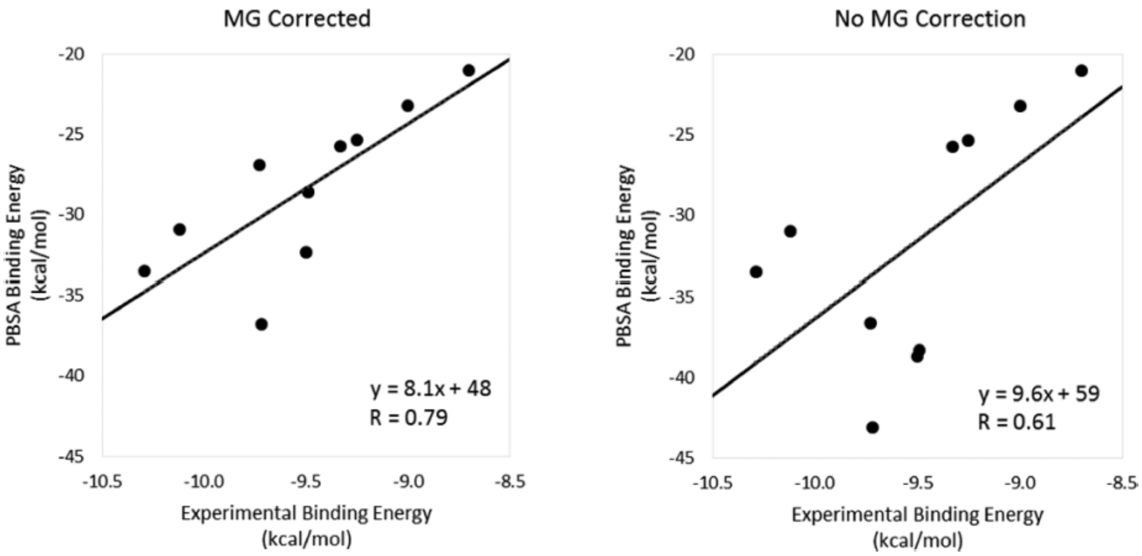


FIGURE 2.6. Absolute binding free energy (ΔG) correlation plots (inp = 2, epsin = 20, and epsmem = 4). The plot that was corrected for the removal of the magnesium ion appears on the left while the plot that did not take into account the magnesium correction appears on the right.

FIGURE 2.6 compares the absolute binding free energies (ΔG) between our calculated MMPBSA values and the experimental results while **FIGURE 2.7** provides an analogous comparison using the relative binding free energies ($\Delta\Delta G$).

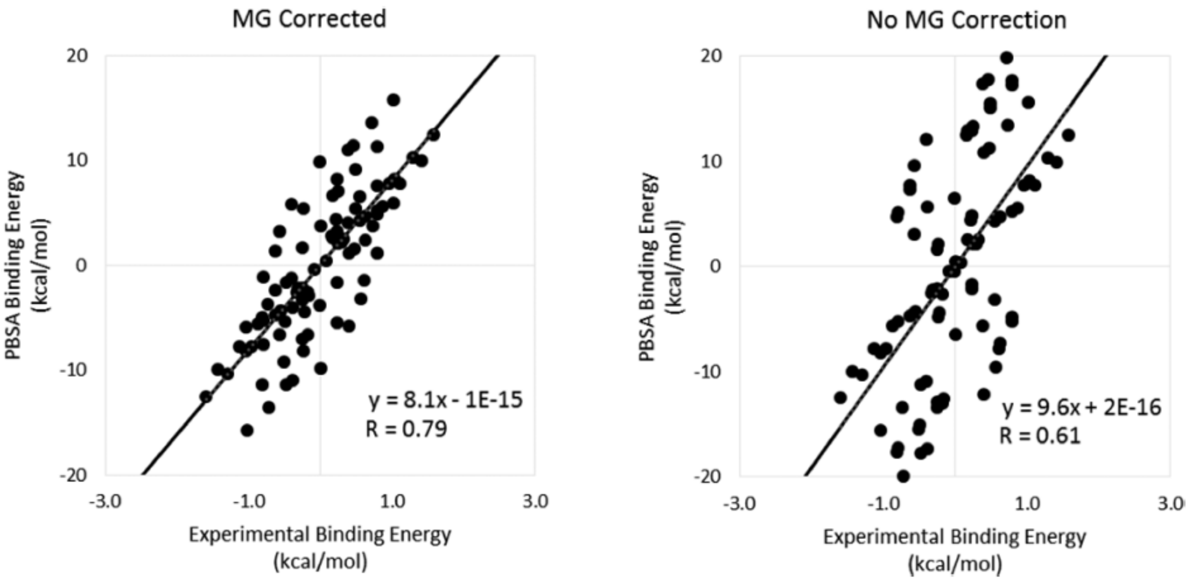


FIGURE 2.7. Relative binding free energy (ΔΔ*G*) correlation plots (inp = 2, epsin = 20, and epsmem = 4). The plot that was corrected for the removal of the magnesium ion appears on the left while the plot that did not take into account the magnesium correction appears on the right.

The large systematic bias inherent in MMPBSA binding affinity calculations manifests itself as both a large slope and a large y-intercept in **FIGURE 2.6**. Using the relative free energy values in **FIGURE 2.7** removes the large y-intercept while keeping the slope and overall correlation the same. It is clear that the rather large error in the uncorrected set is due to the binding data obtained from 2MeSATP and 2MeSADP. When the correction is applied, the agreement improves specifically because it counteracts the extremely favorable electrostatic interactions with these two ligands in the standard

MMPBSA calculation. Note that this is so even when a very high apparent protein dielectric constant is used.

Our hypothesis that the bound magnesium ion is lost upon 2MeSATP and 2MeSADP binding to $P2Y_{12}R$ is based on inspection of the complex crystal structures, and the fact that the binding assay was conducted in the presence of 10mM $MgCl_2$[122]. This is supported by our computational modeling of the binding reactions, which utilizes standard setups of the widely used MMPBSA method. It should be noted that our binding affinity modeling was conducted without the normal mode entropy analysis. This analysis usually does not contribute favorably to the overall agreement with experiment due to the approximation used[6, 8, 145]. Additionally, it is possible that binding-induced conformational changes cannot be fully taken into account by the widely used single-trajectory approach. The more extensive multi-trajectory approach, or more high-end free energy simulation methods with enhanced sampling, will be explored in the future.

Nevertheless, additional structural analysis shows that $P2Y_{12}R$ is indeed an interesting case. **FIGURE 2.8** compares the electrostatic potential at the binding site for $P2Y_{12}R$ bound to 2MeSATP[123] (**FIGURE 2.8A**, PDB ID: 4PY0) with two other proteins which have a magnesium ion bound to ATP in their binding sites: G protein-coupled receptor kinase 1, GRK1[147] (**FIGURE 2.8B**, PDB ID: 3C4X), and Flagella-related protein H, FlaH[148] (**FIGURE 2.8C**, PDB ID: 4YDS). It can be seen in **FIGURES 2.8B** and 2.**8C** that a negatively charged region (red) is present in both GRK1 and FlaH to stabilize the magnesium ion in the binding site. However, no comparable region appears that would stabilize a magnesium ion in $P2Y_{12}R$ as **FIGURE 2.8A** shows that the ligand binding site for $P2Y_{12}R$ is electropositive (blue). The electropositive nature of the $P2Y_{12}R$ active site might explain why the $P2Y_{12}R$ crystal structures, with bound 2MeSATP or 2MeSADP, were obtained without a magnesium ion to begin with.

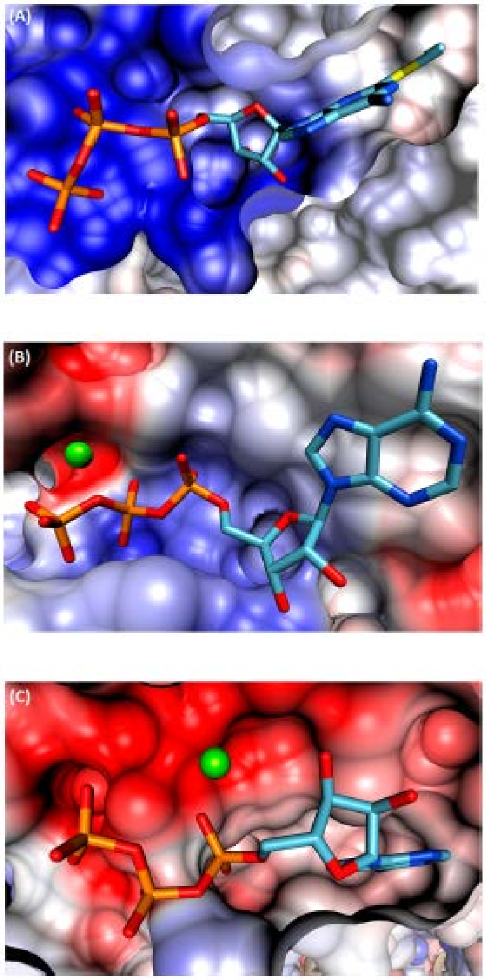


FIGURE 2.8: Cross-sectional comparison of the ATP binding site for A) $P2Y_{12}R$, B) GRK1, and C) FlaH. Blue and red coloring reflects a net positive or negative charge, carbon bonds are shown in cyan, phosphorus bonds are depicted in orange, and the magnesium ions are shown in green.

On the other hand, it is also reasonable to question whether the binding poses observed in the crystal structures are the active forms in the binding assay. Indeed, the authors of the crystal structure study speculated that their P2Y$_{12}$R-2MeSADP structure might be in an “agonist-bound inactive state” based on the positions of certain helices in comparison with other structures.[123]

However, the binding poses observed in the crystal structure active sites do make physical sense in that the electrostatic interactions between the highly negatively charged ligands and the many positively charged side chains in the P2Y$_{12}$R active site are favorable for binding. Absence of additional reactants or additional experimental measurements (i.e. NMR), it is difficult to justify a mechanism that contradicts the binding poses observed in the existing high quality structural data.

2.4 CONCLUDING REMARKS AND FUTURE DIRECTIONS

In this study, the widely used Amber/MMPBSA procedure was extended to model membrane protein-ligand binding affinities; in particular, it was used to model the binding of several ligands to the human purinergic platelet receptor, P2Y$_{12}$R. A good agreement with experimental binding affinities was observed. A detailed investigation of simulation parameters was conducted to assess their impact on the accuracy of the MMPBSA results. The reported optimization procedure also illustrated the various details that should be considered while applying the MMPBSA method in studies of other membrane protein systems.

Testing of the non-polar solvent model indicates that the modern dispersion/cavity method (inp = 2), which separately models dispersion and hydrophobic interactions, yields improved agreement with experiment all around as compared with the classical, but simpler, linear response model (inp = 1). These results suggest that the modern approach is preferred when performing binding free energy calculations for membrane protein-ligand systems, or at least it is preferred for ligand binding cavities that are fully exposed to water, an observation that is consistent with studies of globular proteins[8].

An investigation of protein dielectric constants led to an optimized value of 20. This is due to the highly charged and partially exposed nature of the active site of P2Y$_{12}$R, and this is also consistent with many previous studies of globular proteins[6, 8, 145, 146]. The introduction of a higher protein/solute dielectric constant is a reasonable, but crude, treatment of the screening effect of electrostatic interactions due to polarization of electronic, orientational, and solvent-exchange origins. The screening effect reduces otherwise very favorable electrostatic interactions, rendering these interactions comparable to hydrophobic interactions in most biochemical systems.

A similar investigation of membrane dielectric constants led to an optimized value of 4. This is within the 1-4 range that is indicative of highly hydrophobic regions of membranes, and it agrees well with other values reported in the literature[30, 117, 119, 149]. It is worth noting that an epsmem value of 1, 2, or even 7 can be used without sacrificing much accuracy as the effect of changing the membrane dielectric constant is more subtle in comparison to changes in the protein dielectric constant. This result seems reasonable for our system. An active site exposed on the protein surface should be somewhat separated from the hydrophobic, non-polar region in the surrounding lipid bilayer, and therefore, the effect of the membrane dielectric constant on the binding free energy should be more modest. Finally, it is worth noting that the magnesium correction had no influence on the trends observed in the analysis of simulation parameters. If the uncorrected data is analyzed on its own in **FIGURES 2.3-2.5** and **TABLES 2.1-2.3**, we see the same general trends that we see when we analyze the corrected data on its own.

We hypothesized that the bound magnesium ion is lost when 2MeSATP or 2MeSADP bind to P2Y$_{12}$R. This was based on inspection of the complex crystal structures, and the fact that the binding assay was conducted in the presence of 10mM MgCl$_2$[122]. This is supported by our modeling study of the binding reactions, which uses standard setups from a widely employed method. Additional structural analysis shows that P2Y$_{12}$R is indeed an interesting case. The electropositive nature of its active site might explain why the P2Y$_{12}$R complex structures were obtained without magnesium ions. Nevertheless, the approach we have taken above is just one possible way to resolve the anomaly in the 2MeSATP and 2MeSADP binding affinities. Another possibility is that the bound magnesium ion may remain in the P2Y$_{12}$R binding site upon 2MeSATP and 2MeSADP binding. This is akin to reducing the net charges of the two highly charged ligands. However, in this situation the remaining charges of the ligands are still very high, and the electrostatic interactions remain highly favorable, even when a high apparent protein dielectric constant is used, leading to poor overall agreement with experiment (data not shown). Of course, this possibility also contradicts the magnesium-free binding of 2MeSATP and 2MeSADP observed in the P2Y$_{12}$R crystal structures.

2.5 REFERENCES

1. Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham, T. E., 3rd; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W., et al. *Amber 16*; University of California, San Francisco: San Francisco, CA **2016**.

2. Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham, T.E., 3rd; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W., et al. *AmberTools 16*; University of California, San Francisco: San Francsico, CA **2016.**

3. Srinivasan, J.; Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate - DNA Helices. *J. Am. Chem. Soc.* **1998,** *120*, 9401-9409.

4. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W., et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889-897.

5. Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004,** *25*, 238-250.

6. Yang, T. Y.; Wu, J. C.; Yan, C. L.; Wang, Y. F.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Y. Virtual Screening Using Molecular Simulations. *Proteins: Struct., Funct., Bioinf.* **2011,** *79*, 1940-1951.

7. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314-3321.

8. Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B. N.; Wang, H.; Ren, P.; Luo, R. Calculating Protein–Ligand Binding Affinities With MMPBSA: Method and Error Analysis. *J. Comput. Chem.* **2016**, 37, 2436-2446.

9. Perutz, M. F. Electrostatic Effects in Proteins. *Science* **1978,** *201*, 1187-1191.

10. Davis, M. E.; Mccammon, J. A. Electrostatics in Biomolecular Structure and Dynamics. *Chem. Rev.* **1990,** *90*, 509-521.

11. Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995,** *268*, 1144-1149.

12. Honig, B.; Sharp, K.; Yang, A. S. Macroscopic Models Of Aqueous-Solutions - Biological And Chemical Applications. *J. Phys. Chem.* **1993,** *97*, 1101-1109.

13. Beglov, D.; Roux, B. Solvation of Complex Molecules in a Polar Liquid: An Integral Equation Theory. *J. Chem. Phys.* **1996,** *104*, 8678-8689.

14. Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999,** *99*, 2161-2200.

15. Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000,** *51*, 129-152.

16. Baker, N. A. Improving Implicit Solvent Simulations: A Poisson-Centric View. *Curr. Opin. Struct. Biol.* **2005,** *15*, 137-143.

17. Chen, J. H.; Im, W. P.; Brooks, C. L. Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. *J. Am. Chem. Soc.* **2006,** *128*, 3728-3736.

18. Feig, M.; Chocholousova, J.; Tanizaki, S. Extending the Horizon: Towards the Efficient Modeling of Large Biomolecular Complexes in Atomic Detail. *Theor. Chem. Acc.* **2006,** *116*, 194-205.

19. Koehl, P. Electrostatics Calculations: Latest Methodological Advances. *Curr. Opin. Struct. Biol.* **2006,** *16*, 142-151.

20. Im, W.; Chen, J. H.; Brooks, C. L. in Advances in Protein Chemistry; Baldwin, R., Baker, D.J., Eds.; Academic Press: **2006**; Vol. *72*, pp 173-198.

21. Lu, B. Z.; Zhou, Y. C.; Holst, M. J.; McCammon, J. A. Recent Progress in Numerical Methods for the Poisson-Boltzmann Equation in Biophysical Applications. *Commun. Comput. Phys.* **2008,** *3*, 973-1009.

22. Wang, J.; Tan, C. H.; Tan, Y. H.; Lu, Q.; Luo, R. Poisson-Boltzmann Solvents in Molecular Dynamics Simulations. *Commun. Comput. Phys.* **2008,** *3*, 1010-1031.

23. Altman, M. D.; Bardhan, J. P.; White, J. K.; Tidor, B. Accurate Solution of Multi-Region Continuum Biomolecule Electrostatic Problems Using the Linearized Poisson-Boltzmann Equation with Curved Boundary Elements. *J. Comput. Chem.* **2009,** *30*, 132-153.

24. Cai, Q.; Wang, J.; Hsieh, M. J.; Ye, X.; Luo, R. In *Annual Reports in Computational Chemistry*; Ralph, A. W., Ed.; Elsevier: **2012**, Vol. 8, pp 149-162.

25. Xiao, L.; Wang, C.; Luo, R. Recent Progress in Adapting Poisson–Boltzmann Methods to Molecular Simulations. *J. Theor. Comput. Chem.* **2014,** *13*, 1430001.

26. Botello-Smith, W. M.; Cai, Q.; Luo, R. Biological Applications of Classical Electrostatics Methods. *J. Theor. Comput. Chem.* **2014,** *13*, 1440008.

27. Forsten, K. E.; Kozack, R. E.; Lauffenburger, D. A.; Subramaniam, S. Numerical-Solution of the Nonlinear Poisson-Boltzmann Equation for a Membrane-Electrolyte System. *J. Phys. Chem.* **1994,** *98*, 5580-5586.

28. Spassov, V. Z.; Yan, L.; Szalma, S. Introducing an Implicit Membrane in Generalized Born/Solvent Accessibility Continuum Solvent Models. *J. Phys. Chem. B* **2002,** *106*, 8726-8738.

29. Im, W.; Feigh, M.; Brooks III, C. L. An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophys. J.* **2003,** *85*, 2900-2918.

30. Tanizaki, S.; Feig, M. A Generalized Born Formalism for Heterogeneous Dielectric Environments: Application to the Implicit Modeling of Biological Membranes. *J. Chem. Phys.* **2005,** *122*, 124706.

31. Tanizaki, S.; Feig, M. Molecular Dynamics Simulations of Large Integral Membrane Proteins with an Implicit Membrane Model. *J. Phys. Chem. B* **2006,** *110*, 548-556.

32. Callenberg, K. M.; Choudhary, O. P.; de Forest, G. L.; Gohara, D. W.; Baker, N. A.; Grabe, M. APBSmem: A Graphical Interface for Electrostatic Calculations at the Membrane. *PloS One* **2010,** *5*.

33. Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E. DelPhi: A Comprehensive Suite for DelPhi Software and Associated Resources. *BMC Biophysics* **2012,** *5*, 9.

34. Schafer, N. P.; Truong, H. H.; Otzen, D. E.; Lindorff-Larsen, K.; Wolynes, P. G. Topological Constraints and Modular Structure in the Folding and Functional Motions of GlpG, An Intramembrane Protease. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113*, 2098-2103.

35. Marassi, F. M.; Ding, Y.; Schwieters, C. D.; Tian, Y.; Yao, Y. Backbone Structure of Yersinia Pestis Ail Determined in Micelles by NMR-Restrained Simulated Annealing with Implicit Membrane Solvation. *J. Biomol. NMR* **2015,** *63*, 59-65.

36. Bajda, M.; Filipek, S. Study of Early Stages of Amyloid Aβ13-23 Formation Using Molecular Dynamics Simulation in Implicit Environments. *Comput. Biol. Chem.* **2015,** *56*, 13-18.

37. Li, C.; Li, L.; Zhang, J.; Alexov, E. Highly Efficient and Exact Method for Parallelization of Grid-Based Algorithms and Its Implementation in DelPhi. *J. Comput. Chem.* **2012,** *33*, 1960-1966.

38. Wang, J.; Cai, Q.; Li, Z. L.; Zhao, H. K.; Luo, R. Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms. *Chem. Phys. Lett.* **2009,** *468*, 112-118.

39. Wang, J.; Luo, R. Assessment of Linear Finite-Difference Poisson-Boltzmann Solvers. *J. Comput. Chem.* **2010,** *31*, 1689-1698.

40. Cai, Q.; Hsieh, M. J.; Wang, J.; Luo, R. Performance of Nonlinear Finite-Difference Poisson-Boltzmann Solvers. *J. Chem. Theory Comput.* **2010,** *6*, 203-211.

41. Wang, J.; Cai, Q.; Xiang, Y.; Luo, R. Reducing Grid Dependence in Finite-Difference Poisson-Boltzmann Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 2741-2751.

42. Botello-Smith, W. M.; Luo, R. Applications of MMPBSA to Membrane Proteins I: Efficient Numerical Solutions of Periodic Poisson-Boltzmann Equation. *J. Chem. Inf. Model.* **2015,** *55*, 2187-2199.

43. Warwicker, J.; Watson, H. C. Calculation of the Electric-Potential in the Active-Site Cleft Due to Alpha-Helix Dipoles. *J. Mol. Biol.* **1982,** *157*, 671-679.

44. Bashford, D.; Karplus, M. Pkas Of Ionizable Groups In Proteins - Atomic Detail From A Continuum Electrostatic Model. *Biochemistry* **1990,** *29*, 10219-10225.

45. Jeancharles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. Electrostatic Contributions To Solvation Energies - Comparison Of Free-Energy Perturbation And Continuum Calculations. *J. Am. Chem. Soc.* **1991,** *113*, 1454-1455.

46. Gilson, M. K. Theory Of Electrostatic Interactions In Macromolecules. *Curr. Opin. Struct. Biol.* **1995,** *5*, 216-223.

47. Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. Solvation Free Energies of Peptides: Comparison of Approximate Continuum Solvation Models with Accurate Solution of the Poisson-Boltzmann Equation. *J. Phys. Chem. B* **1997,** *101*, 1190-1197.

48. Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann Calculations for Static and Dynamic Systems. *J. Comput. Chem.* **2002,** *23*, 1244-1253.

49. Lu, Q.; Luo, R. A Poisson-Boltzmann Dynamics Method with Nonperiodic Boundary Condition. *J. Chem. Phys.* **2003,** *119*, 11035-11047.

50. Tan, C.; Yang, L.; Luo, R. How Well Does Poisson-Boltzmann Implicit Solvent Agree With Explicit Solvent? A Quantitative Analysis. *J. Phys. Chem. B* **2006,** *110*, 18680-18687.

51. Cai, Q.; Wang, J.; Zhao, H.-K.; Luo, R. On Removal of Charge Singularity in Poisson-Boltzmann Equation. *J. Chem. Phys.* **2009,** *130*, 145101.

52. Ye, X.; Cai, Q.; Yang, W.; Luo, R. Roles of Boundary Conditions in DNA Simulations: Analysis of Ion Distributions with the Finite-Difference Poisson-Boltzmann Method. *Biophys. J.* **2009,** *97*, 554-562.

53. Ye, X.; Wang, J.; Luo, R. A Revised Density Function for Molecular Surface Calculation in Continuum Solvent Models. *J. Chem. Theory Comput.* **2010,** *6*, 1157-1169.

54. Luo, R.; Moult, J.; Gilson, M. K. Dielectric Screening Treatment of Electrostatic Solvation. *J. Phys. Chem. B* **1997,** *101*, 11226-11236.

55. Wang, J.; Tan, C.; Chanco, E.; Luo, R. Quantitative Analysis of Poisson-Boltzmann Implicit Solvent in Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2010,** *12*, 1194-1202.

56. Hsieh, M. J.; Luo, R. Exploring a Coarse-Grained Distributive Strategy for Finite-Difference Poisson-Boltzmann Calculations. *J. Mol. Model.* **2011,** *17*, 1985-1996.

57. Cai, Q.; Ye, X.; Wang, J.; Luo, R. On-the-Fly Numerical Surface Integration for Finite-Difference Poisson-Boltzmann Methods. *J. Chem. Theory Comput.* **2011,** *7*, 3608-3619.

58. Botello-Smith, W. M.; Liu, X.; Cai, Q.; Li, Z.; Zhao, H.; Luo, R. Numerical Poisson-Boltzmann Model for Continuum Membrane Systems. *Chem. Phys. Lett.* **2013**, *555*, 274-281.

59. Liu, X.; Wang, C.; Wang, J.; Li, Z.; Zhao, H.; Luo, R. Exploring a Charge-Central Strategy in the Solution of Poisson's Equation for Biomolecular Applications. *Phys. Chem. Chem. Phys.* **2013**, *15*, 129-141.

60. Wang, C.; Wang, J.; Cai, Q.; Li, Z. L.; Zhao, H.; Luo, R. Exploring High Accuracy Poisson-Boltzmann Methods for Biomolecular Simulations. *Comput. Theor. Chem.* **2013,** *1024*, 34-44.

61. Luo, R.; Head, M. S.; Moult, J.; Gilson, M. K. pK(a) Shifts in Small Molecules and HIV Protease: Electrostatics and Conformation. *J. Am. Chem. Soc.* **1998,** *120*, 6138-6146.

62. Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. Combining Conformational Flexibility and Continuum Electrostatics for Calculating pK(a)s in Proteins. *Biophys. J.* **2002,** *83*, 1731-1748.

63. Nielsen, J. E.; McCammon, J. A. On the Evaluation and Optimization of Protein X-ray Structures for pKa Calculations. *Protein Sci.* **2003,** *12*, 313-326.

64. Warwicker, J. Improved pK(a) Calculations Through Flexibility Based Sampling of a Water-Dominated Interaction Scheme. *Protein Sci.* **2004,** *13*, 2793-2805.

65. Tang, C. L.; Alexov, E.; Pyle, A. M.; Honig, B. Calculation of pK(a)s in RNA: On the Structural Origins and Functional Roles of Protonated Nucleotides. *J. Mol. Biol.* **2007,** *366*, 1475-1496.

66. Shivakumar, D.; Deng, Y. Q.; Roux, B. Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput.* **2009,** *5*, 919-930.

67. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *J. Med. Chem.* **2008,** *51*, 769-779.

68. Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.* **2004,** *86*, 67-74.

69. Bertonati, C.; Honig, B.; Alexov, E. Poisson-Boltzmann Calculations of Nonspecific Salt Effects on Protein-Protein Binding Free Energies. *Biophys. J.* **2007,** *92*, 1891-1899.

70. Brice, A. R.; Dominy, B. N. Analyzing the Robustness of the MM/PBSA Free Energy Calculation Method: Application to DNA Conformational Transitions. *J. Comput. Chem.* **2011,** *32*, 1431-1440.

71. Luo, R.; Gilson, H. S. R.; Potter, M. J.; Gilson, M. K. The Physical Basis of Nucleic Acid Base Stacking in Water. *Biophys. J.* **2001,** *80*, 140-148.

72. David, L.; Luo, R.; Head, M. S.; Gilson, M. K. Computational Study of KNI-272, a Potent Inhibitor of HIV-1 Protease: On the Mechanism of Preorganization. *J. Phys. Chem. B* **1999,** *103*, 1031-1044.

73. Luo, R.; Gilson, M. K. Synthetic Adenine Receptors: Direct Calculation of Binding Affinity and Entropy. *J. Am. Chem. Soc.* **2000,** *122*, 2934-2937.

74. Marshall, S. A.; Vizcarra, C. L.; Mayo, S. L. One- and Two-Body Decomposable Poisson-Boltzmann Methods for Protein Design Calculations. *Protein Sci.* **2005,** *14*, 1293-1304.

75. Hsieh, M. J.; Luo, R. Physical Scoring Function Based on AMBER Force Field and Poisson-Boltzmann Implicit Solvent for Protein Structure Prediction. *Proteins: Struct., Funct., Bioinf.* **2004,** *56*, 475-486.

76. Wen, E. Z.; Luo, R. Interplay of Secondary Structures and Side-Chain Contacts in the Denatured State of BBA1. *J. of Chem. Phys.* **2004,** *121*, 2412-2421.

77. Wen, E. Z.; Hsieh, M. J.; Kollman, P. A.; Luo, R. Enhanced Ab Initio Protein Folding Simulations in Poisson-Boltzmann Molecular Dynamics with Self-Guiding Forces. *J. Mol. Graphics Modell.* **2004,** *22*, 415-424.

78. Lwin, T. Z.; Luo, R. Overcoming Entropic Barrier with Coupled Sampling at Dual Resolutions. *J. Chem. Phys.* **2005,** *123*, 194904.

79. Lwin, T. Z.; Zhou, R. H.; Luo, R. Is Poisson-Boltzmann Theory Insufficient for Protein Folding Simulations? *J. Chem. Phys.* **2006,** *124*, 034906.

80. Lwin, T. Z.; Luo, R. Force Field Influences in Beta-Hairpin Folding Simulations. *Protein Science* **2006,** *15*, 2642-2655.

81. Tan, Y.-H.; Luo, R. Protein Stability Prediction: A Poisson-Boltzmann Approach. *J. Phys. Chem. B* **2008,** *112*, 1875-1883.

82. Tan, Y.; Luo, R. Structural and Functional Implications of p53 Missense Cancer Mutations. *BMC Biophysics* **2009,** *2*, 5.

83. Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.C. Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase. *Biochemistry* **2008,** *47*, 1837-1847.

84. Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum - a Direct Utilization of Abinitio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981,** *55*, 117-129.

85. Hoshi, H.; Sakurai, M.; Inoue, Y.; Chujo, R. Medium Effects on the Molecular Electronic-Structure 1. the Formulation of a Theory for the Estimation of a Molecular Electronic-Structure Surrounded by an Anisotropic Medium. *J. Chem. Phys.* **1987,** *87*, 1107-1115.

86. Zauhar, R. J.; Morgan, R. S. The Rigorous Computation of the Molecular Electric-Potential. *J. Comput. Chem.* **1988,** *9*, 171-187.

87. Rashin, A. A. Hydration Phenomena, Classical Electrostatics, and the Boundary Element Method. *J. Phys. Chem.* **1990,** *94*, 1725-1733.

88. Yoon, B. J.; Lenhoff, A. M. A Boundary Element Method for Molecular Electrostatics with Electrolyte Effects. *J. Comput. Chem.* **1990,** *11*, 1080-1086.

89. Juffer, A. H.; Botta, E. F. F.; Vankeulen, B. A. M.; Vanderploeg, A.; Berendsen, H. J. C. The Electric-Potential Of A Macromolecule In A Solvent - A Fundamental Approach. *J. Comput. Phys.* **1991,** *97*, 144-171.

90. Zhou, H. X. Boundary-Element Solution of Macromolecular Electorstatic-Interaction Energy Between 2 Proteins. *Biophys. J.* **1993,** *65*, 955-963.

91. Bharadwaj, R.; Windemuth, A.; Sridharan, S.; Honig, B.; Nicholls, A. The Fast Multipole Boundary-Element Method for Molecular Electrostatics - an Optimal Approach for Large Systems. *J. Comput. Chem.* **1995,** *16*, 898-913.

92. Purisima, E. O.; Nilar, S. H. A Simple Yet Accurate Boundary-Element Method for Continuum Dielectric Calculations. *J. Comput. Chem.* **1995,** *16*, 681-689.

93. Liang, J.; Subramaniam, S. Computation of Molecular Electrostatics with Boundary Element Methods. *Biophys. J.* **1997,** *73*, 1830-1841.

94. Vorobjev, Y. N.; Scheraga, H. A. A fast Adaptive Multigrid Boundary Element Method for Macromolecular Electrostatic Computations in a Solvent. *J. Comput. Chem.* **1997,** *18*, 569-583.

95. Totrov, M.; Abagyan, R. Rapid Boundary Element Solvation Electrostatics Calculations in Folding Simulations: Successful Folding of a 23-Residue Peptide. *Biopolymers* **2001,** *60*, 124-133.

96. Boschitsch, A. H.; Fenley, M. O.; Zhou, H. X. Fast Boundary Element Method for the Linear Poisson-Boltzmann Equation. *J. Phys. Chem. B* **2002,** *106*, 2741-2754.

97. Lu, B. Z.; Cheng, X. L.; Huang, J. F.; McCammon, J. A. Order N Algorithm for Computation of Electrostatic Interactions in Biomolecular Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 19314-19319.

98. Lu, B.; Cheng, X.; Huang, J.; McCammon, J. A. An Adaptive Fast Multipole Boundary Element Method for Poisson-Boltzmann Electrostatics. *J. Chem. Theory Comput.* **2009,** *5*, 1692-1699.

99. Bajaj, C.; Chen, S. C.; Rand, A. An Efficient Higher-Order Fast Multipole Boundary Element Solution For Poisson-Boltzmann-Based Molecular Electrostatics. *SIAM J. Sci. Comput.* **2011,** *33*, 826-848.

100. Cortis, C. M.; Friesner, R. A. Numerical Solution of the Poisson-Boltzmann Equation Using Tetrahedral Finite-Element Meshes. *J. Comput. Chem.* **1997,** *18*, 1591-1608.

101. Holst, M.; Baker, N.; Wang, F. Adaptive Multilevel Finite Element Solution of the Poisson-Boltzmann Equation I. Algorithms and Examples. *J. Comput. Chem.* **2000,** *21*, 1319-1342.

102. Baker, N.; Holst, M.; Wang, F. Adaptive Multilevel Finite Element Solution of the Poisson-Boltzmann Equation II. Refinement at Solvent-Accessible Surfaces in Biomolecular Systems. *J. Comput. Chem.* **2000,** *21*, 1343-1352.

103. Shestakov, A. I.; Milovich, J. L.; Noy, A. Solution of the Nonlinear Poisson-Boltzmann Equation Using Pseudo-Transient Continuation and the Finite Element Method. *J. Colloid Interface Sci.* **2002,** *247*, 62-79.

104. Chen, L.; Holst, M. J.; Xu, J. C. The Finite Element Approximation of the Nonlinear Poisson-Boltzmann Equation. *SIAM J. Numer.Anal.* **2007,** *45*, 2298-2320.

105. Xie, D.; Zhou, S. A New Minimization Protocol for Solving Nonlinear Poisson–Boltzmann Mortar Finite Element Equation. *BIT Numer. Math.* **2007,** *47*, 853-871.

106. Bond, S. D.; Chaudhry, J. H.; Cyr, E. C.; Olson, L. N. A First-Order System Least-Squares Finite Element Method for the Poisson-Boltzmann Equation. *J. Comput. Chem.* **2010,** *31*, 1625-1635.

107. Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. Focusing of Electric Fields in the Active Site of Copper-Zinc Superoxide Dismutase Effects of Ionic Strength and Amino Acid Modification. *Proteins: Struct., Funct., Genet.* **1986,** *1*, 47-59.

108. Davis, M. E.; McCammon, J. A. Solving The Finite-Difference Linearized Poisson-Boltzmann Equation - A Comparison Of Relaxation And Conjugate-Gradient Methods. *J. Comp. Chem.* **1989,** *10*, 386-391.

109. Nicholls, A.; Honig, B. A Rapid Finite-Difference Algorithm, Utilizing Successive Over-Relaxation To Solve The Poisson-Boltzmann Equation. *J. Comp. Chem.* **1991,** *12*, 435-445.

110. Luty, B. A.; Davis, M. E.; McCammon, J. A. Solving the Finite-Difference Nonlinear Poisson-Boltzmann Equation. *J. Comput. Chem.* **1992,** *13*, 1114-1118.

111. Holst, M.; Saied, F. Multigrid Solution of the Poisson-Boltzmann Equation. *J. Comput. Chem.* **1993,** *14*, 105-113.

112. Holst, M. J.; Saied, F. Numerical-Solution Of The Nonlinear Poisson-Boltzmann Equation - Developing More Robust And Efficient Methods. *J. Comput. Chem.* **1995,** *16*, 337-364.

113. Bashford, D. An Object-Oriented Programming Suite for Electrostatic Effects in Biological Molecules. *Lect. Notes Comput. Sci.* **1997,** *1343*, 233-240.

114. Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: Computation of Electrostatic Forces from Numerical Solutions to the Poisson-Boltzmann Equation. *Comput. Phys. Commun.* **1998,** *111*, 59-75.

115. Rocchia, W.; Alexov, E.; Honig, B. Extending the Applicability of the Nonlinear Poisson-Boltzmann equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **2001,** *105*, 6507-6514.

116. Im, W.; Feig, M.; Brooks, C. L. An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophys. J.* **2003,** *85*, 2900-2918.

117. Nymeyer, H.; Zhou, H.X. A Method to Determine Dielectric Constants in Nonhomogeneous Systems: Application to Biological Membranes. *Biophys. J.* **2008,** *94*, 1185-1193.

118. Tanizaki, S.; Feig, M. A Generalized Born Formalism for Heterogeneous Dielectric Environments: Application to the Implicit Modeling of Biological Membranes. *The Journal of chemical physics* **2005,** *122*, 124706.

119. Stern, H. A.; Feller, S. E. Calculation of the Dielectric Permittivity Profile for a Nonuniform System: Application to a Lipid Bilayer Simulation. *J. Chem. Phys.* **2003,** *118*, 3401.

120. Ulmschneider, M. B.; Ulmschneider, J. P.; Sansom, M. S.; Di Nola, A. A Generalized Born Implicit-Membrane Representation Compared to Experimental Insertion Free Energies. *Biophys. J.* **2007,** *92*, 2338-2349.

121. Tan, C.; Tan, Y. H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007,** *111*, 12263-74.

122. Zhang, K.; Zhang, J.; Gao, Z. G.; Zhang, D.; Zhu, L.; Han, G. W.; Moss, S. M.; Paoletta, S.; Kiselev, E.; Lu, W. Structure of the Human P2Y12 Receptor in Complex with an Antithrombotic Drug. *Nature* **2014,** *509*, 115-118.

123. Zhang, J.; Zhang, K.; Gao, Z. G.; Paoletta, S.; Zhang, D.; Han, G. W.; Li, T.; Ma, L.; Zhang, W.; Müller, C. E., et al. Agonist-Bound Structure of the Human P2Y12 Receptor. *Nature* **2014,** *509*, 119-122.

124. Savi, P.; Zachayus, J. L.; Delesque-Touchard, N.; Labouret, C.; Hervé, C.; Uzabiaga, M. F.; Pereillo, J. M.; Culouscou, J. M.; Bono, F.; Ferrara, P., et al. The Active Metabolite of Clopidogrel Disrupts P2Y12 Receptor Oligomers and Partitions Them Out of Lipid Rafts. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103*, 11069-11074.

125. Algaier, I.; Jakubowski, J. A.; Asai, F.; Von Kügelgen, I. Interaction of the Active Metabolite of Prasugrel, R-138727, with Cysteine 97 and Cysteine 175 of the Human P2Y12 Receptor. *J. Thromb. Haemostasis* **2008,** *6*, 1908-1914.

126. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinf.* **2016,** *54*, 5.6.1-5.6.37.

127. Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. MultiSeq: Unifying Sequence and Structure Data for Evolutionary Analysis. *BMC Bioinf.* **2006,** *7*, 382.

128. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996,** *14*, 33-38, 27-28.

129. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.

130. Buchanan, A.; Newton, P.; Pehrsson, S.; Inghardt, T.; Antonsson, T.; Svensson, P.; Sjögren, T.; Öster, L.; Janefeldt, A.; Sandinge, A. S., et al. Structural and Functional Characterization of a Specific Antidote for Ticagrelor. *Blood* **2015,** *125*, 3484-3490.

131. *MarvinSketch*, Version *16.3.2*; Software for Chemical Visualization; ChemAxon: Cambridge, MA, **2016**.

132. *OpenEye Toolkits*, Software Toolkits for Cheminformatics and Modeling; OpenEye: Santa Fe, NM, **2016**.

133. Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina From the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013,** *53*, 1893-1904.

134. Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of Polyphosphate Parameters for Use with the AMBER Force Field. *J. Comput. Chem.* **2003,** *24*, 1016-1025.

135. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber, an Accessory Software Package for Molecular Mechanical Calculations. *J. Comput. Chem.* **2005**, 25, 1157-1174.

136. Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007,** *2*, e880.

137. Biró, E.; Akkerman, J. W.; Hoek, F. J.; Gorter, G.; Pronk, L. M.; Sturk, A.; Nieuwland, R. The Phospholipid Composition and Cholesterol Content of Platelet-Derived Microparticles: a Comparison with Platelet Membrane Fractions. *J. Thromb. Haemostasis* **2005,** *3*, 2754-2763.

138. Roe, D. R.; Cheatham, T. E., 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

139. Tan, C.; Tan, Y. H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007,** *111*, 12263-12274.

140. Liao, J. C.; Sun, S.; Chandler, D.; Oster, G. The Conformational States of Mg-ATP in Water. *Eur. Biophys. J.* **2004,** *33*, 29-37.

141. Alberty, R. A. Standard Gibbs Free Energy, Enthalpy, and Entropy Changes as a Function of pH and pMg for Several Reactions Involving Adenosine Phosphates. *J. Biol. Chem.* **1969,** *244*, 3290-3302.

142. Gout, E.; Rébeillé, F.; Douce, R.; Bligny, R. Interplay of Mg2+, ADP, and ATP in the Cytosol and Mitochondria: Unravelling the Role of Mg2+ in Cell Respiration. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*, E4560-E4567.

143. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.

144. Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986,** *319*, 199-203.

145. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011,** *51*, 69-82.

146. Talley, K.; Ng, C.; Shoppell, M.; Kundrotas, P.; Alexov, E. On the Electrostatic Component of Protein-Protein Binding Free Energy. *PMC Biophys.* **2008,** *1*, 1-23.

147. Singh, P.; Wang, B.; Maeda, T.; Palczewski, K.; Tesmer, J. J. Structures of Rhodopsin Kinase in Different Ligand States Reveal Key Elements Involved in G Protein-Coupled Receptor Kinase Activation. *J. Biol. Chem.* **2008,** *283*, 14053-14062.

148. Chaudhury, P.; Neiner, T.; D'Imprima, E.; Banerjee, A.; Reindl, S.; Ghosh, A.; Arvai, A. S.; Mills, D. J.; van der Does, C.; Tainer, J. A., et al. The Nucleotide-Dependent Interaction of FlaH and FlaI is Essential for Assembly and Function of the Archaellum Motor. *Mol. Microbiol.* **2016,** *99*, 674-685.

149. Dias, R. P.; Li, L.; Soares, T. A.; Alexov, E. Modeling the Electrostatic Potential of Asymmetric Lipopolysaccharide Membranes: the MEMPOT Algorithm Implemented in DelPhi. *J. Comput. Chem.* **2014,** *35*, 1418-1429.

CHAPTER 3

Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies

Reprinted (adapted) with permission from Greene, D.; Qi, R.; Nguyen, R.; Qiu, T.; Luo, R. Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies. *J. Chem. Inf. Model.* **2019**, *59*, 3041-3056. Copyright 2019 American Chemical Society.

ABSTRACT

Membrane-bound protein receptors are a primary biological drug target, but the computational analysis of membrane proteins has been limited. In order to improve Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) binding free energy calculations for membrane protein-ligand systems, we have optimized a new heterogeneous dielectric implicit membrane model, with respect to free energy simulations in explicit membrane and explicit water, and implemented it into the Amber software suite. This new model supersedes our previous uniform, single dielectric implicit membrane model by allowing the dielectric constant to vary with depth within the membrane. We calculated MMPBSA binding free energies for the Human Purinergic Platelet Receptor (P2Y$_{12}$R) and two of the muscarinic acetylcholine receptors (M2R and M3R) bound to various antagonist ligands using both membrane models, and we found that the heterogeneous dielectric membrane model has a stronger correlation with experimental binding affinities compared to the older model under otherwise identical conditions. This improved membrane model increases the utility of MMPBSA calculations for the rational design and improvement of future drug candidates.

3.1 INTRODUCTION

Membrane proteins connect the internal environment of a cell to the surrounding medium. In doing so, they act as receptors of extracellular ligands, transporters that allow the passage of small molecules, enzymes that catalyze chemical reactions, and more[1]. The accessibility of membrane-bound protein receptors on the cell surface has made them a primary biological drug target. It is estimated that nearly 40% of all drug targets are either class I G-protein coupled receptors or nuclear receptors while 60% of all drug targets appear on the cell surface[2].

In rational drug design efforts, the computational analysis of protein-ligand binding affinities has become an important tool. Efficient, implicit solvent-based methods for modeling globular protein-ligand binding have been employed for many years, but the extension of such methods to membrane protein-ligand binding is far less developed despite its pharmacological importance. A complication that has held back the development of implicit solvent models for membrane proteins is that they are much more difficult to implement; the surrounding membrane must be taken into account along with the aqueous environment in the solvation model. Another factor is that the total number of membrane protein crystal structures available in the RCSB Protein Data Bank (PDB) is less than 3% of the total number of proteins deposited in the database[3, 4]. This reflects the experimental difficulties in obtaining viable structures of membrane bound proteins[5]. Nevertheless, over 92% of current drug targets are reported to be similar to known proteins in the PDB, which suggests a strong structural coverage of druggable membrane proteins for use in rational drug design[2]. In addition, methodological improvements to cryo-electron microscopy (cryo-EM) within the last few years have opened up the possibility that several high-profile membrane proteins which have resisted crystallization will be available for study in the near future[6].

Recently, our lab implemented an implicit membrane model into the Amber 16 software suite[7, 8] that allowed users to calculate the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) binding free energy of a membrane protein-ligand system[9-12]. PBSA based implicit solvent models have been used in a wide variety of biological applications including the prediction of pKa shifts[13, 14], solvation free energies[15-18], protein folding[19-26], and binding free energies[27-30]. In PBSA calculations, the electrostatic interaction term is calculated using the Poisson-Boltzmann equation (PBE)[31-36], which is frequently approximated as the linear PBE:

$$\nabla \cdot \varepsilon \nabla \phi = -4\pi\rho_0 + \varepsilon_v \kappa^2 \phi \qquad (1)$$

where $\kappa^2 = \frac{8\pi e^2 I}{\varepsilon_v k_B T}$. In eq. (1) v refers to the solvent, and the ionic strength is given as $I = z^2 c$. It can be seen in eq. (1) that the PBE has a dependency on the dielectric constant (ε), and an accurate selection of the dielectric constant will therefore have a direct impact on the accuracy of the overall PBSA calculation.

In Amber 16, our implicit membrane was modeled as a uniform slab with a single, low dielectric constant. While an important first step, a single dielectric implicit membrane model cannot effectively reconcile the low dielectric membrane core region, which consists of long non-polar hydrocarbon tails, with the high dielectric membrane periphery region, which contains polar and charged headgroups[37-39]. The hydrophobic core of the membrane has been described as having a low dielectric constant of around 2 while the polar headgroup region of the membrane has been reported as having a much higher dielectric constant, with estimates in the range of 80-1000[39, 40]. It has been demonstrated that these different dielectric properties can affect the interactions of small molecules with the membrane at different depths within the membrane. For instance, this difference in membrane structure can lead to qualitatively different permeabilities at different depths within the membrane when compared to membranes consisting of more uniformly structured alkane components[38, 41-45]. This is especially a concern for our intended application as many binding interactions take place at the interface between water and protein at the periphery, in the vicinity of phospholipid headgroups, as opposed to in the low dielectric membrane interior.

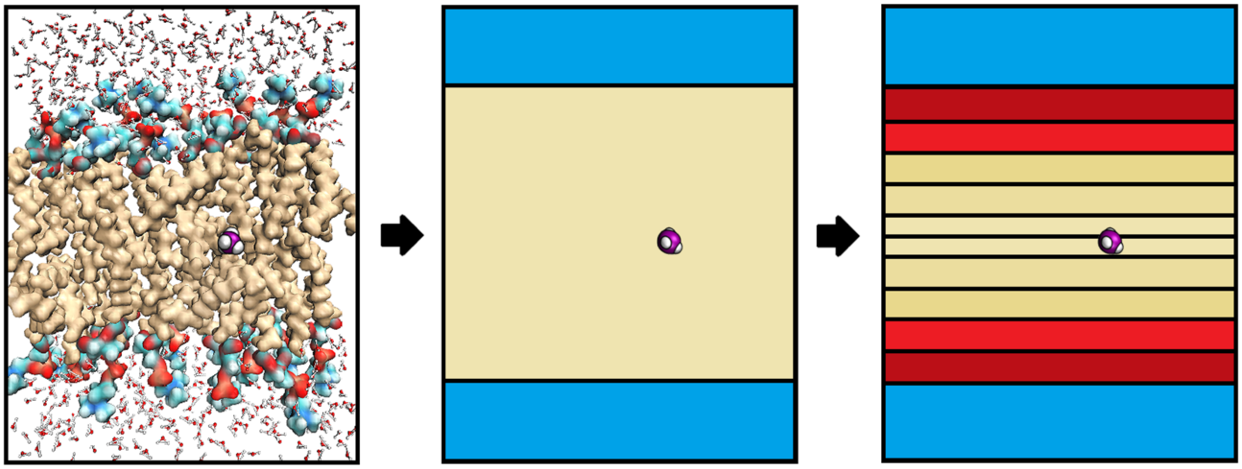


FIGURE 3.1. The evolution of a heterogeneous dielectric implicit membrane model. An explicit solvent model (left image) is first approximated by using an implicit solvent model where the solvent is treated as a dielectric continuum surrounding a molecule of interest, such as methane (purple and white). As a first approximation, the implicit solvent model (middle image) may consist of an aqueous phase (blue) with a high dielectric constant and a single, uniform membrane slab region (tan) with a single, low dielectric constant. To better approximate the heterogeneous nature of the explicit phospholipid membrane, the implicit solvent model can be improved by dividing the membrane slab

region into several small layers (right image) with each layer containing a different value of the dielectric constant that has been optimized for that layer.

The inherent inaccuracies of using a single dielectric membrane model can be lessened by implementing a heterogeneous dielectric membrane model that assigns a local dielectric constant that varies with depth within the phospholipid membrane (**FIGURE 3.1**)[46-50]. Here, we introduce such a heterogeneous dielectric implicit membrane model into Amber 18. In our present approach, we calibrate our implicit membrane model by calculating the free energy of de-charging the side chains of all 20 common amino acid residues at a given depth within the membrane. The optimal dielectric constant at a given depth is obtained from the global minimum of the root mean square deviation (RMSD) between the free energy for de-charging that is calculated in both the implicit and explicit membrane. This approach follows the philosophy of modern force field development by focusing on model compounds. In this case, we calibrate a variable dielectric profile in our implicit membrane model using a full set of parameterized amino acid side chains to represent the chemical properties of full-size membrane proteins. Of course, subsequent studies should pay attention to the transferability of the new implicit membrane model to different protein and ligand molecules, akin to what has been done for explicit membrane models. This current approach does allow for a more direct calibration of the membrane compared to prior methods which used a generic test charge[39, 40, 48].

The implementation of a heterogeneous dielectric implicit membrane model into the Amber PBSA framework is aimed at creating a much more accurate membrane environment for MMPBSA binding free energy calculations of membrane protein-ligand systems. To examine the impact of the new membrane model on this calculation for a few relevant test cases, we have calculated MMPBSA binding free energies for two G-protein coupled receptor systems, the human purinergic platelet receptor ($P2Y_{12}R$) and two muscarinic acetylcholine receptors (M2R and M3R) bound to various antagonist ligands, using both the new heterogeneous dielectric membrane model and the old uniform dielectric membrane model.

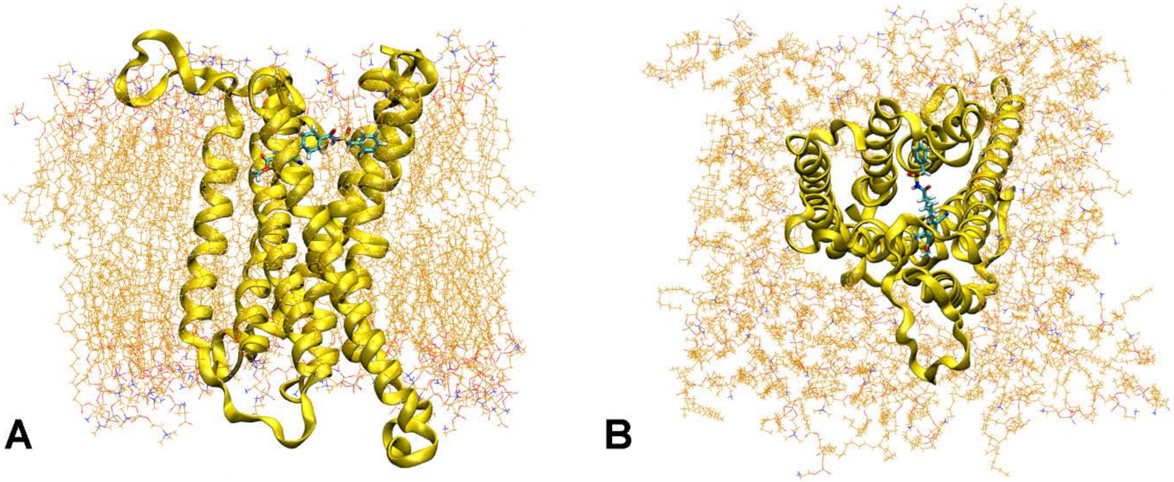


FIGURE 3.2. Model of the antagonist-bound $P2Y_{12}R$ system used in our MD simulations. A) Side view and B) Top down view of a snapshot taken from the production portion of the explicit MD trajectory for the $P2Y_{12}R$ receptor (yellow) bound to the AZD-1283 ligand (cyan). Nitrogen headgroup atoms that were used to calculate the average membrane thickness are shown in blue. Water has been omitted from this image for clarity.

$P2Y_{12}R$ is a prominent membrane bound receptor that is targeted for platelet aggregation inhibition in the treatment of myocardial infarction and stroke. Several antithrombotic drugs target $P2Y_{12}R$, but limitations of these drugs have motivated the development of a new generation of $P2Y_{12}R$ inhibitors[51-53]. The muscarinic acetylcholine receptor M3R is a target for antagonist drugs involved in treating chronic pulmonary disease. A desire to enhance selectivity for M3R over the structurally similar M2R subtype, which modulates heart rate, has fueled a structure-guided approach to antagonist ligand design[54].

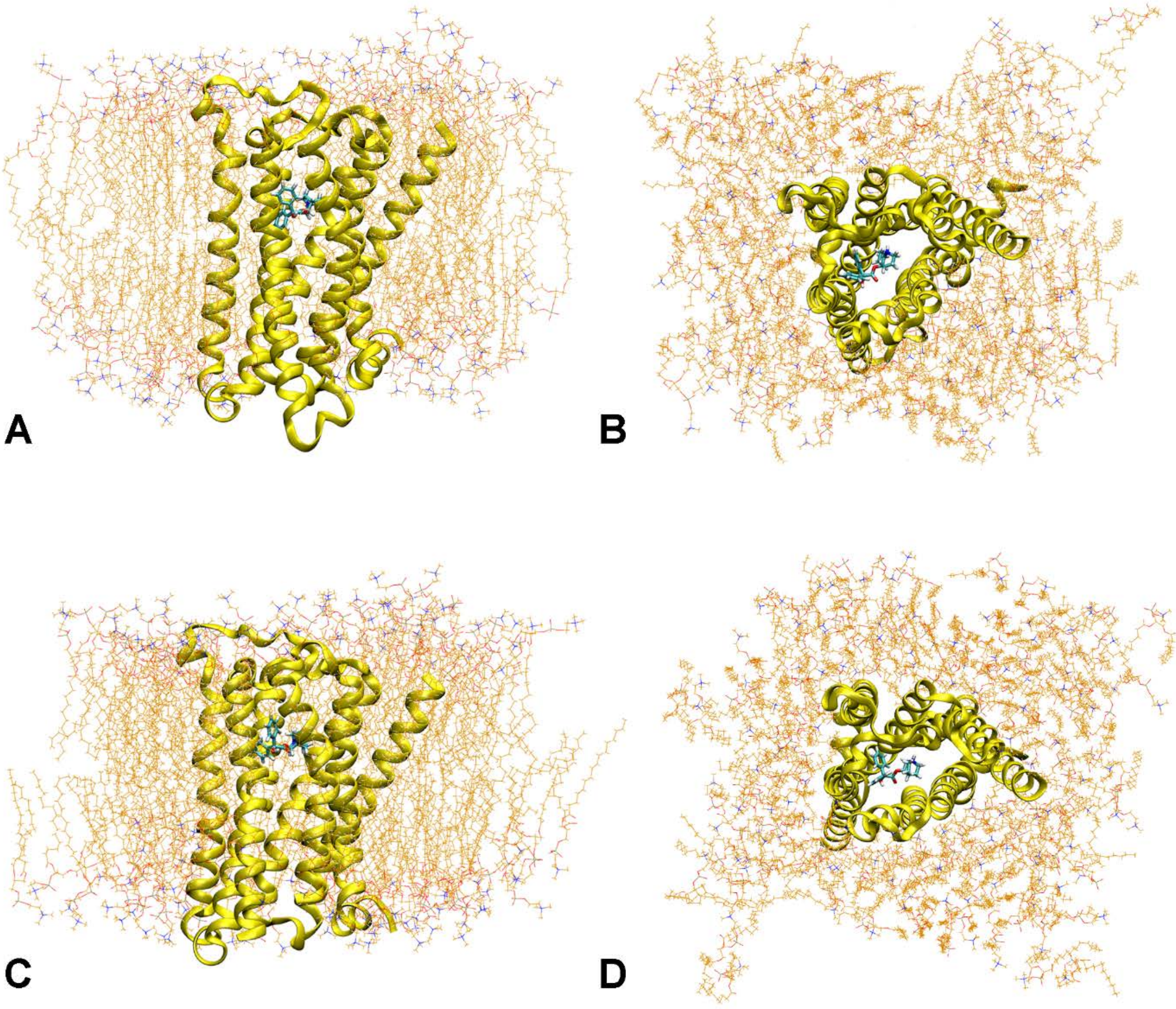


FIGURE 3.3. Model of the antagonist-bound M2R and M3R systems used in our MD simulations. A) Side view and B) Top down view of a snapshot taken from the production portion of the explicit MD trajectory for the M2R receptor (yellow) bound to the QNB

ligand (cyan). C) Side view and D) Top down view of a snapshot taken from the production portion of the explicit MD trajectory for the M3R receptor (yellow) bound to the QNB ligand (cyan). Nitrogen headgroup atoms that were used to calculate the average membrane thickness are shown in blue. Water has been omitted from this image for clarity.

For our membrane comparison, the P2Y$_{12}$R receptor has an open binding site that is exposed to water (**FIGURE 3.2**) whereas the closed binding site for M2R and M3R is buried a bit deeper into the membrane (**FIGURE 3.3**). These systems allow us to test the performance of our MMPBSA calculations at different depths within our membrane models. In the future, computational modeling may play a key role in the drug design process, and here we demonstrate how our improved membrane model increases the accuracy of MMPBSA calculations which can be used in the rational design of future drug candidates.

3.2 METHODS

3.2.1 Preparation of amino acid sidechains

Our first step towards obtaining a heterogeneous dielectric profile was to calculate the free energy difference for de-charging an amino acid side chain, from 100% of its original charge to 0% of its original charge, at a given depth within an explicit all-atom membrane model. With the exception of glycine and proline, each amino acid in the standard library in Amber was edited using xleap to remove the common peptide backbone from the amino acid, isolating each amino acid side chain. For each isolated amino acid side chain, a proton was added to the beta carbon. The charge on the added proton was adjusted to match that of other beta carbon protons, and the charge on the beta carbon itself was adjusted to give the side chain as a whole the correct total net charge of the natural side chain at pH = 7[55]. Glycine and proline were unedited due to the lack of a beta carbon for glycine, and due to the cyclization of the side chain and backbone in the case of proline. These two amino acids were used to test the effect of de-charging the peptide backbone.

3.2.2 Z-restraint, de-charging MD simulations in explicit solvent

Each amino acid side chain was placed into a model membrane to perform z-restraint simulations where the center of mass of the side chain was restrained by a square potential at a certain height on the z-axis[38, 39, 43, 56]. Our explicit membrane model consisted of 36 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) phospholipids in each leaflet of the bilayer for a total of 72 phospholipids used in the membrane model as a whole (**FIGURE C.1**). Although the overall length of the DMPC phospholipid is slightly smaller than the more common 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) phospholipids, DMPC membrane models have been shown to have very similar properties when compared to membrane models that use DPPC phospholipids[57-59]. The explicit DMPC membrane model itself was generated using the CHARMM-GUI membrane builder with the Amber Lipid 14 force field as described in the Amber lipid membrane tutorial[56, 60-62]. The membrane was flanked by 1440 TIP3P water molecules on each side of the membrane to fit inside an approximately 50x50x80 Å^3 box[56]. Potassium ions were added to neutralize the system if necessary.

The center of mass of the bilayer membrane was calculated relative to the N31 atom in the phosphatidylcholine headgroups of each phospholipid in the bilayer, and this center of mass was subsequently set as z = 0 Å. An amino acid side chain was then placed at a certain height on the z-axis relative to the center of mass of the bilayer[56]. We chose to place our side chain at 5 Å intervals along the z-axis (at z = 0, 5, 10, 15, 20, and 25 Å) to split the membrane into 6 regions for sampling that roughly corresponded to the midpoint of: the hydrocarbon core region (split up into three sub-regions: 0 to 2.5 Å, 2.5 to 7.5 Å, and 7.5 to 12.5 Å), the hydrocarbon and ester interface region (12.5 to 17.5 Å), the phosphatidylcholine headgroup region (17.5 to 22.5 Å), and the bulk water solvent (22.5 to 27.5 Å) (**FIGURE C.1**). After a side chain was placed into the membrane model at a given height, a 10,000 step minimization was carried out using 5,000 steps of steepest descent followed by 5,000 steps of conjugate gradient. The particle mesh Ewald method (PME) was employed with a real space cutoff of 10.0 Å, but otherwise, the default options were used for all solvated simulations. The lipid and amino acid side chain molecules were held fixed as the system was heated in two steps. First, the system was heated from 0 K to approximately 100 K for 5.0 ps using the Langevin thermostat in the NVT ensemble. The system was then heated from 100 K to 303 K for 100 ps using the Langevin thermostat in the NPT ensemble with anisotropic pressure scaling and a pressure relaxation time of 2.0 ps.

After heating, an equilibration step using semi-isotropic pressure scaling was performed whereby all restraints on the lipids were removed while a center of mass restraint was applied to the amino acid side chain at the appropriate z-value. This restraint was a harmonic potential (V) of the form:

$$V = k(R - r)^2 \tag{2}$$

where the option was set to apply the restraint only on the z-axis, r was set as the target equilibrium z-value, the range R of the potential was set to -99 Å < R < 99 Å, and the initial value of the spring constant k was set to 2.5 kcal/(mol Å^2). In order to improve the convergence of our free energy calculations, the spring constant was gradually ramped up during the equilibration step. Specifically, we carried out an initial short equilibration for 20 ps with k = 2.5 kcal/(mol Å^2), then an additional short 20 ps equilibration with k = 25 kcal/(mol Å^2), and finally, an extensive equilibration was performed for approximately 100 ns with k = 100 kcal/(mol Å^2). Subsequent free energy calculations demonstrated that, if the results had converged properly within the timeframe of the calculation, ramping up the spring constant in this way did not appreciably alter the calculated free energy values (**TABLE C.1**).

After equilibration, we proceeded to uniformly reduce the charges on all of the atoms in the amino acid side chain from 100% to 0% in successive 10% decrements. For each 10% reduction in charge, MD simulations in the NVT ensemble were carried out in 20 ns intervals. From each 20 ns trajectory, the first 10 ns was discarded, and the second 10 ns was retained for use in our subsequent free energy calculations. For the free energy calculations, a total of 1,000 snapshots, taken at equal intervals from each of the retained 10 ns trajectories, were used for analysis where the energy was extracted from each snapshot using SANDER in Amber 18. 20 ns for each interval was deemed sufficient for such calculations; doubling our data collection to 40 ns for each interval, and using 20 ns for the free energy calculations, did not substantially change our results (**TABLE C.2**). In addition, our inspection of the saved trajectories revealed that bulky side chains such as tryptophan and tyrosine were sampling different orientations during the de-charging runs, which is an apparent concern for studying these molecules in the membrane environment. We did encounter difficulties obtaining converged free energy values at times, for example some data points for charged residues (**TABLE C.2**) had clear deviations when comparing our 20 ns and 40 ns data. Nevertheless, our total simulation times of 300 ns (using 20 ns steps) and 500 ns (using 40 ns steps) were among the longest used for calculations of this type, and even in a few worst-case scenarios, the calculated free energy values converged to within a few kcal/mol.

3.2.3 BAR free energy calculations in explicit solvent

To further assist with the convergence of our free energy calculations, we used the Bennet Acceptance Ratio (BAR) calculation to calculate the free energy difference for each 10% de-charging decrement[63]. The BAR calculation can be described as a self-consistency problem where the value of C is chosen to best satisfy the following relations:

$$\Delta G = \ln \frac{\sum_j f(U_i - U_j + C)}{\sum_i f(U_j - U_i + C)} + C, \tag{3}$$

$$\Delta G = C, \tag{4}$$

$$f(x) = \frac{1}{1 + e^x}, \tag{5}$$

where i refers to the initial state and j to the final state for any two consecutive states in the de-charging procedure, and where an equal number of frames was sampled in each state. Here, $U_i - U_j$ in $\sum_j f\left(U_i - U_j + C\right)$ refers to the difference in energy values calculated for each of the 1,000 frames in the ensemble of trajectory j (trajectory j is the MD trajectory of the amino acid side chain at the final, 10% lower, charge) where the energies were calculated using both sets of charges (U_i representing the energy calculated using the higher charge, and U_j the energy calculated using the 10% lower charge). Similarly, $U_j - U_i$ in $\sum_i f\left(U_j - U_i + C\right)$ refers to the difference in energy values calculated for each of the 1,000 frames in the ensemble of trajectory i (trajectory i is the MD trajectory of the amino acid side chain at the initial, higher charge) with the energies calculated using the initial (U_i) and final (U_j) charges as described previously. C is then a fitting parameter which is chosen in order to satisfy eqs. (3) and (4). The total cost of de-charging an amino acid side chain from 100% to 0% can then be obtained by summing over all of the de-charging decrements in the entire data set:

$$\Delta G_{\text{total}} = \sum_k \Delta G_k \qquad (6)$$

with each ΔG_k representing the change in free energy for a step involving a 10% decrease in the charge of an amino acid side chain.

3.2.4 PBSA-BAR free energy calculations in implicit solvent

Using our uniform, single dielectric implicit membrane model, we calculated the total free energy of de-charging an amino acid side chain from 100% to 0% at a given z-value using the PBSA-BAR procedure[64]. Our goal was to perform this calculation for a large number of different dielectric constants to find out which dielectric constant was optimal at a given z-value. Since the number of individual PBSA calculations that needed to be carried out was large, we reduced the number of frames for each individual calculation from 1,000 to about 50 evenly spaced frames from the original trajectory. In the 50 processed trajectory frames, auto-imaging was used to center the trajectory frames at the C110 atom located in a myristoyl group near the center of the lipid bilayer, and all ions, water, and explicit membrane molecules were removed.

For the PBSA calculations[65-78], a 120x120x160 Å^3 box was constructed with the amino acid side chain placed relative to the previously calculated center of mass of the membrane. In addition to using a large box size, we also increased the ratio between the dimension of the finite-difference grid and that of the solute bounding box to 4.0 for this portion of the analysis, following the recommendation in the Amber manual for small ligand systems. The implicit membrane thickness was obtained by calculating the average explicit membrane thickness over the whole trajectory. This was accomplished by calculating the location of the center of mass of the N31 atoms in the phosphatidyl choline headgroups at the top of the membrane, repeating the same calculation for the bottom of the membrane, and taking the difference between the two. The protein and aqueous solvent dielectric constants were set at 1 and 80 respectively while the membrane dielectric constant was changed for each separate run. The aqueous solvent probe was set to 1.4 Å while the membrane solvent probe was set to 2.70 Å[12].

For each individual amino acid side chain in the bulk water solvent at z = 25 Å, the radii of the side chain atoms were uniformly scaled to give the minimum difference in the total free energy for de-charging the side chain in explicit and implicit solvents. This optimized radius scaling factor was then used for all subsequent PBSA calculations at all other z-values. Using SANDER, the energy from PBSA calculations was extracted from the trajectory snapshots, and eqs. (3)-(6) were used to calculate $\Delta G_{\mathrm{total}}$ as described previously.

3.2.5 RMSD calculations

As a quantitative estimate of the error at a given z-value between the explicit and implicit $\Delta G_{\mathrm{total}}$ results, we calculated the root mean square deviation (RMSD) across our data set of amino acid side chains using the standard formula:

$$\mathrm{RMSD} = \sqrt{\frac{\sum_i (\Delta G_{\mathrm{total(implicit)}\,i} - \Delta G_{\mathrm{total(explicit)}\,i})^2}{N}} \qquad (7)$$

where i represents amino acid i and where N represents the total number of amino acids included in the calculation (taken to be either 20 or 15 depending on whether acidic and basic amino acids were included in the calculation). Eq. (7) was used to calculate the RMSD for our free energy calculations at different membrane dielectric constant values. The specific value of the membrane dielectric constant that gave the minimum RMSD at a given z-value was taken to be the best fit membrane dielectric constant and was used to construct our dielectric profile for the implicit membrane model as a whole.

3.2.6 Implementation of the dielectric profile into PBSA

Given the discrete, z-depth dependent membrane dielectric data set, we performed interpolation using two different schemes to get a smoothed profile. The first scheme was a piecewise cubic Hermite interpolating polynomial (PCHIP) fitting, which passed through all data points[79]. The second scheme was a second-order spline fitting, which generated an overall smoothed fitting but did not necessarily pass through the data points[80]. In addition, to maintain a zero-slope boundary condition in the latter method, we replaced the second-order spline within the fitting range of z = 2.5-5.0 Å with a cubic spline.

The advantage of using a PCHIP fitting over a cubic spline fitting is that the fitted function is smooth enough, but less oscillatory, to capture the monotonic dielectric change inside the membrane. The PCHIP algorithm preserves monotonicity in the interpolation data and does not overshoot if the data is not smooth[79]. However, while the first derivatives are guaranteed to be continuous, the second derivatives may jump at knots. To determine the derivatives f_k' at the points z_k, let $h_k = z_{k+1} - z_k$ and $d_k = (y_{k+1} - y_k)/h_k$ be the slopes at internal points z_k where y_k refers to the dielectric constant. If the signs of d_k and d_{k-1} are different, or either of them equals zero, then $f_k' = 0$. Otherwise, f_k' is given by the weighted harmonic mean:

$$\frac{w_1+w_2}{f_k'} = \frac{w_1}{d_{k-1}} + \frac{w_2}{d_k} \qquad (8)$$

where $w_1 = 2h_k + h_{k-1}$ and $w_2 = h_k + 2h_{k-1}$. The end slopes are generally set using a one-sided scheme where in our case both of the end slopes equal zero[81].

The spline fitting was implemented to get an overall smoothed fitting in a least squares manner. The fitted spline ($spl(z)$) uses a representation in the B-spline basis. In our case, six B-spline basis functions were constructed for the fitting range z = 2.5-20.0 Å. The smoothing condition $s = 0.01$ was used to choose the number of knots that are determined by satisfying:

$$\sum_i w_i \times \left(y_i - spl(z_i)\right)^2 \leq s \tag{9}$$

where w_i is the weight for spline fitting, which in our case was equal for each i, and y_i is the dielectric constant from our test set. The final spline, $spl(z)$, was constructed using a linear combination of the six B-spline basis functions. The full data range used to fit PCHIP was z = 0-25.0 Å while for spline it was z = 2.5-20.0 Å. The fitted piecewise functions were coded into the PBSA program in the Amber 18 package. Note that in practice the membrane thickness can vary depending on the membrane model used in the simulation, and so, to match the overall membrane thickness to the user specified thickness value in the software, we allow the inner membrane core region (originally sampled at a z-depth of 0 Å to 2.5 Å) to be stretchable with a dielectric constant of 1.0. The two fittings were implemented in the programs SciPy 1.1.0 and SymPy 1.2[82, 83].

3.2.7 The natural abundance of each amino acid in different membrane regions

While analyzing our results, we also calculated the natural abundance of each amino acid found in membrane proteins within our specific z-value ranges using a library of structures downloaded from MemProtMD[4, 58]. The purpose for doing this was to examine the abundance of different types of amino acids in each range in order to see the effect of leaving out acidic and basic amino acids from our dielectric profile in the hydrophobic core region. We used a total of 482 MemProtMD structures, representing the bulk of unique

MemProtMD structures that were available at the time of our analysis. Redundant structures that were available on MemProtMD, and a few abnormal structures where the protein had clearly not been placed properly within the membrane, were not included in our analysis.

Each MemProtMD structure was centered at z = 0 Å by subtracting the average z-coordinate of all membrane atoms from the original z-coordinate of each individual atom in the structure file. Individual amino acid residues from the membrane proteins were sorted into 5 bins depending on their location along the z-axis (corresponding to bins of 0 to 2.5 Å, 2.5 to 7.5 Å, 7.5 to 12.5 Å, 12.5 to 17.5 Å, and 17.5 to 22.5 Å), and we counted the number of amino acid residues that appeared in each bin. The probability that a specific amino acid appeared within a given z-value range was taken as the number of times that it appeared divided by the sum of all 20 amino acids that appeared within that range, i.e.:

$$p_i = \frac{m_i}{\sum_i m_i} \tag{10}$$

where p_i represents the probability of finding amino acid i, m_i is the total number of times that amino acid i appeared within the specified z-value range, and the sum is taken over all 20 amino acids.

It should be noted at this point that additional complications to this calculation, such as distinguishing between external amino acids on the surface of the protein and internal amino acids on the protein interior, and separating alpha helical transmembrane proteins from beta barrel transmembrane proteins, were not explicitly taken into account in our analysis[84]; our results were averaged over both external and internal amino acids and across both of the main membrane protein types. However, despite this more simplistic approach, we note that our final results (**FIGURE C.2**) still accurately reflect the observed trends in the distribution of membrane protein amino acids that were reported on previously in the literature[58, 84].

3.2.8 MMPBSA calculations for the P2Y$_{12}$R and M2R/M3R test systems

To test the effect of the heterogeneous dielectric membrane model on MMPBSA calculations[85-90], we carried out MMPBSA calculations for the human purinergic platelet receptor (P2Y$_{12}$R) and two muscarinic acetylcholine receptors (M2R and M3R) bound to various antagonist ligands using the MMPBSA.py module in Amber 18. Models of P2Y$_{12}$R bound to the antagonist ligands: AZD-1283 (AZD), PSB-0739 (PSB), and Ticagrelor (TIQ) were derived from a previously published crystal structure (PDB ID: 4NTJ)[53]. Both the model preparation and the molecular dynamics simulation protocol that we employed for this test system were described in detail in our previous publication[11].

The M2R and M3R protein models with the antagonist ligands 3-quinuclidinyl-benzilate (QNB) and tiotropium were derived from previously available crystal structures (PDB IDs: 4DAJ, 5ZHP, and 3UON). For the M2R model (derived from 3UON), the cytoplasmic domain was cleaved in order to place the center of mass of the protein in close proximity to the center of mass of the model membrane system to facilitate the MMPBSA calculation. The compound 6B antagonist that was used to demonstrate M2R/M3R selectivity by Liu et al. was constructed by modifying the QNB ligand present in the 3UON crystal structure[54]. The ligands were transplanted from one crystal structure to another by performing an alignment of the M2R and M3R crystal structures using UCSF Chimera[91] before making the substitution of one ligand for another. The parameterization of the antagonist ligands, membrane model preparation (for a 64x64 DPPC bilayer), and single trajectory MD simulations were performed in the usual way based on the Amber lipid membrane tutorial and other standard MD and MMPBSA protocols as described previously[11, 61]. For the P2Y$_{12}$R system, we used a 20 ns equilibration step which was followed by a 10 ns production step, and 1,000 evenly spaced frames from the production step were used in our MMPBSA calculations. Due to the use of a docked complex in our M2R and M3R analysis, we increased the equilibration step to 90 ns before carrying out a 10 ns production step. This was followed by the extraction of 1,000 evenly spaced frames from the production step for our MMPBSA analysis.

For the MMPBSA calculations for both systems, we tested three implicit membrane models: the uniform, single dielectric membrane model with the membrane dielectric constant set to 4, the heterogeneous dielectric membrane model using the PCHIP fitting, and the heterogeneous dielectric membrane model using the spline fitting. The average membrane thickness was calculated relative to the N31 atoms in the explicit phospholipid membrane as described above. Other PB settings were as follows: the protein dielectric constant was set to 2, the ionic strength was set to 150 mM, the ratio between the dimension of the finite-difference grid and that of the solute bounding box was set to 1.5, the geometric multigrid solver was used to solve the linear systems[10], electrostatic focusing was switched off, periodic boundary conditions were used, the atom-based cutoff distance to remove short ranged finite difference interactions, and to add pairwise charge-based interactions, was set to 7.0 Å, the atom-based cutoff distance for van der Waals interactions was set to 99.0 Å, the total electrostatic energies and forces were computed using the particle-particle particle-mesh (P3M) procedure[92], Bondi radii from the parameter topology file were used, and the pore searching algorithm was turned off. All other PB parameters were set to their default values in MMPBSA.py for Amber 18. Our binding affinity modeling was carried out without the optional normal mode entropy analysis as this analysis usually does not contribute favorably to the overall agreement with experiment due to the approximation used[89, 93-94]. Our results were compared to experimental binding free energies published in the literature[53, 54].

3.3 RESULTS AND DISCUSSION

3.3.1 BAR results for de-charging amino acid side chains in an explicit membrane

Before turning to our implicit membrane models, we first examine the BAR results obtained from an analysis of our explicit all-atom MD simulations. As calculated using the BAR method, the total free energy change, ΔG_{total}, for de-charging a given amino acid side chain from 100% to 0% as a function of the z-value (with z = 0 Å corresponding to a plane at the center of the membrane bilayer) is given in **TABLES C3** and **C4**. From the data, we see that the ΔG_{total} values were at a minimum at the center of the membrane (z = 0 Å) and reached a maximum in the polar headgroup region (z = 20 Å) or in the bulk aqueous solvent (z = 25 Å) for all amino acids regardless of their type.

Note that the ΔG values given in **TABLES C3** and **C4** include both Coulomb and reaction field contributions. This makes comparisons of reaction field energies between various amino acid side chains difficult. This is not important for our optimization of the membrane model, but focusing on reaction field energies helps us to compare our results to the literature values to assess the accuracy of our approach. To focus on the reaction field energies, we used the data in **TABLES C3** and **C4** to calculate the relative ΔG_{total} values between placing each amino acid side chain at the center of the membrane, which approximates a low dielectric vacuum environment (see **TABLE 3.1**) with a relatively large accessible free volume[95], and at any other z-value using:

$$\Delta\Delta G_{\text{total}\,(\text{z}_i)} = \Delta G_{\text{total}\,(\text{z}_i)} - \Delta G_{\text{total}\,(\text{z}\,=\,0\,\text{Å})} \tag{11}$$

where $\Delta G_{\text{total}\,(\text{z}\,=\,0\,\text{Å})}$ represents the ΔG_{total} value for a given amino acid side chain at z = 0 Å in **TABLE C.3** or **C.4**, and $\Delta G_{\text{total}(\text{z}_i)}$ represents the ΔG_{total} value for that side chain at z = *i* in **TABLE C.3** or **C.4**, where *i* can take on values of z = 0, 5, 10, 15, 20, or 25 Å. $\Delta\Delta G_{\text{total}\,(\text{z}_i)}$ is the relevant quantity to use to analyze reaction field energies.

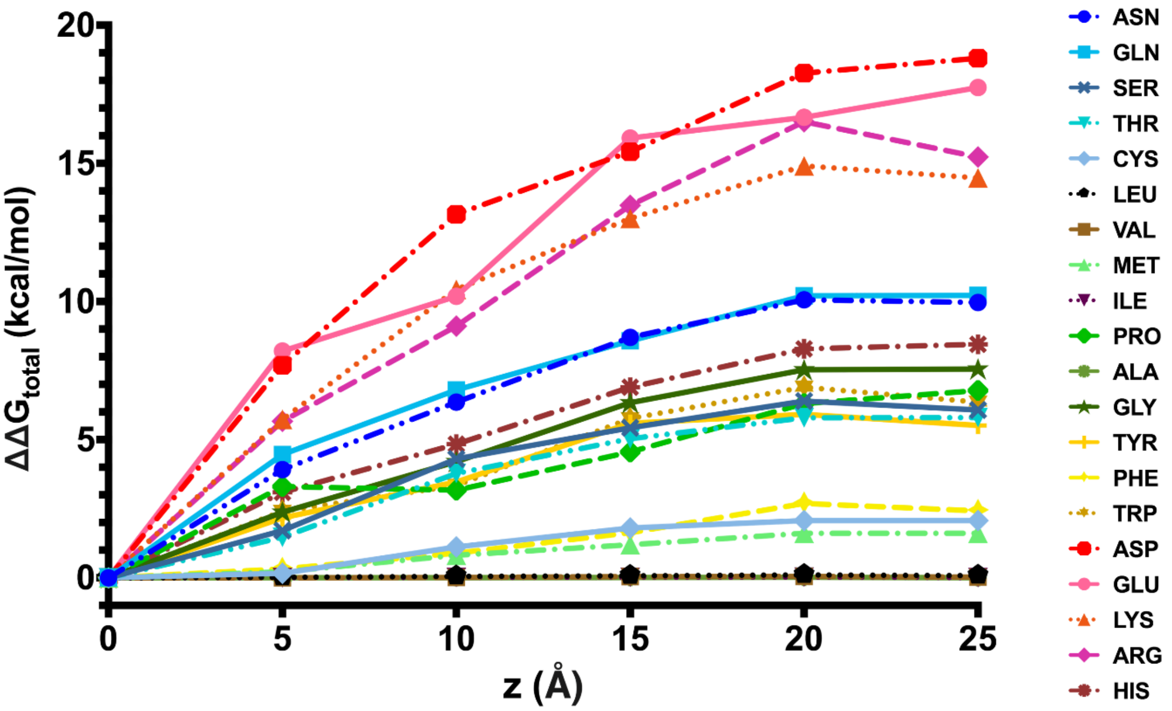


FIGURE 3.4. The difference in ΔG_{total} values for de-charging an amino acid side chain at height z. $\Delta\Delta G_{\text{total}}$ represents the net difference in the de-charging free energy between placing an amino acid side chain at height z compared to placing it at the center of the membrane, at a height of z = 0 Å. The trend for each amino acid sidechain is indicated using the standard three letter amino acid code. The trends for glycine and proline represent both the sidechain and the amino acid backbone as the backbone was not removed in these two cases while the backbone was removed for all other amino acids.

The numerical values obtained by performing this calculation are given in **TABLES C5** and **C6**, and a plot of $\Delta\Delta G_{\text{total}}$ vs. z for each amino acid is given in **FIGURE 3.4**. From **FIGURE 3.4**, we can see that the increase in $\Delta\Delta G_{\text{total}}$ at higher z-values tends to depend on the net overall charge of the amino acid type, with a smaller dependence on the structural differences between individual amino acid side chains. For example, the top four amino acids in **FIGURE 3.4** correspond to the four fully charged amino acids. Other amino acids cluster similarly by type. We note that glycine and proline were both used to test the effect of the amino acid backbone as it was not removed for either residue, and the polarity of the backbone likely earned both residues inclusion into a more polar group.

The overall trend in **FIGURE 3.4** seems clear: the more polar the amino acid side chain, the more expensive the free energy penalty to de-charge that side chain in a more polar environment at higher values of z. This result seems reasonable as de-charging a charged amino acid side chain in water (z = 25 Å) from 100% to 0% effectively means removing favorable interactions between the charged side chain and water while replacing them with less favorable interactions between the de-charged, now non-polar, side chain and water. On the other hand, de-charging an inherently non-polar amino acid side chain in water does not incur the same penalty as the non-polar side chain already has nearly zero net charge to begin with. Placing charged side chains deeper into the membrane environment progressively lessens the penalty for de-charging the side chain in all cases. In this case, the fully de-charged side chain ends up in a more favorable, similarly non-polar membrane environment regardless of its initial charge characteristics.

Our results are in good agreement with amino acid insertion free energy profiles reported in the literature. Bemporad et al. in 2004[43] reported a maximum difference in free energy of insertion of 0-10 kcal/mol for non-polar molecules and 15-25 kcal/mol for polar molecules using the CHARMM force field. Tieleman et al. in 2009[55] reported a maximum difference of approximately 2-5 kcal/mol for non-polar residues, 3-7 kcal/mol for polar and aromatic residues, and 12-20 kcal/mol for charged residues. A more recent, 2017 comparison[96] of free energy insertion profiles using modern versions of the GROMOS, CHARMM, and Amber force fields gave maximum free energy difference estimates of 0-10 kcal/mol for non-polar, aromatic, and polar amino acids and 10-15 kcal/mol for charged amino acids using Amber 14. All of these results are consistent with our maximum $\Delta\Delta G$ results at $z = 25$ Å (see **FIGURE 3.4** and **TABLES C5** and **C6**).

The accuracy of current free energy calculation methods is subject to improvement. For the calibration of our new heterogeneous dielectric membrane model, we employ standard free energy calculation techniques that facilitate comparisons with our past work, but we note that improvements that address issues with these calculations, such as electrostatic artifacts that arise with the use of periodic boundary conditions (PBC) with the PME method are actively being addressed in recent years[97-99]. Here we have tried to minimize the impact of such artifacts by using large box sizes and increasing the size of the fill ratio used in our PBSA calculations. In addition, note that our free energy difference calculations only consider the effect of de-charging the solute; the total free energy change should also take into account cavity formation and the van der Waals interactions. The calibration of these non-polar terms is a subject that will be addressed in a future study.

3.3.2 PBSA results for de-charging amino acid side chains in an implicit membrane

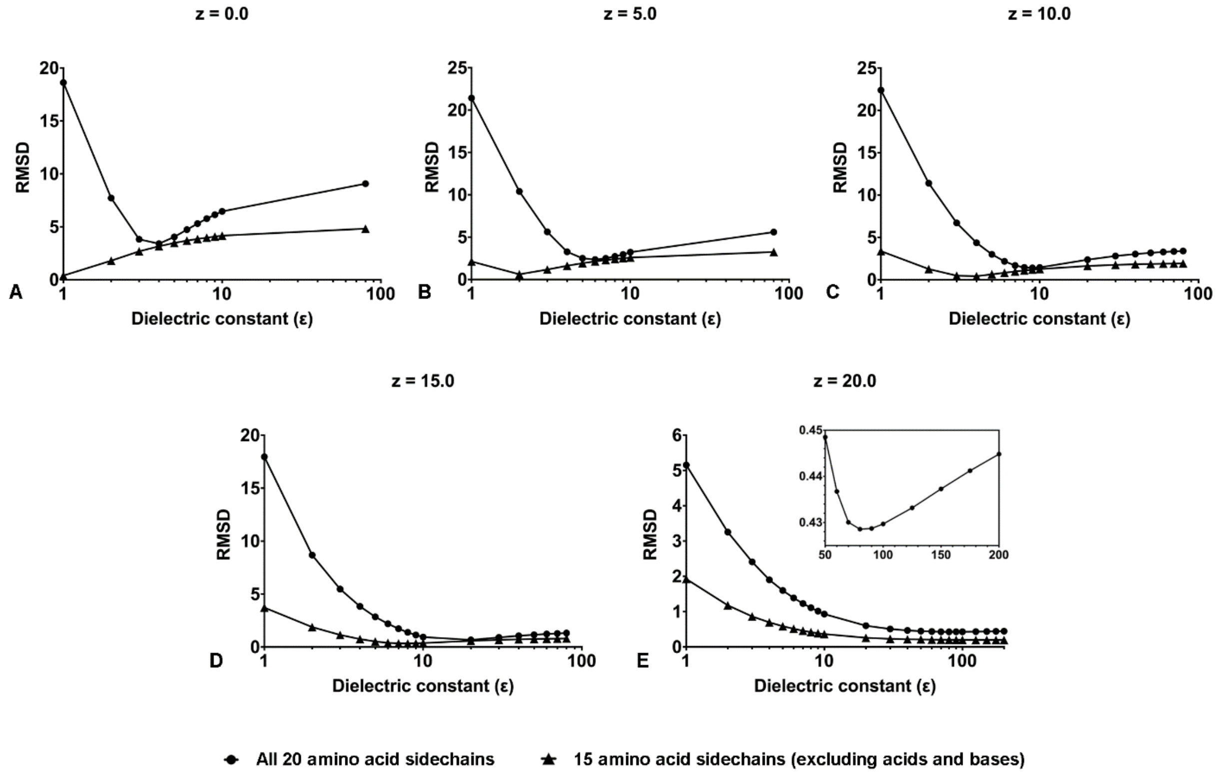


FIGURE 3.5. RMSD between ΔG_{total} values calculated using PBSA-BAR and Explicit BAR for various dielectric constants at different z-values. RMSD values were calculated using eq. (7) over a set of amino acids using the following two schemes: 1) all 20 amino acid side chains (circles), and 2) 15 amino acid side chains, where all acidic and basic amino acids were excluded from the calculation (triangles). All RMSD values were reported in units of kcal/mol. All z-values (plots A-E) were given in units of Å. The x-axis was plotted logarithmically in order to fit a wide range of dielectric constant values on each plot.

After obtaining our MD trajectories from de-charging each amino acid side chain from 100% to 0% at a given z-value, we also calculated ΔG_{total} at that z-value for each amino acid side chain using different values of the membrane dielectric constant while employing our uniform, single dielectric implicit membrane model and the PBSA module. Following this, eq. (7) was used to calculate the RMSD between the implicit PBSA-BAR results and the explicit BAR results at each z-value for each dielectric constant that we used. The results for the RMSD calculation are plotted in **FIGURE 3.5**. From **FIGURE 3.5**, the value of the dielectric constant that led to a minimum in the RMSD plot at a given z-value was taken as the optimum dielectric to use at that z-value. The optimized dielectric values obtained from these plots at the given z-values (**FIGURE 3.5A-3.5E**) are listed in **TABLE 3.1**.

TABLE 3.1. Optimum values of the dielectric constant within the implicit membrane

z (Å)	Method 1	Method 2	Method 3
0	4	1	1
5	6	2	2
10	9	4	4
15	20	8	20
20	80	>200	80

Initially, two dielectric profiles were constructed based on the method for calculating the RMSD using eq. (7): Method 1) where all 20 amino acid side chains were used to find the optimum dielectric constant at each z-value, and Method 2) where 15 amino acid side chains were used to find the optimum dielectric constant, with all acidic and basic amino acids excluded from this calculation at all depths. For the final profile (Method 3), we used the optimized dielectric values from Method 1 for z = 15 and 20 Å, and from Method 2 for z = 0, 5, and 10 Å.

Note that a disparity existed between the optimized dielectric values for the full set of 20 amino acids in the first method (Method 1) versus the values obtained when acidic and basic amino acids were excluded from the calculation (Method 2). Dielectric values in Method 1 were generally higher than Method 2 inside the membrane until the trend reversed in the headgroup region of the membrane. The higher value of the dielectric constants in the membrane interior when charged amino acids were included (Method 1) was likely caused by a larger penetration of water, owing to the well-studied formation of water defects when charged groups are restrained within the membrane (**FIGURE C.3**)[57, 100-102]. A fair amount of water naturally penetrates into the headgroup region of the membrane to solvate the polar and charged portions of the phospholipid that appear there. Here we see that the optimum dielectric in Method 1 for the full set of amino acids converges to a dielectric constant of 80, to match the bulk water solvent as had been assumed in other models of the headgroup region in the past[48].

The optimum dielectric constant for non-polar amino acids was consistently lower within the membrane until it suddenly rose sharply to a value that exceeded 200 in the headgroup region. Such a sharp increase in the optimum dielectric constant to 200 or above in the headgroup region had been previously reported in the literature[39, 40], notably as high as 300-1,000 when a Lennard-Jones test particle with a small net charge was used as a probe molecule[39]. In our data set, amino acids such as alanine and isoleucine, with very small ΔG_{total} values, did not reach a minimum below a dielectric constant of 200 in the headgroup region. However, the choice of a dielectric constant in this region appears to be less critical than in the hydrophobic core region as raising the dielectric constant from 100 to 200 and above had a minimal effect on the accuracy of our calculated RMSD (**FIGURE 3.5E**). This reflects a slower change in the electrostatic energy at higher dielectric values due to the $1/\varepsilon$ dependence on the calculated energy as pointed out previously by Tanizaki and Feig[48]. Therefore, keeping the dielectric constant at 80 in the headgroup region still seems suitable, as doing so gives a greater benefit to lowering the RMSD when charged amino acids are included versus the penalty incurred for using a less-than-optimal dielectric constant for the non-polar side chains in this region.

TABLE 3.2. RMSD results for PBSA-BAR calculations using the heterogeneous dielectric membrane model

z (Å)	RMSD all 20 amino acids		RMSD 15 AAs (no acidic or basic AAs)	
	memopt = 2	memopt = 3	memopt = 2	memopt = 3
0	11.256	11.978	0.579	0.431
5	6.196	6.112	0.796	0.785
10	1.204	1.237	0.732	0.778
15	0.667	0.580	0.631	0.590
20	0.493	0.555	0.251	0.263
25	0.049	0.056	0.015	0.013

The results of PBSA-BAR calculations using the heterogeneous dielectric implicit membrane model with the PCHIP fitting (memopt = 2) and the spline fitting (memopt = 3). RMSD values for this data set were calculated using eq. (7). The first and second columns give the RMSD values for calculations that included all 20 amino acid side chains while the third and fourth columns omit all of the acidic and basic amino acid side chains in the RMSD calculations. All RMSD values are reported in units of kcal/mol.

To assist us in determining which of the optimal dielectric constants to use for the remaining regions, we calculated the natural abundance of amino acids in each of our sampling regions of the membrane using the majority of the membrane protein structures available in the MemProtMD database. **FIGURE C.2** shows that the natural abundance of acidic and basic amino acids in the hydrophobic core region (z = 0-12.5 Å) was low, while it rose noticeably in the polar ester and headgroup regions (z = 12.5-22.5 Å). After considering the large error in the RMSD when charged acidic and basic amino acids were included (see **TABLE 3.2**), we chose to exclude charged amino acids in the hydrophobic core region at z = 0, 5, and 10 Å when determining the optimum dielectric constant. In the polar headgroup region, acidic and basic amino acids were much more abundant, and so we included them in our determination of the optimal dielectric constant at z = 15 and 20 Å. The final dielectric profile that we used is listed in **TABLE 3.1** as Method 3.

3.3.3 Implementation of the heterogeneous dielectric membrane into Amber

Given the discrete z-depth dependent membrane dielectric data set [(0, 1.0), (2.5, 1.0), (5.0, 0.5), (10.0, 0.25), (15.0, 0.05), (20.0, 0.0125), (25.0, 0.0125)], where (z-value, 1/dielectric) forms a coordinate pair with z = 0 Å defined as the membrane center, we performed an interpolation using two different schemes to get a smoothed dielectric profile to implement into Amber 18. The fitted curves are shown in **FIGURE 3.6** while the equations for the fitted profiles are provided in **FIGURE C.4.** As can be seen in **FIGURE 3.6**, the PCHIP approach intends to be true to the actual data, but leads to a busy profile. The spline approach intends to address the uncertainty in the sampled dielectric constants, which is due to the sampling uncertainty in the free energy simulations, and has a smoother profile.

After the fitted piecewise functions were coded into the PBSA program in Amber 18, we repeated our PBSA-BAR calculations for our set of amino acid side chains using the heterogeneous dielectric membrane with the PCHIP fitting or the spline fitting as our membrane model in place of the uniform dielectric membrane model. The results are presented in **TABLE 3.2**. In **TABLE 3.2**, we see that the global RMSD values calculated using 15 out of 20 amino acids (excluding acidic and basic amino acids) were less than 1 kcal/mol at all z-values within the membrane.

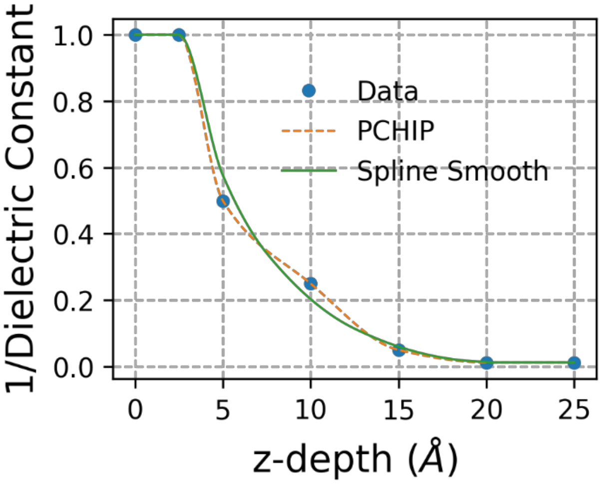


FIGURE 3.6. PCHIP and Spline fittings of a z-depth dependent membrane dielectric profile. Fittings for the dielectric constant are given in terms of the inverse dielectric constant ($1/\varepsilon$) while z-values are given with z = 0 Å corresponding to the center of the membrane and with z = 25 Å corresponding to the bulk aqueous solvent. The PCHIP fitting closely follows the trend established in the data whereas the spline fitting has been smoothed.

When acidic and basic amino acids were included in the calculation, the RMSD values were less than 1 kcal/mol from z = 15-25 Å while it was approximately 1 kcal/mol at z = 10 Å. Below 10 Å, the RMSD rose sharply in the membrane core when acidic and basic amino acids were included in the calculation. Therefore, as mentioned previously, we chose to exclude charged amino acids from the determination of the optimum dielectric constant in the membrane core region. Based on the natural distribution of such amino acids in the hydrophobic core region (**FIGURE C.2**), we expect this rise in the RMSD in the core region for charged amino acids to not be a major issue for the majority of membrane proteins. Nevertheless, this limitation must be kept in mind as our model would not be expected to perform well if the system of interest explicitly involved buried, charged groups in the hydrophobic core. For such cases, more advanced methods that account for membrane deformation and water penetration into the membrane core need to be implemented[103-105]. In the meantime, the effect of increased water penetration into the membrane for buried charged residues can be crudely approximated using our current model by reducing the value of the membrane thickness in the MMPBSA calculation below that of the true average thickness of the membrane. This would shift the high dielectric region of our implicit membrane model further down into the core of the membrane to account for the deeper penetration of water into the membrane for systems that contain buried charged residues.

3.3.4 MMPBSA results for the P2Y$_{12}$R and M2R/M3R test systems

To assess the performance of our new heterogeneous membrane model in comparison to our old uniform membrane model, we carried out MD simulations followed by MMPBSA calculations to calculate the binding free energies for two relevant drug design test systems. The numerical results of our MMPBSA calculations are provided in **TABLES 3.3** and **3.4**, correlation plots that compare our results with experimental binding affinities are provided in **FIGURES 3.7** and **3.8**, and convergence plots for our MMPBSA calculations are available in the Supporting Information (**FIGURES C.5-C.11**).

TABLE 3.3. Results of MMPBSA calculations for antagonist ligands bound to $P2Y_{12}R$

Structure	ΔG (MMPBSA)			ΔG (Exp.)
	memopt = 1	memopt = 2	memopt = 3	
$P2Y_{12}R$ WT (AZD)	-21.1	-16.0	-15.8	-10.16
$P2Y_{12}R$ D294N (AZD)	-17.6	-15.6	-15.7	-10.10
$P2Y_{12}R$ WT (PSB)	-10.5	-11.0	-10.6	-9.35
$P2Y_{12}R$ D294N (PSB)	-6.8	-8.2	-7.6	-9.34
$P2Y_{12}R$ WT (TIQ)	-14.8	-8.3	-8.4	-8.96
$P2Y_{12}R$ D294N (TIQ)	-16.3	-11.6	-11.5	-9.25

MMPBSA calculations were carried out using SANDER in MMPBSA.py for the $P2Y_{12}R$ test system using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). For the uniform, single dielectric membrane model, the membrane dielectric constant was set to 4 while in all three models the protein dielectric constant was set to 2. The experimental binding free energies were obtained from the literature[53].WT refers to the wildtype structure of $P2Y_{12}R$ while D294N refers to the mutant structure of $P2Y_{12}R$. The antagonist ligand bound to $P2Y_{12}R$ is listed in parentheses. All binding free energies are reported in units of kcal/mol.

For the human purinergic platelet receptor ($P2Y_{12}R$) bound to various antagonist ligands, **TABLE 3.3** and **FIGURE 3.7** reveal that the heterogeneous dielectric membrane models, memopt = 2 and memopt = 3, produced a tighter grouping of data points and a higher correlation with experiment, with R = 0.92 and R = 0.91 respectively, when compared to the uniform, single dielectric membrane model, memopt = 1, which had a correlation of R = 0.59.

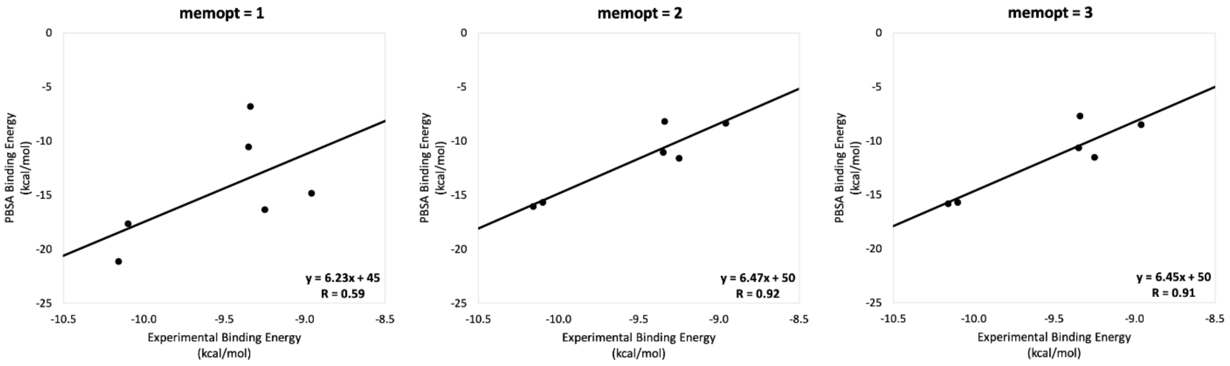


FIGURE 3.7. Binding free energy correlation plots for the P2Y$_{12}$R test system. PBSA binding free energy correlation plots are provided for the antagonist-bound P2Y$_{12}$R system using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). For the uniform, single dielectric membrane model, the membrane dielectric constant was set to 4 while in all three models the protein dielectric constant was set to 2. The experimental binding free energies were obtained from the literature[53].

We note that the deviation of the slopes from unity and the large values of the y-intercepts represent the large systematic errors inherent in MMPBSA free energy calculations when they are compared to experimental binding affinity data. We believe they should be significantly reduced after entropy can be accurately estimated[90, 93]. The lack of electronic polarization can also be similarly addressed with residue-specific variable dielectric protein models[106], and both protein models can be combined together when using the heterogeneous dielectric membrane model in the future.

For the M2R and M3R receptors bound to a variety of antagonist ligands, **TABLE 3.4** and **FIGURE 3.8** reveal similar trends in that five out of the six data points using the new model cluster closer together leading to a higher correlation with experiment of R = 0.72 and R = 0.73, using memopt = 2 and 3 respectively, when compared to the uniform, single dielectric membrane model, memopt = 1, which had a correlation of R = 0.49. One data point for the M2R receptor bound to compound 6B was excluded from **FIGURE 3.8** as its calculated MMPBSA binding free energy was a far outlier in comparison to the rest of the trend (see **TABLE 3.4**).

TABLE 3.4. Results of MMPBSA calculations for antagonist ligands bound to M2R and M3R

Structure	**ΔG (MMPBSA)**			**ΔG (Exp.)**
	memopt = 1	**memopt = 2**	**memopt = 3**	
M3R (QNB)	-28.9	-35.3	-35.7	-13.6
M3R (TIO)	-31.6	-32.3	-32.4	-13.3
M3R (C6B)	-33.4	-36.0	-36.3	-13.3
M2R (QNB)	-46.7	-43.9	-43.9	-13.7
M2R (TIO)	-36.8	-36.8	-36.9	-13.5
***M2R (C6B)**	-38.4	-36.2	-36.5	-10.5

MMPBSA calculations were carried out using SANDER in MMPBSA.py for the M2R and M3R test system using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). For the uniform, single dielectric membrane model, the membrane dielectric constant was set to 4 while in all three models the protein dielectric constant was set to 2. The experimental binding free energies were obtained from the literature[54]. The antagonist ligand bound to M2R or M3R is listed in parentheses. All binding free energies are reported in units of kcal/mol. * was excluded from **FIGURE 3.8**.

Our MMPBSA calculation predicted that compound 6B should bind slightly stronger to M2R than M3R whereas the experimental result has compound 6B binding to M3R with a K_i of 0.2 nM and to M2R with a much weaker binding affinity, indicated by a K_i of 21 nM. However, we note that experimentally the drug compound 1B (which was a designed precursor to compound 6B) had a binding affinity for M2R that was greater than it was for M3R with Ki values of 0.19 nM and 0.63 nM respectively, closer to what we observed in our MMPBSA calcualtions[54].

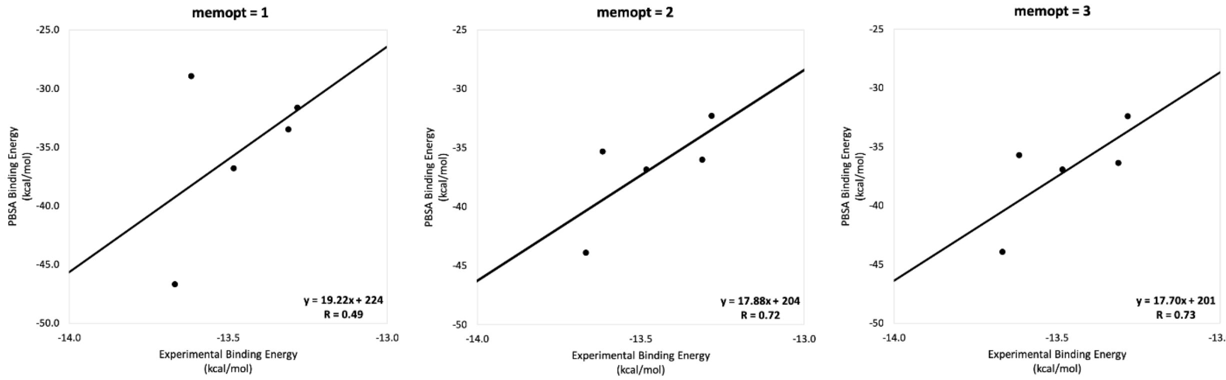


FIGURE 3.8. Binding free energy correlation plots for the M2R/M3R test system. PBSA binding free energy correlation plots are provided for the antagonist-bound

M2R/M3R test system using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). For the uniform, single dielectric membrane model, the membrane dielectric constant was set to 4 while in all three models the protein dielectric constant was set to 2. Note that the trend shown only contains five out of the six data points for the M2R/M3R test system as the M2R structure bound to C6B was a far outlier to the rest of the trend. The experimental binding free energies were obtained from the literature[54].

Li et al. interpreted the difference in the foregoing experimental binding affinity results for compound 1B and compound 6B as having to do with the flexibility of an aromatic ring in compound 1B that could adopt a conformation that interacted via pi stacking with a phenylalanine in the active site of M2R. Compound 6B was designed to be conformationally restricted in order to prevent this interaction in M2R, and this was a factor that led to a large decrease in binding affinity for M2R when comparing the two compounds[54]. The focus on a specific phenylalanine residue was based on the observation that the active site of M3R is structurally almost identical to M2R except for the replacement of this phenylalanine residue with a leucine residue in M3R.

We did not see such an effect lowering the binding affinity of compound 6B for M2R in comparison to M3R in our MMPBSA binding free energy calculations. One reason may have to do with one of the underlying assumptions when using the single trajectory MD method for our MMPBSA calculations, which is that the sampled states of the bound conformations of both the protein and ligand in the complex are assumed to be similar to the conformations of the apo protein and ligand in solution[107, 108]. It seems reasonable to suggest that this assumption would probably not hold up well in a case where the ligand suffers from a conformational restriction that apparently only manifests itself in the M2R, but not in the, structurally very similar, M3R active site.

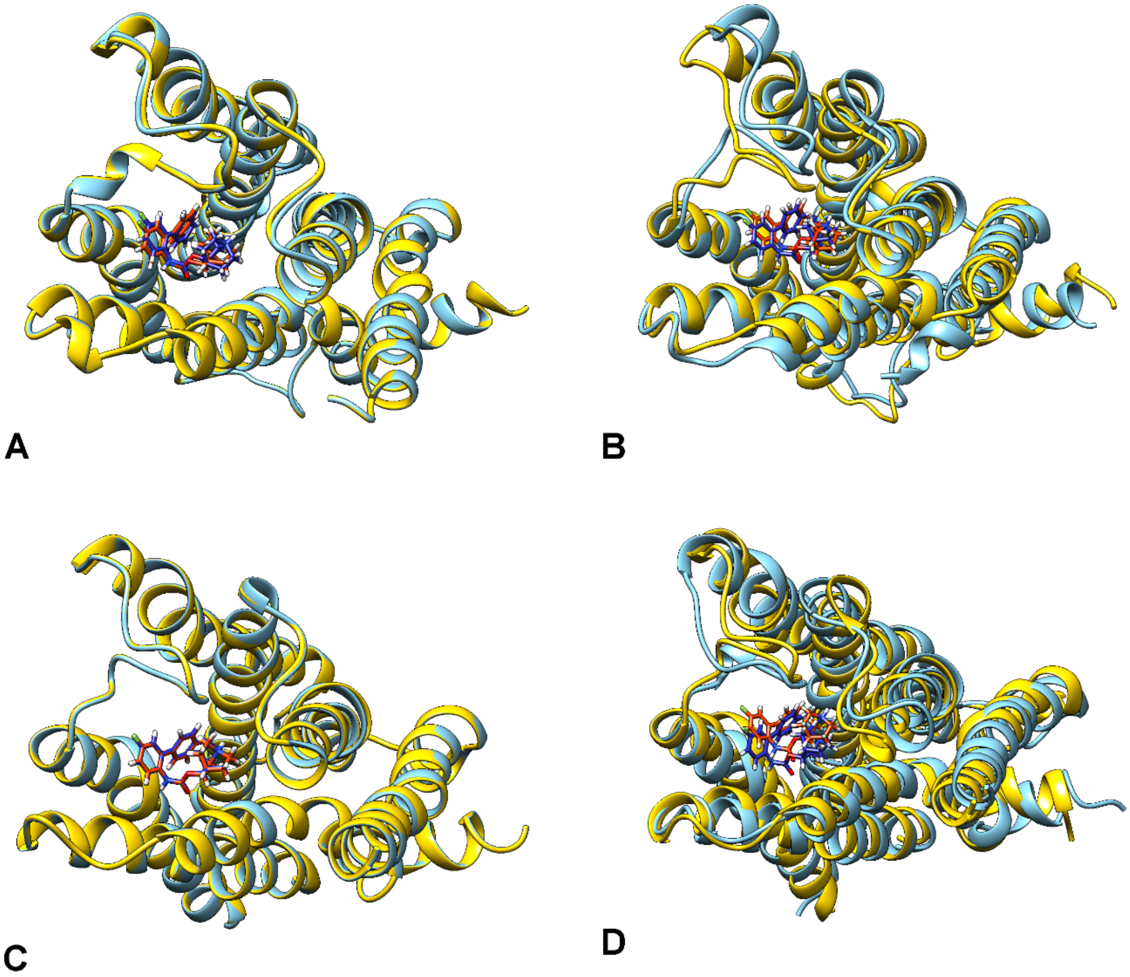


FIGURE 9. Snapshots of compound 6B bound to the M2R and M3R receptors from our MD simulations. Aligned, superimposed snapshots of compound 6B bound to the M2R and M3R receptors were taken just before the heating step (image A for M2R and image C for M3R) and the very last frame of the production run (image B for M2R and image D for

M3R) from the explicit MD trajectory. When the initial binding pose in the starting structure was based on the pose of QNB from the 3UON crystal structure, the receptors were shown in yellow and compound 6B was shown in orange. When the initial binding pose of compound 6B was determined by minimizing compound 6B in a vacuum environment before redocking it to the protein surface, the receptor was shown in cyan while compound 6B was given in dark blue. The structures were aligned with each other for this comparison using UCSF Chimera[91].

Perhaps another limitation of our study is that we started our MD simulations using the pre-fitted active site conformations for both the protein and ligand geometries that were observed in the original crystal structures. We did not have the true binding pose of compound 6B available from our crystal structures, and so we constructed it using the known binding pose of QNB from the crystal structure 3UON. This is a reasonable starting point if the crystal structure represents the "true" binding pose of the ligand in the active site, but it also introduces a bias in that the system will have a tendency to fluctuate about the conformation present in the original crystal structure while perhaps being kinetically excluded from sampling different conformations that might be relevant in the true binding interaction.

In an attempt to test the effect of introducing a potential conformational bias into our analysis, we erased the "QNB memory" of the active site binding pose for compound 6B by minimizing the conformation of compound 6B in a protein-free, vacuum environment. We then redocked compound 6B into the active site of M2R and M3R using Autodock Vina/SMINA[109]. In **FIGURE 3.9**, the initial binding pose is seen to be qualitatively similar using either method when viewed after minimization and prior to the start of our MD simulations. However, by the end of our production run, it can be seen that noticeable differences in the final binding pose are apparent when comparing the two methods. In **FIGURE 3.9**, noticeable changes in the receptor structures are also apparent as the loops above the active site are shifted depending on which method was used to place compound 6B into the active site at the start of the simulation.

The MMPBSA results after redocking compound 6B, given in **TABLE C.7**, also revealed a noticeable difference in binding affinity compared to our earlier approach, but the results overall were still somewhat inconclusive. On one hand, the MMPBSA binding free energy for the redocked compound 6B bound to M2R increased by about 39% for memopt = 1 and 21% for memopt = 2 and 3 over the values given when using QNB as a reference for the initial pose, clearly showing that using even a slightly different initial binding pose at the start of the MD simulation can have an effect on our final calculated result. In addition, the increase in binding free energy for compound 6B bound to M2R brought the calculated MMPBSA value into much closer agreement with the rest of the trend established in **FIGURE 3.8**. However, the MMPBSA binding free energy for the redocked compound 6B bound to M3R also increased similarly, by 33% for memopt = 1 and 19% for memopt = 2 and 3. We note that this change was slightly more pronounced for M2R than it was for M3R. This difference restored the proper order of binding between the two ligands when using the new membrane model as compound 6B was now shown to be binding at a higher affinity to M3R in comparison to M2R as expected from the experimental results (**TABLE C.7**). However, the numerical values of the two MMPBSA calculations were still within 1 kcal/mol of each other after redocking using the new membrane model. Even though their relative binding order had switched, these close numerical values still predict that the binding affinity of compound 6B to M2R and M3R should be similar based on the rest of our trend in **FIGURE 3.8**. This does not agree with the much larger gap in affinity that was observed experimentally. Our results do suggest that the binding affinity is very sensitive to the positioning of compound 6B in the active site, and that subtle differences in the ability of M2R and M3R to accommodate the shape of compound 6B may lead to noticeable structural differences in the positioning of loop regions near the active site of the receptor.

For the present purpose of comparing the performance of our heterogeneous dielectric and uniform dielectric membrane models using the standard single trajectory MD and MMPBSA approach, this one data point was a far outlier using either membrane model. Omitting it from our analysis should not grossly alter the present trends that we have observed in the comparison of our two membrane models in both **FIGURE 3.7** and **FIGURE 3.8**, which contained eleven other data points spread across our three model systems. In the future, more extensive studies will have to be carried out to assess the performance of these two membrane models on a wider variety of systems.

3.4 CONCLUSION

The use of MMPBSA binding free energy calculations for membrane protein-ligand systems in rational drug design is likely to increase in the near future. This is due to the pharmacological importance of membrane-bound protein receptors as drug targets and to improved methodology in cryo-EM that will rapidly contribute to the repository of membrane protein structures available for computational study. In anticipation of an increasing demand for MMPBSA calculations for membrane proteins, we have implemented a heterogeneous dielectric implicit membrane model into the Amber software suite. We obtained our depth-dependent dielectric profile for this new model by minimizing the RMSD between the results of explicit and implicit free energy calculations at various depths within the membrane model. This model supersedes our previous uniform, single dielectric implicit membrane model with a more physical definition of a depth-dependent dielectric profile, and we showed that our new heterogeneous dielectric membrane model can lead to an improved correlation for a series of antagonist ligands bound to the $P2Y_{12}R$ and M2R/M3R test systems when compared to the correlation obtained using the old uniform dielectric membrane model under otherwise identical conditions.

Our heterogeneous membrane model has addressed a few major shortcomings in our prior approach using the single dielectric membrane model[11]. Although we had obtained a good agreement with experimental binding affinities for our $P2Y_{12}R$ test system in our prior study, our simple membrane model was unphysical in that our ligands had to be placed outside of the membrane environment to simulate the lack of a polar headgroup region in our single dielectric membrane model. If the user was to use the actual average thickness of the membrane as a parameter in the MMPBSA calculations, we observed that the agreement with experimental binding affinities decreased, since in reality our ligands actually did bind within the membrane in the polar headgroup and ester regions. In our old model, inserting our ligand within the membrane region placed it into a low dielectric environment of 1-4 instead of in an environment of 20-80 to match the true dielectric environment of the ligand when bound in the active site.

In addition to creating a more realistic and physical depth-dependent model, this new model also avoids complications with the old model that can greatly influence the accuracy of MMPBSA free energy calculations, such as fluctuations in the z-axis positions of atoms that may appear within the sampling trajectory of the bound ligand. In the old model, the dielectric constant would change sharply from a high dielectric constant of 80 to a low dielectric constant of 1-4 at the interface between membrane and water whereas in the new heterogeneous model this transition gradually takes place over a distance of about 10 Å.

The potential for developing an automated process for carrying out MMPBSA calculations using the new membrane model will likely prove to be important for rational drug design efforts. To increase the utility of the method, the user needs to be able to place a ligand in an accurate environment for MMPBSA calculations without having to do a lot of manual adjusting of the local dielectric environment while still obtaining physical, meaningful results. Our current model is a step in that direction, and this model can be further improved by calibrating the non-polar contribution to the free energy calculations in a depth-dependent manner and treating other aspects such as entropy and polarization in the future.

3.5 REFERENCES

1. Almen, M. S.; Nordstrom, K. J.; Fredriksson, R.; Schioth, H. B. Mapping the Human Membrane Proteome: A Majority of the Human Membrane Proteins can be Classified According to Function and Evolutionary Origin. *BMC Biol.* **2009**, *7*, 50.

2. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets are There? *Nat. Rev. Drug Discov.* **2006**, *5*, 993-996.

3. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.

4. White, S. Membrane Proteins of Known 3D Structure. http://blanco.biomol.uci.edu/mpstruc/ (accessed Nov 26, **2018**).

5. Seddon, A. M.; Curnow, P.; Booth, P. J. Membrane Proteins, Lipids and Detergents: Not Just a Soap Opera. *Biochim. Biophys. Acta* **2004**, *1666*, 105-117.

6. Rawson, S.; Davies, S.; Lippiat, J. D.; Muench, S. P. The Changing Landscape of Membrane Protein Structural Biology Through Developments in Electron Microscopy. *Mol. Membr. Biol.* **2016**, *33*, 12-22.

7. Case, D. A.; Cheatham, T. E. 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668-1688.

8. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., 3rd; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L; York, D. M.; Kollman, P. A. *AMBER 18*; University of California, San Francisco, **2018**.

9. Botello-Smith, W. M.; Liu, X.; Cai, Q.; Li, Z.; Zhao, H.; Luo, R. Numerical Poisson-Boltzmann Model for Continuum Membrane Systems. *Chem. Phys. Lett.* **2013**, *555*, 274-281.

10. Botello-Smith, W. M.; Luo, R. Applications of MMPBSA to Membrane Proteins I: Efficient Numerical Solutions of Periodic Poisson-Boltzmann Equation. *J. Chem. Inf. Model.* **2015**, *55*, 2187-2199.

11. Greene, D.; Botello-Smith, W. M.; Follmer, A.; Xiao, L.; Lambros, E.; Luo, R. Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor. *J. Phys. Chem. B* **2016**, *120*, 12293-12304.

12. Xiao, L.; Diao, J.; Greene, D.; Wang, J.; Luo, R. A Continuum Poisson–Boltzmann Model for Membrane Channel Proteins. *J. Chem. Theory Comput.* **2017**, *13*, 3398-3412.

13. Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. Combining Conformational Flexibility and Continuum Electrostatics for Calculating pK(a)s in Proteins. *Biophys. J.* **2002**, *83*, 1731-1748.

14. Luo, R.; Head, M. S.; Moult, J.; Gilson, M. K. pK(a) Shifts in Small Molecules and HIV Protease: Electrostatics and Conformation. *J. Am. Chem. Soc.* **1998**, *120*, 6138-6146.

15. Shivakumar, D.; Deng, Y.; Roux, B. Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput.* **2009**, *5*, 919-930.

16. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *J. Med. Chem.* **2008**, *51*, 769-779.

17. Tan, C.; Yang, L.; Luo, R. How Well Does Poisson-Boltzmann Implicit Solvent Agree with Explicit Solvent? A Quantitative Analysis. *J. Phys. Chem. B* **2006**, *110*, 18680-18687.

18. Tan, C.; Tan, Y. H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007**, *111*, 12263-12274.

19. Wen, E. Z.; Luo, R. Interplay of Secondary Structures and Side-Chain Contacts in the Denatured State of BBA1. *J. Chem. Phys.* **2004**, *121*, 2412-2421.

20. Wen, E. Z.; Hsieh, M. J.; Kollman, P. A.; Luo, R. Enhanced Ab Initio Protein Folding Simulations in Poisson-Boltzmann Molecular Dynamics with Self-Guiding Forces. *J. Mol. Graphics Modell.* **2004**, *22*, 415-424.

21. Lwin, T. Z.; Zhou, R. H.; Luo, R. Is Poisson-Boltzmann Theory Insufficient for Protein Folding Simulations? *J. Chem. Phys.* **2006**, *124*, 034902.

22. Lwin, T. Z.; Luo, R. Force Field Influences in Beta-Hairpin Folding Simulations. *Protein Sci.* **2006**, *15*, 2642-2655.

23. Lwin, T. Z.; Luo, R. Overcoming Entropic Barrier with Coupled Sampling at Dual Resolutions. *J. Chem. Phys.* **2005**, *123*, 194904.

24. Tan, Y. H.; Chen, Y. M.; Ye, X.; Lu, Q.; Tretyachenko-Ladokhina, V.; Yang, W.; Senear, D. F.; Luo, R. Molecular Mechanisms of Functional Rescue Mediated by P53 Tumor Suppressor Mutations. *Biophys. Chem.* **2009**, *145*, 37-44.

25. Wang, J.; Tan, C.; Tan, Y. H.; Lu, Q.; Luo, R. Poisson-Boltzmann Solvents in Molecular Dynamics Simulations. *Commun. Comput. Phys.* **2008**, *3*, 1010-1031.

26. Lu, Q.; Tan, Y. H.; Luo, R. Molecular Dynamics Simulations of p53 DNA-Binding Domain. *J. Phys. Chem. B* **2007**, *111*, 11538-11545.

27. Swanson, J. M.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.* **2004**, *86*, 67-74.

28. Bertonati, C.; Honig, B.; Alexov, E. Poisson-Boltzmann Calculations of Nonspecific Salt Effects on Protein-Protein Binding Free Energies. *Biophys. J.* **2007**, *92*, 1891-1899.

29. Luo, R.; Gilson, M. K. Synthetic Adenine Receptors: Direct Calculation of Binding Affinity and Entropy. *J. Am. Chem. Soc.* **2000**, *122*, 2934-2937.

30. Luo, R.; Head, M. S.; Given, J. A.; Gilson, M. K. Nucleic Acid Base-Pairing and N-Methylacetamide Self-Association in Chloroform: Affinity and Conformation. *Biophys. Chem.* **1999**, *78*, 183-193.

31. Warwicker, J.; Watson, H. C. Calculation of the Electric Potential in the Active Site Cleft Due to Alpha-Helix Dipoles. *J. Mol. Biol.* **1982**, *157*, 671-679.

32. Jeancharles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. Electrostatic Contributions to Solvation Energies - Comparison of Free-Energy Perturbation and Continuum Calculations. *J. Am. Chem. Soc.* **1991**, *113*, 1454-1455.

33. Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. Solvation Free Energies of Peptides: Comparison of Approximate Continuum Solvation Models with Accurate Solution of the Poisson-Boltzmann Equation. *J. Phys. Chem. B* **1997**, *101*, 1190-1197.

34. Luo, R.; Moult, J.; Gilson, M. K. Dielectric Screening Treatment of Electrostatic Solvation. *J. Phys. Chem. B* **1997**, *101*, 11226-11236.

35. Cai, Q.; Wang, J.; Zhao, H. K.; Luo, R. On Removal of Charge Singularity in Poisson-Boltzmann Equation. *J. Chem. Phys.* **2009**, *130*, 145101.

36. Wang, J.; Tan, C. H.; Chanco, E.; Luo, R. Quantitative Analysis of Poisson-Boltzmann Implicit Solvent in Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1194-1202.

37. Egberts, E.; Marrink, S. J.; Berendsen, H. J. C. Molecular Dynamics Simulation of a Phospholipid Membrane. *Eur. Biophys. J.* **1994**, *22*, 423-436.

38. Marrink, S. J.; Berendsen, H. J. C. Simulation of Water Transport through a Lipid-Membrane. *J. Phys. Chem.* **1994**, *98*, 4155-4168.

39. Nymeyer, H.; Zhou, H. X. A Method to Determine Dielectric Constants in Nonhomogeneous Systems: Application to Biological Membranes. *Biophys*. *J*. **2008**, *94*, 1185-1193.

40. Stern, H. A.; Feller, S. E. Calculation of the Dielectric Permittivity Profile for a Nonuniform System: Application to a Lipid Bilayer Simulation. *J*. *Chem*. *Phys*. **2003**, *118*, 3401-3412.

41. Pohorille, A.; Wilson, M. A. Excess Chemical Potential of Small Solutes Across Water-Membrane and Water-Hexane Interfaces. *J*. *Chem*. *Phys*. **1996**, *104*, 3760-3773.

42. Marrink, S. J.; Berendsen, H. J. C. Permeation Process of Small Molecules Across Lipid Membranes Studied by Molecular Dynamics Simulations. *J*. *Phys*. *Chem*. **1996**, *100*, 16729-16738.

43. Bemporad, D.; Essex, J. W.; Luttmann, C. Permeation of Small Molecules Through a Lipid Bilayer: A Computer Simulation Study. *J*. *Phys*. *Chem*. *B* **2004**, *108*, 4875-4884.

44. Bemporad, D.; Luttmann, C.; Essex, J. W. Computer Simulation of Small Molecule Permeation Across a Lipid Bilayer: Dependence on Bilayer Properties and Solute Volume, Size, and Cross-Sectional Area. *Biophys*. *J*. **2004**, *87*, 1-13.

45. Pratt, L. R.; Pohorille, A. Hydrophobic Effects and Modeling of Biophysical Aqueous Solution Interfaces. *Chem*. *Rev*. **2002**, *102*, 2671-2692.

46. Im, W.; Feig, M.; Brooks, C. L., 3rd. An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophys*. *J*. **2003**, *85*, 2900-2918.

47. Feig, M.; Im, W.; Brooks, C. L., 3rd. Implicit Solvation Based on Generalized Born Theory in Different Dielectric Environments. *J*. *Chem*. *Phys*. **2004**, *120*, 903-911.

48. Tanizaki, S.; Feig, M. A Generalized Born Formalism for Heterogeneous Dielectric Environments: Application to the Implicit Modeling of Biological Membranes. *J. Chem. Phys.* **2005**, *122*, 124706.

49. Tanizaki, S.; Feig, M. Molecular Dynamics Simulations of Large Integral Membrane Proteins with an Implicit Membrane Model. *J. Phys. Chem. B* **2006**, *110*, 548-556.

50. Dutagaci, B.; Sayadi, M.; Feig, M. Heterogeneous Dielectric Generalized Born Model with a Van der Waals Term Provides Improved Association Energetics of Membrane-Embedded Transmembrane Helices. *J. Comput. Chem.* **2017**, *38*, 1308-1320.

51. Savi, P.; Zachayus, J. L.; Delesque-Touchard, N.; Labouret, C.; Herve, C.; Uzabiaga, M. F.; Pereillo, J. M.; Culouscou, J. M.; Bono, F.; Ferrara, P.; Herbert, J.M. The Active Metabolite of Clopidogrel Disrupts P2Y12 Receptor Oligomers and Partitions Them out of Lipid Rafts. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11069-11074.

52. Algaier, I.; Jakubowski, J. A.; Asai, F.; von Kugelgen, I. Interaction of the Active Metabolite of Prasugrel, R-138727, with Cysteine 97 and Cysteine 175 of the Human P2Y12 Receptor. *J. Thromb. Haemostasis* **2008**, *6*, 1908-1194.

53. Zhang, K.; Zhang, J.; Gao, Z.G.; Zhang, D.; Zhu, L.; Han, G.W.; Moss, S.M.; Paoletta, S.; Kiselev, E.; Lu, W.; Fenalti, G.; Zhang, W.; Müller, C.E.; Yang, H.; Jiang, H.; Cherezov, V.; Katritch, V.; Jacobson, K.A.; Stevens, R.C.; Wu, B.; Zhao, Q. Structure of the Human P2Y12 Receptor in Complex with an Antithrombotic Drug. *Nature* **2014**, *509*, 115-118.

54. Liu, H.; Hofmann, J.; Fish, I.; Schaake, B.; Eitel, K.; Bartuschat, A.; Kaindl, J.; Rampp, H.; Banerjee, A.; Hübner, H.; Clark, M.J.; Vincent, S.G.; Fisher, J.T.; Heinrich, M.R.; Hirata, K.; Liu, X.; Sunahara, R.K.; Shoichet, B.K.; Kobilka, B.K.; Gmeiner, P. Structure-Guided Development

of Selective M3 Muscarinic Acetylcholine Receptor Antagonists. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 12046-12050.

55. MacCallum, J. L.; Bennett, W. F.; Tieleman, D. P. Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **2008**, *94*, 3393-3404.

56. Dickson, C. J. Amber-Umbrella COM restraint tutorial. https://github.com/callumjd/AMBER-Umbrella_COM_restraint_tutorial (accessed Nov 26, **2018**).

57. Li, L. B.; Vorobyov, I.; Allen, T. W. The Role of Membrane Thickness in Charged Protein-Lipid Interactions. *Biochim. Biophys. Acta* **2012**, *1818*, 135-145.

58. Stansfeld, P. J.; Goose, J. E.; Caffrey, M.; Carpenter, E. P.; Parker, J. L.; Newstead, S.; Sansom, M. S. MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **2015**, *23*, 1350-1361.

59. Hogberg, C. J.; Nikitin, A. M.; Lyubartsev, A. P. Modification of the CHARMM Force Field for DMPC Lipid Bilayer. *J. Comput. Chem.* **2008**, *29*, 2359-2369.

60. Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865-879.

61. Madej, B.; Walker, R. An Amber Lipid Force Field Tutorial: Lipid14 Edition. http://ambermd.org/tutorials/advanced/tutorial16/ (accessed Nov 26, **2018**).

62. Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Davila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J.B.; Im, W. CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014**, *35*, 1997-2004.

63. Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245-268.

64. Wang, C.; Ren, P.; Luo, R. Ionic Solution: What Goes Right and Wrong with Continuum Solvation Modeling. *J. Phys. Chem. B* **2017,** *121*, 11169-11179.

65. Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann Calculations for Static and Dynamic Systems. *J. Comput. Chem.* **2002,** *23*, 1244-1253.

66. Lu, Q.; Luo, R. A Poisson-Boltzmann Dynamics Method with Nonperiodic Boundary Condition. *J. Chem. Phys.* **2003,** *119*, 11035-11047.

67. Wang, J.; Cai, Q.; Li, Z. L.; Zhao, H. K.; Luo, R. Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms. *Chem. Phys. Lett.* **2009,** *468*, 112-118.

68. Wang, J.; Luo, R. Assessment of Linear Finite-Difference Poisson-Boltzmann Solvers. *J. Comput. Chem.* **2010,** *31*, 1689-1698.

69. Cai, Q.; Hsieh, M. J.; Wang, J.; Luo, R. Performance of Nonlinear Finite-Difference Poisson-Boltzmann Solvers. *J. Chem. Theory Comput.* **2010,** *6*, 203-211.

70. Ye, X.; Wang, J.; Luo, R. A Revised Density Function for Molecular Surface Calculation in Continuum Solvent Models. *J. Chem. Theory Comput.* **2010,** *6*, 1157-1169.

71. Cai, Q.; Ye, X.; Wang, J.; Luo, R. On-the-Fly Numerical Surface Integration for Finite-Difference Poisson-Boltzmann Methods. *J. Chem. Theory Comput.* **2011,** *7*, 3608-3619.

72. Wang, J.; Cai, Q.; Xiang, Y.; Luo, R. Reducing Grid-Dependence in Finite-Difference Poisson-Boltzmann Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 2741-2751.

73. Wang, C.; Wang, J.; Cai, Q.; Li, Z.; Zhao, H. K.; Luo, R. Exploring Accurate Poisson-Boltzmann Methods for Biomolecular Simulations. *Comput. Theor. Chem.* **2013,** *1024*, 34-44.

74. Xiao, L.; Cai, Q.; Ye, X.; Wang, J.; Luo, R. Electrostatic Forces in the Poisson-Boltzmann Systems. *J. Chem. Phys.* **2013,** *139*, 094106.

75. Wang, C. H.; Xiao, L.; Luo, R. Numerical Interpretation of Molecular Surface Field in Dielectric Modeling of Solvation. *J. Comput. Chem.* **2017,** *38*, 1057-1070.

76. Qi, R.; Botello-Smith, W. M.; Luo, R. Acceleration of Linear Finite-Difference Poisson-Boltzmann Methods on Graphics Processing Units. *J. Chem. Theory Comput.* **2017,** *13*, 3378-3387.

77. Wei, H.; Luo, R.; Qi, R. An Efficient Second-Order Poisson-Boltzmann Method. *J. Comput. Chem.* **2019,** *40*, 1257-1269.

78. Qi, R.; Luo, R. Robustness and Efficiency of Poisson-Boltzmann Modeling on Graphics Processing Units. *J. Chem. Inf. Model.* **2019,** *59*, 409-420.

79. Fritsch, F. N.; Carlson, R. E. Monotone Piecewise Cubic Interpolation. *SIAM Journal on Numerical Analysis* **1980**, *17*, 238-246.

80. de Boor, C. *A Practical Guide to Splines*. Springer-Verlag: New York, **1978**.

81. Moler, C. B. *Numerical Computing with Matlab*. Society for Industrial and Applied Mathematics: Philadelphia, **2004**.

82. Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10-20.

83. Meurer, A.; Smith, C.P.; Paprocki, M.; Čertík, O.; Kirpichev, SB.; Rocklin, M.; Kumar, A.; Ivanov, S.; Moore, J.K.; Singh, S.; Rathnayake, T.; Vig, S.; Granger, B.E.; Muller, R.P.; Bonazzi, F.; Gupta, H.; Vats, S.; Johansson, F.; Pedregosa, F.; Curry, M.J.; Terrel, A.R.; Roučka, S.; Saboo, A.; Fernando, I.; Kulal, S.; Cimrman, R.; Scopatz, A. SymPy: Symbolic Computing in Python. *PeerJ Computer Science* **2017**, 3.

84. Ulmschneider, M. B.; Sansom, M. S. Amino Acid Distributions in Integral Membrane Protein Structures. *Biochim. Biophys. Acta* **2001**, *1512*, 1-14.

85. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T. E., 3rd. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889-897.

86. Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004,** *25*, 238-250.

87. Srinivasan, J.; Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate - DNA Helices. *J. Am. Chem. Soc.* **1998,** *120*, 9401-9409.

88. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314-3321.

89. Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B.; Wang, H.; Ren, P.; Luo, R. Calculating Protein-Ligand Binding Affinities with MMPBSA: Method and Error Analysis. *J. Comput. Chem.* **2016**, *37*, 2436-2446.

90. Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **2017**, *4*, 87.

91. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.

92. Xiao, L.; Wang, C.; Ye, X.; Luo, R. Charge Central Interpretation of the Full Nonlinear PB Equation: Implications for Accurate and Scalable Modeling of Solvation Interactions. *J. Phys. Chem. B* **2016**, *120*, 8707-8721.

93. Yang, T. Y.; Wu, J. C.; Yan, C. L.; Wang, Y. F.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Y. Virtual Screening Using Molecular Simulations. *Proteins: Struct., Funct., Bioinf.* **2011,** *79*, 1940-1951.

94. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011,** *51*, 69-82.

95. Marrink, S. J.; Sok, R. M.; Berendsen, H. J. C. Free Volume Properties of a Simulated Lipid Membrane. *J. Chem. Phys.* **1996**, *104*, 9090-9099.

96. Sandoval-Perez, A.; Pluhackova, K.; Bockmann, R. A., Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface. *J. Chem. Theory Comput.* **2017**, *13*, 2310-2321.

97. Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hunenberger, P. H. Calculating the Binding Free Energies of Charged Species Based on Explicit-Solvent Simulations Employing Lattice-Sum Methods: An Accurate Correction Scheme for Electrostatic Finite-Size Effects. *J. Chem. Phys.* **2013**, *139*, 184103.

98. Lin, Y. L.; Aleksandrov, A.; Simonson, T.; Roux, B. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2690-2709.

99. Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J. Chem. Theory Comput.* **2018**, *14*, 6346-6358.

100. Bennett, W. F.; Tieleman, D. P. The Importance of Membrane Defects-Lessons From Simulations. *Acc. Chem. Res.* **2014**, *47*, 2244-2251.

101. Marrink, S. J.; de Vries, A. H.; Tieleman, D. P. Lipids on the Move: Simulations of Membrane Pores, Domains, Stalks and Curves. *Biochim. Biophys. Acta* **2009**, *1788*, 149-168.

102. Gurtovenko, A. A.; Anwar, J.; Vattulainen, I. Defect-Mediated Trafficking Across Cell Membranes: Insights From In Silico Modeling. *Chem. Rev.* **2010**, *110*, 6077-6103.

103. Lazaridis, T.; Leveritt, J. M., 3rd; PeBenito, L. Implicit Membrane Treatment of Buried Charged Groups: Application to Peptide Translocation Across Lipid Bilayers. *Biochim. Biophys. Acta* **2014**, *1838*, 2149-2159.

104. Miloshevsky, G. V.; Hassanein, A.; Partenskii, M. B.; Jordan, P. C. Electroelastic Coupling Between Membrane Surface Fluctuations and Membrane-Embedded Charges: Continuum Multidielectric Treatment. *J. Chem. Phys.* **2010**, *132*, 234707.

105. Panahi, A.; Feig, M. Dynamic Heterogeneous Dielectric Generalized Born (DHDGB): An Implicit Membrane Model with a Dynamically Varying Bilayer Thickness. *J. Chem. Theory Comput.* **2013**, *9*, 1709-1719.

106. Ravindranathan, K.; Tirado-Rives, J.; Jorgensen, W. L.; Guimaraes, C. R. Improving MM-GB/SA Scoring Through the Application of the Variable Dielectric Model. *J. Chem. Theory Comput.* **2011**, *7*, 3859-3865.

107. Lee, M. S.; Olson, M. A. Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophys. J.* **2006,** *90*, 864-877.

108. Shirts, M. R.; Mobley, D.; Brown, S.P. Free-Energy Calculations in Structure-Based Drug Design. in *Drug Design: Structure and Ligand-based Approaches*; Merz, K.M.; Ringe,S.; Reynolds, C.H., Eds.; Cambridge University Press, **2010**; pp. 61-86.

109. Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J*. *Chem*. *Inf*. *Model*. **2013,** *53*, 1893-1904.

CHAPTER 4

Calibration of the Non-Polar Terms for the Heterogeneous Dielectric Implicit Membrane Model

4.1 INTRODUCTION

The Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method allows users of the Amber software to calculate protein-ligand binding free energies using:

$$\Delta G_{bind,solv} = \Delta G_{bind,vac} + \Delta G_{solv} \tag{1}$$

where $\Delta G_{bind,solv}$ is the binding free energy for the association of a protein with a ligand in a solvated environment, $\Delta G_{bind,vac}$ is the binding free energy for the association in a vacuum environment, and ΔG_{solv} accounts for the solvation free energy. In MMPBSA, the solvation free energy is calculated in an implicit solvent environment to speed up the calculation[1-7]. To serve this purpose, the solvation free energy term is further subdivided into polar and non-polar components:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np}. \tag{2}$$

ΔG_{elec} accounts for the free energy of solvation due to electrostatic solvent-solute interactions and is calculated using the Poisson-Boltzmann equation[8-14]. ΔG_{np} is an empirical term that represents the free energy of solvation due to hydrophobic effects[15].

Recently, we have implemented an implicit solvent model into the Amber software suite which allows user to calculate MMPBSA binding free energies for membrane protein-ligand systems[16-19]. The use of an implicit solvent model for membrane proteins is complicated by the fact that the membrane must be modeled along with the surrounding aqueous environment. To calibrate an implicit membrane to better reproduce electrostatic effects, our first implicit membrane model used a single, low dielectric constant to parameterize a rectangular slab representing the membrane, which was then surrounded by an implicit aqueous solvent environment with a high dielectric constant[16-17]. Next, we introduced a heterogeneous dielectric implicit membrane model which allowed the dielectric constant to vary with depth inside the membrane slab[19]. This change provided a more accurate implicit representation of a biological membrane, which consists of a low dielectric membrane interior that transitions to a high dielectric headgroup region on the membrane periphery.

The next step in improving our MMPBSA binding free energy calculations for membrane protein-ligand systems is to calibrate the non-polar terms for this new heterogeneous membrane model in a depth-dependent manner[20]. Properly parameterizing non-polar terms for an implicit solvent is known to be a key issue, especially for phenomenon that take place deep within the hydrophobic membrane core. For example, to overcome energetic inaccuracies from a study of transmembrane alpha helices when using polar and vacuum implicit solvent models, a non-polar implicit solvent model was developed by Efremov et al. that mimicked the hydrocarbon core of a membrane[21-23]. More recently, several non-polar implicit membrane models have been featured in a wide range of applications such as membrane protein native state discrimination[24], peptide translocation[25], protein folding[26], NMR structure calculations[27], and transmembrane helix association[20].

Two different non-polar implicit solvation models are currently available in Amber[7]. The first of these is a classical method. This classical method has its roots in a study by Lee and Richards who found that the accessibility of solvent molecules to the protein surface is proportional to the surface area of the protein[28]. This finding was exploited by later investigators to simplify non-polar free energy calculations. For example, Cramer and Truhlar grouped the effects of cavity formation, dispersion interactions, and solute-induced restructuring of water into a single semi-empirical cavity surface tension term in their calculations[29].

Typically, the non-polar component of the solvation free energy is approximated using a linear relationship involving the Solvent-Accessible Surface Area (SASA)[30-31]. The SASA is obtained by rolling the center of a solvent probe sphere over the solute surface[32]. The expression used to calculate the non-polar solvation free energy is then given by:

$$\Delta G_{np} = \gamma(\text{SASA}) + \text{C} \tag{3}$$

where ΔG_{np} is the non-polar solvation free energy, γ is a surface tension coefficient, and C is an offset that accounts for the solvation free energy when considering a point solute (i.e. when SASA $= 0$).

While the classical model is widely employed, the correlation between solvation free energies calculated with implicit and explicit solvation using this method was found to be poor[15, 33]. A second, more modern, method was developed based on the work of Gallicchio et al. who replaced the single SASA term with two terms that split the calculation of the non-polar solvation free energy into repulsive and attractive effects[34-35]:

$$\Delta G_{np} = \Delta G_{rep} + \Delta G_{att}. \tag{4}$$

In eq. (4), ΔG_{np} is the non-polar solvation free energy, ΔG_{rep} accounts for solvent-solute repulsion, and ΔG_{att} accounts for solvent-solute attraction due to dispersion forces. One way to carry out this calculation is to employ the Weeks-Chandler-Anderson (WCA) scheme[36]. In this approach, ΔG_{rep} is described using the SASA relationship given in eq. (3)

while ΔG_{att} is approximately obtained by integrating the attractive part of the Lennard-Jones potential over the solvent-occupied volume[15, 20, 34].

Both of the non-polar models in Amber are currently parameterized for globular protein-ligand systems in an aqueous implicit solvent environment. Here, we aim to further improve the accuracy of our MMPBSA calculations for membrane protein-ligand systems by parameterizing the non-polar terms within a membrane implicit solvent environment. Our approach builds off of the method that we used to calibrate the dielectric constant in our heterogeneous dielectric implicit membrane model. In that method, representative side chains of amino acid residues were held fixed at various depths within an explicit membrane during a molecular dynamics (MD) simulation[19-20, 37]. The free energy for de-charging the amino acid side chain at a certain depth, from 100% to 0% of its original charge, was calculated using the BAR method. The explicit solvent was then replaced with an implicit solvent, and the previous procedure was repeated where the PBSA module in Amber was used to calculate the de-charging free energy. The best fit dielectric constant at each membrane depth was then selected to minimize the root mean square deviation (RMSD) between the explicit and implicit results.

Here, we start with a test set of neutralized amino acid side chains that were obtained at the end of the de-charging step in our prior method, and we carry out a softcore thermodynamic integration (TI) method to obtain an estimate of the non-polar contribution to the free energy of solvation[38-40]. As before, we repeat the procedure to calculate the free energy using our implicit membrane model and obtain scaling parameters for our non-polar terms that minimize the RMSD between the two results. At the time of this writing, our softcore TI methodology in Amber is currently being improved and revised, and so our results here should be treated as preliminary. With that caveat in mind, we aim to demonstrate that our present approach will allow us to calibrate the non-polar terms in our implicit membrane model in a depth-dependent manner using the softcore TI method.

4.2 METHODS

In order to validate our approach to parameterizing our implicit membrane solvent to best match the energies calculated using an explicit membrane solvent, we carried out explicit MD simulations for several amino acid side chains using an updated version of the TI softcore method (pmemd.sc) in Amber that was still under development at the time of this study. 11 amino acids in total were used in our current test set although only 9 were used in the highly polar/aqueous regions for z = 20-25 Å (see **TABLE 4.1**). Details on how our explicit membrane model was set up and how our amino acid side chains were prepared, restrained within the membrane at various depths, and de-charged during MD simulations can be found in our prior publication[19]. In this setup, z = 0 Å corresponds to the center of the hydrophobic membrane core while the polar headgroup region of the membrane transitions to the aqueous solvent on the periphery of the membrane between z = 15-25 Å. The initial parameter topology and coordinate files for our present study were obtained from the endpoint of our de-charging protocol from our prior study. All λ values in pmemd.sc were sampled using MD simulations as λ progressed from 0 to 1 for a total of 16 sampled trajectory windows. Initially, MD simulations were carried out for λ step sizes of 10 ns and 20 ns to check for convergence, and we used 20 ns in all of our subsequent analysis.

After the MD simulations, our trajectories were post processed to extract the energy from each frame. This was performed both in the explicit solvent and in the gas phase where all water, ions, and lipids had been removed. The Bennet Acceptance Ratio (BAR) method was employed to assist in the convergence of our calculated free energy values[41]. The free energy of solvation, $\Delta G_{explicit,solv}$, was taken as the difference between the fully solvated and vacuum free energy values:

$$\Delta G_{explicit,solv} = \Delta G_{explicit} - \Delta G_{gas}. \tag{5}$$

We then proceeded to calculate free energy values, $\Delta G_{implicit,solv}$, for the gas phase MD run using our initial structure ($\lambda = 0$) placed inside both of our implicit membrane models (inp = 1 for the classical model and inp = 2 for the modern model) with the PBSA module in Amber. A parameter α was also introduced to scale the calculated $\Delta G_{implicit,solv}$ values. For inp = 1, α is applied to scale the overall free energy, $\Delta G_{implicit,solv}$, but for inp = 2, we can calibrate each of the two terms in eq. (4) separately to fine tune the agreement. In this work, we scaled the ΔG_{rep} term using an α parameter, and then we compared the overall $\Delta G_{implicit,solv}$ afterwards to $\Delta G_{explicit,solv}$. For inp = 2, the dispersion term ΔG_{att} can also be adjusted to maximize agreement with the explicit solvent results, but adjustments to this term will be carried out at a later point. For each value of α, the RMSD between the explicit and implicit results was calculated using:

$$\mathrm{RMSD} = \frac{\sum_i^N (\Delta G_{explicit,solv} - \Delta G_{implicit,solv})_i}{N}. \tag{6}$$

The sum in eq. (6) is over each amino acid i out of the N amino acid side chains that were used in our test set at a given depth.

4.3 RESULTS AND DISCUSSION

In order to demonstrate the suitability of our method for parameterizing the non-polar terms in our implicit membrane model, we must first examine the BAR results obtained for our test set of amino acid side chains using an explicit solvent. **TABLE 4.1** demonstrates the convergence for our calculated $\Delta G_{explicit,solv}$ values using a λ step size of 10 ns or 20 ns in our softcore TI protocol.

TABLE 4.1 Convergence data for $\Delta G_{explicit,solv}$

	z = 0 Å			z = 5 Å			z = 10 Å			z = 15 Å			z = 20 Å			z = 25 Å		
AA	**10 ns**	**20 ns**	**Err.**	**10 ns**	**20 ns**	**Err.**	**10 ns**	**20 ns**	**Err.**	**10 ns**	**20 ns**	**Err.**	**10 ns**	**20 ns**	**Err.**	**10 ns**	**20 ns**	**Err.**
A	0.81	0.77	-0.04	0.80	0.81	0.01	0.86	0.85	0.00	0.73	0.71	-0.03	0.34	0.28	0.06	0.25	0.28	-0.03
Q	3.49	3.38	-0.10	3.81	3.67	-0.14	3.95	4.02	0.07	3.77	3.79	0.01	2.87	2.73	0.14	2.50	2.39	0.11
G	1.75	1.80	0.06	1.90	1.88	-0.03	1.94	1.93	-0.01	1.63	1.61	-0.02	N/A	N/A	N/A	0.68	0.64	0.04
H	4.73	4.73	0.00	5.05	5.12	0.07	5.40	5.44	0.04	5.19	5.39	0.21	5.01	5.08	-0.07	3.42	3.54	-0.12
L	1.25	1.26	0.02	1.38	1.37	-0.02	1.47	1.37	-0.10	0.99	1.10	0.10	0.59	0.43	0.15	0.46	0.44	0.02
M	2.52	2.55	0.02	2.71	2.73	0.01	2.83	2.91	0.07	2.54	2.60	0.06	N/A	N/A	N/A	N/A	N/A	N/A
F	4.75	4.75	0.00	5.14	5.14	0.00	5.42	5.12	-0.30	3.74	3.47	-0.27	0.57	0.41	0.16	0.40	0.47	-0.07
S	1.33	1.34	0.01	1.45	1.45	0.00	1.51	1.52	0.01	1.38	1.30	-0.08	1.00	0.96	0.04	0.77	0.76	0.01
T	1.63	1.66	0.04	1.66	1.76	0.10	1.81	1.88	0.07	1.53	1.61	0.08	1.45	1.47	-0.02	0.84	0.80	0.04
W	7.65	7.53	-0.11	8.41	8.42	0.01	9.06	8.89	-0.16	3.96	3.74	-0.22	5.00	4.51	0.49	N/A	N/A	N/A
V	1.01	1.07	0.05	1.15	1.20	0.06	1.16	1.13	-0.03	0.81	0.96	0.15	0.44	0.32	0.12	0.38	0.34	0.04

Calculated $\Delta G_{explicit,solv}$ free energy values are reported in units of kcal/mol for λ step sizes of 10 ns or 20 ns in our softcore TI protocol. The error (Err.) is reported as the difference between the 20 ns and 10 ns values. The identity of each amino acid (AA) is indicated using the single letter code. Data is marked not available (N/A) for amino acids that were not included in our z = 20 Å and z = 25 Å data sets.

TABLE 4.1 shows that the errors between the $\Delta G_{explicit,solv}$ values that were calculated using a λ step size of 10 ns or 20 ns were generally small. The highest errors appeared for bulky amino acid side chains such as tryptophan, phenylalanine, and glutamine, which also had three out of the four largest $\Delta G_{explicit,solv}$ values. The relatively small errors across the entire set indicated good convergence, and 20 ns was seen to be an acceptable size to use for λ in our softcore TI method.

TABLE 4.2 gives the results for the optimization of our classical non-polar model (inp = 1). Within the hydrophobic membrane core region (z = 0-10 Å), we see that the RMSD between $\Delta G_{explicit,solv}$ and $\Delta G_{implicit,solv}$ $(\alpha = 1)$ in Amber can be lowered by scaling the α parameter to a higher value, between 2.50-3.00.

TABLE 4.2 RMSD values for a given value of the scaling parameter α using the classical non-polar implicit membrane model

z = 0 Å		**z = 5 Å**		**z = 10 Å**		**z = 15 Å**		**z = 20 Å**		**z = 25 Å**	
α	**RMSD**	**α**	**RMSD**	**α**	**RMSD**	**α**	**RMSD**	**α**	**RMSD**	**α**	**RMSD**
1.00	2.45	1.00	2.78	1.00	2.98	1.00	1.80	1.00	1.77	1.00	1.02
2.62	1.53	2.85	1.71	2.98	1.83	2.14	1.19	1.63	1.60	1.00	1.02

The RMSD between $\Delta G_{explicit,solv}$ and $\Delta G_{implicit,solv}$ using the classical non-polar method (inp = 1) is shown for a given value of the scaling parameter α. $\alpha = 1$ corresponds to the current parameterization in Amber while the other value corresponds to the value of α that gives the lowest RMSD for our test set of 11 amino acids (for z = 0-15 Å) and 9 amino acids (for z = 20 and 25 Å).

TABLE 4.2 also shows that the trend of scaling α upward to improve the RMSD reverses at higher values of z as the membrane transitions into the polar headgroup region (z = 15-20 Å), and finally into the bulk aqueous solvent (z = 25 Å). In particular, we note that at z = 25 Å, the current parameterization is already seen as optimized (i.e., $\alpha = 1$ is already the optimal value). This makes sense since the non-polar terms are indeed optimized for an aqueous solvent at present in Amber. It should therefore be expected that no major optimization is needed in the aqueous solvent at z = 25 Å for the non-polar terms, and that this is so further serves to validate our present approach for parameterizing the non-polar terms in both the aqueous and membrane implicit solvent environments.

TABLE 4.3 RMSD values for a given value of the scaling parameter α using the modern non-polar implicit membrane model

z = 0 Å		**z = 5 Å**		**z = 10 Å**		**z = 15 Å**		**z = 20 Å**		**z = 25 Å**	
α	**RMSD**	α	**RMSD**	α	**RMSD**	α	**RMSD**	α	**RMSD**	α	**RMSD**
1.00	2.60	1.00	2.90	1.00	3.09	1.00	2.03	1.00	2.41	1.00	1.86
1.12	2.19	1.14	2.34	1.15	2.46	1.07	1.87	1.02	2.17	0.94	1.72

The RMSD between $\Delta G_{explicit,solv}$ and $\Delta G_{implicit,solv}$ using the modern non-polar method (inp = 2) is shown for a given value of the scaling parameter α. $\alpha = 1$ corresponds to the current parameterization in Amber while the other value corresponds to the value of α that gives the lowest RMSD for our test set of 11 amino acids (for z = 0-15 Å) and 9 amino acids (for z = 20 and 25 Å).

The results for the preliminary optimization of our modern non-polar implicit membrane model (inp = 2) are given in **TABLE 4.3**. We see that in general the same trends appear as for the classical model in that the optimal value of the scaling parameter increases at higher depths within the hydrophobic core region (z = 0-10 Å) before the optimal α decreases down to a value slightly below unity in the aqueous solvent (z = 25 Å). It is interesting to note that the improvement in the RMSD is much more noticeable when scaling the classical model as opposed to the modern model. In the modern model, the optimized scaling parameters are all close to the default of $\alpha = 1$ at every membrane depth.

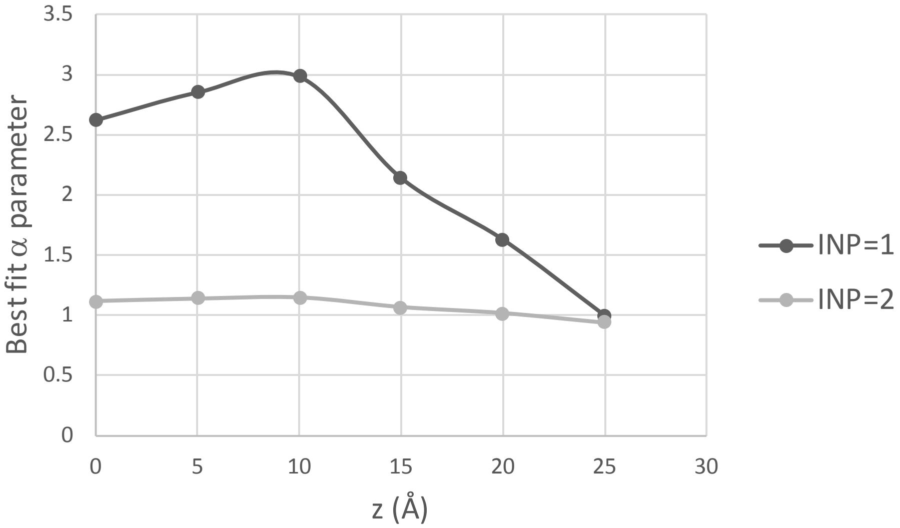


FIGURE 4.1. Optimal α scaling parameters as a function of z. The optimal scaling parameter values for α using both the classical non-polar model (inp = 1) and the modern non-polar model (inp = 2) are given along the z-axis. The hydrophobic membrane center is located at z = 0 Å, which extends to the membrane periphery at z = 20 Å, which is surrounded by the bulk water solvent at z = 25 Å.

As we mentioned in the **METHODS** section, we have yet to treat the dispersion term in eq. (4), so our current results for the modern model reflect the scaling of the repulsion term only. Nevertheless, we see in **TABLE 4.3** that adjusting α for the ΔG_{rep} term alone in our modern model is enough to lower the RMSD, albeit not to the same extent that we saw for our classical model in **TABLE 4.2**. We expect that we will be able to lower the RMSD further by adjusting both ΔG_{rep} and ΔG_{att} for the final calibration of the modern non-polar implicit membrane model. The scaling trends in **TABLES 4.2** and **4.3** for α as a function of z depth are summarized in **FIGURE 4.1**.

4.4 CONCLUSION

To further improve our MMPBSA calculations for membrane protein-ligand systems, we have demonstrated a method that allowed us to re-parameterize the non-polar terms within our implicit membrane solvent environment. The approach involved restraining single amino acid sidechains at a certain depth within an explicit membrane environment, carrying out softcore TI on the side chain to give an estimate of the non-polar free energy as $\Delta G_{explicit,solv}$, turning on our implicit membrane model to calculate $\Delta G_{implicit,solv}$, scaling $\Delta G_{implicit,solv}$ with a parameter α, and finally, calculating the RMSD to find the value of α that minimizes the difference between the explicit and implicit calculated free energy results. We found that this approach allowed us to lower the RMSD at every sampled depth within the implicit membrane for both the classical and modern non-polar implicit membrane models in Amber.

Our current results are preliminary, as our softcore TI method is undergoing improvement and our approach for improving the modern implicit membrane model is incomplete as the calibration of the dispersion term ΔG_{att} has yet to be addressed. Nevertheless, we have shown that the current approach can be used to calibrate the non-polar terms in our implicit membrane models, which upon completion will improve the utility of MMPBSA for binding free energy calculations in studies involving membrane protein-ligand systems in the future.

4.5 REFERENCES

1. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889-897.

2. Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005,** *48*, 4040-4048.

3. Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004,** *25*, 238-250.

4. Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B.; Wang, H.; Ren, P.; Luo, R. Calculating Protein-Ligand Binding Affinities with MMPBSA: Method and Error Analysis. *J. Comput. Chem.* **2016,** *37*, 2436-2446.

5. Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **2017,** *4*, 87.

6. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.

7. Case, D. A.; Cheatham, T. E, 3rd; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P. ; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.;

Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 17*, University of California, San Francisco: **2017**.

8. Wang, J.; Cai, Q.; Li, Z. L.; Zhao, H. K.; Luo, R. Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms. *Chem. Phys. Lett.* **2009,** *468*, 112-118.

9. Wang, J.; Tan, C.; Chanco, E.; Luo, R. Quantitative Analysis of Poisson-Boltzmann Implicit Solvent in Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2010,** *12*, 1194-1202.

10. Wang, J.; Luo, R. Assessment of Linear Finite-Difference Poisson-Boltzmann Solvers. *J. Comput. Chem.* **2010,** *31*, 1689-1698.

11. Cai, Q.; Ye, X.; Wang, J.; Luo, R. On-the-Fly Numerical Surface Integration for Finite-Difference Poisson-Boltzmann Methods. *J. Chem. Theory Comput.* **2011,** *7*, 3608-3619.

12. Wang, J.; Cai, Q.; Xiang, Y.; Luo, R. Reducing Grid-Dependence in Finite-Difference Poisson-Boltzmann Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 2741-2751.

13. Wang, C.; Wang, J.; Cai, Q.; Li, Z.; Zhao, H. K.; Luo, R. Exploring Accurate Poisson-Boltzmann Methods for Biomolecular Simulations. *Comput. Theor. Chem.* **2013,** *1024*, 34-44.

14. Xiao, L.; Cai, Q.; Ye, X.; Wang, J.; Luo, R. Electrostatic Forces in the Poisson-Boltzmann Systems. *J. Chem. Phys.* **2013,** *139*.

15. Tan, C.; Tan, Y. H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007,** *111*, 12263-12274.

16. Botello-Smith, W. M.; Luo, R. Applications of MMPBSA to Membrane Proteins I: Efficient Numerical Solutions of Periodic Poisson-Boltzmann Equation. *J. Chem. Inf. Model* **2015,** *55*, 2187-2199.

17. Greene, D.; Botello-Smith, W. M.; Follmer, A.; Xiao, L.; Lambros, E.; Luo, R. Modeling Membrane Protein-Ligand Binding Interactions: The Human Purinergic Platelet Receptor. *J. Phys. Chem. B* **2016,** *120*, 12293-12304.

18. Xiao, L.; Diao, J.; Greene, D.; Wang, J.; Luo, R. A Continuum Poisson-Boltzmann Model for Membrane Channel Proteins. *J. Chem. Theory Comput.* **2017,** *13*, 3398-3412.

19. Greene, D.; Qi, R.; Nguyen, R.; Qiu, T.; Luo, R. Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies. *J. Chem. Inf. Model* **2019,** *59*, 3041-3056.

20. Dutagaci, B.; Sayadi, M.; Feig, M. Heterogeneous Dielectric Generalized Born Model with a van der Waals Term Provides Improved Association Energetics of Membrane-Embedded Transmembrane Helices. *J. Comput. Chem.* **2017,** *38*, 1308-1320.

21. Nolde, D. E.; Arseniev, A. S.; Vergoten, G.; Efremov, R. G. Atomic Solvation Parameters for Proteins in a Membrane Environment. Application to Transmembrane Alpha-Helices. *J. Biomol. Struct. Dyn.* **1997,** *15*, 1-18.

22. Efremov, R. G.; Nolde, D. E.; Vergoten, G.; Arseniev, A. S. A Solvent Model for Simulations of Peptides in Bilayers. I. Membrane-Promoting Alpha-Helix Formation. *Biophys. J.* **1999,** *76*, 2448-59.

23. Efremov, R. G.; Nolde, D. E.; Vergoten, G.; Arseniev, A. S. A Solvent Model for Simulations of Peptides in Bilayers. II. Membrane-Spanning Alpha-Helices. *Biophys. J.* **1999,** *76*, 2460-71.

24. Yuzlenko, O.; Lazaridis, T., Membrane Protein Native State Discrimination by Implicit Membrane Models. *J. Comput. Chem.* **2013,** *34*, 731-738.

25. Lazaridis, T.; Leveritt, J. M., 3rd; PeBenito, L. Implicit Membrane Treatment of Buried Charged Groups: Application to Peptide Translocation Across Lipid Bilayers. *Biochim. Biophys. Acta* **2014,** *1838*, 2149-2159.

26. Bordner, A. J.; Zorman, B.; Abagyan, R. Efficient Molecular Mechanics Simulations of the Folding, Orientation, and Assembly of Peptides in Lipid Bilayers Using an Implicit Atomic Solvation Model. *J. Comput. Aided Mol. Des.* **2011,** *25*, 895-911.

27. Tian, Y.; Schwieters, C. D.; Opella, S. J.; Marassi, F. M. A Practical Implicit Membrane Potential for NMR Structure Calculations of Membrane Proteins. *Biophys. J.* **2015,** *109*, 574-585.

28. Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971,** *55*, 379-400.

29. Cramer, C. J.; Truhlar, D. G. An SCF Solvation Model for the Hydrophobic Effect and Absolute Free Energies of Aqueous Solvation. *Science* **1992,** *256*, 213-217.

30. Masuda, T.; Jikihara, T.; Nakamura, K.; Kimura, A.; Takagi, T.; Fujiwara, H. Introduction of Solvent-Accessible Surface Area in the Calculation of the Hydrophobicity Parameter log P from an Atomistic Approach. *J. Pharm. Sci.* **1997,** *86*, 57-63.

31. Raschke, T. M.; Tsai, J.; Levitt, M. Quantification of the Hydrophobic Interaction by Simulations of the Aggregation of Small Hydrophobic Solutes in Water. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 5965-5969.

32. Ali, S. A.; Hassan, M. I.; Islam, A.; Ahmad, F. A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States. *Curr. Protein Pept. Sci.* **2014,** *15*, 456-476.

33. Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute-Solvent Interaction Energy. *J. Am. Chem. Soc.* **2003,** *125*, 9523-9530.

34. Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002,** *23*, 517-529.

35. Gallicchio, E.; Levy, R. M. AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling. *J. Comput. Chem.* **2004,** *25*, 479-499.

36. Shivakumar, D.; Deng, Y.; Roux, B. Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput.* **2009,** *5*, 919-930.

37. Lazaridis, T. Effective Energy Function for Proteins in Lipid Membranes. *Proteins* **2003,** *52*, 176-192.

38. Gapsys, V.; Seeliger, D.; de Groot, B. L. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 2373-2382.

39. Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011,** *32*, 3253-3263.

40. Su, P. C.; Johnson, M. E. Evaluating Thermodynamic Integration Performance of the New Amber Molecular Dynamics Package and Assess Potential Halogen Bonds of Enoyl-ACP Reductase (FabI) Benzimidazole Inhibitors. *J. Comput. Chem.* **2016,** *37*, 836-847.

41. Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976,** *22*, 245-268.

SUMMARY AND CONCLUSIONS

To combat the rising cost of new drug development, the biopharmaceutical industry has increasingly employed computational methods to assist in lead generation and lead optimization. A major focus is on the application and development of methods that rank prospective leads by their ability to bind to a selected target with high affinity and/or selectivity. While fast, empirical scoring methods are typically used for lead generation due to the high number of potential compounds that must be screened, more rigorous, physics-based binding free energy calculations are typically preferred for lead optimization. Thermodynamic Integration (TI) and Free Energy Perturbation (FEP) methods are the most accurate methods available for binding free energy calculations, but they are also slow and difficult to implement. Other approximation methods are available that sacrifice some accuracy in order to speed up the calculation process. Here we have focused on the application and development of the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method for its use in rational drug design efforts. The MMPBSA method allows users to calculate the binding free energy for the formation of a protein-ligand complex via a thermodynamic cycle where the effect of the surrounding solvent is modeled implicitly to speed up the calculation.

In our first study, the MMPBSA method was applied to the computational analysis of anti-amyloid beta (Aβ) antibodies. We showed that fragment-based docking using single amino acid residues could predict the emergence of the common EFRH epitope observed for anti-Aβ antibodies. In particular, phenylalanine emerged as a dominant anchor residue for all 10 anti-Aβ antibodies that we studied. To test the importance of phenylalanine in the binding of Aβ to an anti-Aβ antibody, we recreated an experimental study by Gardberg et al. where we calculated the MMPBSA binding free energies of several $A\beta_{2-7}$ peptide variants bound to the protofibril antibody 1 (PFA1) antibody. Our MMPBSA calculations correlated well with experimental binding affinities, and an analysis of the Molecular Dynamics (MD) trajectories gave insight as to why the loss of phenylalanine in $A\beta_{2-7}$ abolished its binding to PFA1.

The MMPBSA method was further applied to another experimental scenario involving $A\beta$ epitope cross binding. While some anti-$A\beta$ antibodies are known to exclusively bind to the N-terminal epitope in fibers (such as bapinezumab) or to the central epitope for smaller oligomeric species (such as solanezumab), other antibodies are believed to bind to both epitopes to some extent. We examined the possibility that epitope cross binding may occur for two recent anti-$A\beta$ antibodies, gantenerumab and crenezumab. For gantenerumab, our calculated MMPBSA binding affinities predicted that the N-terminal epitope was strongly preferred over the central epitope. Experimentally, gantenerumab was known to bind to fibers via the N-terminal epitope, but unlike solanezumab, it did not bind to water soluble amyloid species via the central epitope. For crenezumab, the difference in calculated binding affinity was less extreme. Experimentally, crenezumab was known to bind to water soluble amyloid species via the central epitope, but surprisingly, it also exhibited sporadic binding to amyloid fibers. This was puzzling as the central epitope was believed to be sequestered away in mature fibers. Our MMPBSA results suggested that crenezumab may still bind to the N-terminal epitope, but the binding interaction would be predicted to be much weaker compared to the central epitope. Due to the high effective concentration of N-terminal epitope targets in a solution containing mature fibers, a weak binding to N-terminal species is one possible way to account for the fiber binding patterns that were observed experimentally for crenezumab.

Aside from using MMPBSA calculations to assist in rationalizing experimental observations, we also demonstrated their use in predicting the effectiveness of proposed mutations to assist in the improvement of anti-Aβ antibodies. We rationally introduced mutations to the PFA1 antibody in an effort to increase its calculated binding affinity for the pE3-Aβ_{3-8} epitope. Two out of our four proposed mutations were predicted to stabilize binding to pE3-Aβ_{3-8}. These mutations were also predicted to improve the binding affinity for Aβ_{1-8} while suffering a modest sacrifice in affinity for Aβ_{2-7}. An extension of such methods may yield a series of PFA1 variants that can be tested in the lab to experimentally confirm the predicted increase in binding affinity. Taken together, these results demonstrated the utility of the MMPBSA method in the study of anti-Aβ antibodies and suggested that it can play a key role in both the analysis and optimization of future anti-Aβ drug candidates.

In our second study, we switched our focus to the application of MMPBSA to the calculation of binding free energies for membrane protein-ligand systems. Here we applied our single dielectric implicit membrane model in the analysis of the human purinergic platelet receptor bound to several agonist and antagonist ligands. The effect on the correlation of the calculated MMPBSA binding free energies with experimental binding affinities was studied with respect to changing the non-polar solvation model, the protein dielectric constant, and the membrane dielectric constant. We found that the modern non-polar solvation model clearly outperformed the classical non-polar solvation model. For the protein dielectric constant, we found a peak optimization at a high value of 20. This was likely due to the presence of the two heavily charged nucleotide agonists; raising the dielectric constant to such a high value helped to compensate for the lack of polarization treatment in our current version of Amber. The effect on the membrane dielectric constant was more modest. While we did find an optimum value, we noted that other values of the membrane dielectric constant still worked well. The insensitivity to the membrane dielectric constant likely had to do with the exposure of the active site to water in this study.

Even with the above three parameters optimized, the overall correlation between the agonists and antagonists was still low, with the agonists falling far below the linear trend established by the antagonist ligands. One possible explanation was that we had to model the removal of magnesium ion from the agonists. The experimental binding affinities were obtained in the presence of magnesium ions while magnesium was noticeably absent in the agonist crystal structures that we used in our binding free energy calculations. Correcting for the removal of a magnesium ion from the agonists generally improved the correlation for our full data set while its application did not alter any of the optimized parameter values that we discussed previously. This study demonstrated that a high correlation can be achieved for a membrane protein-ligand system when using our single dielectric implicit membrane model.

In our third study, we developed a new heterogeneous dielectric implicit membrane model and applied it to two relevant G-protein coupled receptor (GPCR) membrane protein-ligand systems. This model improved upon our single dielectric implicit membrane model by introducing a membrane dielectric constant that varied with depth inside the implicit membrane. This new model captured the heterogenous nature of a phospholipid membrane, which features a low dielectric hydrophobic core that transitions to a high dielectric headgroup region on the membrane periphery. Our heterogeneous model was obtained by de-charging the sidechains of all 20 amino acids in explicit solvent, calculating the free energy change using the Bennet Acceptance Ratio (BAR) method, and comparing the result to a free energy calculation using the PBSA-BAR method with our implicit membrane model. The best fit dielectric constant was chosen at each sampled membrane depth based on the value that gave the lowest RMSD value across the entire data set. A dielectric profile was then constructed to interpolate the data across the membrane as a function of depth, and this was implemented into Amber 18. We tested the performance of our new heterogeneous dielectric implicit membrane model and our older single dielectric implicit membrane model using two GPCR test systems bound to several antagonist ligands. The new heterogeneous model outperformed the older single dielectric model in both cases.

In our final study, we gave a first glimpse at future developments that are aimed at improving our implicit membrane model even further. Here we examined the potential to calibrate the non-polar terms for our membrane model in a depth-dependent fashion. Parameterizing the non-polar terms has the potential to improve the accuracy of MMPBSA calculations, especially when it comes to ligands that bind to an active site deep in the membrane core.

The MMPBSA method is already a popular technique for calculating the binding free energy for globular protein-ligand systems. With the rise of cryogenic electron microscopy (cryo-EM) as an experimental technique, structures of membrane-bound proteins should soon become more available for use in drug discovery programs. Large biological systems, such as membrane protein receptors, can benefit greatly from using an implicit solvent to speed up the calculation time in binding free energy calculations. Here we have chronicled our further development of an implicit membrane model for use in MMPBSA calculations. Having improved the treatment of the dielectric constant in this work, the non-polar terms will next be configured to optimize our membrane model in a depth-dependent manner. Other efforts aimed at improving the implicit membrane model, such as addressing polarization in the Amber force field, are also ongoing. As the accuracy of the calculation improves, the MMPBSA method will become an increasingly viable option for calculating binding free energies in drug discovery efforts for the foreseeable future.

APPENDIX A

TABLE A.1. Percent occupancy results for N-terminal and central Aβ peptides bound to gantenerumab

	Aβ residue	gant residue	Chain	% Occupancy
gant	R5	Y93	L	100
gant	D7	N94	L	94.2
gant	E3	R57	H	100
gant	E11	R57	H	72.3
gant	F4	F113	H	96.1
gant (forward)	K16	Y93	L	0.0
gant (forward)	E22	N94	L	65.4
gant (forward)	D23	R57	H	45.5
gant (forward)	E22	R57	H	85.6
gant (forward)	F19	F113	H	99.9

The residue-to-residue percent occupancy was calculated between each pair of residues using a 10 Å cutoff over 5000 frames collected at equally spaced intervals from the 50 ns production run. Aβ residues were labeled to match the standard Aβ_{1-42} numbering sequence while gantenerumab residues and chain designations were labeled to match the original PDB file (PDB ID: 5CSZ). The N-terminal Aβ peptide was bound in its original pose from the crystal structure in gant while the most stable central Aβ peptide bound to gantenerumab (1.8 kcal/mol) was used for comparison in gant (forward).

TABLE A.2. Percent occupancy results for N-terminal and central Aβ peptides bound to crenezumab

	Aβ residue	cren residue	Chain	% Occupancy
cren	E11	R54	L	100
cren	F19	V94	L	99.0
cren	E22	N52	H	100
cren	E22	N53	H	100
cren	D23	N52	H	100
cren	D23	N53	H	36.9
cren	K16	D101	H	100
cren (reverse2)	D8	N28	L	100
cren (reverse2)	H6	Y27D	L	100
cren (reverse2)	F4	V94	L	100
cren (reverse2)	D1	N52	H	100
cren (reverse2)	D1	N53	H	86.3
cren (reverse2)	E3	N52	H	92.3
cren (reverse2)	E3	N53	H	1.5
cren (reverse2)	R5	D101	H	0.0

The residue-to-residue percent occupancy was calculated between each pair of residues using a 10 Å cutoff over 5000 frames collected at equally spaced intervals from the 50 ns production run. Aβ residues were labeled to match the standard Aβ_{1-42} numbering sequence while crenezumab residues and chain designations were labeled to match the original PDB file (PDB ID: 5VZY). The central Aβ peptide was bound in its original pose from the crystal structure in cren while the most stable N-terminal Aβ peptide bound to crenezumab (-3.1 kcal/mol) was used for comparison in cren (reverse2).

TABLE A.3. Percent occupancy results for pE3-Aβ_{3-8} bound to PFA1 wild type and mutant antibodies

	Aβ residue	PFA1 residue	Chain	% Occ. (WT)	% Occ. (N60A)	% Occ. (Y59A)	% Occ. (H93K)	% Occ. (S92K)
pE3-Aβ 3-8	PCA3	H27D	L	19.7	97.9	25.8	0	1.0
pE3-Aβ 3-8	H6	H27D	L	74.9	100	100	98.3	99.9
pE3-Aβ 3-8	H6	Y32	L	62.5	100	100	100	100
pE3-Aβ 3-8	H6	S92	L	32.2	100	100	100	100
pE3-Aβ 3-8	PCA3	H93	L	44.0	99.9	62.4	9.8	98.2
pE3-Aβ 3-8	PCA3	L96	L	17.3	7.1	22.4	100	76.1
pE3-Aβ 3-8	F4	L96	L	35.0	100	100	100	100
pE3-Aβ 3-8	R5	D54	H	100	100	100	100	100
pE3-Aβ 3-8	PCA3	S58	H	50.9	8.6	81.0	100	76.6

The residue-to-residue percent occupancy was calculated between each pair of residues using a 10 Å cutoff over 5000 frames collected at equally spaced intervals from the 50 ns production run. Aβ residues were labeled to match the standard Aβ_{1-42} numbering sequence while PFA1 residues and chain designations were labeled to match the original PDB file (PDB ID: 3EYS).

TABLE A.4. Percent occupancy results for A$\boldsymbol{\beta}_{1-8}$ bound to PFA1 wild type and mutant antibodies

	Aβ residue	PFA1 residue	Chain	% Occupancy (WT)	% Occupancy (N60A)	% Occupancy (Y59A)
Aβ 1-8	D1	N27	L	79.1	99.9	83.8
Aβ 1-8	E3	H27D	L	100	100	100
Aβ 1-8	H6	H27D	L	100	100	100
Aβ 1-8	H6	Y32	L	100	100	100
Aβ 1-8	H6	S92	L	100	100	100
Aβ 1-8	E3	H93	L	100	100	100
Aβ 1-8	E3	L96	L	0.1	0	0
Aβ 1-8	F4	L96	L	100	100	100
Aβ 1-8	R5	D54	H	100	100	100
Aβ 1-8	E3	S58	H	0	0	0

The residue-to-residue percent occupancy was calculated between each pair of residues using a 10 Å cutoff over 5000 frames collected at equally spaced intervals from the 50 ns production run. Aβ residues were labeled to match the standard Aβ_{1-42} numbering sequence while PFA1 residues and chain designations were labeled to match the original PDB file (PDB ID: 2IPU).

TABLE A.5. Percent occupancy results for Aβ_{2-7} bound to PFA1 wild type and mutant antibodies

	Aβ Residue	PFA1 Residue	Chain	% Occupancy (WT)	% Occupancy (N60A)	% Occupancy (Y59A)
Aβ 2-7	E3	H27D	L	100	100	100
Aβ 2-7	H6	H27D	L	100	100	100
Aβ 2-7	H6	Y32	L	100	100	100
Aβ 2-7	H6	S92	L	100	100	100
Aβ 2-7	E3	H93	L	100	100	100
Aβ 2-7	E3	L96	L	0.1	0	0
Aβ 2-7	F4	L96	L	100	100	100
Aβ 2-7	R5	D54	H	100	100	100
Aβ 2-7	E3	S58	H	0	0	0

The residue-to-residue percent occupancy was calculated between each pair of residues using a 10 Å cutoff over 5000 frames collected at equally spaced intervals from the 50 ns production run. Aβ residues were labeled to match the standard Aβ_{1-42} numbering sequence while PFA1 residues and chain designations were labeled to match the original PDB file (PDB ID: 2IPU).

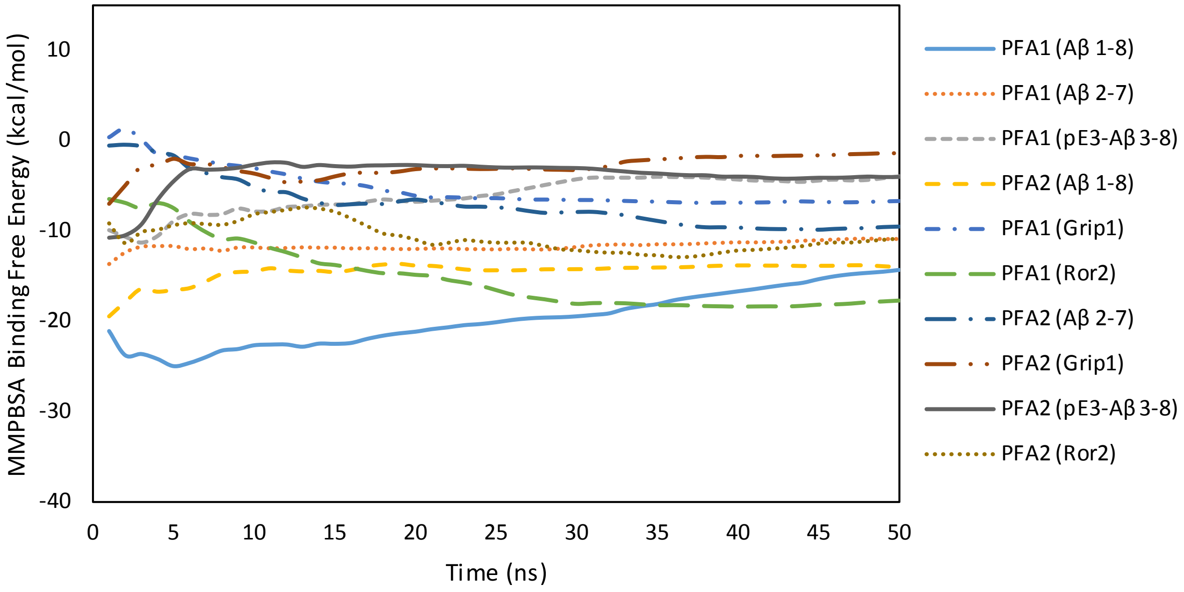


FIGURE A.1. Convergence plots of MMPBSA calculations for various A*β* peptides bound to the antibodies PFA1 and PFA2. Cumulative MMPBSA binding free energies were calculated at each 1 ns time step during the 50 ns production run of the MD simulation in order to validate our MD protocol.

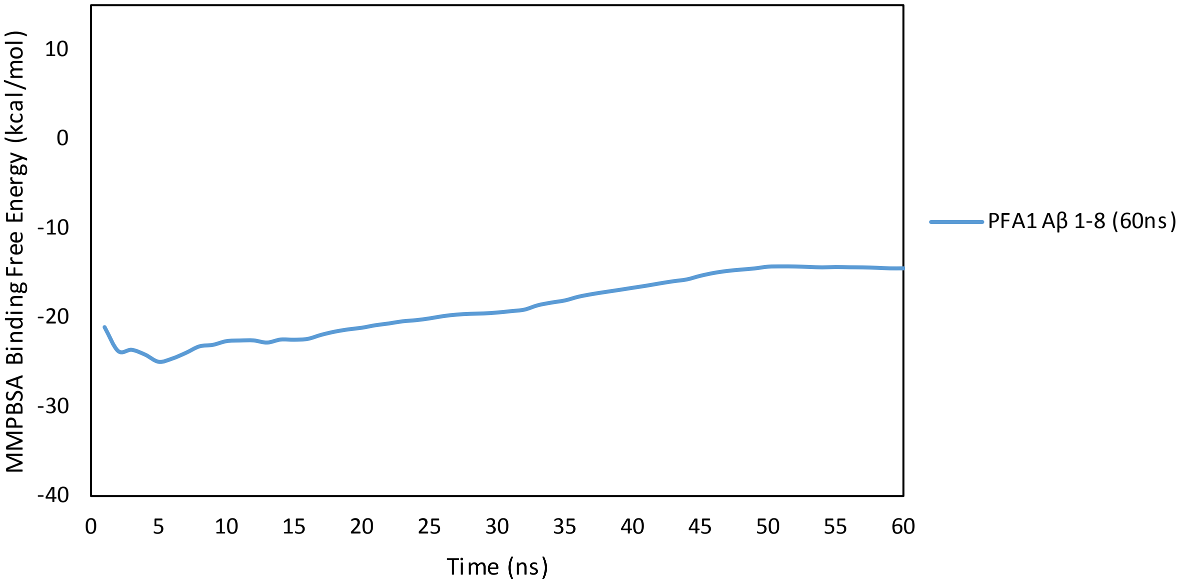


FIGURE A.2. MMPBSA convergence plot for the production run of A$\boldsymbol{\beta}_{\mathbf{1-8}}$ bound to PFA1. The convergence for A$\boldsymbol{\beta}_{\mathbf{1-8}}$ bound to PFA1 appeared questionable in **FIGURE A1**, and so we extended the total simulation time to 310 ns and ran a 60 ns MMPBSA calculation to verify that our result had indeed converged. The MMPBSA binding free energy for A$\boldsymbol{\beta}_{\mathbf{1-8}}$ bound to PFA1 is shown to converge just before 50 ns.

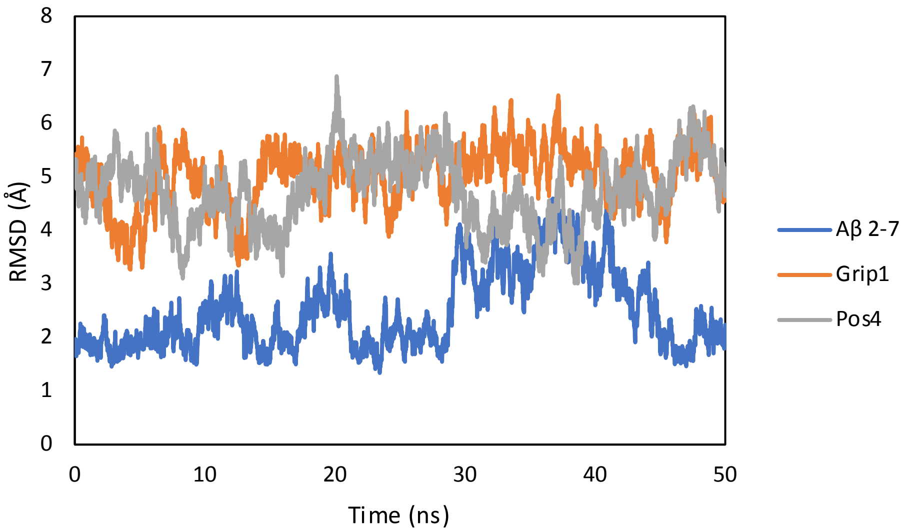


FIGURE A.3. RMSD plots for $A\beta_{2-7}$ peptides bound to PFA1. The Grip1 and Pos4 mutants of $A\beta_{2-7}$ both have larger RMSD values on average in comparison to PFA1 bound to the normal $A\beta_{2-7}$ peptide. The RMSD plots shown above comprise the entire 50 ns production run and were calculated with reference to the initial structure/frame of the 300 ns MD simulation. The RMSD values are reported in units of Angstroms (Å).

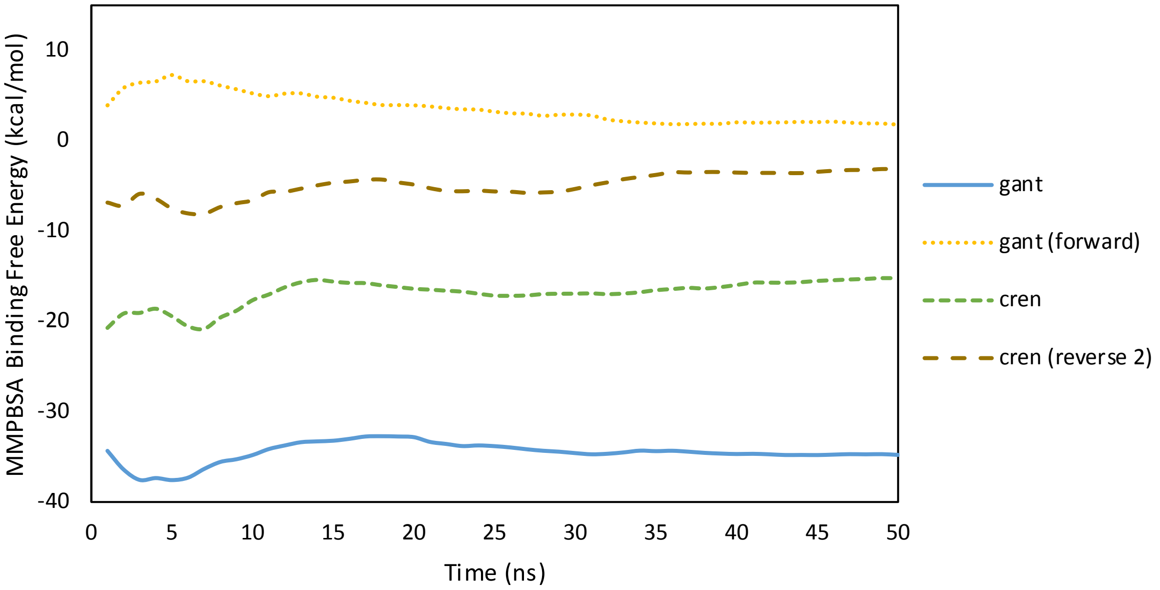


FIGURE A.4. Convergence plots of MMPBSA calculations for N-terminal and central Aβ peptides bound to gantenerumab and crenezumab. Cumulative MMPBSA binding free energies were calculated at each 1 ns time step during the 50 ns production run of the MD simulation.

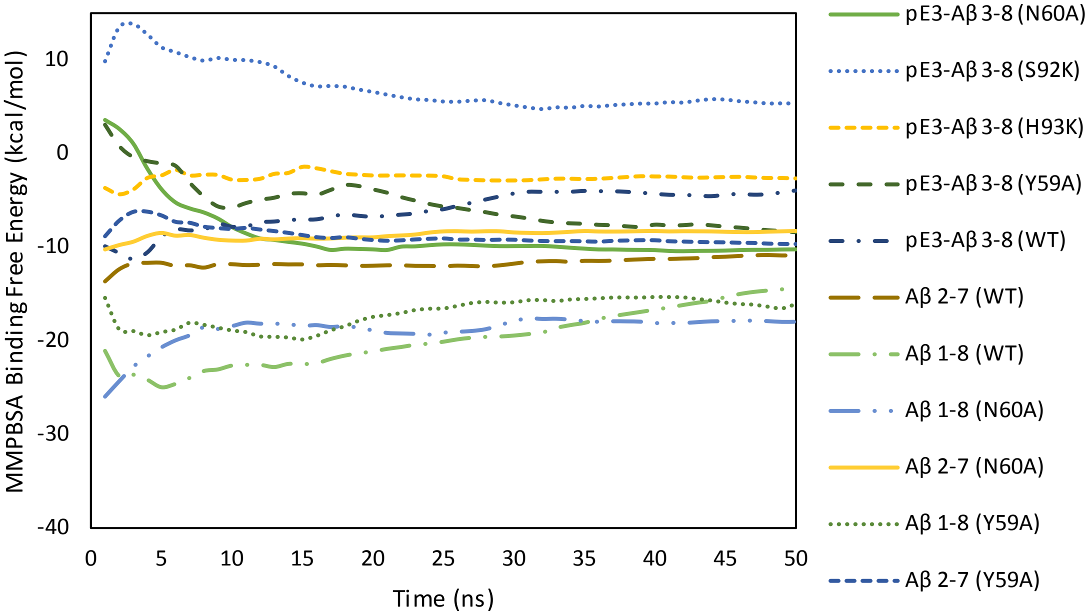


FIGURE A.5. Convergence plots of MMPBSA calculations for $A\beta_{1-8}$, $A\beta_{2-7}$, and pE3-$A\beta_{3-8}$ bound to wild type and mutant forms of PFA1. Cumulative MMPBSA binding free energies were calculated at each 1 ns time step during the 50 ns production run of the MD simulation.

APPENDIX B

TABLE B.1. Raw data that were used for TABLES 2.1 & 2.3

INP=2						INP=1
		EPSIN=20	EPSIN=20	EPSIN=20	EPSIN=20	EPSIN=20
	Ref. 118	EPSMEM=1	EPSMEM=2	EPSMEM=4	EPSMEM=7	EPSMEM=4
4PXZ_WT	-9.49	-43 ± 5	-42 ± 5	-38 ± 6	-39 ± 5	-75 ± 5
4PXZ_D294N	-9.73	-43 ± 4	-42 ± 4	-37 ± 5	-39 ± 4	-76 ± 5
4PY0_WT	-9.50	-48 ± 6	-46 ± 6	-39 ± 6	-42 ± 6	-84 ± 6
4PY0_D294N	-9.72	-50 ± 5	-48 ± 5	-43 ± 6	-43 ± 5	-83 ± 5
4NTJ_WT	-10.29	-37 ± 4	-35 ± 4	-34 ± 4	-32 ± 3	-71 ± 3
4NTJ_D294N	-10.12	-31 ± 3	-31 ± 3	-31 ± 3	-30 ± 3	-67 ± 3
4NTJ_WT_PSB	-9.00	-22 ± 3	-23 ± 3	-23 ± 3	-24 ± 3	-62 ± 3
4NTJ_D294N_PSB	-9.33	-26 ± 4	-26 ± 4	-26 ± 3	-26 ± 3	-67 ± 3
4NTJ_WT_TIQ	-8.70	-22 ± 3	-21 ± 3	-21 ± 3	-21 ± 3	-61 ± 3
4NTJ_D294N_TIQ	-9.25	-27 ± 3	-26 ± 3	-25 ± 3	-25 ± 3	-65 ± 3
2MeSADP_MG	N/A	-10 ± 4	-10 ± 4	-10 ± 4	-10 ± 4	-9 ± 4
2MeSATP_MG	N/A	-6 ± 4	-6 ± 4	-6 ± 4	-6 ± 4	-5 ± 4

All values shown are free energy values (ΔG) for the binding affinity in units of kcal/mol.

TABLE B.2. Raw data that were used for TABLE 2.2

INP=2										
		EPSIN=1	EPSIN=2	EPSIN=4	EPSIN=6	EPSIN=8	EPSIN=12	EPSIN=16	EPSIN=20	EPSIN=24
	Ref. 118	EPSMEM=4	EPSMEM=4	EPSMEM=4	EPSMEM=4	EPSMEM=4	EPSMEM=4	EPSMEM=4	EPSMEM=4	EPSMEM=4
4PXZ_WT	-9.49	-220 ± 17	-135 ± 9	-73 ± 15	-69 ± 5	-59 ± 5	-46 ± 5	-44 ± 5	-38 ± 6	-38 ± 5
4PXZ_D294N	-9.73	-147 ± 16	-102 ± 9	-56 ± 13	-61 ± 5	-55 ± 5	-47 ± 4	-43 ± 4	-37 ± 5	-38 ± 4
4PY0_WT	-9.50	-252 ± 42	-156 ± 20	-79 ± 13	-79 ± 8	-67 ± 7	-55 ± 6	-48 ± 6	-39 ± 6	-41 ± 6
4PY0_D294N	-9.72	-189 ± 15	-128 ± 8	-79 ± 11	-73 ± 5	-64 ± 5	-54 ± 5	-49 ± 5	-43 ± 6	-43 ± 5
4NTJ_WT	-10.29	-1 ± 4	-19 ± 3	-31 ± 7	-30 ± 3	-31 ± 3	-32 ± 3	-32 ± 3	-34 ± 4	-32 ± 3
4NTJ_D294N	-10.12	-8 ± 4	-21 ± 3	-28 ± 4	-29 ± 3	-29 ± 3	-30 ± 3	-30 ± 3	-31 ± 3	-30 ± 3
4NTJ_WT_PSB	-9.00	15 ± 7	-3 ± 5	-13 ± 4	-17 ± 3	-19 ± 3	-21 ± 3	-22 ± 3	-23 ± 3	-24 ± 3
4NTJ_D294N_PSB	-9.33	19 ± 7	-4 ± 4	-16 ± 4	-20 ± 4	-22 ± 3	-24 ± 3	-25 ± 3	-26 ± 3	-26 ± 3
4NTJ_WT_TIQ	-8.70	3 ± 5	-10 ± 3	-17 ± 3	-19 ± 3	-20 ± 3	-21 ± 3	-21 ± 3	-21 ± 3	-21 ± 3
4NTJ_D294N_TIQ	-9.25	0 ± 6	-14 ± 4	-21 ± 3	-23 ± 3	-24 ± 3	-25 ± 3	-25 ± 3	-25 ± 3	-25 ± 3
6AD_MG	N/A	-54 ± 3	-31 ± 2	-19 ± 3	-15 ± 3	-13 ± 3	-11 ± 3	-10 ± 3	-10 ± 4	-9 ± 4
6AT_MG	N/A	-92 ± 5	-47 ± 2	-24 ± 3	-17 ± 3	-13 ± 4	-9 ± 4	-7 ± 4	-6 ± 4	-6 ± 4

All values shown are free energy values (ΔG) for the binding affinity in units of kcal/mol.

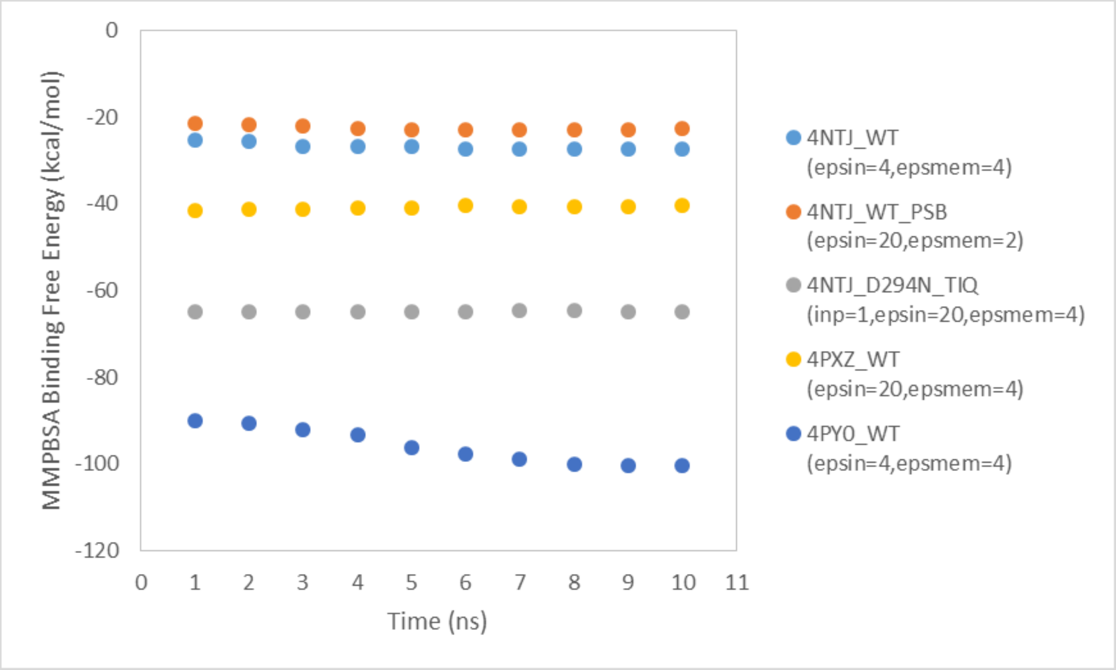


FIGURE B.1. Representative convergence plots of MMPBSA binding free energy (ΔG) values versus the simulation time of the production run.

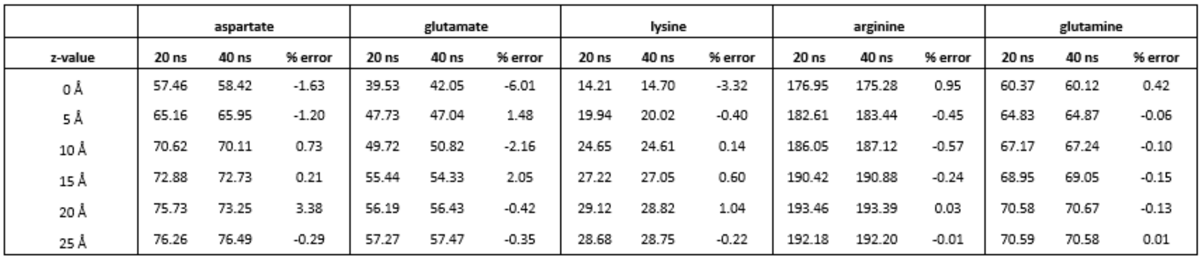
APPENDIX C

TABLE C.1. The effect of ramping up the spring constant on de-charging ΔG calculations

spring constant (equilibration protocol)	**lysine**	**arginine**
$\boldsymbol{k}$ **(kcal/(mol Å^2))**	$\Delta \boldsymbol{G}_{\mathbf{100-90\%}}$ **(kcal/mol)**	$\Delta \boldsymbol{G}_{\mathbf{100-90\%}}$ **(kcal/mol)**
2.5 (100 ns)	5.57	37.38
2.5 (20 ps), 25 (100 ns)	5.62	37.30
2.5 (20 ps), 25 (20 ps),100 (100 ns)	5.57	37.18

Three separate runs were conducted where lysine and arginine were decharged from 100% to 90% at z = 25 Å (in the bulk aqueous solvent) following an approximately 100 ns equilibration step. In the first run, the spring constant *k* was set to 2.5 kcal/(mol Å^2) during the equilibration and was held constant for the remainder of the MD simulation. In the second run, the spring constant was set to 2.5 kcal/(mol Å^2) for 20 ps and was subsequently ramped up to 25 kcal/(mol Å^2) and held there for the remainder of the MD simulation. In the third run, the spring constant was set to 2.5 kcal/(mol Å^2) for the first 20 ps, 25 kcal/(mol Å^2) for an additional 20 ps, and then was set to 100 kcal/(mol Å^2) for the remainder of the MD simulation. Following the equilibration step, the simulation was extended in two further 20 ns steps with the amino acid side chain at 100% and 90% of its original charge respectively. The last 10 ns of each step was used to calculate the free energy difference for de-charging ($\Delta G_{100-90\%}$) using the BAR method.

TABLE C.2. The convergence of explicit BAR free energy difference calculations for de-charging amino acid side chains when placed at a given z-value in a DMPC membrane



	aspartate			glutamate			lysine			arginine			glutamine		
z-value	20 ns	40 ns	% error	20 ns	40 ns	% error	20 ns	40 ns	% error	20 ns	40 ns	% error	20 ns	40 ns	% error
0 Å	57.46	58.42	-1.63	39.53	42.05	-6.01	14.21	14.70	-3.32	176.95	175.28	0.95	60.37	60.12	0.42
5 Å	65.16	65.95	-1.20	47.73	47.04	1.48	19.94	20.02	-0.40	182.61	183.44	-0.45	64.83	64.87	-0.06
10 Å	70.62	70.11	0.73	49.72	50.82	-2.16	24.65	24.61	0.14	186.05	187.12	-0.57	67.17	67.24	-0.10
15 Å	72.88	72.73	0.21	55.44	54.33	2.05	27.22	27.05	0.60	190.42	190.88	-0.24	68.95	69.05	-0.15
20 Å	75.73	73.25	3.38	56.19	56.43	-0.42	29.12	28.82	1.04	193.46	193.39	0.03	70.58	70.67	-0.13
25 Å	76.26	76.49	-0.29	57.27	57.47	-0.35	28.68	28.75	-0.22	192.18	192.20	-0.01	70.59	70.58	0.01

The results of BAR free energy difference calculations for de-charging four charged amino acids and one polar amino acid from 100% to 0% (ΔG_{total}) at the stated z-value are shown in the table above. Two different sets of data were reported; the first calculation used the last 10 ns of 20 ns of trajectory data (1,000 frames) while the second used the last 20 ns of 40 ns of trajectory data (2,000 frames) for each individual step involving a 10% reduction in the charge of the amino acid. The center of the membrane corresponds to a z-value of 0 Å, and a z-value of 25 Å corresponds to the bulk aqueous solvent. These results for the four charged amino acids represent the agreement for the worst-case examples of convergence from our data set (owing to the placement of a fully charged amino acid side chain within the non-polar membrane environment) while the polar glutamine represents the agreement for the next worst-case example. ΔG_{total} values are reported in units of kcal/mol.

TABLE C.3. ΔG_{total} values for de-charging non-polar and aromatic amino acid side chains in explicit solvent

	non-polar, aliphatic							aromatic		
z (Å)	**GLY**	**ALA**	**PRO**	**VAL**	**LEU**	**ILE**	**MET**	**PHE**	**TYR**	**TRP**
0	-17.1	0.0	-9.5	11.5	32.1	-0.2	-0.7	-0.5	18.7	-9.3
5	-14.7	0.0	-6.2	11.5	32.1	-0.2	-0.4	-0.1	20.9	-6.9
10	-12.8	0.0	-6.4	11.5	32.2	-0.2	0.2	0.4	22.2	-6.0
15	-10.7	0.0	-5.0	11.5	32.2	-0.2	0.5	1.2	24.3	-3.6
20	-9.5	0.0	-3.2	11.5	32.2	-0.2	1.0	2.2	24.6	-2.4
25	-9.5	0.0	-2.8	11.5	32.2	-0.2	1.0	2.0	24.2	-3.0

ΔG_{total} values represent the total free energy change (including both the Coulomb and reaction field contributions) for uniformly de-charging an amino acid side chain from 100% to 0% at the given z-value. z-values are measured vertically relative to a plane at the center of the membrane bilayer that is set at z = 0 Å. The membrane region corresponds to values of z within the range 0-20 Å while 25 Å corresponds to the bulk aqueous solvent region. All ΔG_{total} values have units of kcal/mol.

TABLE C.4. ΔG_{total} values for de-charging polar, acidic, and basic amino acid side chains in explicit solvent

	polar, uncharged					acidic		basic		
z (Å)	**SER**	**THR**	**CYS**	**ASN**	**GLN**	**ASP**	**GLU**	**LYS**	**HIS**	**ARG**
0	-5.1	13.7	-6.3	87.9	60.4	57.5	39.5	14.2	10.9	176.9
5	-3.4	15.2	-6.2	91.8	64.8	65.2	47.7	19.9	14.0	182.6
10	-0.8	17.5	-5.2	94.3	67.2	70.6	49.7	24.6	15.7	186.1
15	0.4	18.7	-4.5	96.6	68.9	72.9	55.4	27.2	17.8	190.4
20	1.3	19.5	-4.3	98.0	70.6	75.7	56.2	29.1	19.2	193.5
25	1.0	19.5	-4.2	97.9	70.6	76.3	57.3	28.7	19.4	192.2

ΔG_{total} values represent the total free energy change (including both the Coulomb and reaction field contributions) for uniformly de-charging an amino acid side chain from 100% to 0% at the given z-value. z-values are measured vertically relative to a plane at the center of the membrane bilayer that is set at z = 0 Å. The membrane region corresponds to values of z within the range 0-20 Å while 25 Å corresponds to the bulk aqueous solvent region. All ΔG_{total} values have units of kcal/mol.

TABLE C.5. $\Delta\Delta G_{\text{total}}$ values for de-charging non-polar and aromatic amino acid side chains in explicit solvent

	non-polar, aliphatic						aromatic			
z (Å)	**GLY**	**ALA**	**PRO**	**VAL**	**LEU**	**ILE**	**MET**	**PHE**	**TYR**	**TRP**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	2.4	0.0	3.3	0.0	0.0	0.0	0.2	0.3	2.2	2.4
10	4.2	0.0	3.2	0.0	0.0	0.0	0.8	0.9	3.5	3.3
15	6.3	0.0	4.6	0.0	0.1	0.0	1.2	1.6	5.6	5.8
20	7.5	0.0	6.3	0.1	0.1	0.0	1.6	2.7	5.9	6.9
25	7.6	0.0	6.8	0.0	0.1	0.0	1.6	2.4	5.5	6.4

$\Delta\Delta G_{\text{total}}$ values represent the relative total free energy change (where the z = 0 Å data for each sidechain was subtracted from all z values in **TABLE C.3**) for uniformly de-charging an amino acid side chain from 100% to 0% at the given z-value. z-values are measured vertically relative to a plane at the center of the membrane bilayer that is set at z = 0 Å. The membrane region corresponds to values of z within the range 0-20 Å while 25 Å corresponds to the bulk aqueous solvent region. All $\Delta\Delta G_{\text{total}}$ values have units of kcal/mol.

TABLE C.6. $\Delta\Delta G_{\text{total}}$ values for de-charging polar, acidic, and basic amino acid side chains in explicit solvent

	polar, uncharged					acidic		basic		
z (Å)	**SER**	**THR**	**CYS**	**ASN**	**GLN**	**ASP**	**GLU**	**LYS**	**HIS**	**ARG**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	1.7	1.5	0.2	3.9	4.5	7.7	8.2	5.7	3.1	5.7
10	4.3	3.8	1.1	6.4	6.8	13.2	10.2	10.4	4.8	9.1
15	5.4	5.0	1.8	8.7	8.6	15.4	15.9	13.0	6.9	13.5
20	6.4	5.8	2.1	10.1	10.2	18.3	16.7	14.9	8.3	16.5
25	6.1	5.8	2.1	10.0	10.2	18.8	17.7	14.5	8.5	15.2

$\Delta\Delta G_{\text{total}}$ values represent the relative total free energy change (where the z = 0 Å data for each sidechain was subtracted from all z values in **TABLE C.4**) for uniformly de-charging an amino acid side chain from 100% to 0% at the given z-value. z-values are measured vertically relative to a plane at the center of the membrane bilayer that is set at z = 0 Å. The membrane region corresponds to values of z within the range 0-20 Å while 25 Å corresponds to the bulk aqueous solvent region. All $\Delta\Delta G_{\text{total}}$ values have units of kcal/mol.

TABLE C.7. Results of MMPBSA calculations for compound 6B ligands bound to M2R and M3R

Structure	**ΔG (MMPBSA)**			**ΔG (Exp.)**	**K_i (nM)**
	memopt = 1	**memopt = 2**	**memopt = 3**		
M3R (C6B)	-33.4	-36.0	-36.3	-13.3	0.20
M2R (C6B)	-38.4	-36.2	-36.5	-10.5	21
M3R (Redocked C6B)	-22.3	-29.2	-29.6	-13.3	0.20
M2R (Redocked C6B)	-23.6	-28.5	-28.8	-10.5	21

MMPBSA calculations were carried out using SANDER in MMPBSA.py for the M2R and M3R test system using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). For the uniform, single dielectric membrane model, the membrane dielectric constant was set to 4 while in all three models the protein dielectric constant was set to 2. The experimental binding free energies were obtained from the literature[40]. The antagonist ligand bound to M2R or M3R is listed in parentheses. The original compound 6B (C6B) binding pose was constructed using the binding pose of QNB from the crystal structure as a base whereas the redocked compound 6B (Redocked C6B) binding pose was constructed by minimizing compound C6B in a vacuum environment before docking it to the active site using Autodock Vina/SMINA. All binding free energies are reported in units of kcal/mol.

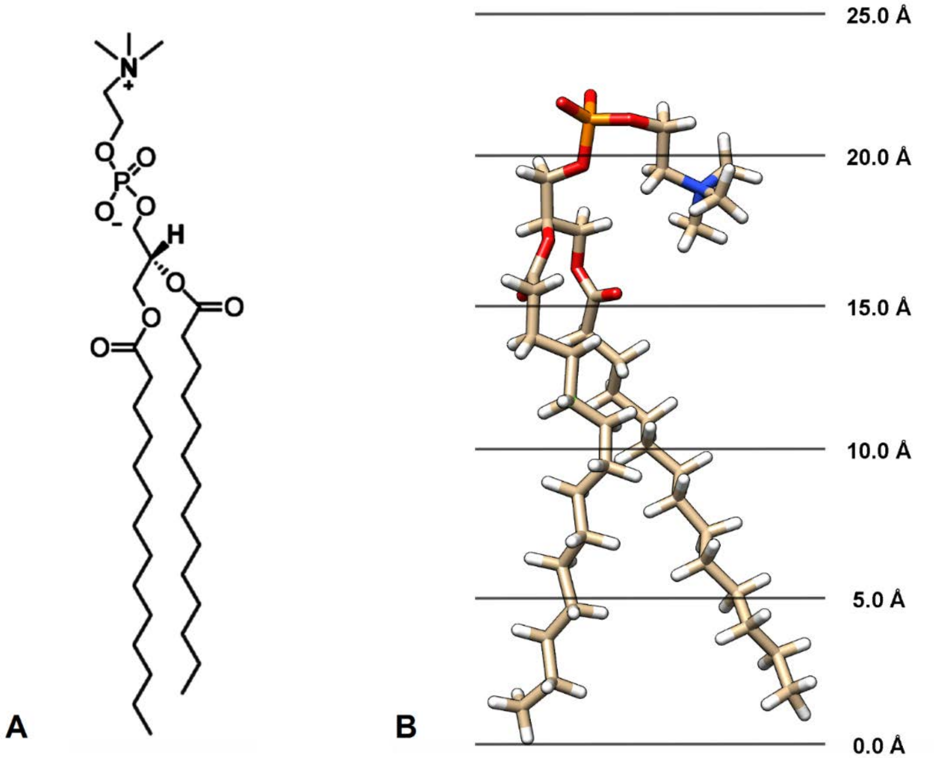


FIGURE C.1. DMPC phospholipid membrane model. We chose to place our amino acid side chain at 5 Å intervals along the z-axis (at z = 0, 5, 10, 15, 20, and 25 Å) for sampling within our DMPC membrane model to split the membrane into 6 regions that roughly corresponded to the midpoint of: the hydrocarbon core region (split up into three subregions: 0 to 2.5 Å, 2.5 to 7.5 Å, and 7.5 to 12.5 Å), the hydrocarbon and ester interface region (12.5 to 17.5 Å), the phosphatidylcholine headgroup region (17.5 to 22.5 Å), and the bulk water solvent (22.5 to 27.5 Å). Note that a traditional depiction of DMPC as shown in (A) does not take into account typical distortions in the structure due to variations in lipid packing and compression of the headgroup region as a result of intermolecular choline-phosphate interactions. A sample frame of a DMPC molecule from one of our MD simulations is shown in (B) to illustrate these common characteristics.

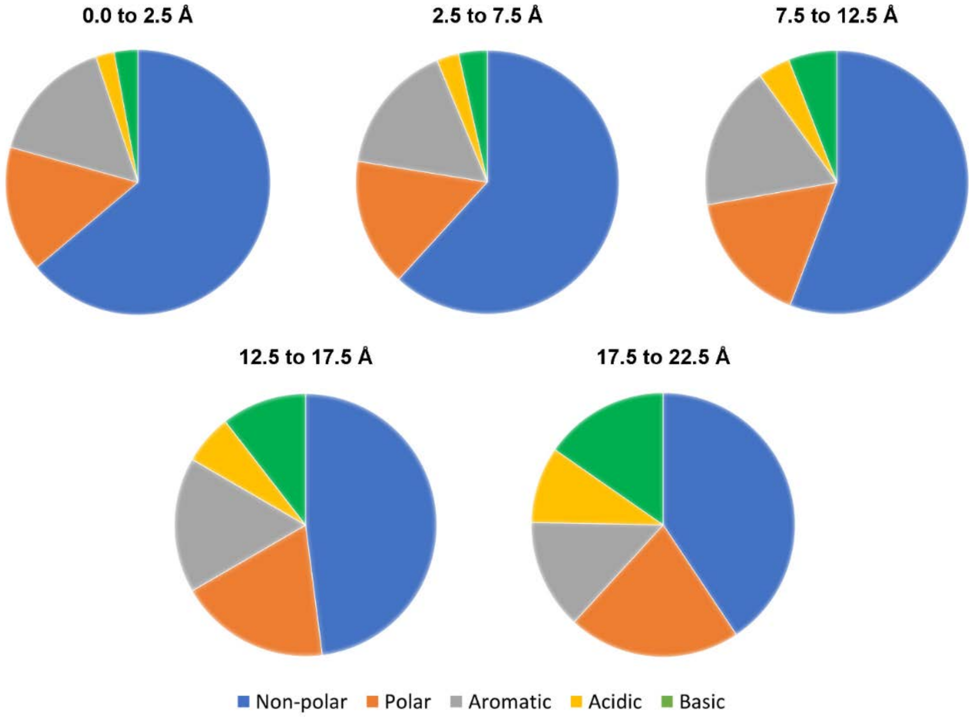


FIGURE C.2. Natural distribution of amino acids within membrane proteins at a given z-value range inside a membrane bilayer. The distribution of amino acids within membrane proteins in a membrane bilayer was calculated using 482 membrane protein structures that were downloaded from the MemProtMD database. Given that z = 0 Å corresponded to a plane at the center of the membrane bilayer, the number of amino acid residues that appeared within a certain range along the z-axis were counted (corresponding to bins of 0 to 2.5 Å, 2.5 to 7.5 Å, 7.5 to 12.5 Å, 12.5 to 17.5 Å, and 17.5 to 22.5 Å). Charged acidic and basic amino acids become noticeably more abundant as the surface of the membrane was approached at higher values of z.

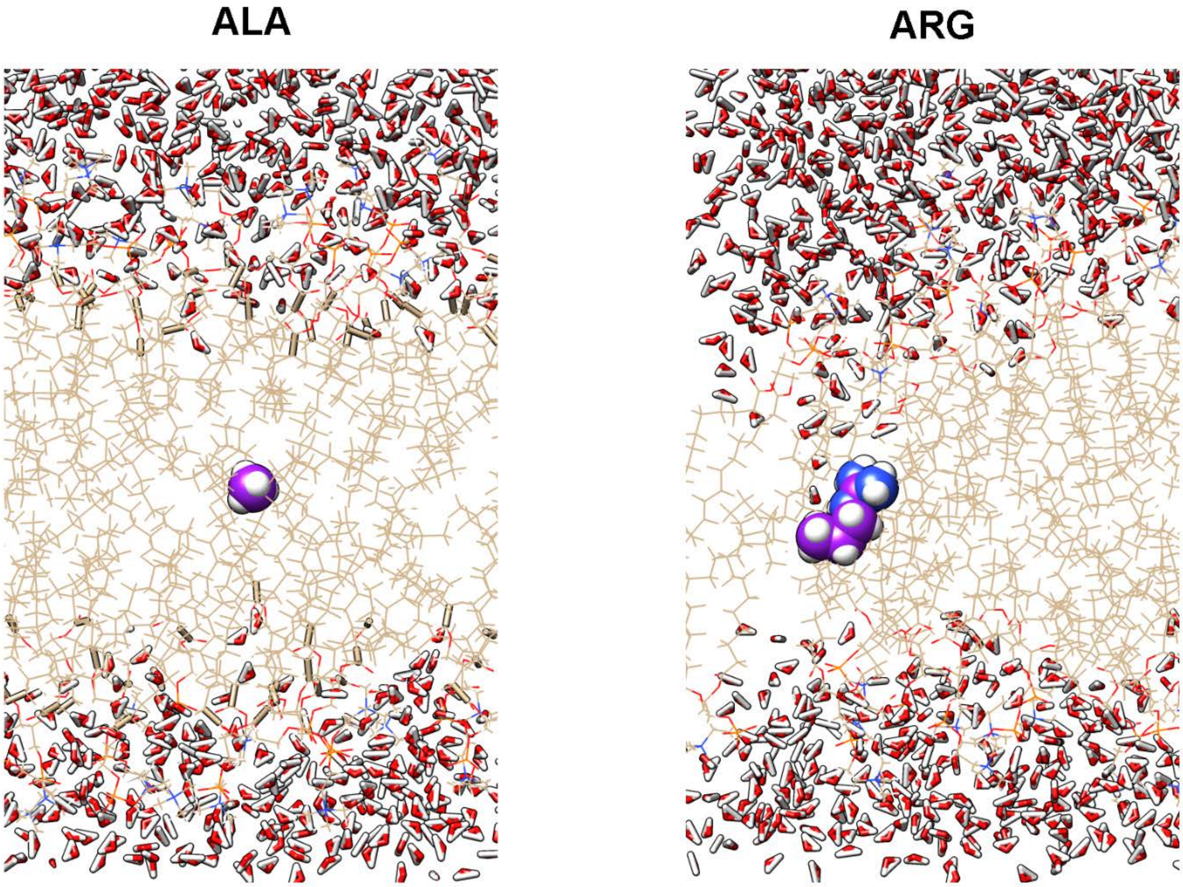


FIGURE C.3. The formation of a membrane defect due to placing a charged amino acid in the hydrophobic core of a phospholipid membrane. At one extreme, such as the non-polar methyl side chain in alanine (left), the hydrophobic core of the phospholipid bilayer remains free of water, leading to a lower estimate of the dielectric constant in the core region. However, at the other extreme, such as when the side chain of arginine is placed in the core region (right), a water defect forms in the membrane to stabilize the charge. The greater penetration of water into the core region can lead to higher estimates of the dielectric constant in the core region when charged species are used as a probe molecule.

PCHIP

$$f(z) =$$

$$\begin{cases} 1.0, & z < 2.5 \\ (5.03 \cdot 10^{-2})z^3 - (5.83 \cdot 10^{-1})z^2 + 1.97z - 1.07, & 2.5 \le z < 5.0 \\ -(1.21 \cdot 10^{-3})z^3 + (3.13 \cdot 10^{-2})z^2 - (3.08 \cdot 10^{-1})z + 1.41, & 5.0 \le z < 10.0 \\ (9.17 \cdot 10^{-4})z^3 - (3.12 \cdot 10^{-2})z^2 + (3.05 \cdot 10^{-1})z - (5.92 \cdot 10^{-1}), & 10.0 \le z < 15.0 \\ (9.47 \cdot 10^{-5})z^3 - (3.71 \cdot 10^{-3})z^2 + (3.47 \cdot 10^{-2})z + (4.41 \cdot 10^{-2}), & 15.0 \le z < 20.0 \\ 1.25 \cdot 10^{-2}, & z \ge 20 \end{cases}$$

Spline

$$f(z) =$$

$$\begin{cases} 1.0, & z < 2.5 \\ (3.44 \cdot 10^{-2})z^3 - (4.12 \cdot 10^{-1})z^2 + (1.41)z - (4.99 \cdot 10^{-1}), & 2.5 \le z < 5.0 \\ (1.16 \cdot 10^{-2})z^2 - (2.40 \cdot 10^{-1})z + 1.48, & 5.0 \le z < 7.5 \\ (3.94 \cdot 10^{-3})z^2 - (1.24 \cdot 10^{-1})z + 1.05, & 7.5 \le z < 12.5 \\ (1.66 \cdot 10^{-3})z^2 - (6.72 \cdot 10^{-2})z + (6.95 \cdot 10^{-1}), & 12.5 \le z < 20.0 \\ (1.25 \cdot 10^{-2}), & z \ge 20 \end{cases}$$

FIGURE C.4. Equations for the fitted membrane dielectric profiles. The fitted membrane dielectric profiles, $f(z)$, as a function of depth (z) are displayed using both the PCHIP and spline approaches. Using the heterogeneous dielectric membrane model for MMPBSA calculations, the PCHIP fitting can be turned on by setting the membrane option (memopt) to 2 while the spline fitting can be turned on by setting the membrane option to 3. A value of z = 0 Å refers to the membrane center where z = 25 Å corresponds to the bulk water solvent. For spline fitting, the range z = 2.5-5.0 Å was replaced with a cubic spline to get a first derivative of zero at the point z = 2.5 Å while still maintaining the same derivative with the second-order spline fitting on the right side at the point z = 5.0 Å. The coefficients in both of the fitted profiles have been rounded to three digits for clarity.

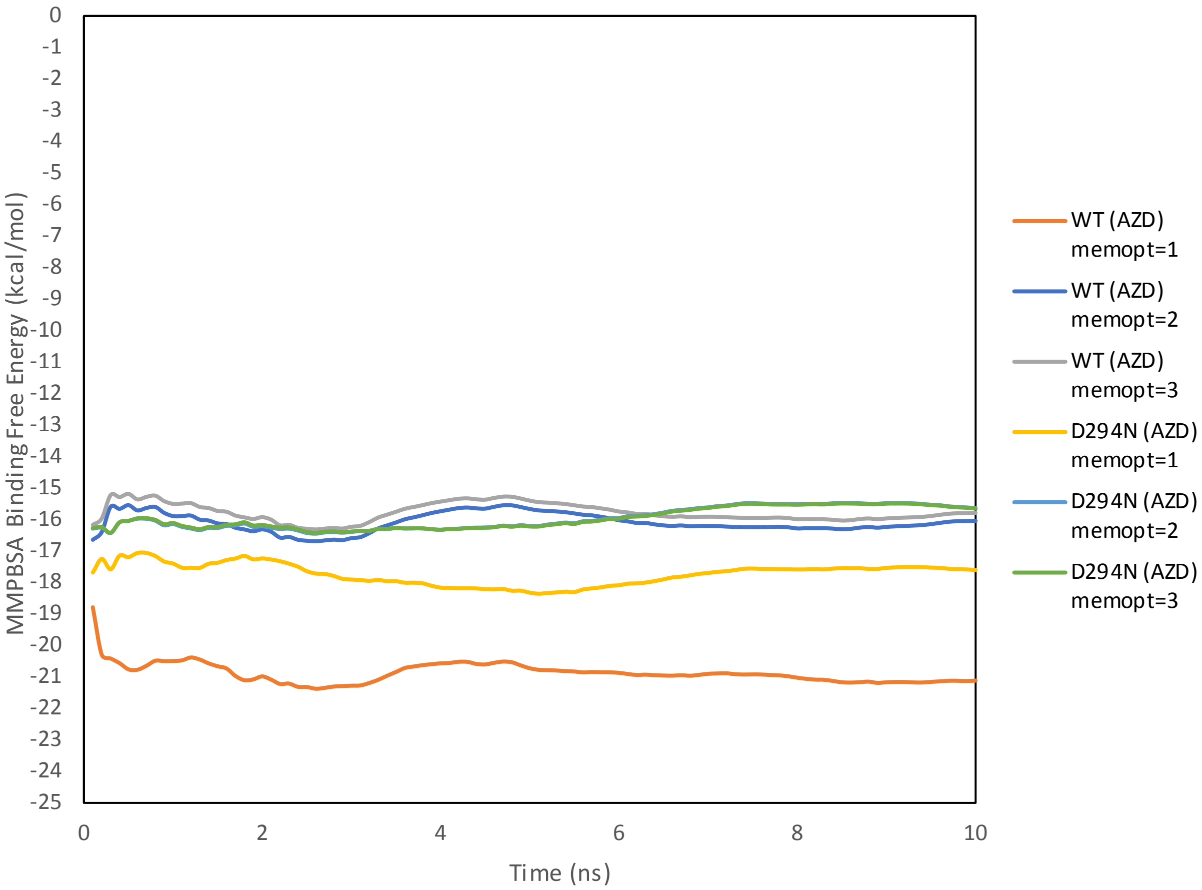


FIGURE C.5. Convergence plots of MMPBSA calculations for AZD-1283 bound to wild type and mutant forms of P2Y$_{12}$R. Cumulative MMPBSA binding free energies were calculated at each 0.1 ns time step increment during the 10 ns production run of the MD simulation. MMPBSA calculations were carried out using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). Note that for AZD-1283 bound to D294N, the curves for memopt = 2 and 3 overlap very closely to the point where the curve for memopt = 2 is barely visible underneath the curve for memopt = 3.

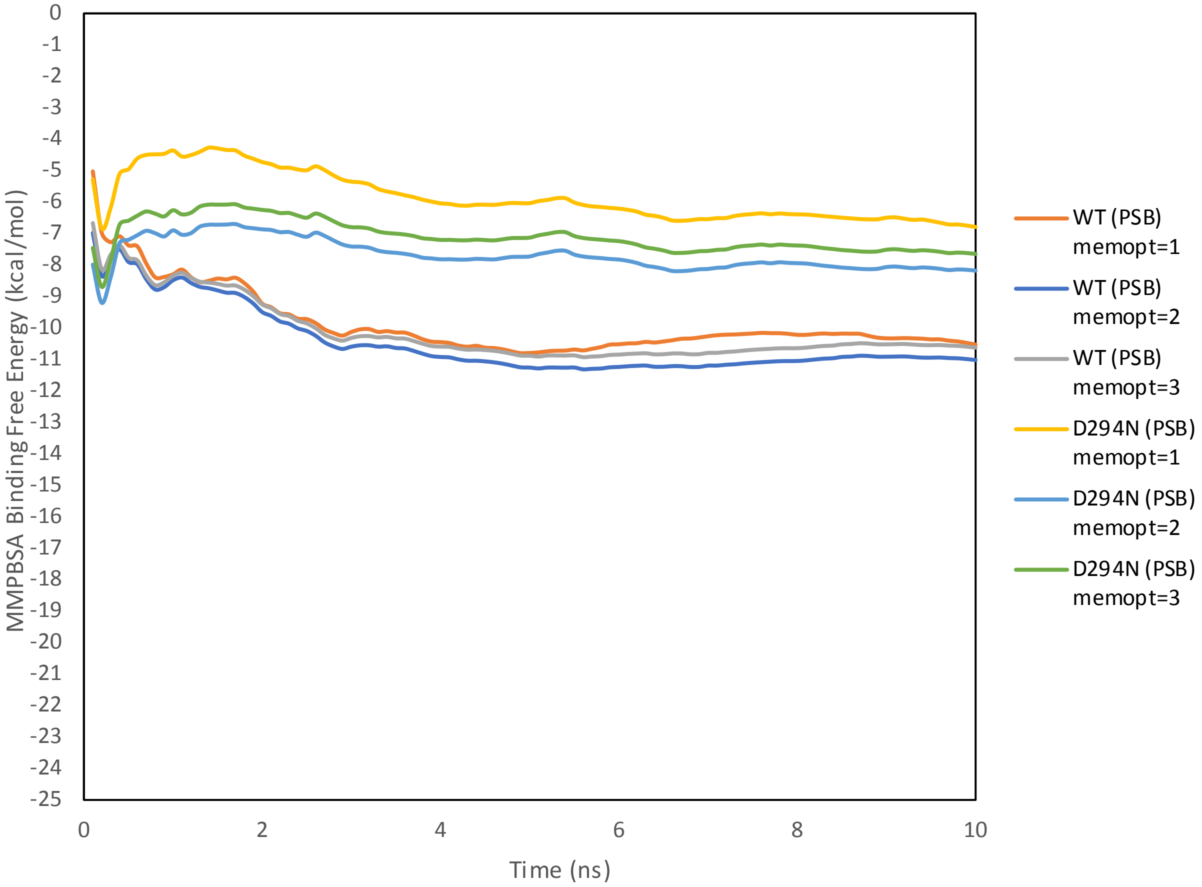


FIGURE C.6. Convergence plots of MMPBSA calculations for PSB-0739 bound to wild type and mutant forms of P2Y$_{12}$R. Cumulative MMPBSA binding free energies were calculated at each 0.1 ns time step increment during the 10 ns production run of the MD simulation. MMPBSA calculations were carried out using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3).

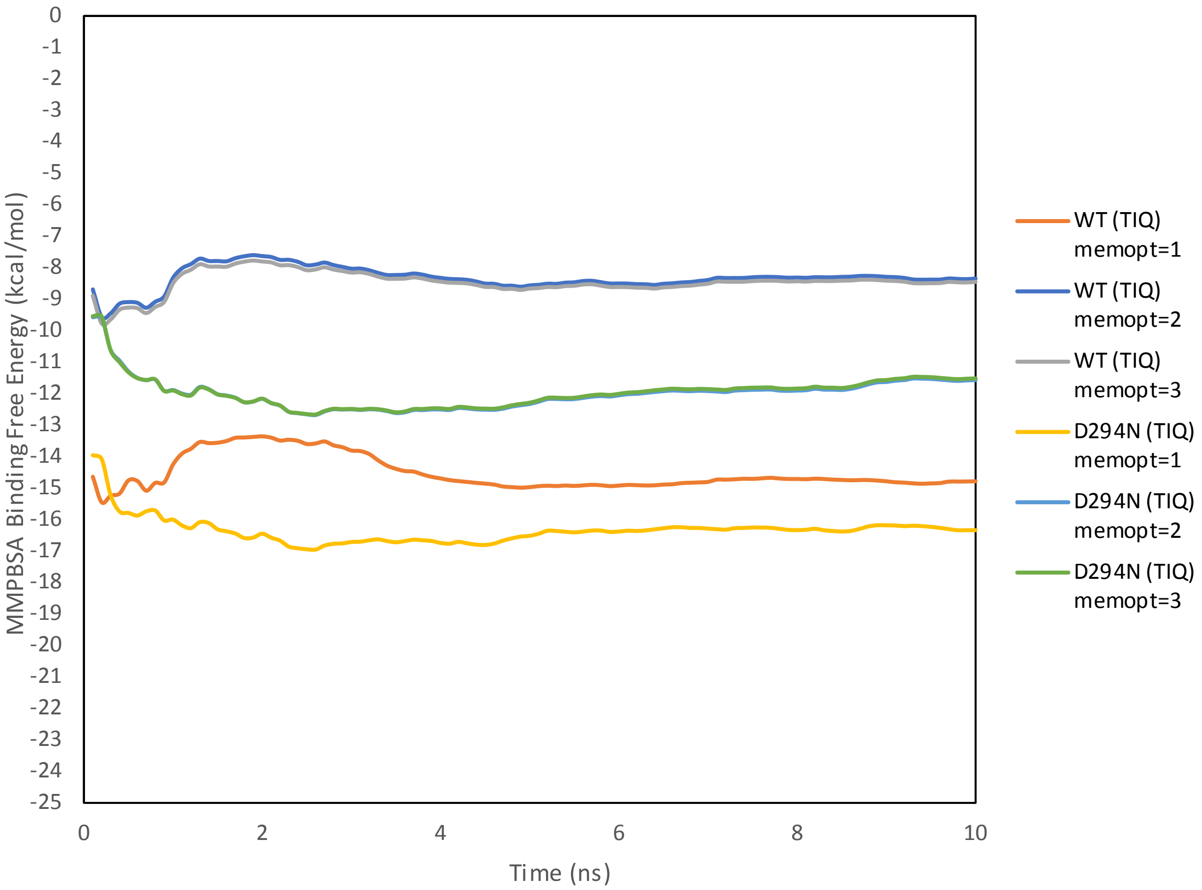


FIGURE C.7. Convergence plots of MMPBSA calculations for Ticagrelor bound to wild type and mutant forms of P2Y$_{12}$R. Cumulative MMPBSA binding free energies were calculated at each 0.1 ns time step increment during the 10 ns production run of the MD simulation. MMPBSA calculations were carried out using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). Note that for ticagrelor bound to D294N, the curves for memopt = 2 and 3 overlap very closely to the point where the curve for memopt = 2 is barely visible underneath the curve for memopt = 3.

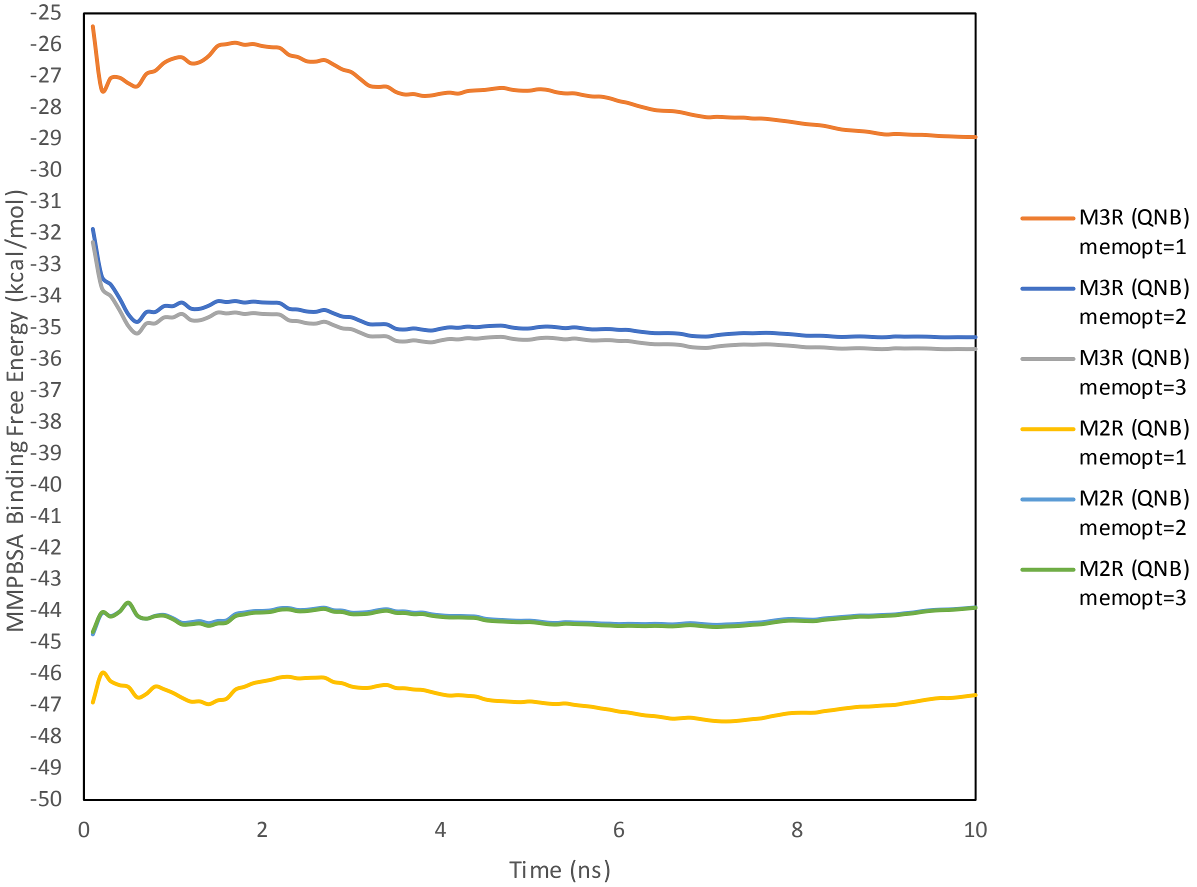


FIGURE C.8. Convergence plots of MMPBSA calculations for 3-quinuclidinyl-benzilate bound to M2R and M3R. Cumulative MMPBSA binding free energies were calculated at each 0.1 ns time step increment during the 10 ns production run of the MD simulation. MMPBSA calculations were carried out using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). Note that for 3-quinuclidinyl-benzilate bound to M2R the curves for memopt = 2 and 3 overlap very closely to the point where the curve for memopt = 2 is barely visible underneath the curve for memopt = 3.

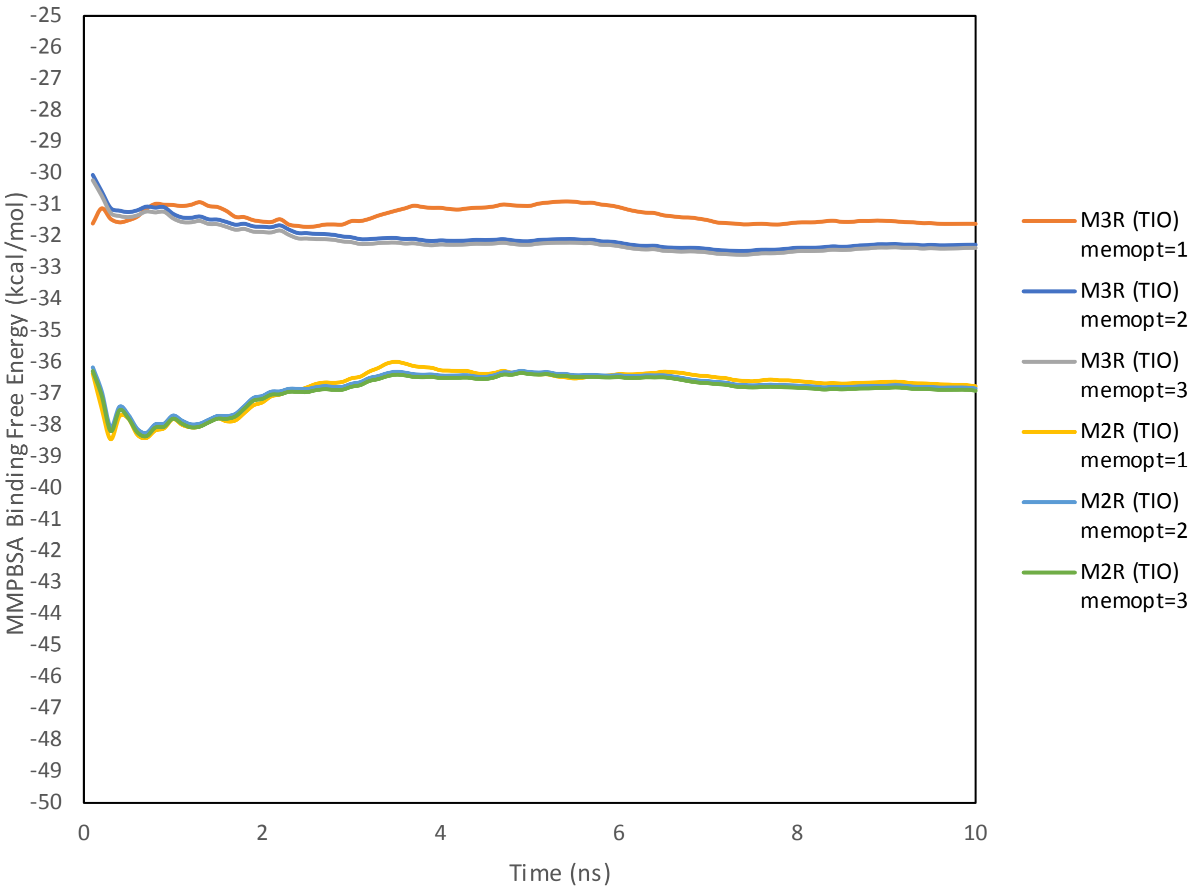


FIGURE C.9. Convergence plots of MMPBSA calculations for tiotropium bound to M2R and M3R. Cumulative MMPBSA binding free energies were calculated at each 0.1 ns time step increment during the 10 ns production run of the MD simulation. MMPBSA calculations were carried out using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3). Note that for tiotropium bound to M2R the curves for memopt = 2 and 3 overlap very closely to the point where the curve for memopt = 2 is barely visible underneath the curve for memopt = 3.

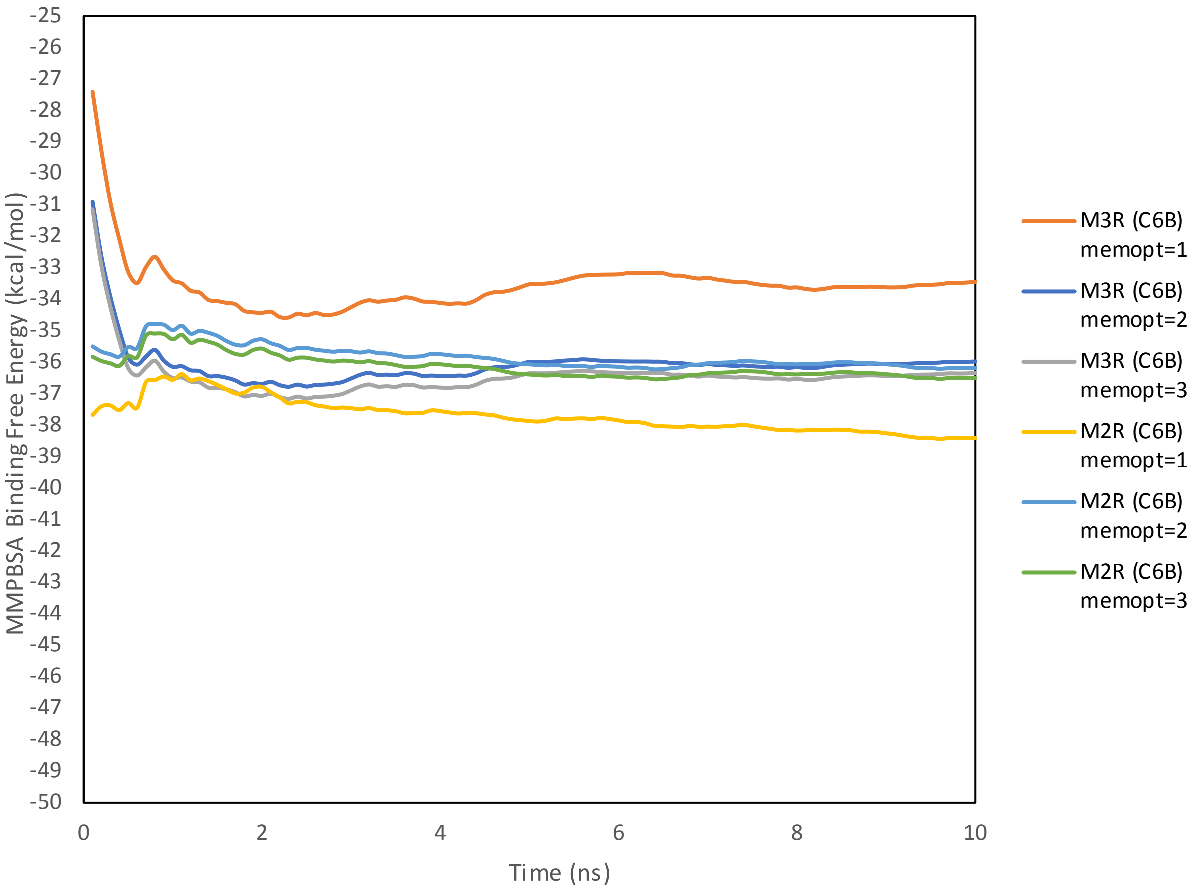


FIGURE C.10. Convergence plots of MMPBSA calculations for compound 6B bound to M2R and M3R. Cumulative MMPBSA binding free energies were calculated at each 0.1 ns time step increment during the 10 ns production run of the MD simulation. MMPBSA calculations were carried out using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3).

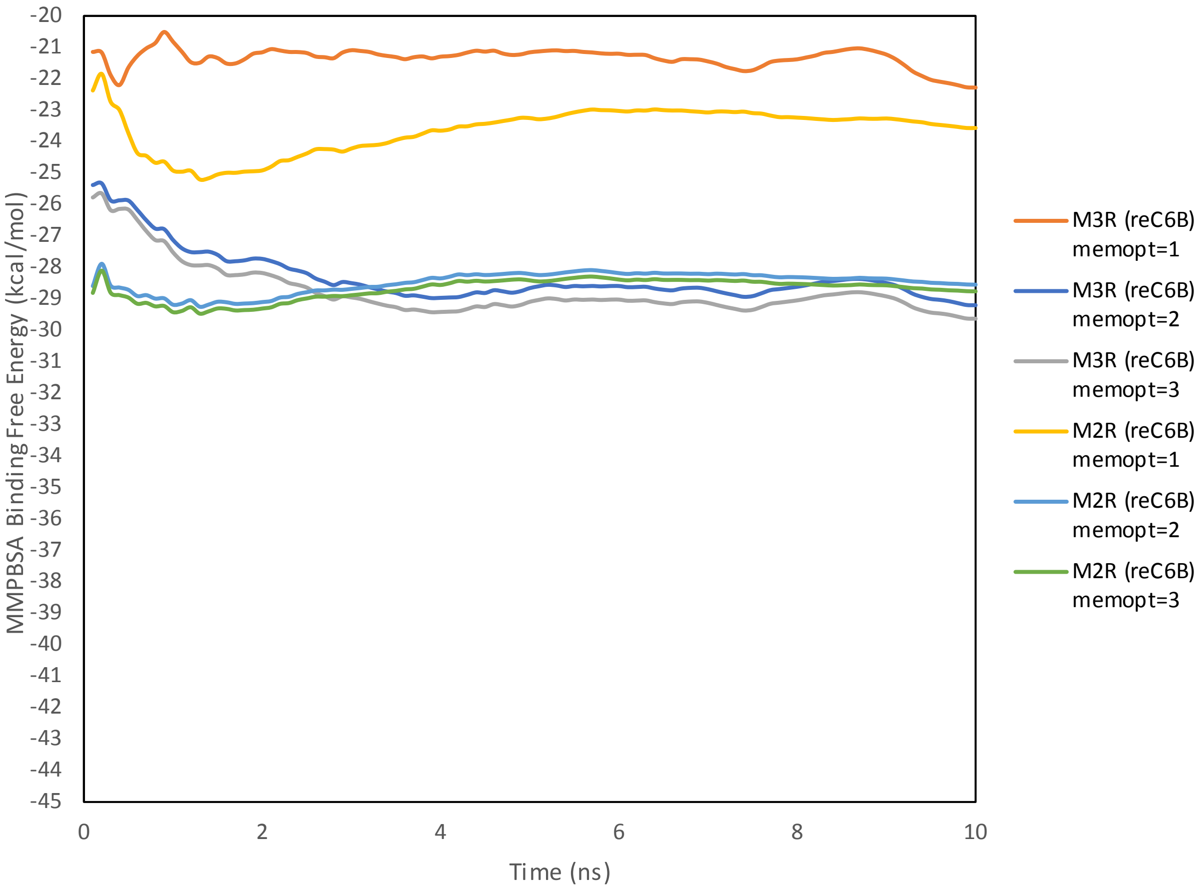


FIGURE C.11. Convergence plots of MMPBSA calculations for the redocked compound 6B bound to M2R and M3R. Cumulative MMPBSA binding free energies were calculated at each 0.1 ns time step increment during the 10 ns production run of the MD simulation. MMPBSA calculations were carried out using the uniform, single dielectric membrane model (memopt = 1), the heterogeneous dielectric membrane model with the PCHIP fitting (memopt = 2), and the heterogeneous dielectric membrane model with the spline fitting (memopt = 3).