Summer 7-16-2021

Development of Computational Antibiotic Screening Platform Across Bacterial Outer Membrane Proteins

Yinghui Dai
Syracuse University

Abstract

Antibiotics are medicines used to treat bacterial infections by either killing bacteria or stopping them from reproducing. Throughout the use of antibiotics, bacteria has developed a variety of defense mechanisms against antibiotics and thus diminishing their effectiveness. Antibiotic resistance is a growing threat and becomes a global crisis as it is able to constantly evolve and rapidly spread. In the face of increasing bacterial resistance to all known antibiotics, there is an urgent need to accelerate the antibiotic discovery pipeline and discover new classes of antibiotics. A major bottleneck in the discovery of novel antibiotics is the limited permeability of potent drug molecules across the bacterial envelope to reach their target, and thus hindering their activities *in vivo*. With the aid of state-of-the-art computational methods and tools, we developed a computational platform to automate and study the translocation of small molecule drugs across bacterial outer membrane proteins, with a goal of accelerating the antibiotic discovery process. We applied all-atom and coarse-grained molecular modeling, enhanced sampling techniques, and a parallel computing environment to maximize the performance. We further demonstrate the efficacy of this platform with a comprehensive study of a benchmark case. Key findings include free energy profile, translocation kinetics and thermodynamics, and molecular interactions between drug molecules and protein residues. Ultimately, this approach is designed to screen small molecule libraries with a fast turnaround time to yield structure-property relationships to discover antibiotics with high permeability. Furthermore, this work is expected to provide insights in inverse engineering and mutation design during drug development.

Development of Computational Antibiotic Screening Platform Across Bacterial Outer Membrane Proteins

Yinghui Dai

B.S., University of Rochester, 2014

M.S., University of California, Berkeley, 2015

Dissertation

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering

Syracuse University

July 2021

Acknowledgements

I would love to take this opportunity to thank many people for the help and support throughout my PhD journey.

I would first like to thank my advisor, Dr. Shikha Nangia, who has been constantly providing insightful guidance to my research work. Her deep passion towards research, keen eye for detail and willingness to help at all times have brought my work to the next level. She introduced me to the world of computational biophysics and since then broadened the scope of my visions. Dr. Nangia has taught me so much in both of my academic and personal life. She is truly a mentor, and I will always be grateful to be her student.

I wish to show my sincere appreciation to my committee members: Dr. Jesse Bond, Dr. Julie Hasenwinkel, Dr. Era Jain, Dr. Davoud Mozhdehi, and Dr. Dacheng Ren for their availability and participation. I feel extremely honored to present my work in front of them who have always dedicated to state-of-the-art research and are truly great scientists in their field of study. I would like to express my sincere gratitude to the Biomedical & Chemical Engineering Department at Syracuse University for all faculty and staff working hard to create opportunities and support us students.

I would like to thank Dr. Davoud Mozhdehi for our collaboration work. I have gained plenty of experience by working cross-functionally with his team. This is an extraordinary opportunity for me to bridge understanding between micro- to macro-scale molecular behavior.

I would like to thank our lab - Nangia research group for the wonderful members and working atmosphere. I thank Dr. Huilin Ma for mentoring me in the beginning and shaping my research plan. Dr. Nandhini Rajagopal offered me countless valuable help in troubleshooting

issues I encountered during work. I thank Kathryn Piston, Jingjing Ji, Allyson Karmazyn and all the graduate and undergraduate students that I have worked with. I enjoyed the learning experience by helping each other.

I would love to express my deepest gratitude to my dear husband, Dr. Jianli Zhang, who has always believed in me and provided me with as much support and encouragement as I needed. I always go to him for his invaluable suggestions when I am lost. I cannot be more grateful to have him in my life. I would like to give special thanks to my son, Peter, who truly is a blessing. I am motivated and fearless to thrive with him in my life. I would like to express my appreciation to my parents and my parents-in-law for their unconditional love and support. They have taught me to work hard for everything I aspire to accomplish.

Yinghui Dai

05/12/2021

Table of Contents

LIST OF FIGURES

CHAPTER 1.

INTRODUCTION

1.1 Background

Antibiotics are medicines used to treat bacterial infections by either killing the bacteria or slowing down their reproduction. Throughout the use of antibiotics, bacteria has developed different resistance mechanisms to fight against anti-infective drugs (Figure 1-1).[1] There are many defense strategies, including but not limited to, producing pumps to get rid of antibiotics, developing protein channels with low permeability to restrict access, destroying or modifying drugs with enzymes, mutating target structure to prevent the binding of antibiotics, and expressing alternative proteins required for cell functions.[1-2]

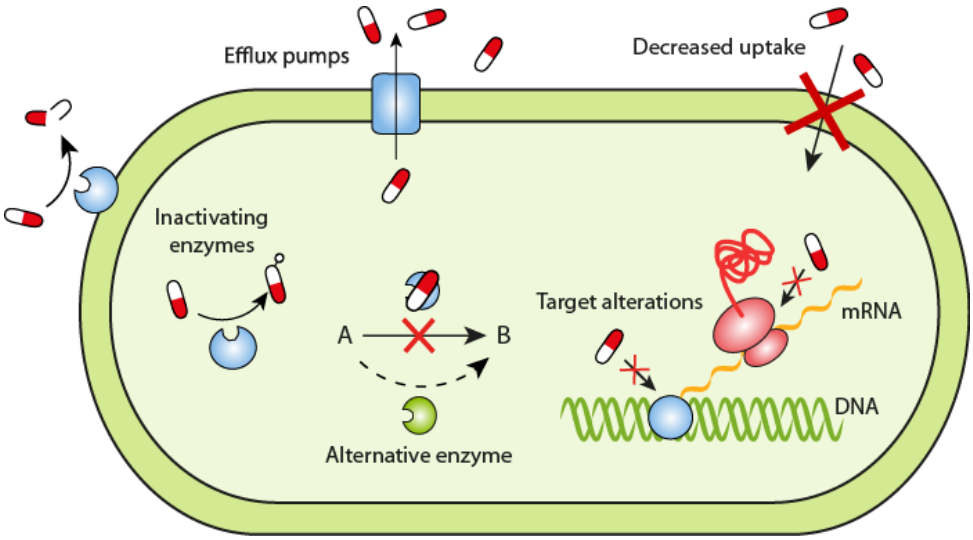


Figure 1-1 Antibiotic resistance strategies in bacteria

Antibiotic resistant germs are able to quickly spread everywhere considering the fact that they are very small living organisms and they constantly evolve at a staggering rate. A variety of factors have mixed and exacerbated the antimicrobial resistance. They spread within and across all settings in the community that are associated with the entire environment. Moreover, antibiotic resistance is rapidly spreading all over the world due to overprescription and misuse of drugs. For example, there is a large-scale of antibiotic consumption in agriculture and livestock farming, and

animals that carry drug resistant pathogens will consequently contaminate the food supply. The antibiotic resistance is endangering both individual and society. It has been reported that there are approximately 3 million people in the United States infected by antibiotic resistant pathogens every year, leading to more than 35000 deaths.[2-3]

In addition to the rapid spreading, stranded discovery pipeline has become exhausted. The drug discovery pipeline is at a virtual standstill due to limited profits and regulatory bottlenecks. It is estimated that over $1 billion is needed for developing an antibiotic.[3] As a result, many pharmaceutical companies have abandoned the antibiotic discovery market in favor of pursuing more profitable drug discovery line, such as cancer drugs. Infectious diseases are now on the verge of acquiring effective treatments using novel antibiotics. Antibiotic discovery requires revolutionary methodology to accelerate the process and reduce overall costs.

One of the bottlenecks in antibiotic discovery is to overcome the inability of drug candidates to penetrate through bacterial membrane to eventually reach their target.[3] Comparing to Gram-positive bacteria, Gram-negative bacteria has developed liposaccharide-rich, negatively charged outermembrane, which serves as a highly effective barrier against most small molecule drugs. In addition, a variety of size-selective, substrate-specific protein channels (or porins) are expressed, which are embedded in the outermembrane to further decrease the uptake (Figure 1-2).[4] Understanding the porin facilitated translocation of small molecules at a molecular scale is particularly imperative to drug discovery.

In our work, we focused on *Pseudomonas aeruginosa*, which is a notoriously multi-drug resistant Gram-negative bacterium. According to the 2019 Antibiotic Resistance Threats Report published on the Centers for Disease Control and Prevention website, *P. aeruginosa* was listed as one of the serious threats for its infections in the U.S.[5] *P. aeruginosa* has an extra outer membrane

as opposed to Gram-positive bacteria. The discovery of new antibiotics against especially Gram-negative bacteria is a major challenge, and primarily because of a low hit rate during screening of compound libraries. In the case of *P. aeruginosa*, the rate can be up to 1000-fold lower than in the case of Gram-positive bacteria, and the major reason for such a low hit rate is caused by the low permeability barrier of this extra outer membrane.[13] In addition, *P. aeruginosa* has developed many small-channel porins such that becoming intrinsically resistant towards most antibiotics.

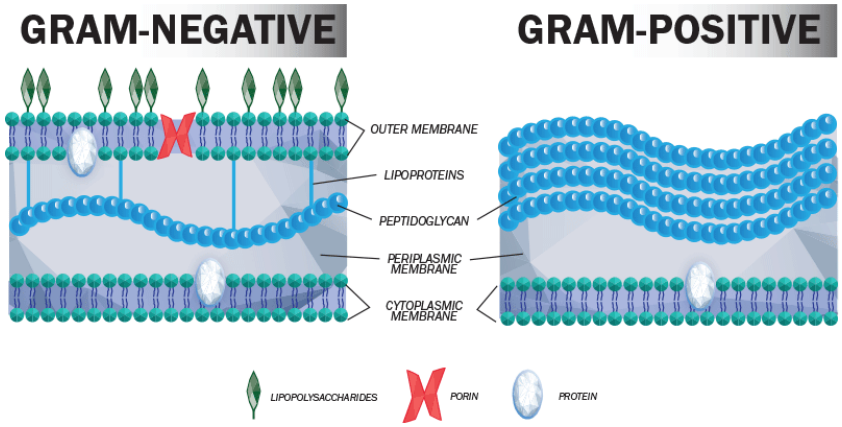


Figure 1-2 Cell wall structure in gram-negative vs. gram-positive bacteria

Among the *P. aeruginosa* porins, the most important is the Outer membrane Carboxylate Channel (Occ) family that are dedicated to the uptake of small, water-soluble, carboxylate containing solutes.[6] The Occ family comprise 2 subfamilies, OccD and OccK, which differ in pore size and structure dynamics. The OccD subfamily has 8 members and the OccK subfamily has 11 members as these structures have been experimentally resolved. The majority of them are available on Protein Data Bank database and are rendered in Figure 1-3. In general, the OccD porins have a smaller pore size compared to the ones from OccK subfamily with the exception of some OccD proteins showing similar sizes as OccK proteins, which suggests that these two subfamilies have different substrate selectivity. Besides the variation in pore size, it is reported that OccD structures

are more dynamic as they can stabilize in either closed or open state, whereas the OccK porins have a more rigid channel.[6-7]

As the first member of the Occ family, OccD1 porin has been studied extensively via both experimental and computational approaches. OccD1 is the least specific channel among the Occ family, and thus creating a less restricted access and allowing more efficient permeation of relatively larger molecules such as medium-chain fatty acids. In particular, it shows that small structural modifications of known antibiotics have caused significant impacts on their permeation through OccD1 porin.[8] Therefore, understanding the porin facilitated translocation behavior is of huge importance in drug candidate optimization. Interestingly, OccD3 is structurally similar as OccD1, which is a monomeric protein in β-barrel shape and serves as an uptake channel of arginine and dipeptide in *P. aeruginosa*. As recent works showed that OccD3 plays a key role in carbapenem translocation in *P. aeruginosa*, it attracts great attention in exploring OccD3 in-depth in the field of antibiotic resistance study.

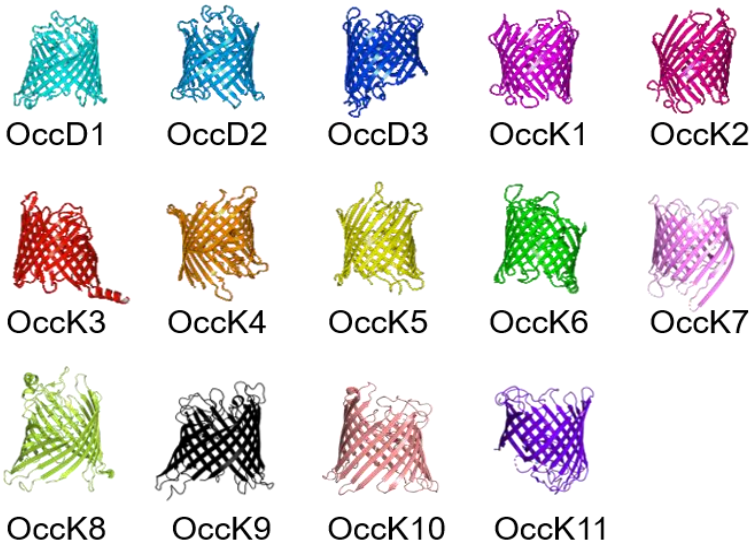


Figure 1-3 Occ porins in the order from left to right in cartoon representation: OccD1, OccD2, OccD3, OccK1, OccK2, OccK3, OccK4, OccK5, OccK6, OccK7, OccK8, OccK9, OccK10, and OccK11.

1.2 Computational Methods

Computational methods possess significant advantages in the study of understanding biological systems and processes, such as drug molecule-protein residue interactions, as they provide direct insights at the molecular level. The detailed system initiation, Molecular Dynamics simulation and umbrella sampling protocol, and python and shell scripting implemented scripts for analyzing simulation trajectories are distributed and discussed in-depth in future chapters. Here is a list of advanced computational methods and tools that have been used throughout my work.

1.2.1 Molecular Dynamics simulation

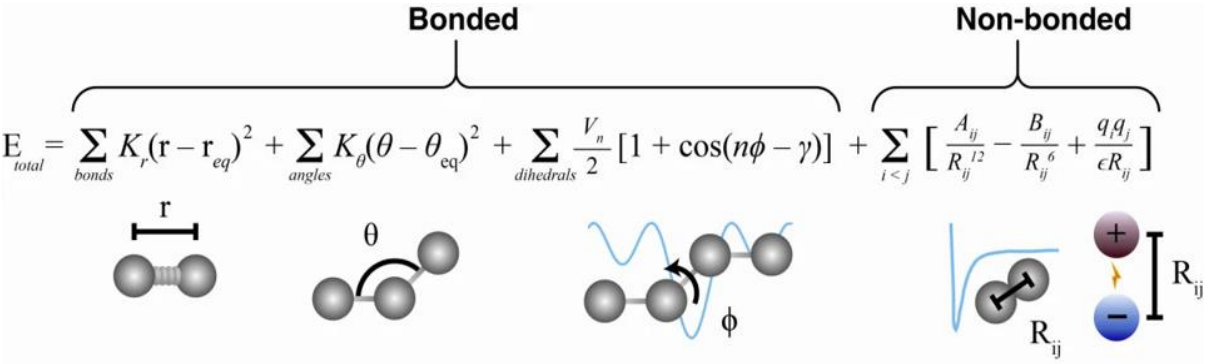


Figure 1-4 An example of a typical equation used to approximate the atomic forces that govern molecular movement

Molecular Dynamics (MD) is a computational simulation method that mimics atomic and molecular movements in real life by solving Newton's equations of motion. From a knowledge of the force acting on each atom, it is able to determine the acceleration of each atom in the system, and therefore obtaining its position and velocity over a period of time and predicting the future behavior of the system.

MD simulation is able to approximate the interaction energy of each atom within itself (bonded interaction) and with the surrounding atoms (non-bonded interaction). Figure 1-4

demonstrates a typical equation used in calculating total interaction energy by incorporating the two types, bonded and nonbonded interaction energy.[9] Both can be further broken down into specific terms: bonded interaction includes bond-stretch, angle-bend, and rotate-along-bond terms; non-bonded interaction comprises of van der Waal's and electrostatic components.

With the aid of a number of well parameterized force fields, we are able to model and simulate a wide variety of biomolecules, including amino acids, proteins, lipids, organic compounds, polymers, and solvents. We conducted MD simulations by using a powerful MD engine, GROMACS software suite, which has grown out of GROMOS force field and supports the majority of other well-formulated force fields.

1.2.2 Multiscale modeling

In our research, we have adopted multiscale MD modeling, which combines the all-atom (AA) and the coarse-grained (CG) resolutions. In atomistic simulations, each atom is calculated to determine its position and velocity over a period of time. Among many available force fields, we often use CHARMM force field when we need to perform atomistic equilibrations before moving forward to coarse-grained simulations. The CHARMM potential energy function is applicable to a variety of atoms, including but not limited to carbon, oxygen, nitrogen, sulfur, and phosphorous, and therefore, has been widely used in simulating complex biomolecules, e.g., protein-membrane systems.[14]

Coarse grain method groups several atoms into one bead to decrease the amount of calculation while preserving the molecular properties. Therefore, coarse grained resolution is able to enlarge the simulation system size and extend simulation timescale comparing to the atomistic precision. We apply MARTINI approach for coarse graining, which has the advantage of

transferability of the force field parameters to a wide range of biomolecules and on-demand reverse mapping to atomistic resolution.[10] MARTINI force field maps four adjoining heavy atoms into one CG bead, in the meantime it is also flexible to provide three-to-one or five-to-one mapping scheme if required. The beads are distinguished by their charges, polarities, and structures.

1.2.3 Enhanced sampling

The bottleneck of traditional MD simulation is that there are high-energy barriers separating two regions of phase or configuration space, which can result in poor sampling as transitions between them can take a long time to or rarely happen. It is necessary to overcome the energy barrier to save both simulation time and computational costs. Among many advanced techniques that have been developed over the past few decades, umbrella sampling has been widely used to achieve enhanced sampling.[11] It works by decreasing the energy barrier between different structures or conformations so that the transition between these regions is more accessible and hence, enhancing the sampling.

Umbrella sampling is the first collective variable (or reaction coordinate) based sampling methods. In umbrella sampling simulations, the collective variable space is divided into many windows, and time-independent bias potentials are added to each window to obtain the umbrella potential and estimate the equilibrium probability distribution. Next, the probability distribution from all windows are calculated iteratively and concatenated using the well-established weighted histogram analysis method (WHAM).[12] Finally, with the total equilibrium probability distribution constructed, free energies as a function of the collective variable can be calculated.

Umbrella sampling is the fundamental machinery behind our computational method, making the MD simulation more powerful and robust.

1.3 References

1. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature*. **2000**, *406***,** 775–781
2. Antibiotic resistance threats in the united states. *Centers for Disease Control and Prevention*. **2019**
3. Ribeiro da Cunha, B., Fonseca, L. P., & Calado, C. Antibiotic Discovery: Where Have We Come from, Where Do We Go?. *Antibiotics (Basel, Switzerland)*. **2019,** 8(2), 45
4. Gram-Positive vs. Gram-Negative. *biologydictionary.net*. **2021**
5. Biggest Threats and Data. *Centers for Disease Control and Prevention*. **2019**
6. Eren E, Vijayaraghavan J, Liu J, Cheneke BR, Touw DS, et al. Substrate Specificity within a Family of Outer Membrane Carboxylate Channels. *PLOS Biology*. **2012**, 10(1): e1001242
7. Sylvie Chevalier, Emeline Bouffartigues, Josselin Bodilis, Olivier Maillot, Olivier Lesouhaitier, Marc G. J. Feuilloley, Nicole Orange, Alain Dufour, Pierre Cornelis. Structure, function and regulation of *Pseudomonas aeruginosa* porins, *FEMS Microbiology Reviews*, Volume 41, Issue 5, September **2017**, 698–722
8. Iyer, R., Sylvester, M. A., Velez-Vega, C., Tommasi, R., Durand-Reville, T. F., and Miller, A. A. Whole-Cell-Based Assay To Evaluate Structure Permeation Relationships for Carbapenem Passage through the Pseudomonas aeruginosa Porin OprD. *ACS Infect. Dis*. **2017**, 3 (4), 310−319
9. Durrant, J.D., McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol*, **2011**, 9, 71

10. Wassenaar, T. A.; Pluhackova, K.; Bockmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *Journal of Chemical Theory and Computation* **2014,** *10* (2), 676-690
11. Yi Isaac Yang, Qiang Shao, Jun Zhang, Lijiang Yang, and Yi Qin Gao. Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics*, **2019**, 151, 070902
12. Davide Bochicchio, Emanuele Panizon, Riccardo Ferrando, Luca Monticelli, and Giulia Rossi. Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *The Journal of Chemical Physics*, **2015**, 143, 144108
13. Fair RJ, Tor Y. Antibiotics and bacterial resistance in the 21st century. *Perspect Medicin Chem*, **2014**, 6, 25-64
14. Judith A. Harrison, J. David Schall, Sabina Maskey, Paul T. Mikulski, M. Todd Knippenberg, and Brian H. Morrow. Review of force fields and intermolecular potentials used in atomistic computational materials research. *Applied Physics Reviews*, **2018**, 5, 031104

CHAPTER 2.

DEVELOPMENT OF THE COMPUTATIONAL ANTIBIOTIC SCREENING PLATFORM (CLASP) TO AID IN THE DISCOVERY OF NEW ANTIBIOTICS

Reprint with permission from "Dai, Y.; Ma, H.; Wu, M.; Welsch, T. A.; Vora, S. R.; Ren, D.; Nangia, S. Development of the computational antibiotic screening platform (CLASP) to aid in the discovery of new antibiotics. Soft Matter, 2021, 17, 2725-2736.
DOI: 10.1039/D0SM02035D" Copyright The Royal Society of Chemistry 2021.
Author Contributions: Y. D. and H. M. contributed equally to this article.

2.1 Abstract

Bacterial colonization of biotic and abiotic surfaces and antibiotic resistance are grand challenges with paramount societal impacts. However, in the face of increasing bacterial resistance to all known antibiotics, efforts to discover new classes of antibiotics have languished, creating an urgent need to accelerate the antibiotic discovery pipeline. A major deterrent in the discovering of new antibiotics is the limited permeability of molecules across the bacterial envelope. Notably, the Gram-negative bacteria have nutrient specific protein channels (or porins) that restrict the permeability of non-essential molecules, including antibiotics. Here, we have developed the ComputationaL Antibiotic Screening Platform (CLASP) for screening of potential drug molecules through the porins. The CLASP takes advantage of coarse grain (CG) resolution, advanced sampling techniques, and a parallel computing environment to maximize its performance. The CLASP yields comprehensive thermodynamic and kinetic output data of a potential drug molecule within a few hours of wall-clock time. Its output includes the potential of mean force profile, energy barrier, the rate constant, and contact analysis of the molecule with the pore-lining residues, and the orientational analysis of the molecule in the porin channel. In our first CLASP application, we report the transport properties of six carbapenem antibiotics—biapenem, doripenem, ertapenem, imipenem, meropenem, and panipenem—through OccD3, a major channel for carbapenem uptake in Pseudomonas aeruginosa. The CLASP is designed to screen small molecule libraries with a fast turnaround time to yield structure-property relationships to discover antibiotics with high permeability. The CLASP will be freely distributed to enable accelerated antibiotic drug discovery.

2.2 Introduction

Antibiotic resistant bacteria are endangering the efficacy of known antibiotics, which have saved millions of lives since the middle of the last century.[2-5] In the past few decades, antibiotics have become the most prescribed drugs in human medicine. However, over prescription and misuse of antibiotics have led to the upsurge in bacterial resistance.[6] To make matters worse, the drug discovery pipeline is at a virtual standstill due to limited profits and regulatory bottlenecks.[7] The development of a new antibiotic is a prolonged process requiring years of drug screening, *in vitro* testing, clinical trials, and billions of dollars in investment. Therefore, there is an urgent need to accelerate the screening of drug candidates and develop new antibiotics.[8-11]

A bottleneck in a drug molecule's penetration into a bacterial cell is the multilayered bacterial envelope.[12] Both Gram-positive and Gram-negative bacteria have evolved a sophisticated and complex cell wall that protects them from hostile chemical environments. For Gram-negative bacteria in particular, their liposaccharide-rich, negatively charged outer membranes with narrow, water-filled protein channels (or porins) allow selective uptake of nutrients, while preventing the passage of toxins into the cell.[13-14] The chemical nature (lipophilic versus hydrophilic) of the antibiotic molecule determines whether the permeation occurs via the lipid-mediated or the porin-mediated pathway.[14]

A critical step in antibiotic discovery is to identify chemical compounds with high permeability through Gram-negative bacterial outer membranes. The bacterial outer membrane porins are passive transporters of water-soluble molecules with a size limit of 600-700 Da.[14, 15-16] Besides mediating uptake of nutrients for survival, porins fortuitously allow the diffusion of antibiotics across the outer membrane.[17-18] For example, β-lactams, including cephalosporins,

penicillins, and carbapenems, are known to penetrate the outer membrane through the porins.[12] These small molecule portals make the bacteria more susceptible to new classes of antibiotics.

Understanding the porin-mediated uptake of small molecules at a molecular-level can spur the identification of natural compounds or aid in the rational design of new classes of antibiotics. Screening for penetration experimentally is challenging due to the low-throughput nature of related procedures. The use of computational approaches can provide molecular insights that are undetectable via currently available experimental methods.[19-25] For example, advanced simulation approaches, such as metadynamics,[21,26-27] steered molecular dynamics,[20,22] umbrella sampling,[28-29] and accelerated molecular dynamics simulations[30] have been used to simulate translocation through bacterial porins. However, these and other computational methods do not provide high-throughput thermodynamic and kinetic data of uptake of substrate molecules, precluding the high-throughput screening of libraries of natural compounds as possible antibiotics.

To advance screening of molecular libraries for the identification of potential antibiotics, in this work, we report the development of the Computational Antibiotic Screening Platform (CLASP). The CLASP takes advantage of coarse-grained (CG) resolution, advanced sampling techniques, and a parallel computing environment to maximize performance. The CLASP yields comprehensive thermodynamic and kinetic data of a solute's transport through a bacterial outer membrane porin. Its output includes the potential of mean force profile, energy barrier, rate constant of the molecule's transport, contact analysis of the molecule with the pore-lining residues, and the orientational analysis of the molecule in the porin channel. The CLASP automated workflow yields result in less than two hours of wall-clock time. The output turnaround times can be further improved with an increase in computer resources.

Our first CLASP application focuses on antibiotic uptake by *Pseudomonas aeruginosa*, a multi-drug resistant Gram-negative bacterium. Compared to other common Gram-negative bacterial species, the membrane permeability of *P. aeruginosa* is about one order of magnitude lower, primarily due to the presence of substrate-specific porins with sub-nanometer diameters.[31-32] As a result, the development of new drugs to treat infections caused by *P. aeruginosa* has been a major challenge.[11] In the 2019 Antibiotic Resistance report, published by the Centers for Disease Control and Prevention, *P. aeruginosa* was identified as a serious threat for infections in the United States.[33]

Among the *P. aeruginosa* porins, the most important is the Outer membrane Carboxylate Channel (Occ) family that are dedicated to the uptake of small, water-soluble, carboxylate containing solutes. In recent years, several members of the Occ family have been studied, including OccD1 and OccD3 that facilitate permeation of arginine, dipeptides, and carbapenem antibiotics.[34-35] As the first member of the Occ family, OccD1 porin has been studied extensively via experimental[23, 36-39] and computational[20-22] approaches. The closely related OccD1 ortholog, OccD3 has also been implicated as a channel for carbapenem uptake.[34-35, 40] We focused our investigation on OccD3, which has high structural similarity and higher ion conductance than OccD1.[35] The OccD3 is a β-barrel porin formed by 18 β-strands that are connected by large extracellular loops and short turns.[40] The porin's lumen is hourglass-shaped, which has a 3.7 Å diameter bottleneck region (Figure 2-1).

We investigated six members of the carbapenem family of antibiotics—biapenem, doripenem, ertapenem, imipenem, meropenem, and panipenem. The carbapenem family is one of the limited classes of antibiotics that can treat *P. aeruginosa* infections.[41-45] The carbapenems are

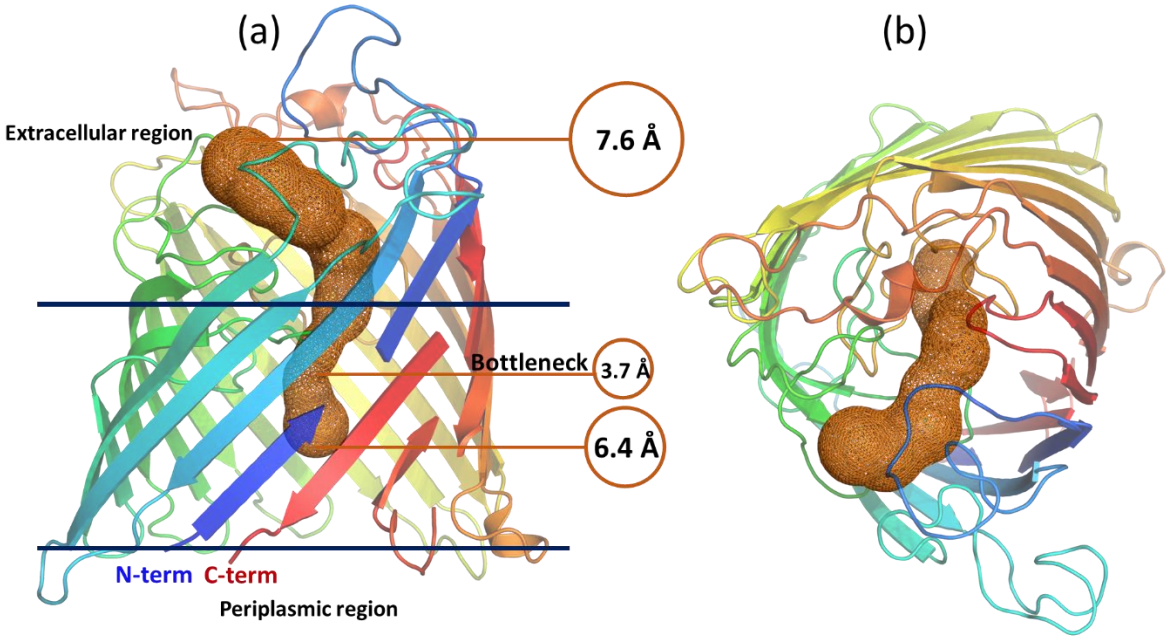


Figure 2-1. Molecular structure of OccD3 porin. The (a) side-view and (b) top-view of OccD3 in cartoon representation (colored blue to red from N-term to C-term) along with the porin channel (in orange surface representation). The diameter of the channel is marked in circles. The two horizontal lines show the membrane embedded region of the porin.

effective because they have a lower propensity to enzymatic degradation inside the bacterial cell, and are therefore, commonly prescribed for the treatment of infections caused by *P. aeruginosa*.[45] Understanding the mechanism of carbapenems' transport and is imperative to develop new drugs and reduce the threat of *P. aeruginosa*.

The results provide the details of the CLASP application and comprehensive analysis of the translocation profiles of the six carbapenems through OccD3. The free energy profiles and the energy barriers for carbapenem translocation are consistent with previous reports in the literature. Besides, we computed the contact frequencies between the carbapenem molecules and each OccD3 protein residue, which identified the key pore-lining residues and revealed the bottleneck

conformation of the carbapenem molecule during translocation. Additionally, we reverse mapped the CG structures of carbapenems to atomistic resolution in the bottleneck region to further illustrate the role of pore-lining residues in the translocation process. Finally, guided by the high-frequency contacts, we mutated two bulky pore-lining residues Y217A and F334A, to create OccD3m. The effect of mutations in OccD3m provided vital insight into the mechanism of carbapenem translocation.

2.3 CLASP Workflow

Molecular understanding of the transport of drugs through biological barriers is fundamental to the future of biomedicine. There is a need to provide robust thermodynamic and kinetic data for the permeation of drugs that spans nano- to macro-scale. The key is to provide mechanistic, structural, orientational, conformational, and mutational details of the drug permeation process that can inform the development of new drugs and accelerate the sluggish drug discovery pipeline. To cover the spatial and temporal scales involved in the drug transport, we have adopted a multiscale approach that takes advantage of different resolutions. We use the state-of-the-art umbrella sampling method to compute the potential of mean force (PMF) of permeation of drugs through bacterial porins.

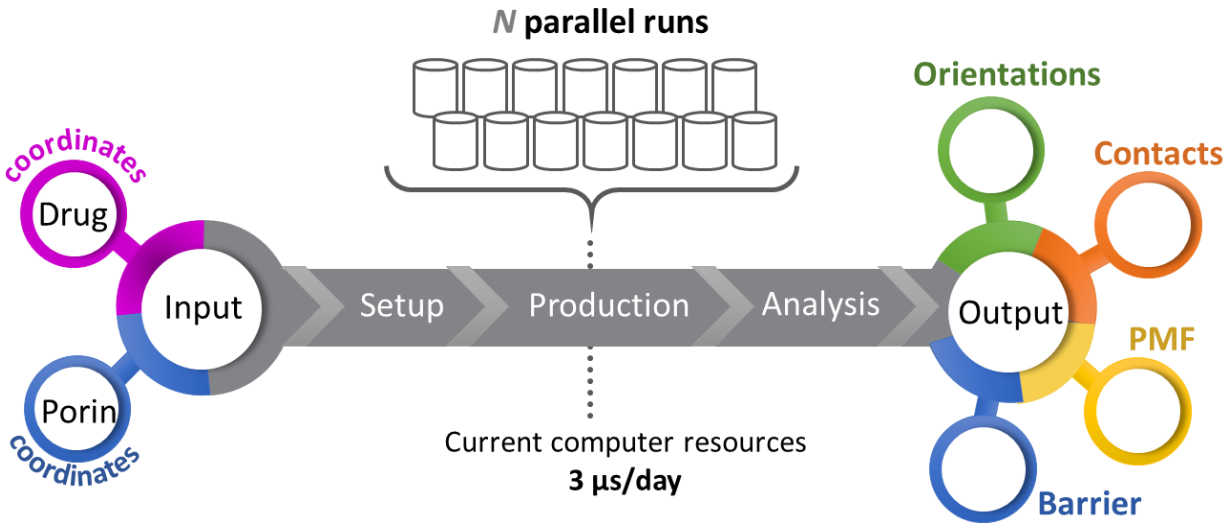


Figure 2-2. The CLASP workflow. The systematic representation of CLASP inputs and outputs. The porin and solute coordinates are the required inputs, which then are processed by the C_setup.py python script to generate *N* parallel umbrella sampling runs. The post simulation analysis scripts combine the *N* trajectories to generate the free energy profile, solute-protein residue contact map, and solute orientational analysis.

Umbrella sampling (US) simulations[46] are often laborious and typically require frequent user interventions at multiple steps. To streamline US simulations by eliminating the need for manual interventions, CLASP employs an efficient algorithm for the permeation of small molecules through bacterial porins. To maintain fast turnaround time in the current implementation of CLASP algorithm, we adopted coarse-grain (CG) resolution. Coarse-grain representation affords larger system sizes and longer timescale simulations by reducing the slow degree of freedom essential in all-atom simulations. Mapping the atomistic system to CG allows use of a time step that is one order of magnitude larger; plus, the CG resolution reduces the number of particles to be tracked in a simulation by at least one additional order of magnitude. Therefore, CG resolution is often two orders of magnitude more efficient than atomistic systems. The loss of atomic resolution in the CG simulations can be recovered by reverse mapping the output to atomistic representation.

Among the several CG approaches available, MARTINI force field has been used for a diverse set of biomolecules, including lipids, proteins, solvents, and ions.[19, 47-54] In our previous work, we developed MARTINI force field parameters for outer membrane lipids for several pathogenic and nonpathogenic Gram-negative bacterial species.[19,53-54] In CLASP, we retained the molecular descriptors of the proteins and substrate (here the carbapenems) at the level of MARTINI CG that recommends mapping four adjoining nonhydrogen atoms to one CG bead. This approach, popularly employed in biomolecular simulations, preserves the bonded and nonbonded interactions of the atomistic system and circumvents oversimplification of the molecular properties. The MARTINI CG has the advantage of transferability of the force field parameters to a wide range of biomolecules and on-demand reverse mapping to atomistic resolution.[52] In addition, we employed the PyCGTOOL,[55] which semi-automates the parametrization process for obtaining CG models for small molecules.

The CLASP implementation is compatible with the GROMACS 5.0 simulation package.[56] The program uses a series of python scripts for building the input files, setting up parallel simulation runs, and performing post-simulation analysis. The workflow enables highly-automated simulations (Figure 2-2). The CLASP workflow and implementation used are described as follows.

CLASP input. The CLASP requires two inputs, the three-dimensional structure and coordinates of the bacterial outer membrane protein and the potential antibiotic or the solute molecule.

Solute molecule. The MARTINI force field parameters from the literature will be adopted for small molecules when available. Otherwise, we recommend employing the PyCGTOOL.[55] This python script generates the molecule's CG model parameters based on its atomistic dynamical

data. The automated script provides molecular topologies and CG parameters with high chemical fidelity to the atomistic structure.

Bacterial outer membrane protein. The structure of the OccD3 bacterial protein is obtained from the Protein Data Bank. Missing residues are built via homology modeling using the standalone YASARA software[57] or online SWISS-MODEL server.[58] Next, the protein is embedded in a bacterial membrane and thermally equilibrated in the all-atom representation using standard MD packages or the CHARMM-GUI server.[59] The equilibrated protein is then converted to CG resolution using the MARTINI approach.

CLASP set up and production. In the setup stage, all systems are in the CG representation. The CG protein is embedded in CG bacterial lipids using Bacterial Outermembrane

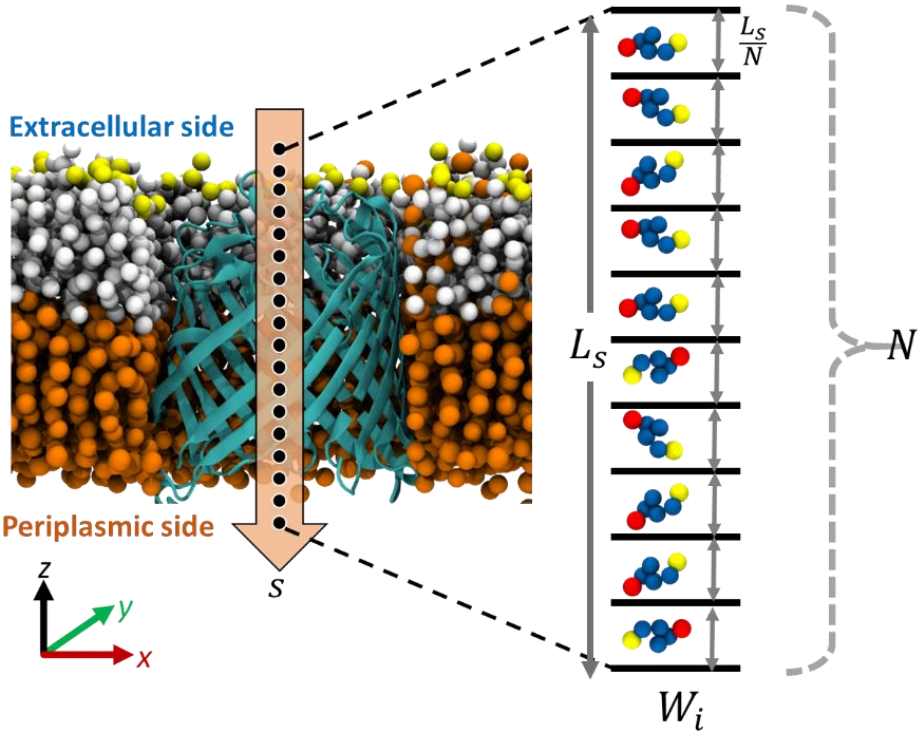


Figure 2-3. The CLASP simulation setup. The membrane embedded porin channel oriented parallel to the z-axis is divided into N equidistant umbrella sampling windows. The probe molecule is inserted into each W_i window.

Builder script (*BOB.py*),[53-54] an in-house modified *insane.py* script[60] that includes CG parameters for bacterial outer membranes. The protein is embedded in an asymmetric patch of the outer membrane lipids and is solvated with standard MARTINI water. The system's energy is then minimized, followed by equilibration in the isothermal-isochoric (*NVT*) and isothermal-isobaric (*NPT*) conditions. The output of the *NPT* run is the starting structure for CLASP.

In the next step, bacterial membrane is generated in the *xy*-plane and the porin is embedded in the membrane such that the porin channel is along the *z*-axis. The molecule's permeation coordinate (s) is parallel to the z-axis, and the length of the porin channel from the extracellular region to the periplasmic region along s is denoted by L_s (Figure 2-3). The free energy profile along s is generated by dividing L_s into N umbrella sampling windows (W_i) with uniform spacing of L_s/N nm. The solute molecule is inserted in each W_i window, where the insertion coordinates are computed relative to the center-of-mass (COM) of the porin. Given, the hourglass shape of the porin, the COM lies in the middle of the porin channel. The location of the COM is considered $W_{N/2}$, with an equal number of windows on either side.

Next, the CLASP script generates N separate folders, each with a unique location of the substrate molecule defined within W_i, totaling the length L_s along s. Each folder is set up to execute energy minimization and US simulations independently. The N jobs run concurrently on separate computer cluster nodes for maximum efficiency and shortest completion time.

CLASP analysis. The CLASP outputs are four primary outputs from a CLASP run:

Free energy profile. The potential of mean force (PMF) profile of the solute molecule's transport from the extracellular to the periplasmic region is computed along the one-dimensional translocation coordinate (s). The *C_PMF.py* script combines the trajectories from the N folders to

generate the PMF profile of the solute molecule. The PMF average profile with standard deviations were obtained using a GROMACS built-in utility. The detailed usage of the script is provided in the Supporting Information.

Permeation barrier. the barrier height (*BH*) for the solute's permeation is calculated by taking the difference between the maximum and the minimum energy points along s. Using the transition state theory as a first-order approximation of the solute translocation, we defined rate constant as $k = (k_B T/h) exp^{(-BH/RT)}$, where k_B is the Boltzmann constant, *R* is the gas constant, and *T* is the temperature.

Orientational analysis. The orientation of the solute molecule is quantified by defining the interatomic vector connecting head bead to the tail bead of the solute molecule, $\vec{r} = \overrightarrow{r_T} - \overrightarrow{r_H}$. Based on the angle θ that the unit vector $\hat{d} = \frac{\vec{r}}{|r|}$ makes with the z-axis, we used $d_z = |d| \cos\theta$ to determine in the orientation of the molecule, where $d_z \in [-1, 1]$. If $d_z = 1$, the molecule is orientated headfirst towards the translocation direction, and $d_z = -1$ for tail first. If $d_z = 0$, then the molecule lies in the *xy*-plane, perpendicular to the *z*-axis.

Contact analysis. The interaction of the solute molecule with the protein residues is performed for all N windows. A contact is defined to have formed between the solute and protein residue if the distance between any of their beads is within a 1.2 nm cut off. A contact receives a value of either 1 when formed or 0 when the contact is broken. The cumulative number of contacts made by the solute with each protein residue in N windows is calculated. The residues are sorted from smallest to largest based on the total number of residue-solute contacts to identify the highest contacting residues (99^{th} percentile).

Reverse mapping. The on-demand CG to atomistic reverse mapping is available to determine the atomistic level interactions of the solute molecule with the pore-lining residues.[52]

2.4 Methods

Carbapenem coarse graining. The MARTINI CG parameters for the carbapenem molecules were developed based on many-to-one mapping. PyCGTOOL was used to generate the CG coordinates and topologies for all solute molecules. Benchmarking of the CG parameters against the atomistic models was performed using the PyCGTOOL. The CG mapping of the carbapenem molecules are shown in Figure 2-4. The force field parameters of all six carbapenems are provided in Tables 2-2 – 2-7.

Porin structure. The x-ray crystallographic structure of OccD3 was obtained from the Protein Data Bank (PDB ID: 3syb), and the missing residues were built using the SWISS-MODEL server. The structure was then optimized in atomistic representation using the CHARMM36 force field.[61] The *NVT* and *NPT* simulations were performed for 200 ns each at T = 320 K. The point mutations in the OccD3 structure were generated using the CHARMM-GUI webserver.[59] The mutated structures were then optimized using the same approach outlined above for the OccD3.

Porin-membrane system. The CG mapping of the OccD3 was performed using *martinize.py* script. The MARTINIv2.1 force field parameters[50] were used along with the ELNeDyn network[49] constraint to maintain the porin's secondary structure. The porin was inserted in a 10×10 nm^2 patch of the *P. aeruginosa* membrane using the *BOB.py* script. The membrane's outer leaflet consisted of *P. aeruginosa*'s lipid A and 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (DPPE) in 9:1 ratio and pure DPPE in the lower leaflet. The membrane was placed in a 10×10×12 nm^3 simulation box and solvated with MARTINI water (9:1 of W:WF) and 150 mM $CaCl_2$. The molecular composition of the simulation box is provided in Table 2-8.

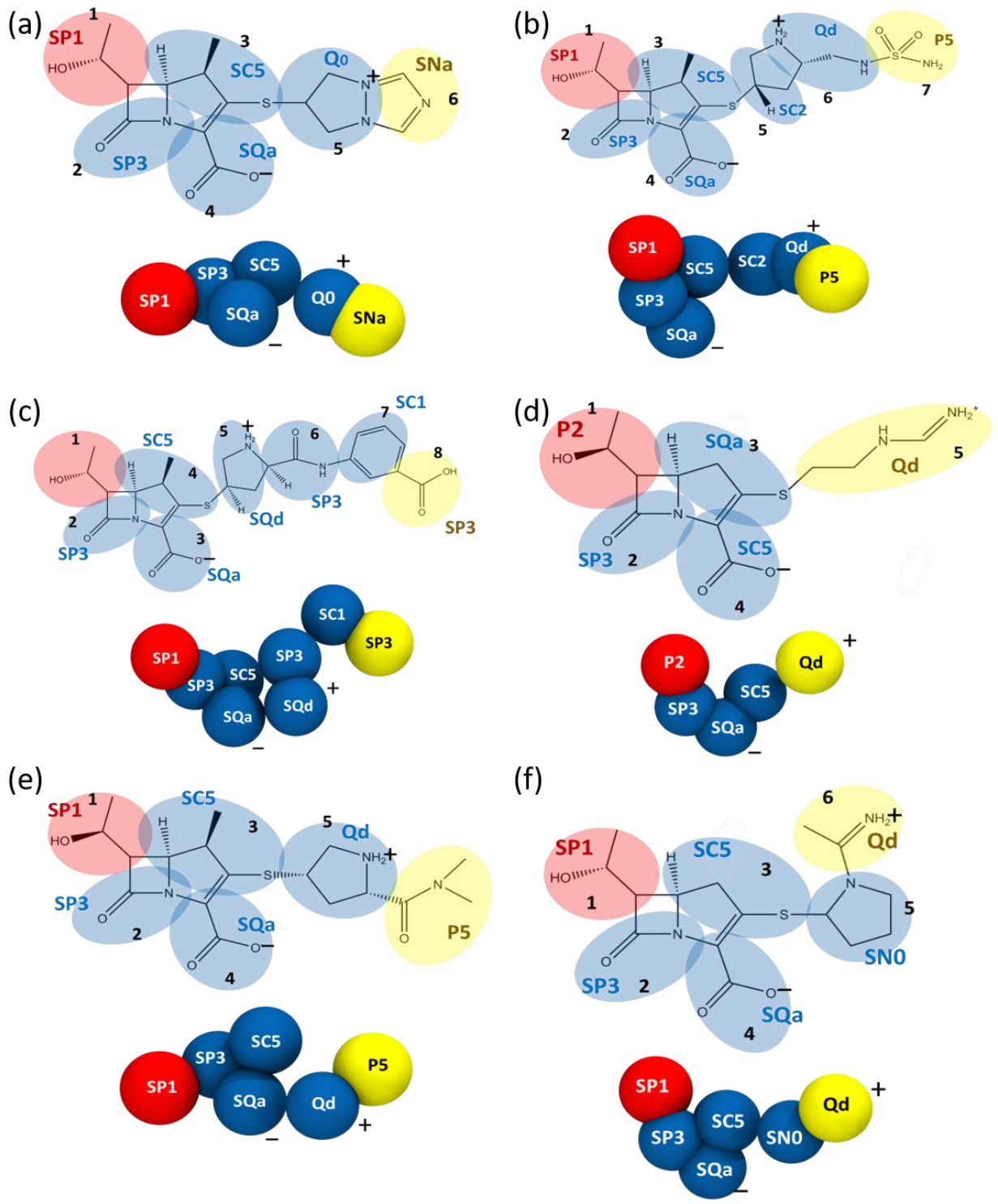


Figure 2-4. Atomistic to CG mapping for (a) biapenem, (b) doripenem, (c) ertapenem, (d) imipenem, (e) meropenem, and (f) panipenem showing the MARTINI bead type assignments along with positive and negative charges. The red bead, which is part of the β-lactam ring, denotes the head and the yellow bead, the farthest bead from the β-lactam ring, denotes the tail.

CG MD Simulation. The GROMACS 5.1.2 package[56] was used for CG MD simulations. The system was energy minimized using the steepest-decent algorithm until the maximum force on any bead was below the tolerance parameter of 10 $kJmol^{-1}nm^{-1}$. This was followed by *NVT* and *NPT* equilibration for 0.02 μs and 2 μs, respectively. A 20 fs time step was used for the equilibration and production runs. The temperature was set to 320 K for all systems using the v-rescale thermostat with τ_t = 1 ps. A semi-isotropic pressure coupling of 1 bar was maintained using Berendsen barostat[62] with τ_p = 4 ps. Both the nonbonded van der Waals and electrostatics interaction cut offs were set to 1.2 nm. Periodic boundary conditions were applied in all three dimensions.

CLASP production. A total of *N* independent simulations were perfomed concurrently on separate computer nodes (Table 2-9). Within each window the simulations were performed in two steps. The first US run was performed for 0.2 μs using a 20 fs time step. The OccD3 protein and the solute molecule were position-restrained with a 1000 $kJmol^{-1}nm^{-2}$ force constant. In the second step, the US simulation was run for 1 μs with a 20 fs time step. The position restraint of the drug molecule was removed in this step with all other parameters maintained. The harmonic potential with a force constant of 3000 $kJmol^{-1}nm^{-2}$ was used. A cut off of 1.1 nm was used for both the long-range electrostatic and the nonbonded van der Waals interactions. The potential-shift-Verlet algorithm was applied to shift the van der Waals interactions beyond the cut off. The Coulombic interactions were calculated using the reaction-field algorithm. The temperature was maintained at 300 K using the v-rescale thermostat with τ_t = 1.0 ps.

Analysis. Post simulation analyses were performed using CLASP python scripts. We computed the histograms of the US window along the translocation path (Figures S1-S6). The

CAVER plugin[63] available in PyMOL was used for OccD3 channel analysis. Molecular visualization and graphics were generated using VMD,[64] PyMOL,[65] and YASARA.[57]

2.5 Results and Discussion

The CLASP simulations were employed to characterize and compare the permeabilities of six carbapenems— biapenem, doripenem, ertapenem, imipenem, meropenem, and panipenem— through the wildtype OccD3 porin. The role of pore-lining residues involved in carbapenem translocation was further evaluated by site-directed mutagenesis of Y217 and F334 residues to alanine in the double mutant OccD3 (OccD3m) porin. The results provided detailed insight into the uptake mechanism and demonstrated the robustness of the CLASP method in capturing the effects of point mutations.

2.5.1 Carbapenem uptake through OccD3 channel with characteristic PMF

The PMF profile of each carbapenem molecule's transport from the extracellular to the periplasmic region of the OccD3 porin was studied in the $s \in [0, 11]$ nm range. For clarity, in this discussion, the s coordinate is subdivided into three sections—the extracellular region with the porin vestibule, $s \in [0 - 4.2]$ nm; the porin constriction region, $s \in [4.2 - 7.4]$ nm; and the periplasmic region, $s \in [7.4 - 11.0]$ nm.

In the extracellular region, the carbapenem molecule interacts freely with the porin loops and shows an energy minimum at the pore vestibule $s = 4.2$ nm. Each carbapenem is different in molecular size (Table 2-1) and functional groups attached to the β-lactam ring (Tables 2-2 – 2-7), so free energy variations among the carbapenems are expected; however, stabilization before penetrating the porin channel is observed in all six cases (Figure 2-5). The free energy profile of all six carbapenems except doripenem show approximately 10 kcal mol^{-1} stabilization at $s = 4.5$

nm. Beyond the extracellular vestibule, the free energy increases for all carbapenems until they reach the bottleneck.

Table 2-1 Molecular weight (MW), barrier height (BH), and rate constant (k) of six carbapenems through OccD3 and OccD3m.

Antibiotic	MW (Da)	OccD3		OccD3m	
		BH (kcal mol^{-1})	*k* (s^{-1})	*BH* (kcal mol^{-1})	*k* (s^{-1})
Biapenem	350.4	27±2.5	1.4×10^{-7}	15±2.1	7.6×10^{1}
Doripenem	420.5	20±3.2	1.7×10^{-2}	16±2.2	1.4×10^{1}
Ertapenem	474.5	31±2.8	6.0×10^{-12}	10±2.5	2.1×10^{-5}
Imipenem	299.4	22±2.1	1.1×10^{-4}	16±1.6	1.4×10^{1}
Meropenem	383.5	21±1.8	3.3×10^{-3}	14±1.6	4.1×10^{2}
Panipenem	339.4	26±2.3	7.5×10^{-7}	12±2.2	1.2×10^{4}

A comparison of the free energy barrier for translocation through the OccD3 pore shows a range between 20-31 kcal mol^{-1} (Table 2-1). In OccD1, the translocation barriers for positively charged amino acid substrates vary between 5-10 kcal mol^{-1}.[21] Energy barriers are three to four times higher for permeation through OccD3 than OccD1, which is expected because OccD3 has more constrained channel and contains a unique N-terminal extension loop.[34]

We observe a weak correlation between the magnitude of the energy barrier and the molecular size of the carbapenem molecule. Ertapenem is the largest carbapenem and has the highest *BH* = 31±2.8 kcal mol^{-1}. Imipenem, which is the smallest carbapenem, has one of the smallest *BH* = 22±2.1 kcal mol^{-1}. The similarly sized, biapenem and panipenem, have energy barriers of *BH* = 27±2.5 and 26±2.3 kcal mol^{-1}, respectively. Meanwhile, doripenem, which is the second largest in terms of MW, has the lowest *BH* = 20±3.2 kcal mol^{-1}. The results indicate that

the molecular size of the carbapenems is not the only criteria that determines the magnitude of the energy barrier. In a previous study, Soundararajan *et al*., reported that meropenem interacts more strongly with the OccD3 porin than imipenem using ion conductance measurements[35]; however Table 1 shows that both molecules have similar barrier heights. To investigate the differences in energy barriers, we evaluated the interactions of carbapenem molecule throughout the porin channel with emphasis on the bottleneck region. A detailed discussion of the carbapenem conformation at the bottleneck region of the pore is provided in Section 2.5.3.

The rate constant (s^{-1}) of the six carbapenems are shown in Table 2-1. Rate constants are calculated by assuming a first order transition state theory and are directly related to energy barriers. They are used to distinguish the relative rate and preference of carbapenems permeating through OccD3. Between ertapenem (*BH* = 31±2.8 kcal mol^{-1}) and doripenem (*BH* = 20±3.2 kcal mol^{-1}), which has the largest and smallest energy barriers, the rate constants are different by multiple orders of magnitudes. We further compare the rate constants of mutations, which will be discussed in Section 2.5.4.

It is apparent from the thermodynamic that besides the molecular size of the translocating molecule, the orientation of the molecule in pore's constriction region can also be crucial. For this purpose, we evaluated the orientation of the carbapenem molecule along $s \in [0 - 11.0]$ nm. The orientation profiles of the six carbapenems were somewhat similar in the extracellular and periplasmic regions. In all cases, the carbapenem adopted a wide range of orientations, indicated by the large deviation in the average d_z values in both the extracellular region $s \in [0 - 4.2]$ nm and the periplasmic region, $s \in [7.4 - 11.0]$ nm (Figure 2-6). However, in the constriction zone, the solute molecule is restrained by the channel walls, which was manifest in small deviations in the d_z values.

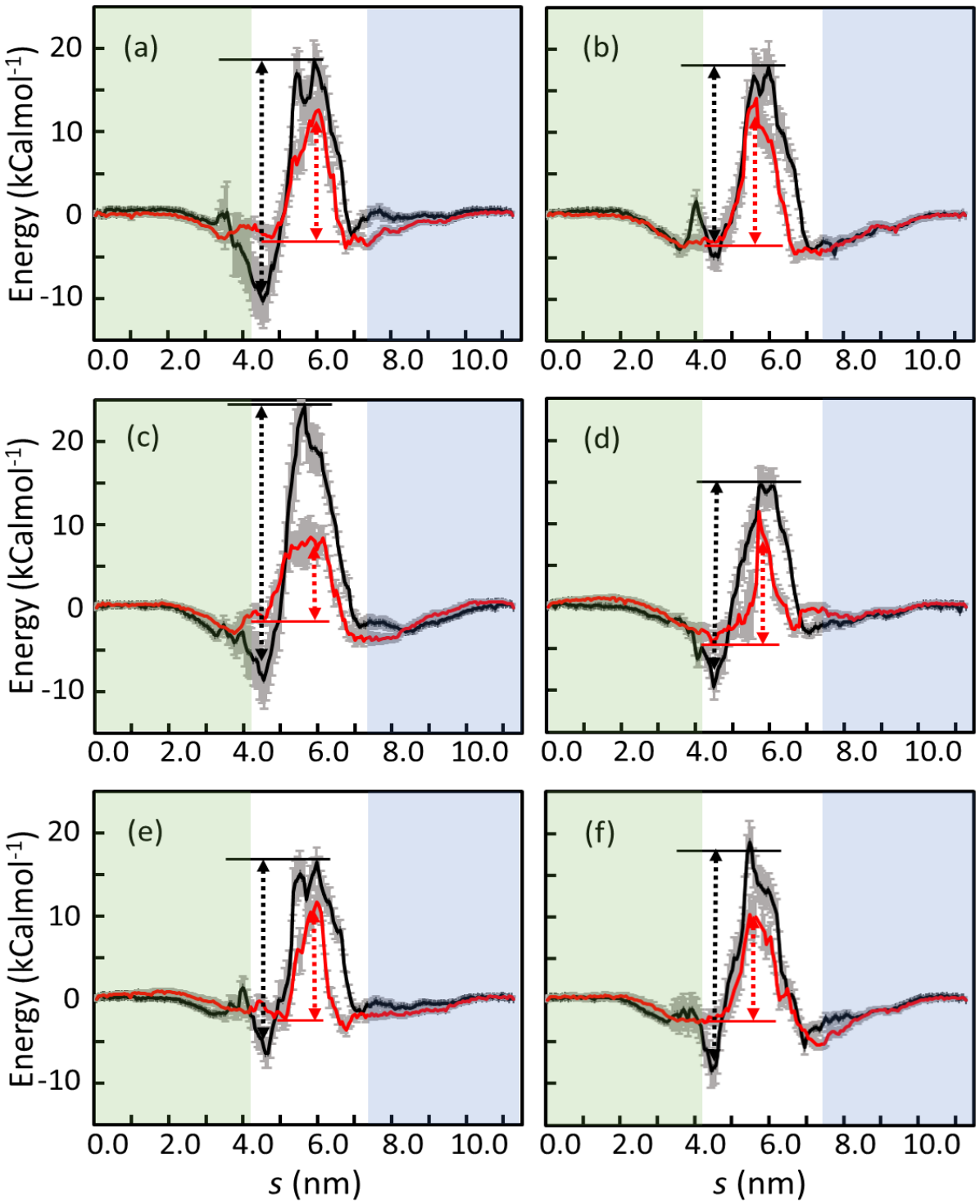


Figure 2-5. The PMF profiles of carbapenems through OccD3 and OccD3m. The comparison of (a) biapenem, (b) doripenem, (c) ertapenem, (d) imipenem, (e) meropenem, and (f) panipenem translocation through OccD3 (black line) and OccD3m (red line). The error bars calculated using the bootstrap analysis are marked in gray. The barrier heights (dashed lines) are marked in each panel. The translocation coordinate (s) is subdivided into outer membrane region (green), constriction zone (white) and periplasmic region (blue).

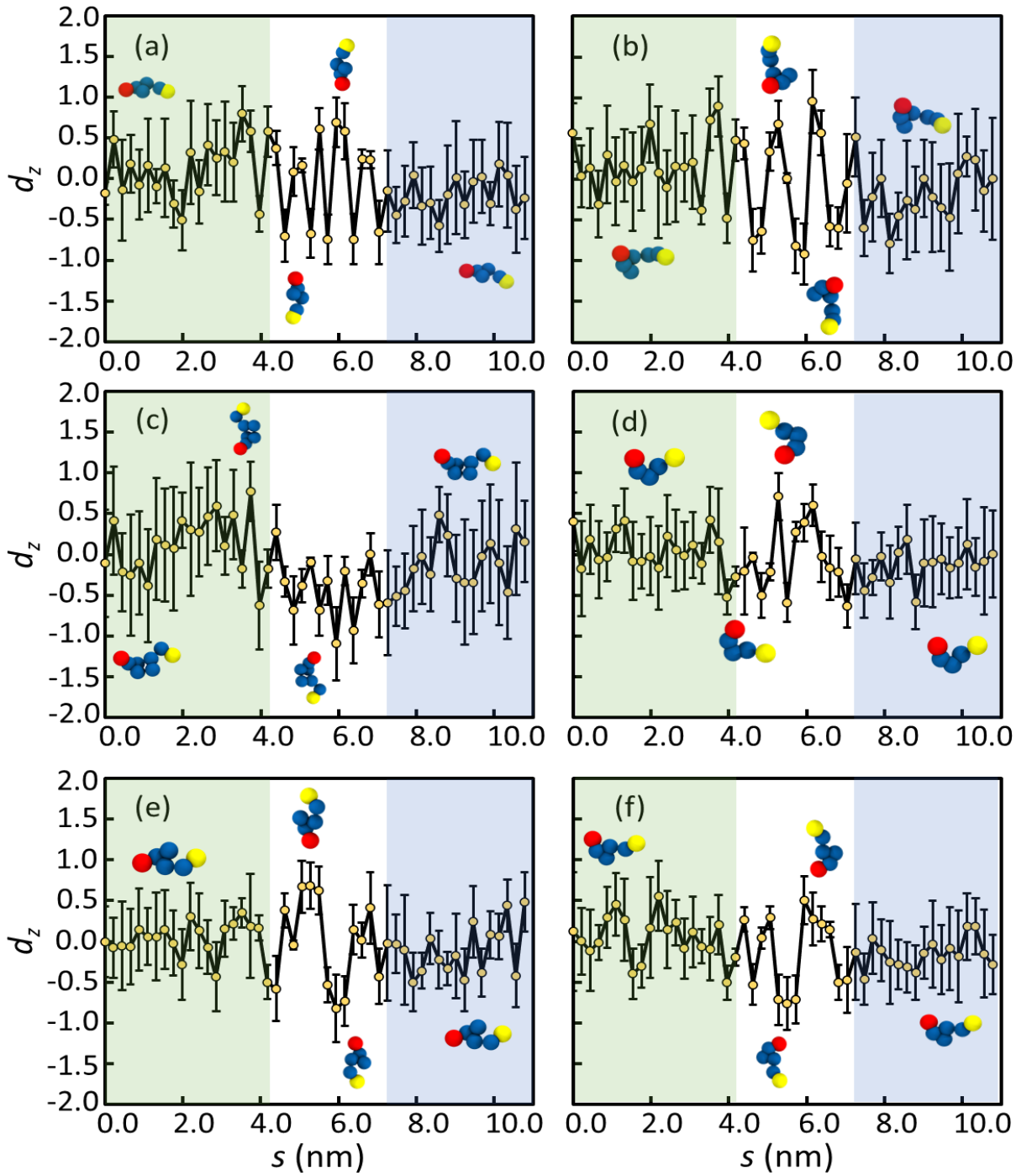


Figure 2-6. Orientational analysis of carbapenems along the translocation coordinate in OccD3. The variation in d_z for (a) biapenem, (b) doripenem, (c) ertapenem, (d) imipenem, (e) meropenem, and (f) panipenem as a function of s. The mean d_z values are denoted by the yellow dots, and the standard deviations are shown by the black bars. The s coordinate is subdivided into outer membrane region (green), constriction region (white) and periplasmic region (blue).

2.5.2 Carbapenems make highest contact with bulky Y217 and F334 OccD3 residues

Using the molecular interaction data of the carbapenem molecules with the OccD3 residues, we generated the contact maps. The contact map provides a percentile-based overview of contacts the molecule makes with each of the 452 OccD3 residues (Figure 2-7). All carbapenems have the same set of highest contact residues, namely Y169, Y217, Y326, H328, and F334 (Figures S7-S12). In fact, comparing the full contact map between all carbapenems, the most contacted residues (>80th percentile) are consistent. This reflects the dominant role of these residues in kinetics of carbapenem uptake. Besides, the importance of the charged D342 and R449 residues was shown by Soundarajan *et al.* in the ion-conductance experiments involving molecular uptake of imipenem and meropenem via OccD3 porin; the authors demonstrated that both mutations D342H (charge reversal) and R449A (charge neutralization) disturbed the translocation of imipenem and meropenem molecules.[35]

Notably, in OccD1 carbapenem transporter, residues Y176, Y282, D307, Y305 and D295 have been reported as the most probable pore-lining residues;[21] the sequence alignment between OccD3 and OccD1 (Table 2-10) shows that Y176 and Y282 in OccD1 match Y217 and Y326 (99th percentile residues) in OccD3, and D307, Y305 and D295 in OccD1 match D342, L340 and S331 (90th percentile residues) in OccD3. These results confirm that both OccD3 and OccD1 are carbapenem transporters.

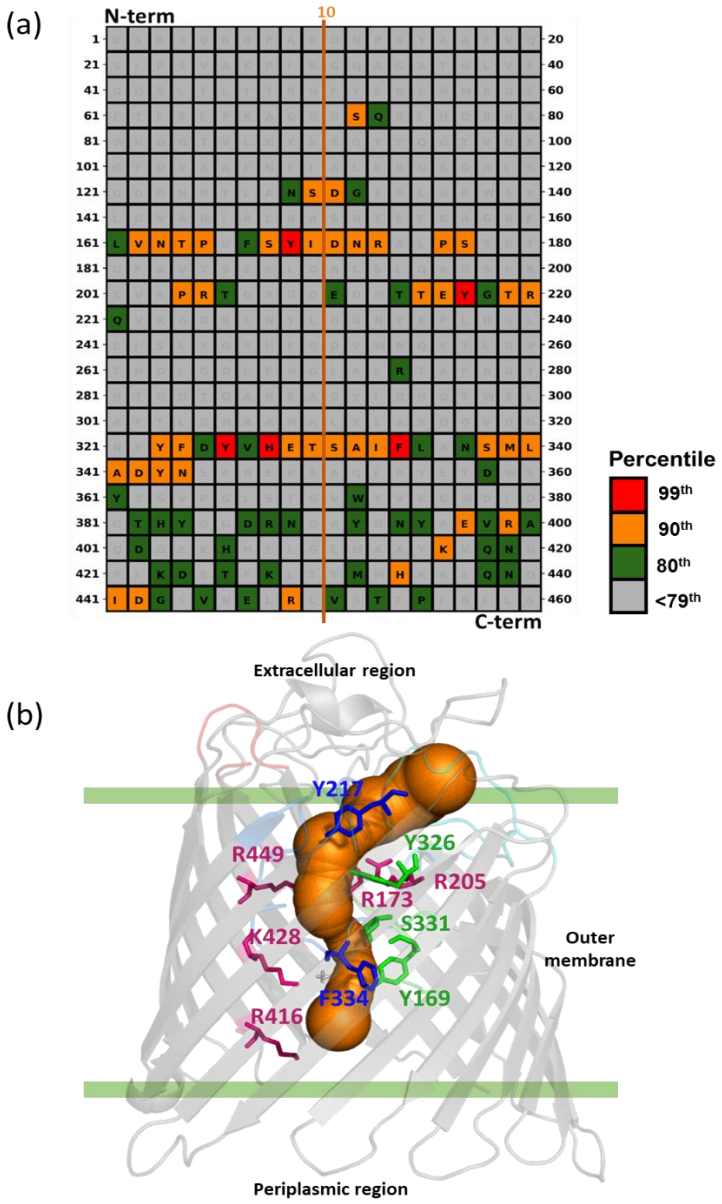


Figure 2-7 Interaction of meropenem with OccD3 porin channel. (a) Meropenem-OccD3 contact map, and (b) structure of the OccD3 (cartoon representation, gray) along with key pore-lining residues (stick model) that line the channel (orange, surface representation).

All six carbapenems make contacts with a series of charged arginine and lysine residues (Figures S7-S12), which constitutes the basic ladder. These observed interactions with the charged residues are consistent with the earlier reports involving uptake of substrates via OccD3 and

OccD1 porins. The presence of the basic ladder is considered crucial in providing the Coulombic interactions for the substrate to translocate through the porin channel.

Previous experimental and computational studies have focused on evaluating the role of charged residues in OccD3 and OccD1 channels.[35, 40] However, the steric effect of bulky pore-lining residues in OccD3 channel have not been investigated. Using CLASP, we explore the effect of two bulky residues—F334 and Y217—that make the first and second highest frequency contacts in all six carbapenems. As shown in Figure 2-7b, Y217 lies at the tip of the constriction zone on the extracellular side. The F334 is the most important hydrophobic residue in OccD3 that contacts all six carbapenems in their highest energy structures. We hypothesized that a double mutation of bulky residues Y217 and F334 to alanine would increase channel diameter and lower the energy barrier to carbapenem permeation. We also expect that these mutations will not cause the carbapenem to adopt a new pathway in the channel.

Remarkably, in OccD3m, the top residue contacts (99^{th} percentile) for all six carbapenems are identical as well (Y169, N172, R173, Y326, S331), and the top 90^{th} percentile contacts are fairly consistent (Figure S7-S12). Contact maps of OccD3 and OccD3m have similar most-contacted residues regions, which confirms the presence of a single pathway through the porin. The double mutation shows that the two bulky residues play a significant role in creating the bottleneck in the porin pathway. The discussion on the mechanism of carbapenem translocation through OccD3m is provided in Section 2.5.4.

2.5.3 Conformation of the carbapenem at the bottleneck in the OccD3 channel influence the energy barrier to translocation

The orientational and contact map analysis of the carbapenem molecules in the OccD3 channel indicates the strong interplay between the carbapenem structure and the pore-lining residues in determining the energy barrier to the translocation. The chemical groups attached to the β-lactam ring in different carbapenems impact the conformational flexibility of the molecule, which can be an important factor at the pore's bottleneck region.

For this purpose, we reverse mapped the carbapenem molecule and porin from CG to atomistic resolution and evaluated their molecular interactions in the bottleneck conformation. In biapenem, the molecule is oriented tail first with $d_z = -1.0$ at $s = 5.4$ nm, and then fliped to $d_z = 0.9$ after the bottleneck (Figure S13). At $s = 5.4$ nm, biapenem showed interactions with Y169, N172, F334, and M339. These interactions were also observed as high frequency contacts (90^{th} percentile) in the contact map (Figure S7).

Doripenem was in a folded-state with the β-lactam ring pointing towards the periplasm at $s = 6.1$ nm, which corresponds to the bottleneck region (Figure S14). The contacting pore residues are N163, P165, Y169, and F334; the former three residues can be found from the high (90th percentile) contacts and F334 belongs to the group of the highest (99th percentile) frequency contacts (Figure S8).

The ertapenem molecule is fully extended and is aligned with its tail pointing towards the periplasm (Figure S15). This orientation is maintained for much of the constriction zone for s from 4.5 to 6.5 nm beyond which the molecule eases out of the fixed orientation as it diffuses out to the periplasmic region. As expected, ertapenem is the largest carbpenem we studied (MW

474.5 Da), and it does experience the highest energy barrier to the translocation. At $s = 5.7$ nm, the molecule makes multiple contacts with pore-lining residues. These residues include the bulky Y169 and F334 that constrict the channel, and polar residues N172, S331, and S338 that provide stabilization.

Imipenem is the smallest of the six carbapenems and is able to translocate through the pore in an oblique orientation with $d_z = -0.5$ (Figure S16). The primary residue contacts at the highest point in energy profile are I170, S331, and F334, which unlike the other five carbapenems, does not include Y169. As a result of the small size of imipenem, fewer bulky contacts are experienced and therefore, reducing the energy barrier to translocation.

Meropenem has an orientational profile similar to imipenem. The contacting residues at $s = 5.6$ nm include the bulky Y169 and F334 residues, nonpolar I170, polar S168, and charged D171 (Figure S17). The orientation of the molecule causes it be in contact with residues that have high variability in their side chain size and polarity. Despite the slightly bigger size of meropenem from the six carbapenems, it has a quite small energy barrier.

Panipenem is a relatively small carbapenem (MW 339 Da) and it underwent only a few orientational flips as it navigated the constriction zone. At the highest energy point, the molecule adopted an L-shaped geometry with $d_z = 1.0$ (Figure S18). The contacting residues include four bulky groups Y169, Y217, Y326, F334, and polar D171.

This work demonstrates that the differences in the carbapenem orientation, contacting residues, and location of the bottleneck along the translocation path are the contributing factors to the varibality in the energy barrier and the translocation rates. This detailed analysis is possible

due to the robustness of the CLASP alogirthm that can provide the on-demand reverse mapped structure of the simualtion system at atomistic-level precision.

2.5.4 Carbapenems have lower permeation barrier through OccD3m

The free energy profile of each carbapenem through OccD3m (Figure 5, red lines) show lower barriers than through OccD3. The barrier heights and rate constants results of carbapenems via OccD3m and OccD3 channels are provided in Table 2-1.

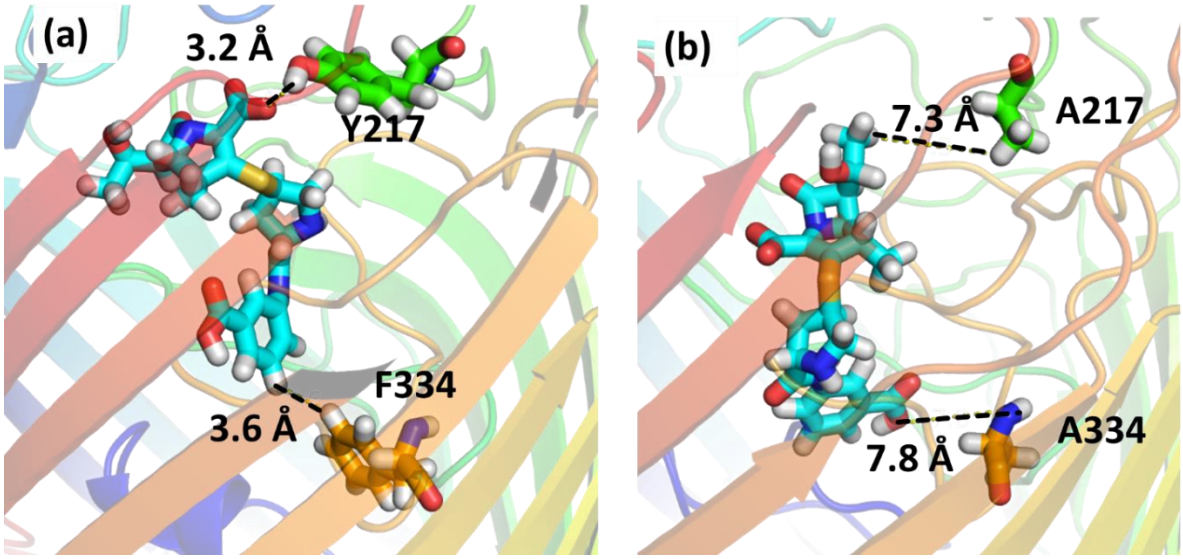


Figure 2-8. Comparison of reserve-mapped atomistic structure and orientation of ertapenem in the bottleneck region, $s = 5.8$ nm of (a) OccD3 and (b) OccD3m channel. The contacts the molecules make with the pore-lining residues in their highest energy conformation. The porin is shown in ribbon representation (blue to red from N-term to C-term) and ertapenem is shown as sticks (C cyan; O red; H white; S gold; and N blue). The pore-lining residues are shown in stick representation Y217 and A217 (C green), and F334 and A334 (C orange); the other atoms have same color scheme as in ertapenem.

Interestingly, the largest change in barrier height for translocation occurred in ertapenem (ΔBH = 21 kcal mol^{-1}), the largest carbapenem in the group studied. In contrast, one of the smallest changes in all six carbapenems occurred with imipenem (ΔBH = 6 kcal mol^{-1}), the smallest carbapenem in the group. A close up view of ertapenem's molecular interactions with OccD3 and OccD3m at $s = 5.8$ nm, the highest energy point, reveals the reason for the change in free energy

(Figure 2-8). The minimum in OccD3m is possible because the large ertapenem molecule is stabilized by the added volume created by concurrent Y217A and F334A mutations; none of the other five carbapenems were large enough to experience this dual volume increase.

Furthermore, the orientational profile comparison of the six carbapenems through OccD3 and OccD3m supports our hypothesis that Y217 and F334 influence the translocation behavior. The carbapenem orientational analysis for translocation through the OccD3m pore shows a higher number of flips in the constriction region (Figure S19). The widening of the pore provides small pockets for the molecules to orient in different conformations as they navigate the translocation coordinate.

2.6 Conclusions

Here we report the developement of a new computational screening platform called CLASP that is designed to accelerate the antibiotic discovery process. This computational platform provides an automated screening of small molecules by quantifying their permeabilities through the bacterial outer membrane porins. The CLASP outputs comprehensive thermodynamic data, which includes, potential of mean force profile, energy barrier, translocation rate constant, contact analysis of the molecule with the pore-lining residues, orientation analysis of the molecule within the porin, and the overall mechanism of the antibiotic uptake. Additionally, the fast turnaround times will enable the establishment of structure-property relationships among molecular species that have high bacterial membrane permeability.

In this first application of CLASP, we demonstrate the permeability profile of six, well-established antibiotics from the carbapenem family through *P. aeruginosa*‘s OccD3 channel. The proof-of-concept study was undertaken to use the known literature and illustrate the permeability

profiles for biapenem, doripenem, ertapenem, imipenem, meropenem, and panipenem. The results show excellent agreement with the barrier heights, the key residues, and the translocation rate constants of the drug molecules. We were able to identify critical pore-lining residues in the OccD3 channel that provide steric gating. Mutation of these residues to less bulky groups enhanced the predicted rates. Moreover, the wall-clock time to simulate one antibiotic molecule was less than two hours based on the current computational resource.

In conclusion, our goal was to advance strategies that will facilitate the identification of new antibiotics. This work has enabled comprehensive characterization of antibiotic uptake through bacterial envelope. In the future, we plan to employ CLASP to screen large natural compound libraries to identify the new potential antibiotics.

2.7 Acknowledgements

This work is supported by the National Science Foundation grant numbers CBET-1706061 and DMR-1757749. Computational resources were provided by Information and Technology Services at Syracuse University (Eric Sedore, Larne Pekowsky, and Michael R. Brady) as well as the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. The authors thank Somya Chakraborty for the help with the manuscript.

2.8 Supporting Information

The contents include description of the CLASP python scripts, coarse grain-mapping of the six antibiotics, simulation setup details, contact maps, and reversed mapped orientational analysis of the antibiotics in OccD3 channel, and orientational analysis of carbapenems along the translocation coordinate in OccD3m.

2.8.1 Description of the CLASP python scripts

The CLASP program is designed to facilitate high-throughput screening of small molecule solutes that have the potential of being antibiotic drugs. The solute molecules can include compounds available in natural compound libraries and other synthetic compounds that can be transported into the bacterial periplasm through the porin channels. The salient features of CLASP program are:

- **automates** the tedious umbrella sampling simulation set up. The user provides the desired number of umbrella sampling windows (N) and length (L_s) of the translocation coordinate, which roughly corresponds to the length of the porin channel.
- **minimizes** the amount of effort to prepare the simulation set up. Only structure of the solute molecule and the porin need to be provided.
- **streamlines** the input and output workflow. All programs are executed by simple command line. No additional programming is required. CLASP offers user-friendly organization of output as separate labeled folders of the simulation runs will be generated and well labeled. All necessary topologies and parameter files needed for the simulation are added automatically to the simulation folders.

- **accelerates** the simulations of transport mechanisms. CLASP reaches 3 μs/day based on the current resources.
- **analyses** simulation results as CLASP provides a variety of outputs, including the potential of mean force, transport barriers, molecular orientations during translocation, and contact map of molecule-protein residues, etc.

The description of the input and output scripts are as follows:

C_setup.py *generates umbrella sampling configurations.*

The C_setup.py script takes the membrane-embedded, solvated, and equilibrated porin structure as input along with two user defined parameters:

(a) length, L_s, of the translocation coordinate (s), which roughly corresponds to one and a half times of the length of the porin channel; default value = 11 nm

(b) N, the number of umbrella sampling (US) windows, default value is 100. The user can change N to check for convergence.

The C_setup.py script then prepares N independent US windows with spacing of L_s/N nm. In each window, W_i, the script inserts the probe substrate molecule using gmx insert-molecules utility available in GROMACS. The coordinates for the insertion of substrate molecule in each window is computed relative to the center-of-mass (COM) of the porin. Given the hourglass shape of the porin, the COM lies in the middle of the porin channel. The location of the COM is considered $W_{N/2}$, with equal number of windows on either side.

Next, the C_setup.py script generates N separate folders, each with unique location of the substrate molecule defined within W_i, totaling the length L_s along the translocation

coordinate. Each folder is set up to execute energy minimization and umbrella sampling simulations independently. The *N* jobs run concurrently on separate nodes for maximum efficiency and shortest run completion times.

C_PMF.py *generates potential of mean force profile along the transport coordinate.*

The C_PMF.py script combines the *N* umbrella sampling simulation run-input files (*.tpr* files) into one file (*tpr.dat*) and combines the *N* umbrella sampling simulation output files (**f.xvg* files) into one file (*fxvg.dat*). The two combined files are used as the two input files with gmx wham, a GROMACS built-in utility, along with the error analysis (bootstrap) to generate the potential of mean force plot. Bootstrap analysis is the error analysis available from the gmx wham program to estimate statistical uncertainty. In this study, 200 bootstraps were used as recommended from the GROMACS manual to calculate the standard deviations.

C_Dz.py *generates the orientational analysis of the molecule along the transport coordinate.*

The C_Dz.py script reads the x-, y- and z-axis coordinates of the solute molecule during the transport process using gmx distance, a built-in Gromacs utility with one of its options, -oxyz. Then the script extracts the z-component data and calculates the average and the standard deviation. Finally, the z component is converted to the molecular orientation, which is explained in the main paper.

C_contact.py *generates the contact map of the molecule with the protein residues along the transport coordinate.*

The C_contact.py script calculates the cumulative number of contacts made by the solute molecule with each protein residue throughout the whole trajectory as described in the main paper. Then the script plots the contact map based on the numbers.

In summary, the CLASP is designed to work with GROMACS 5.0 package and later. The CLASP is developed to support outer membrane porins of all bacterial species with porin, provided the porin structure is available. The porins can be embedded in lipopolysaccharide-rich, asymmetric lipid membrane using the BOB.py script, which currently has MARTINI coarse-grained parameters of ten Gram-negative bacterial lipids. The CLASP can be easily extended to include new bacterial species once their lipid parameters become available.

2.8.2 Supporting Tables

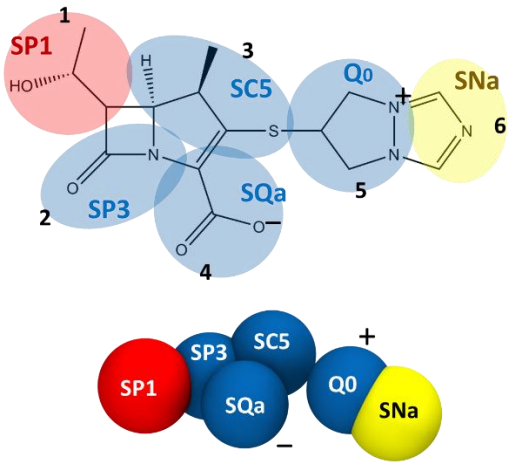


Figure 2-9 Biapenem

Table 2-2 CG mapping and parameters for Biapenem

Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.297	5000
1-3	0.420	5000
2-3	0.332	5000
2-4	0.295	5000
3-4	0.334	5000
3-5	0.464	5000
5-6	0.252	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	145	50
1-3-4	96	50
1-3-5	148	50
2-3-5	112	50
4-3-5	67	50
3-5-6	152	50

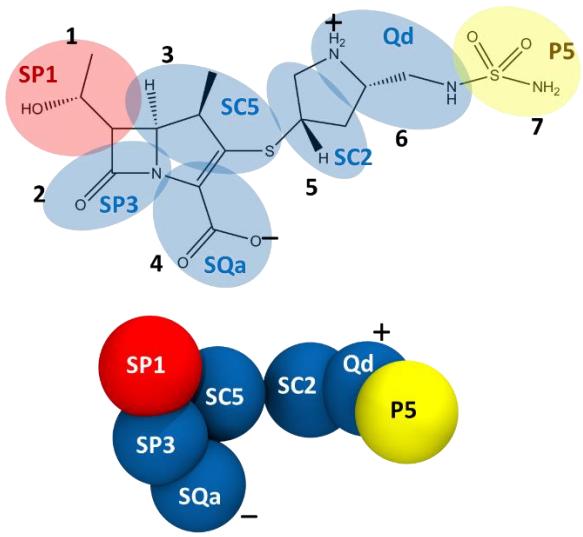


Figure 2-10 Doripenem

Table 2-3 CG mapping and parameters for Doripenem

Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.292	5000
1-3	0.422	5000
2-3	0.329	5000
2-4	0.300	5000
3-4	0.334	5000
3-5	0.371	5000
5-6	0.247	5000
6-7	0.324	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	146	50
1-3-4	96	50
1-3-5	150	50
2-3-5	129	50
4-3-5	84	50
3-5-6	137	50
5-6-7	134	50

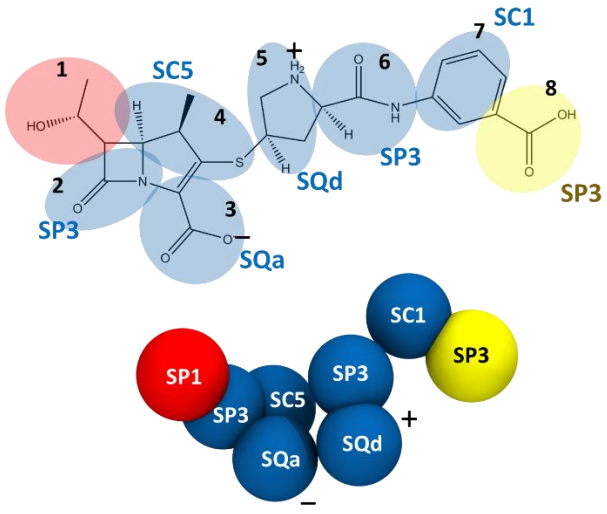


Figure 2-11 Ertapenem

Table 2-4 CG mapping and parameters for Ertapenem

Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.268	5000
1-3	0.566	5000
2-3	0.304	5000
2-4	0.325	5000
3-4	0.339	5000
4-5	0.362	5000
5-6	0.279	5000
6-7	0.385	5000
7-8	0.312	Constrained

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-3	104	50
1-3-4	56	50
2-4-5	128	50
3-4-5	87	50
4-5-6	119	50
5-6-7	134	50
6-7-8	103	50

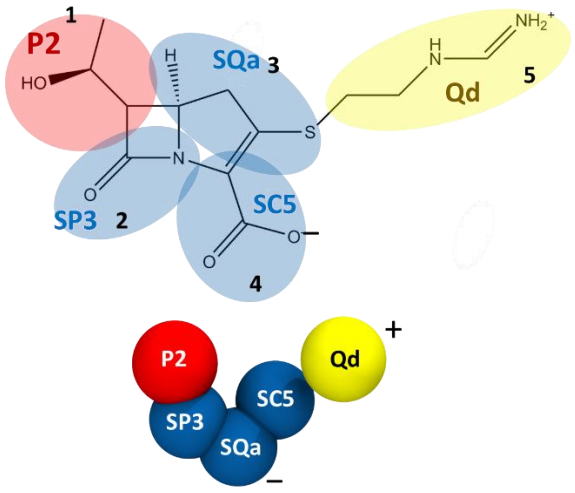


Figure 2-12 Imipenem

Table 2-5 CG mapping and parameters for Imipenem

Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.293	5000
1-3	0.434	5000
2-3	0.316	5000
2-4	0.301	5000
3-4	0.305	5000
3-5	0.494	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	149	50
1-3-4	99	50
1-3-5	134	50
2-3-5	122	50
4-3-5	85	50

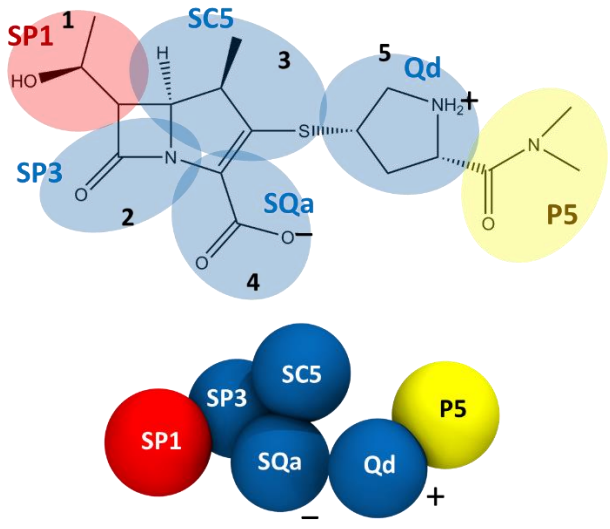


Figure 2-13 Meropenem

Table 2-6 CG mapping and parameters for Meropenem

Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.297	5000
1-3	0.410	5000
2-3	0.329	5000
2-4	0.303	5000
3-4	0.334	5000
3-5	0.494	5000
5-6	0.343	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	143	50
1-3-4	99	50
2-4-5	114	50
3-4-5	55	50
4-5-6	106	50

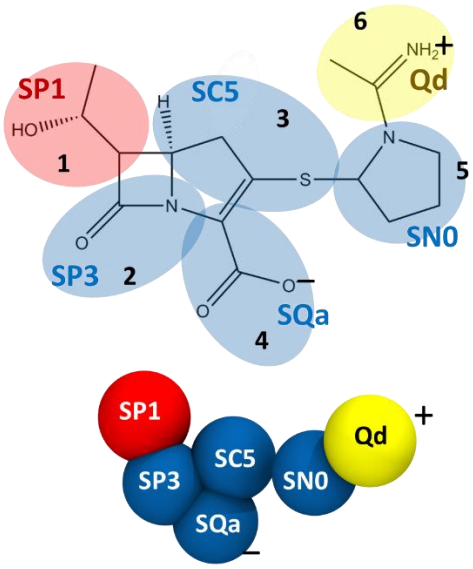


Figure 2-14 Panipenem

Table 2-7 CG mapping and parameters for Panipenem

Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.293	5000
1-3	0.434	5000
2-3	0.316	5000
2-4	0.301	5000
3-4	0.334	5000
3-5	0.403	5000
5-6	0.311	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	146	50
1-3-4	99	50
1-3-5	138	50
2-3-5	143	50
4-3-5	103	50
3-5-6	142	50

Table 2-8 Composition of the system with embedded OccD3 protein

Antibiotic	outer		inner	W		Ions	
	LPS	DPPE	DPPE	W	WF	Ca^{2+}	Na^{+}
Biapenem	54	6	141	5625	625	58	1
Doripenem	54	6	141	5625	625	58	1
Ertapenem	54	6	141	5625	625	58	1
Imipenem	54	6	141	5625	625	58	1
Meropenem	54	6	141	5625	625	58	1
Panipenem	54	6	141	5625	625	58	1

Table 2-9 Values of Umbrella sampling windows (N) used for the six carbapenem simulations

Antibiotic	*N* value
Biapenem	400
Doripenem	400
Ertapenem	400
Imipenem	400
Meropenem	400
Panipenem	400

An optimal number for N depends on both the size of the drug molecule, and the length of the reaction coordinate. A poor sampling will appear as a gap area in the histogram. When N is inadequate, there will be gaps in the histograms. The user needs to adjust the value of N to achieve better sampling along the entire reaction coordinate. In this study, 100, 200, and 400 for N were tested and the data was converged at 200. In the interest of maximum statistical data, the results for N = 400 are provided.

The black dots denote identity. The key pore-lining residues identified in OccD3 are marked in blue outlines.

Table 2-10 Pairwise sequence alignment of OccD3 and OccD1

2.8.3 Supporting Figures

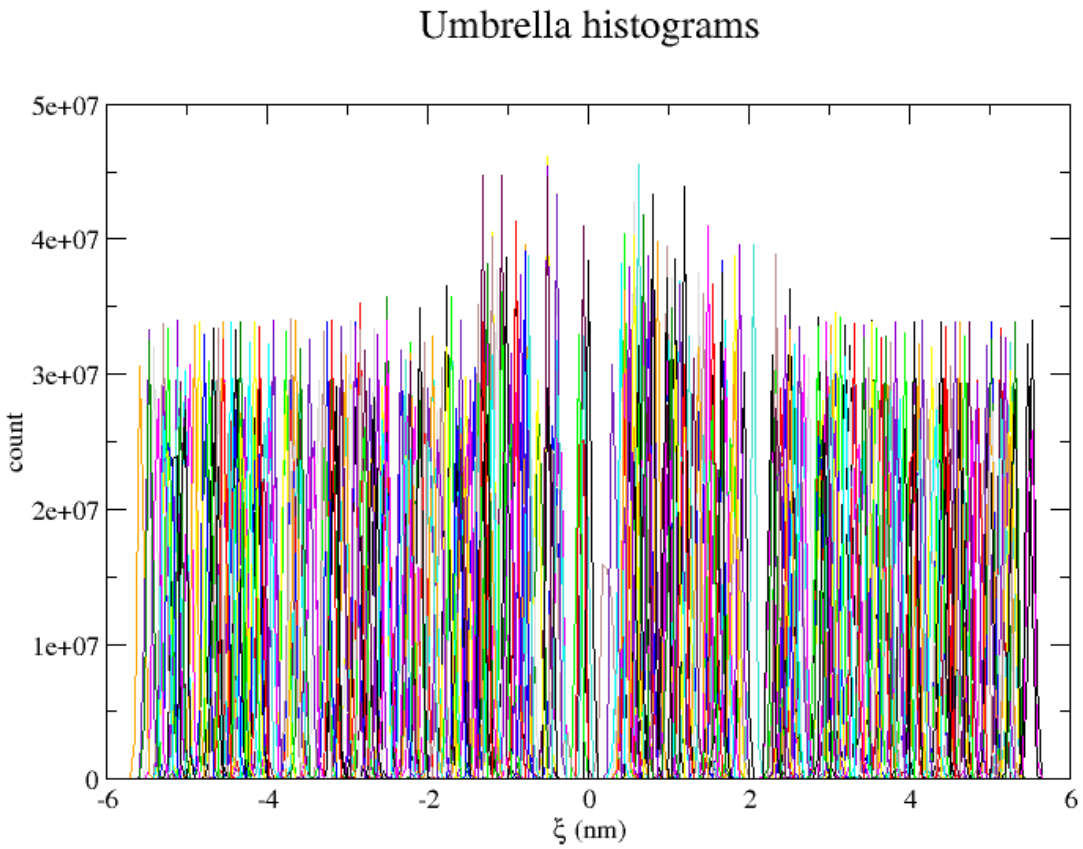


Figure S 1 Histogram of Biapenem with OccD3

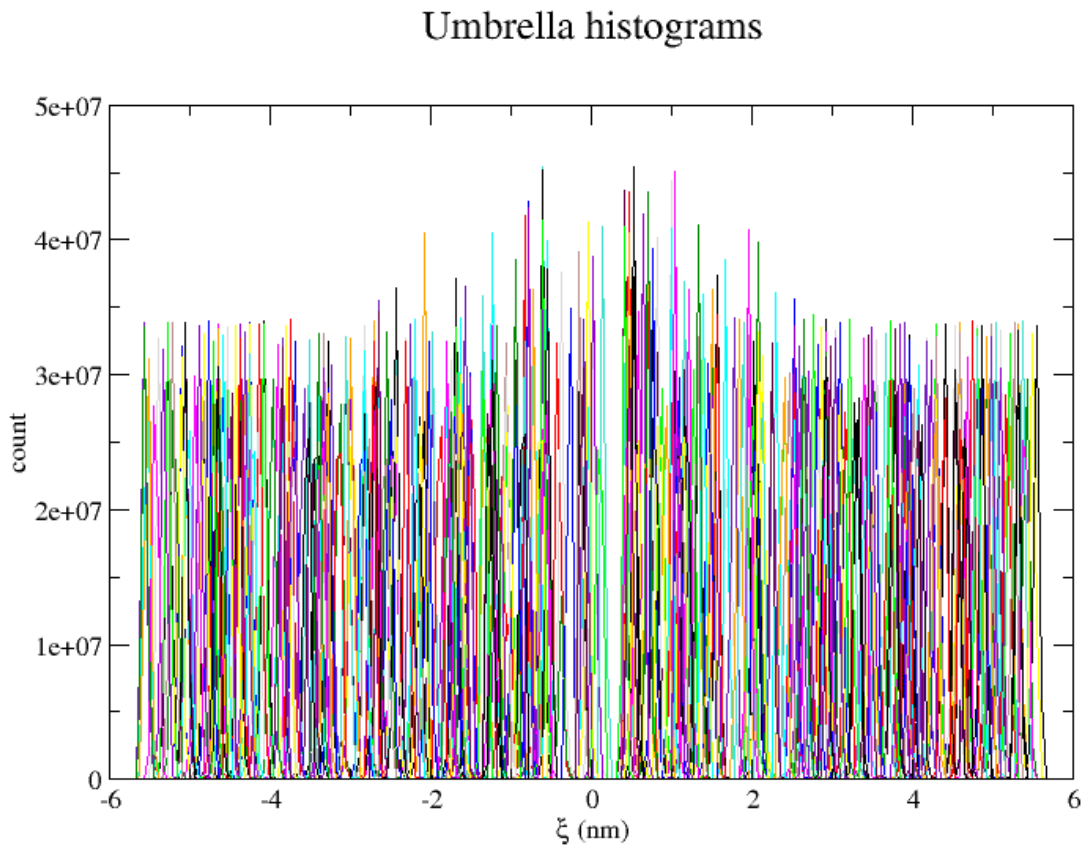


Figure S 2 Histogram of Doripenem with OccD3

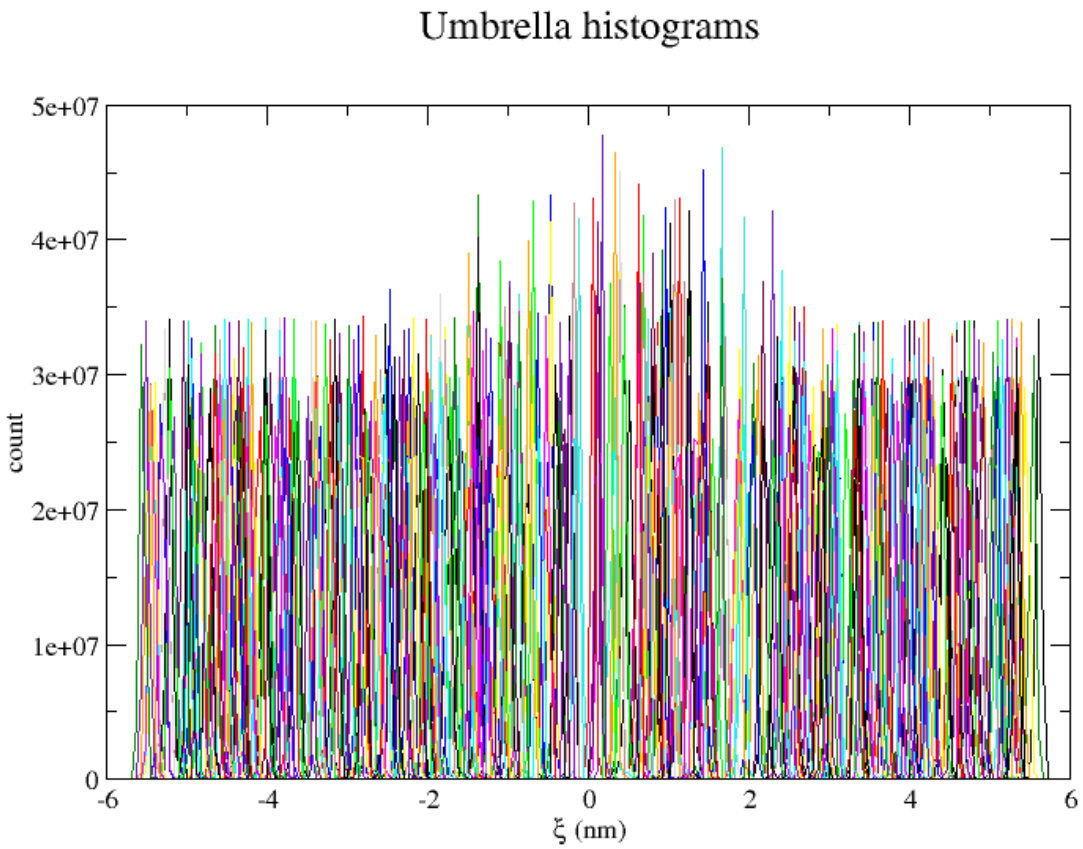


Figure S 3 Histogram of Ertapenem with OccD3

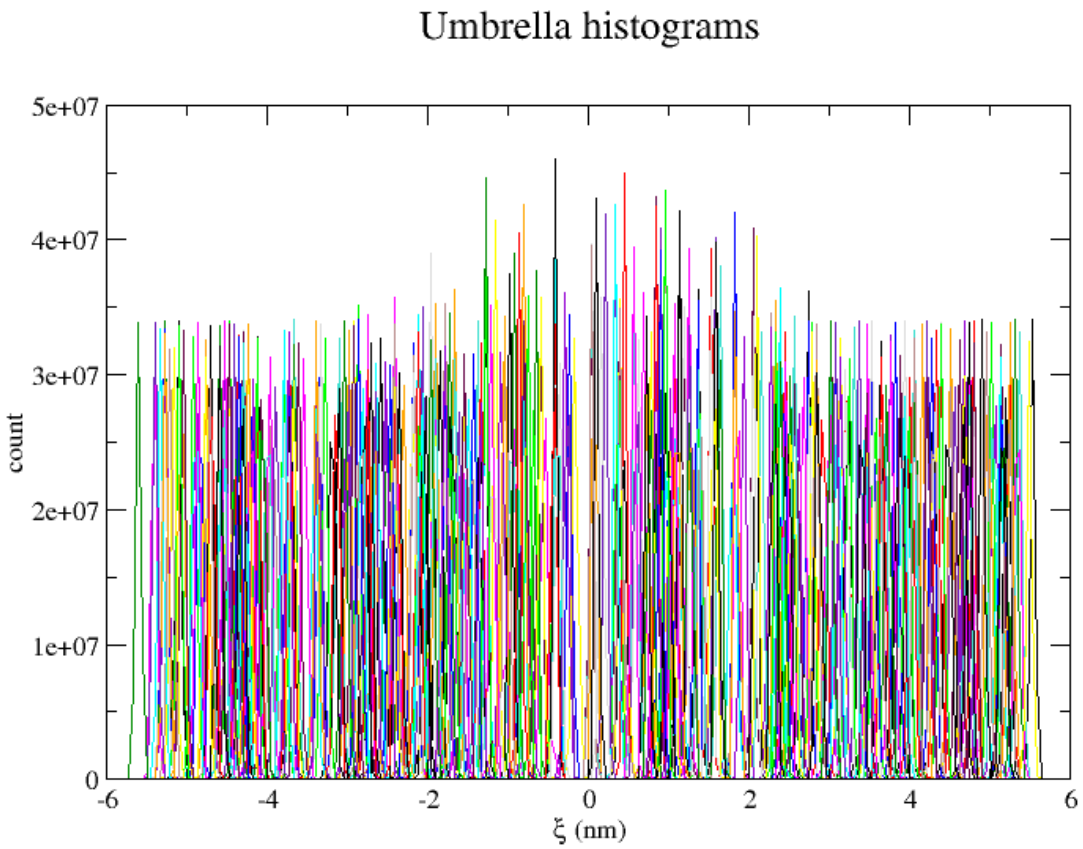


Figure S 4 Histogram of Imipenem with OccD3

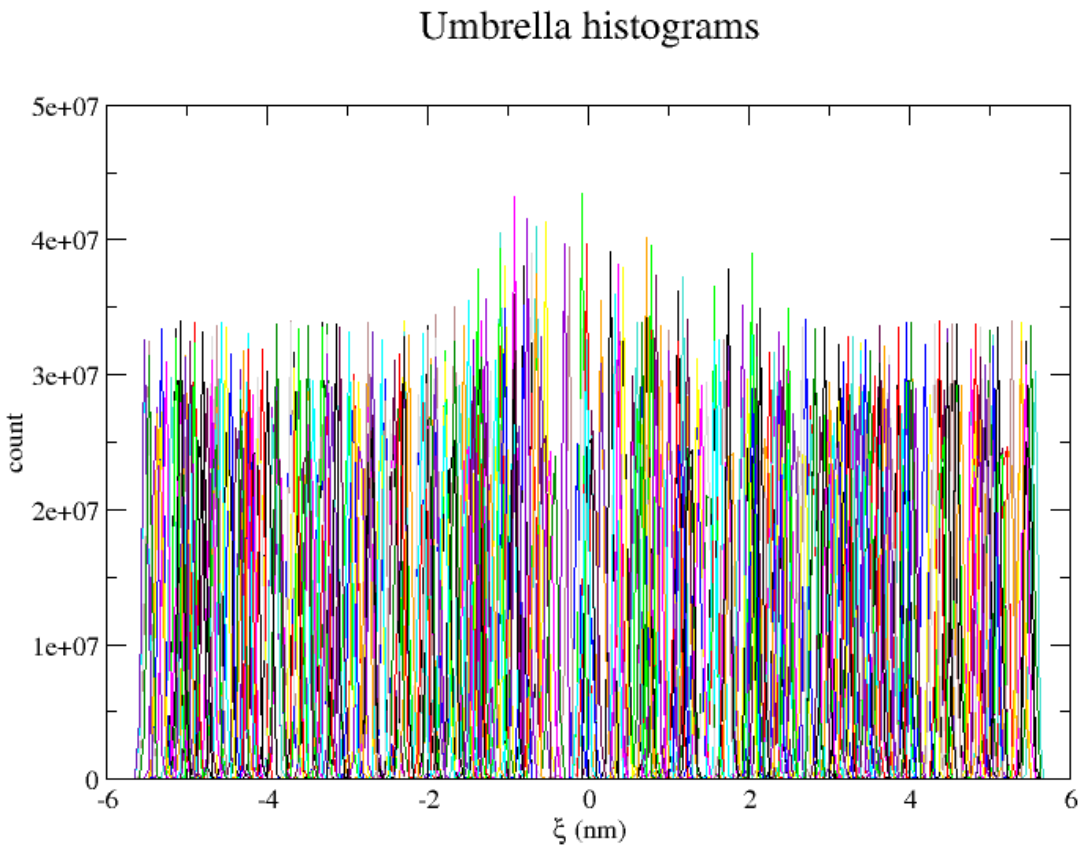


Figure S 5 Histogram of Meropenem with OccD3

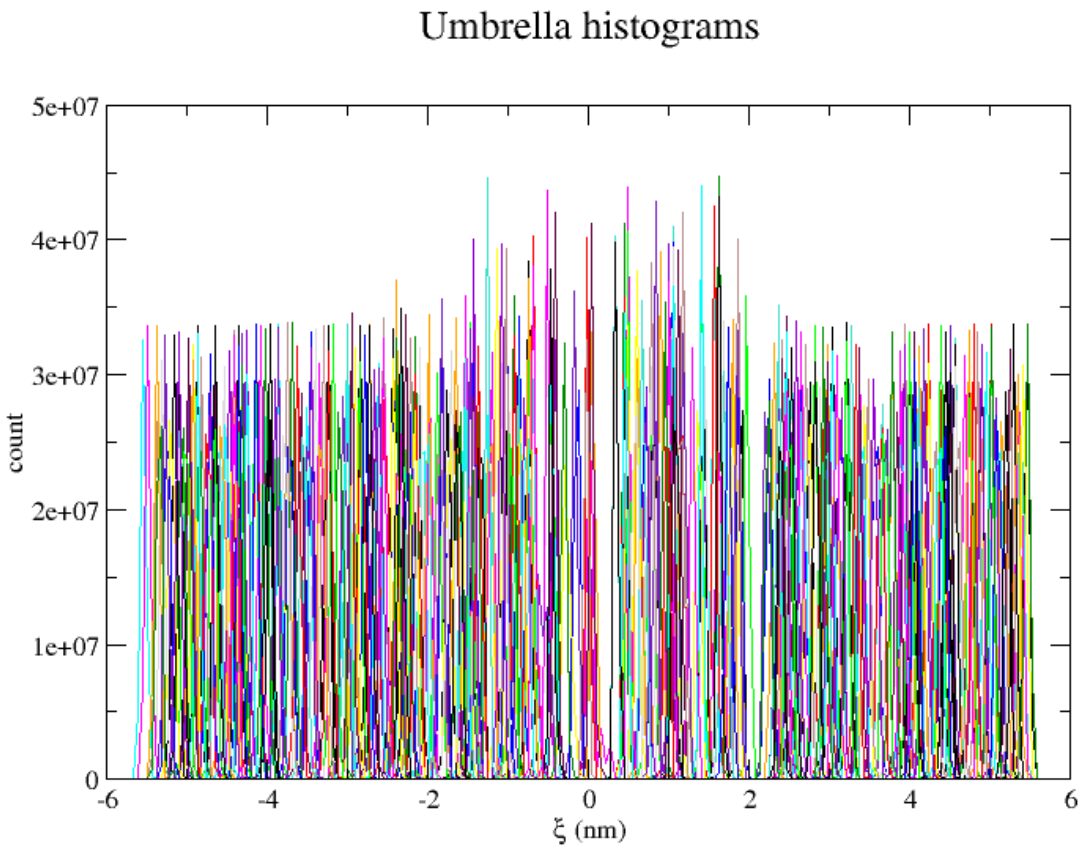


Figure S 6 Histogram of Panipenem with OccD3

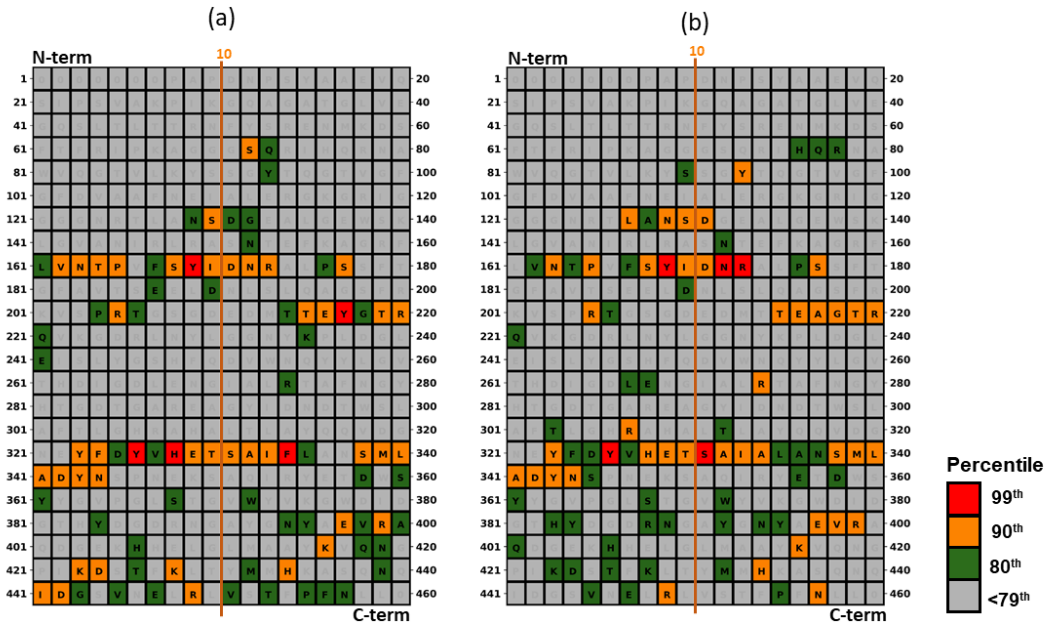


Figure S 7 Contact map of Biapenem with (a) OccD3 and (b) OccD3m

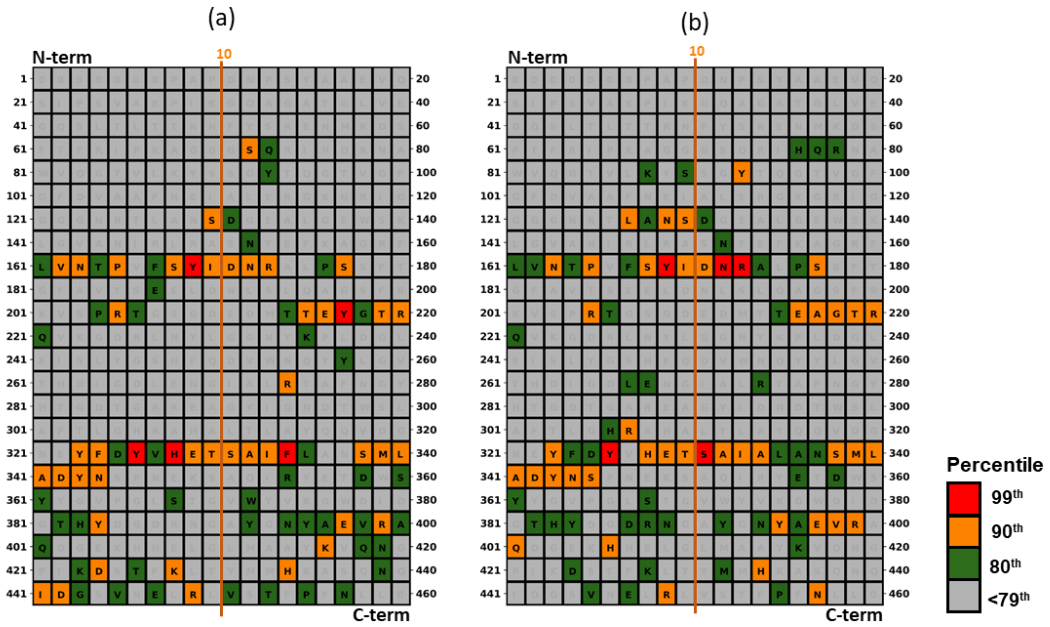


Figure S 8 Contact map of Doripenem with (a) OccD3 and (b) OccD3m

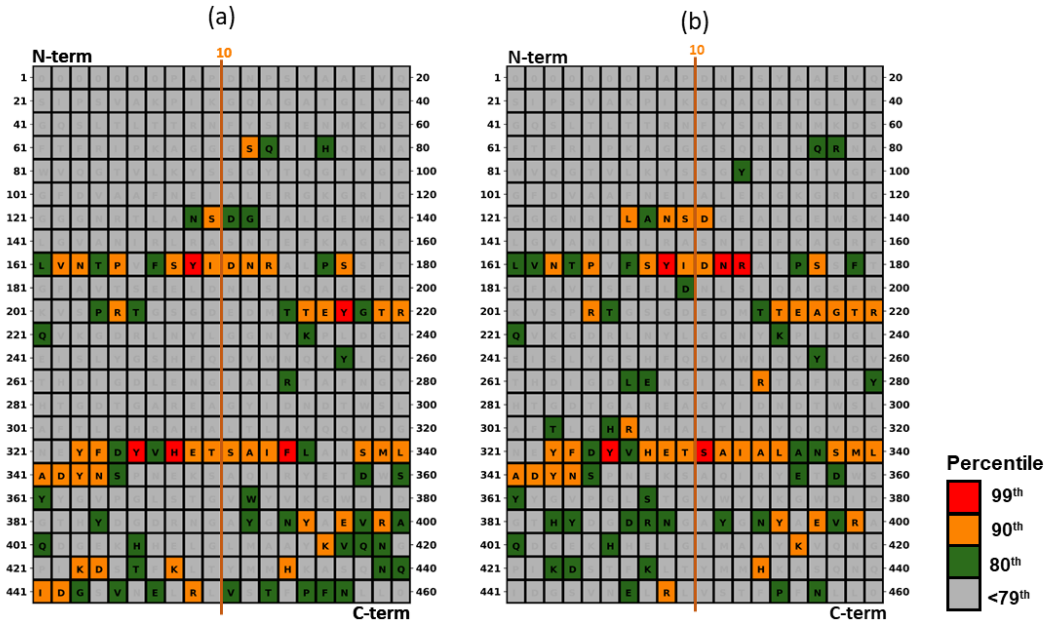


Figure S 9 Contact map of Ertapenem with (a) OccD3 and (b) OccD3m

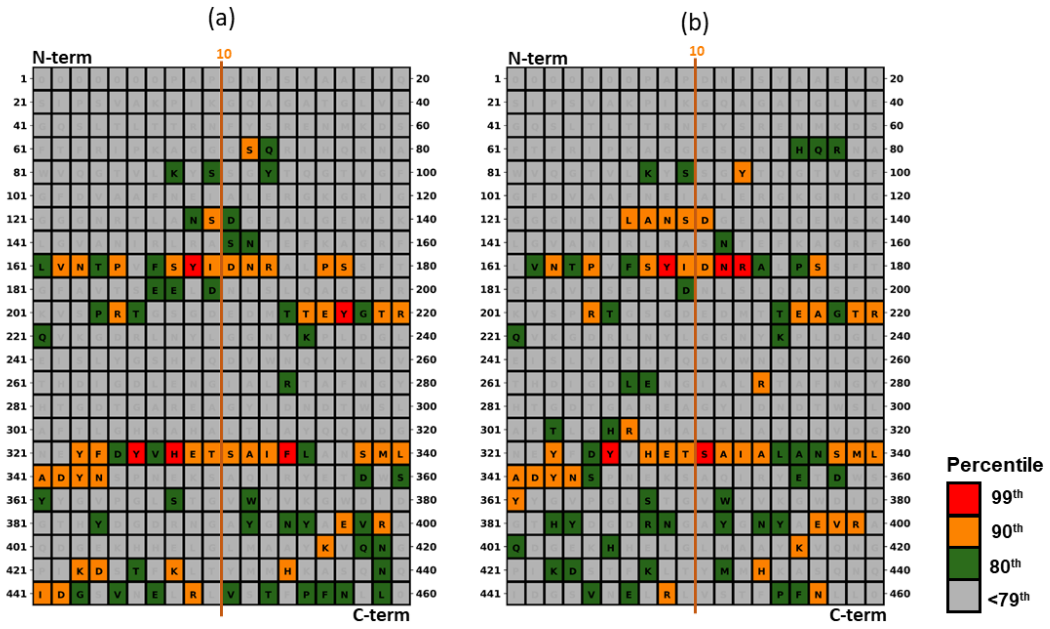


Figure S 10 Contact map of Imipenem with (a) OccD3 and (b) OccD3m

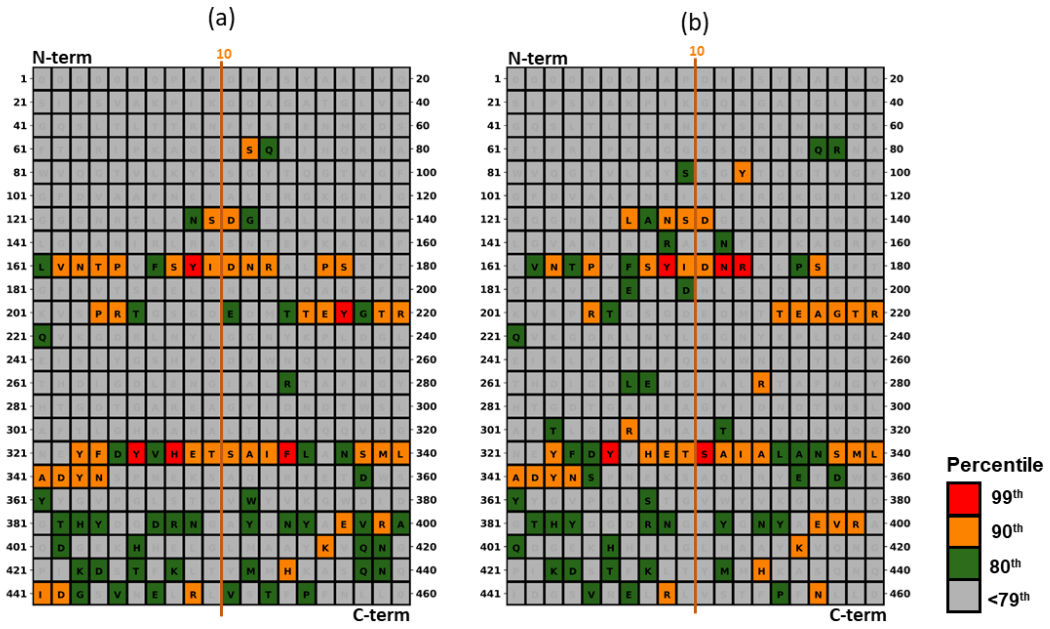


Figure S 11 Contact map of Meropenem with (a) OccD3 and (b) OccD3m

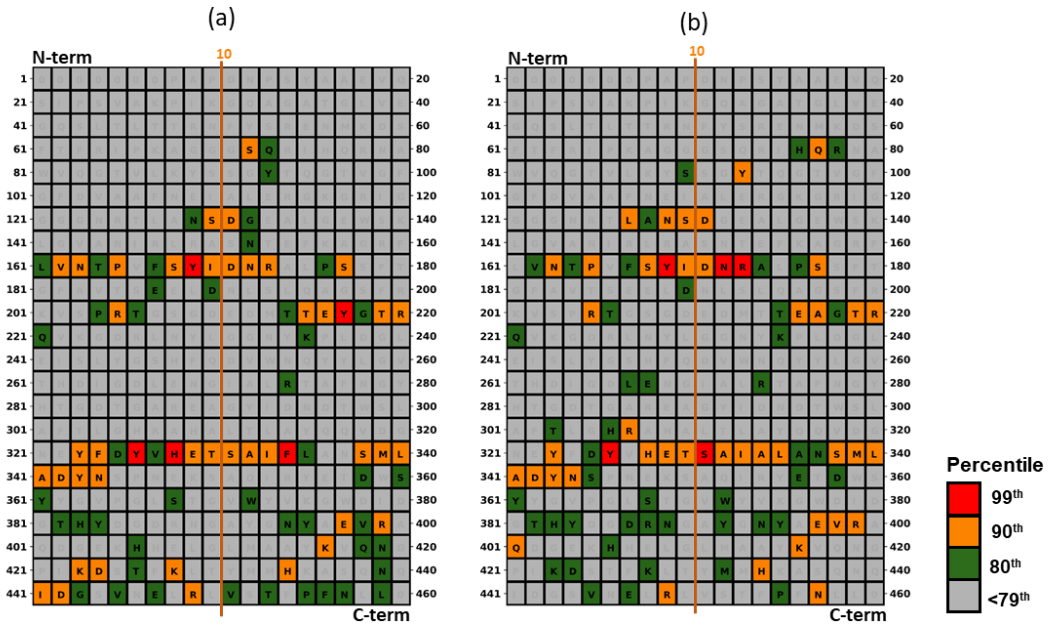


Figure S 12 Contact map of Panipenem with (a) OccD3 and (b) OccD3m

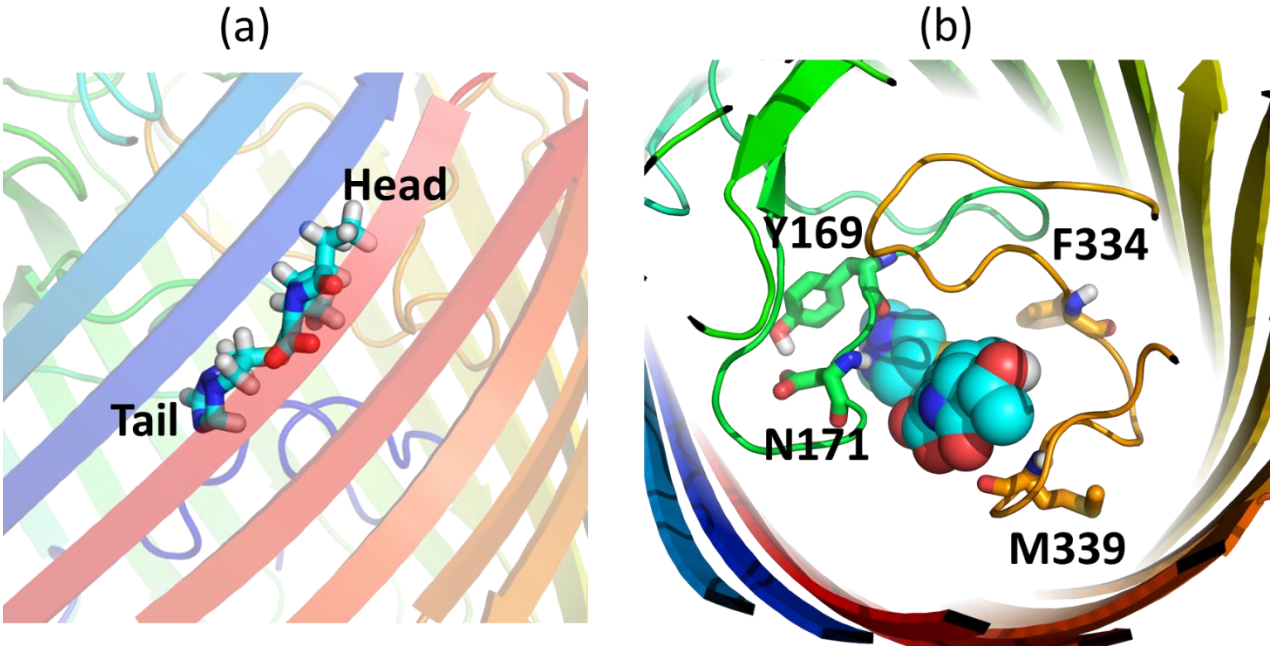


Figure S 13Reserve mapped atomistic structure and orientation of Biapenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at s=5.4 nm.

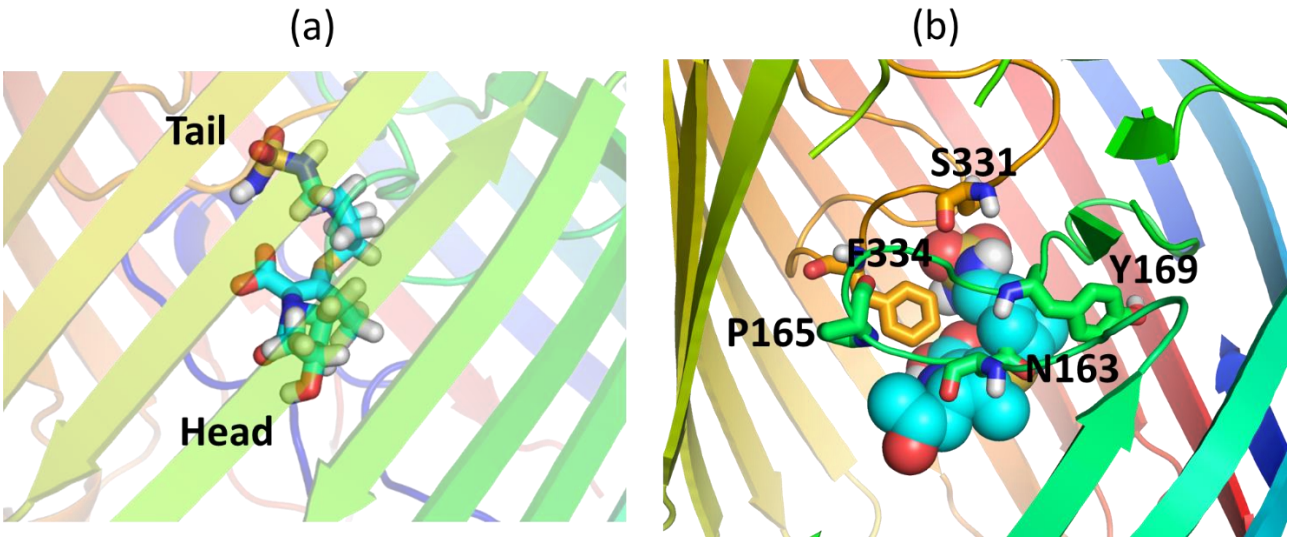


Figure S 14 Reserve mapped atomistic structure and orientation of Doripenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at s=6.1 nm.

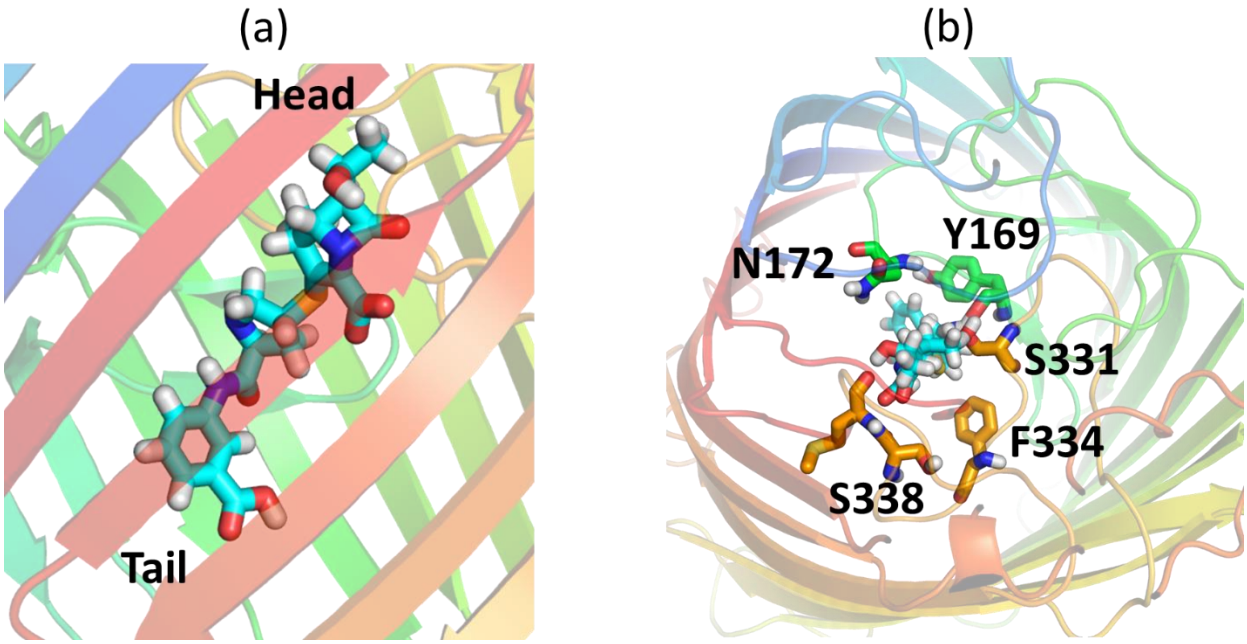


Figure S 15 Reserve mapped atomistic structure and orientation of Ertapenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at s=5.7 nm.

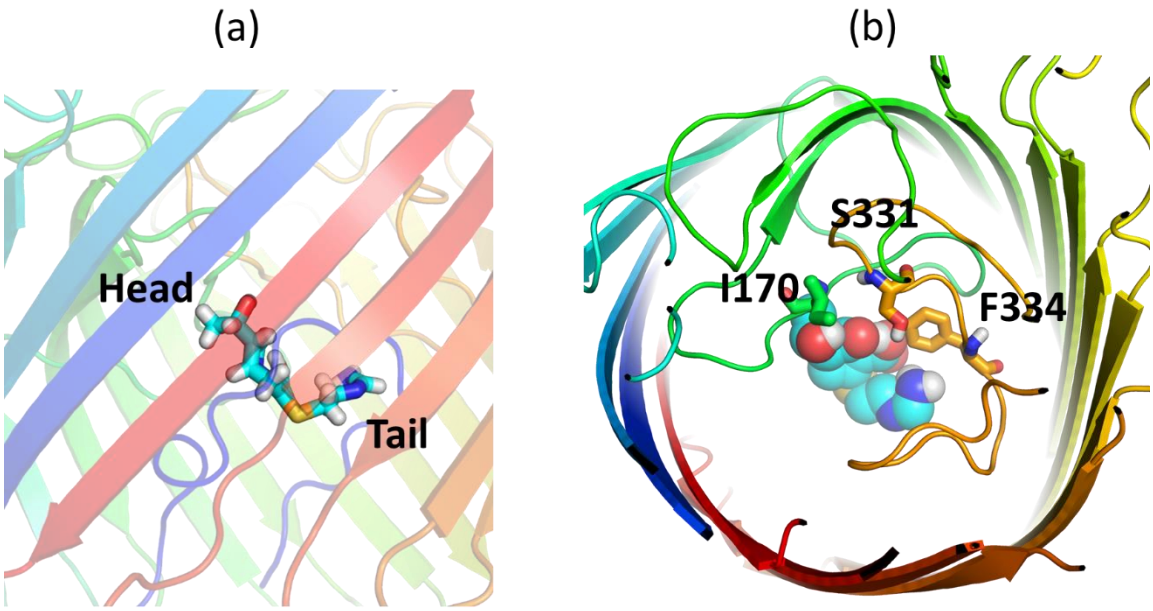


Figure S 16 Reserve mapped atomistic structure and orientation of Imipenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at s=5.9 nm.

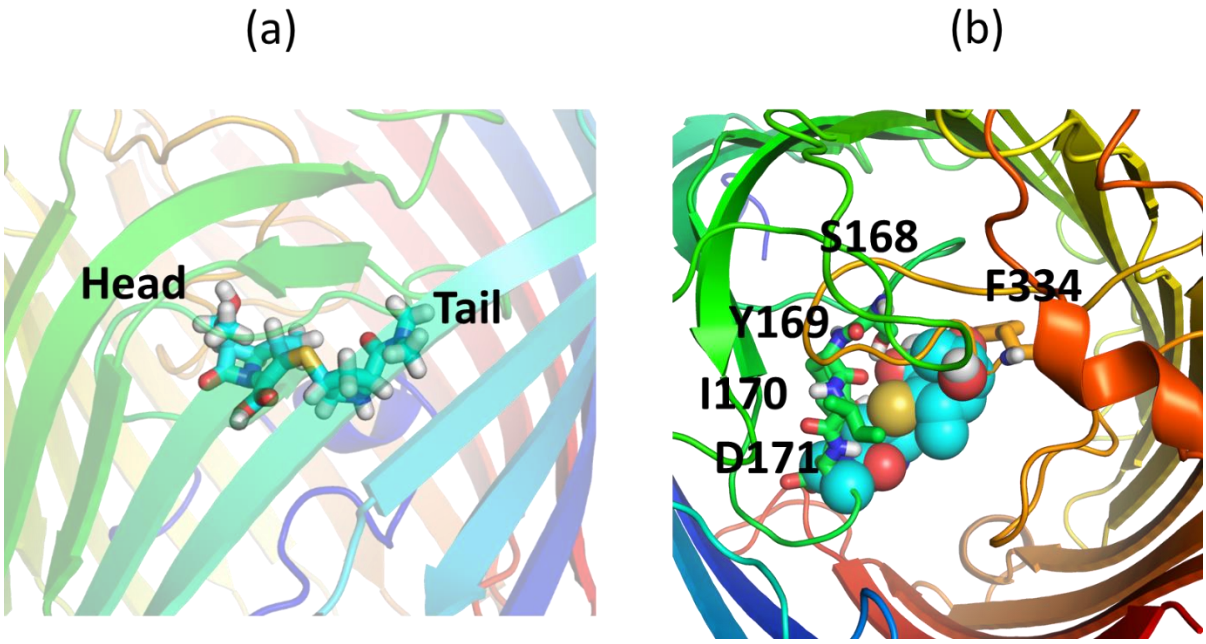


Figure S 17 Reserve mapped atomistic structure and orientation of Meropenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at s=5.6 nm.

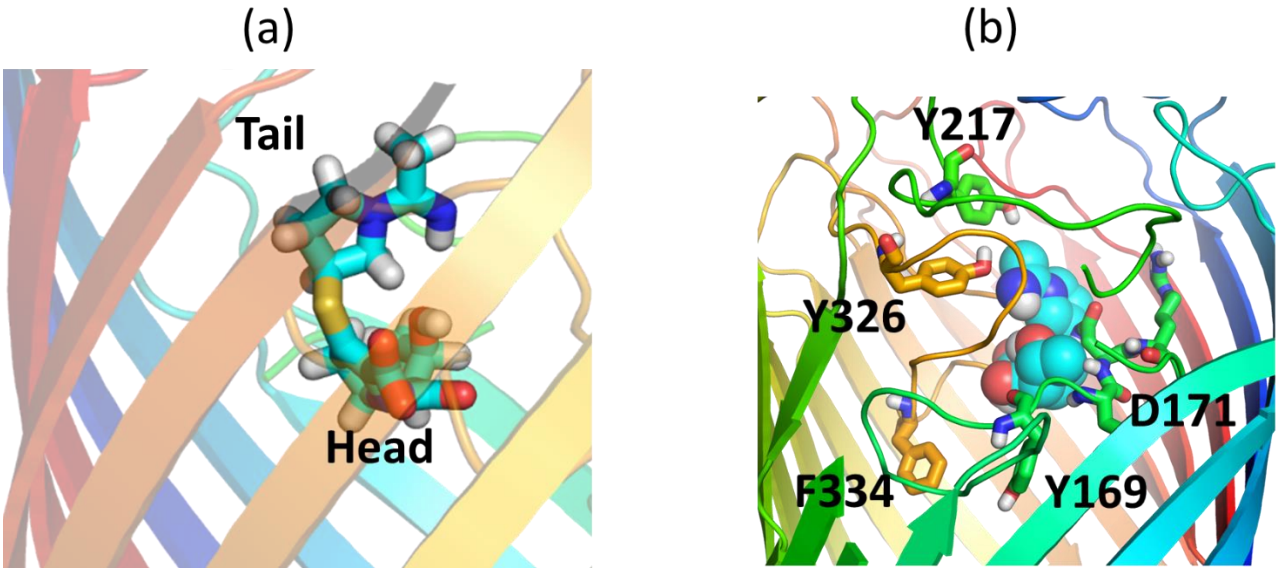


Figure S 18 Reserve mapped atomistic structure and orientation of Panipenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at s=6.0 nm.

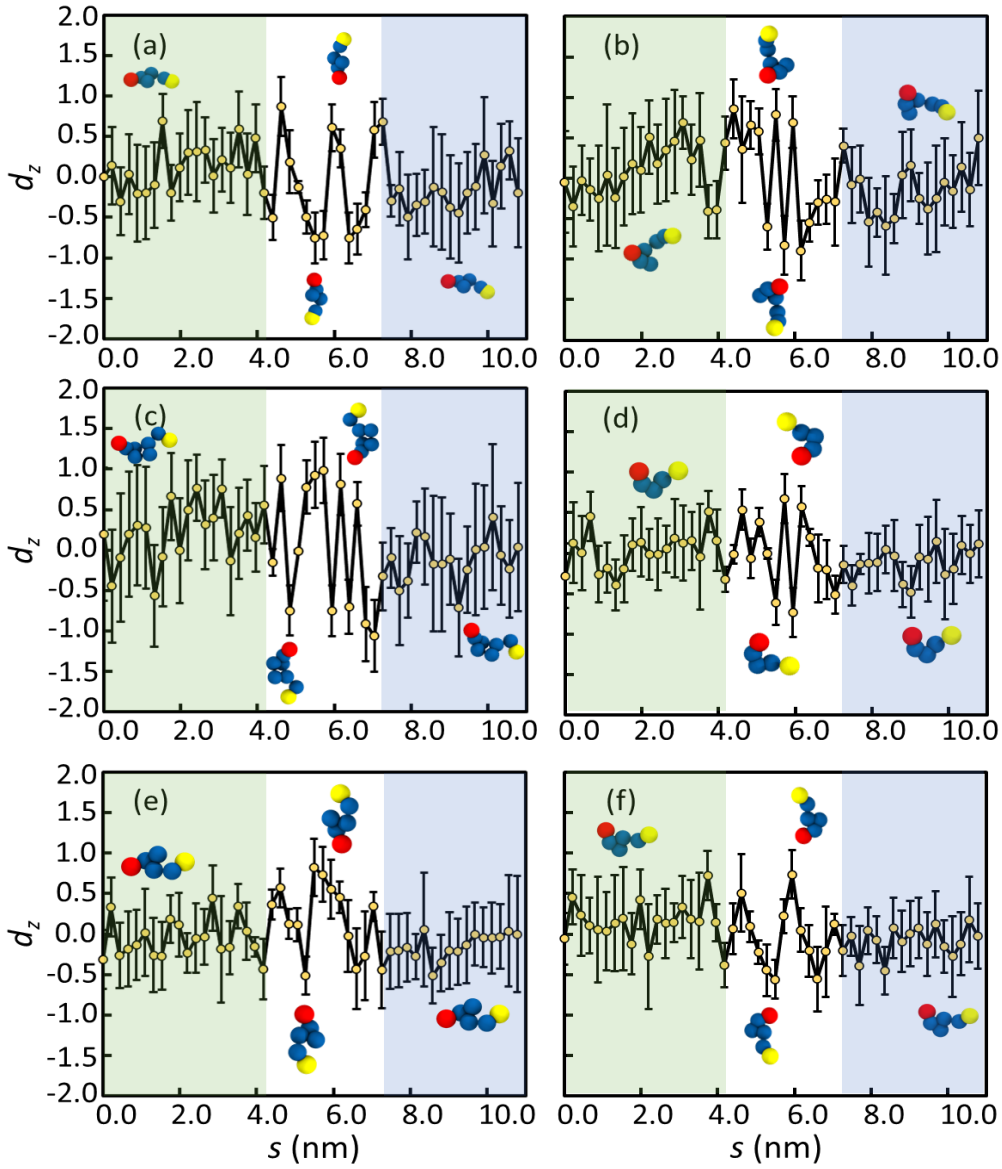


Figure S 19 Orientational analysis of carbapenems along the translocation coordinate in OccD3m. The variation in d_z for (a) biapenem, (b) doripenem, (c) ertapenem, (d) imipenem, (e) meropenem, and (f) panipenem as a function of s. The mean d_z valued are denoted

2.9 References

1. Davies, J.; Davies, D., Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews* **2010,** *74*, 417.

2. Blair, J. M. A.; Webber, M. A.; Baylay, A. J.; Ogbolu, D. O.; Piddock, L. J. V., Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology* **2015,** *13*, 42-51.

3. Frieri, M.; Kumar, K.; Boutin, A., Antibiotic resistance. *Journal of Infection and Public Health* **2017,** *10*, 369-378.

4. Martens, E.; Demain, A. L., The antibiotic resistance crisis, with a focus on the united states. *Journal of Antibiotics* **2017,** *70*, 520-526.

5. Levy, S. B.; Marshall, B., Antibacterial resistance worldwide: Causes, challenges and responses. *Nature Medicine* **2004,** *10*, S122-S129.

6. Luepke, K. H.; Suda, K. J.; Boucher, H.; Russo, R. L.; Bonney, M. W.; Hunt, T. D.; Mohr, J. F., Past, present, and future of antibacterial economics: Increasing bacterial resistance, limited antibiotic pipeline, and societal implications. *Pharmacotherapy* **2017,** *37*, 71-84.

7. Andersson, D. I.; Hughes, D., Antibiotic resistance and its cost: Is it possible to reverse resistance? *Nature Reviews Microbiology* **2010,** *8*, 260-271.

8. Brooks, B. D.; Brooks, A. E., Therapeutic strategies to combat antibiotic resistance. *Advanced Drug Delivery Reviews* **2014,** *78*, 14-27.

9. Rossiter, S. E.; Fletcher, M. H.; Wuest, W. M., Natural products as platforms to overcome antibiotic resistance. *Chemical Reviews* **2017,** *117*, 12415-12474.

10. Tacconelli, E.; Carrara, E.; Savoldi, A.; Harbarth, S.; Mendelson, M.; Monnet, D. L.; Pulcini, C.; Kahlmeter, G.; Kluytmans, J.; Carmeli, Y.; Ouellette, M.; Outterson, K.; Patel, J.; Cavaleri, M.; Cox, E. M.; Houchens, C. R.; Grayson, M. L.; Hansen, P.; Singh, N.; Theuretzbacher, U.; Magrini, N.; Workin, W. H. O. P. P. L., Discovery, research, and development of new antibiotics: The who priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infectious Diseases* **2018,** *18*, 318-327.

11. Delcour, A. H., Outer membrane permeability and antibiotic resistance. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2009,** *1794*, 808-816.

12. Nikaido, H., Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews* **2003,** *67*, 593-+.

13. Vergalli, J.; Bodrenko, I. V.; Masi, M.; Moynié, L.; Acosta-Gutiérrez, S.; Naismith, J. H.; Davin-Regli, A.; Ceccarelli, M.; van den Berg, B.; Winterhalter, M.; Pagès, J.-M., Porins and small-molecule translocation across the outer membrane of gram-negative bacteria. *Nature Reviews Microbiology* **2020,** *18*, 164-176.

14. Liu, J.; Wolfe, A. J.; Eren, E.; Vijayaraghavan, J.; Indic, M.; van den Berg, B.; Movileanu, L., Cation selectivity is a conserved feature in the occd subfamily of pseudomonas aeruginosa. *Biochim Biophys Acta* **2012,** *1818*, 2908-2916.

15. Dam, S.; Pages, J. M.; Masi, M., Stress responses, outer membrane permeability control and antimicrobial resistance in enterobacteriaceae. *Microbiology (Reading, England)* **2018,** *164*, 260-267.

16. Nikaido, H., Porins and specific channels of bacterial outer membranes. *Molecular Microbiology* **1992,** *6*, 435-442.

17. Pages, J. M.; James, C. E.; Winterhalter, M., The porin and the permeating antibiotic: A selective diffusion barrier in gram-negative bacteria. *Nature Reviews Microbiology* **2008,** *6*, 893-903.

18. Ma, H. L.; Khan, A.; Nangia, S., Dynamics of ompf trimer formation in the bacterial outer membrane of escherichia coli. *Langmuir* **2018,** *34*, 5623-5634.

19. Parkin, J.; Khalid, S., Atomistic molecular-dynamics simulations enable prediction of the arginine permeation pathway through occd1/oprd from pseudomonas aeruginosa. *Biophysical Journal* **2014,** *107*, 1853-1861.

20. Samanta, S.; Scorciapino, M. A.; Ceccarelli, M., Molecular basis of substrate translocation through the outer membrane channel oprd of pseudomonas aeruginosa. *Physical Chemistry Chemical Physics* **2015,** *17*, 23867-23876.

21. Samsudin, F.; Khalid, S., Movement of arginine through oprd: The energetics of permeation and the role of lipopolysaccharide in directing arginine to the protein. *Journal of Physical Chemistry B* **2019,** *123*, 2824-2832.

22. Somboon, K.; Niramitranon, J.; Pongprayoon, P., Probing the binding affinities of imipenem and ertapenem for outer membrane carboxylate channel d1 (occd1) from p. Aeruginosa: Simulation studies. *Journal of Molecular Modeling* **2017,** *23*.

23. Eren, E.; Parkin, J.; Adelanwa, A.; Cheneke, B.; Movileanu, L.; Khalid, S.; van den Berg, B., Toward understanding the outer membrane uptake of small molecules by pseudomonas aeruginosa. *The Journal of biological chemistry* **2013,** *288*, 12042-53.

24. Pothula, K. R.; Solano, C. J. F.; Kleinekathöfer, U., Simulations of outer membrane channels and their permeability. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016,** *1858*, 1760-1771.

25. Bajaj, H.; Scorciapino, M. A.; Moynie, L.; Page, M. G.; Naismith, J. H.; Ceccarelli, M.; Winterhalter, M., Molecular basis of filtering carbapenems by porins from beta-lactam-resistant clinical strains of escherichia coli. *The Journal of biological chemistry* **2016,** *291*, 2837-47.

26. D'Agostino, T.; Salis, S.; Ceccarelli, M., A kinetic model for molecular diffusion through pores. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016,** *1858*, 1772-1777.

27. Subramanian, N.; Condic-Jurkic, K.; Mark, A. E.; O'Mara, M. L., Identification of possible binding sites for morphine and nicardipine on the multidrug transporter p-glycoprotein using umbrella sampling techniques. *J Chem Inf Model* **2015,** *55*, 1202-17.

28. Padhi, S.; Priyakumar, U. D., Urea–aromatic stacking and concerted urea transport: Conserved mechanisms in urea transporters revealed by molecular dynamics. *Journal of Chemical Theory and Computation* **2016,** *12*, 5190-5200.

29. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of chemical physics* **2004,** *120*, 11919-29.

30. Hancock, R. E. W.; Brinkman, F. S. L., Function of pseudomonas porins in uptake and efflux. *Annual Review of Microbiology* **2002,** *56*, 17-38.

31. Strateva, T.; Yordanov, D., Pseudomonas aeruginosa - a phenomenon of bacterial resistance. *Journal of Medical Microbiology* **2009,** *58*, 1133-1148.

32. *Antibiotic resistance threats in the united states*; Centers for Disease Control and Prevention: 2019.

33. Isabella, V. M.; Campbell, A. J.; Manchester, J.; Sylvester, M.; Nayar, A. S.; Ferguson, K. E.; Tommasi, R.; Miller, A. A., Toward the rational design of carbapenem uptake in pseudomonas aeruginosa. *Chemistry & biology* **2015,** *22*, 535-547.

34. Soundararajan, G.; Bhamidimarri, S. P.; Winterhalter, M., Understanding carbapenem translocation through occd3 (opdp) of pseudomonas aeruginosa. *Acs Chemical Biology* **2017,** *12*, 1656-1664.

35. Kohler, T.; Michea-Hamzehpour, M.; Epp, S. F.; Pechere, J. C., Carbapenem activities against pseudomonas aeruginosa: Respective contributions of oprd and efflux systems. *Antimicrobial Agents and Chemotherapy* **1999,** *43*, 424-427.

36. El Amin, N.; Giske, C. G.; Jalal, S.; Keijser, B.; Kronvall, G.; Wretlind, B., Carbapenem resistance mechanisms in pseudomonas aeruginosa: Alterations of porin oprd and efflux proteins do not fully explain resistance patterns observed in clinical isolates. *Apmis* **2005,** *113*, 187-196.

37. Tamber, S.; Ochs, M. M.; Hancock, R. E. W., Role of the novel oprd family of porins in nutrient uptake in pseudomonas aeruginosa. *Journal of Bacteriology* **2006,** *188*, 45-54.

38. Biswas, S.; Mohammad, M. M.; Patel, D. R.; Movileanu, L.; van den Berg, B., Structural insight into oprd substrate specificity. *Nature Structural & Molecular Biology* **2007,** *14*, 1108-1109.

39. Eren, E.; Vijayaraghavan, J.; Liu, J.; Cheneke, B. R.; Touw, D. S.; Lepore, B. W.; Indic, M.; Movileanu, L.; van den Berg, B., Substrate specificity within a family of outer membrane carboxylate channels. *PLoS biology* **2012,** *10*, e1001242.

40. Bonfiglio, G.; Russo, G.; Nicoletti, G., Recent developments in carbapenems. *Expert Opinion on Investigational Drugs* **2002,** *11*, 529-544.

41. Shah, P. M.; Isaacs, R. D., Ertapenem, the first of a new group of carbapenems. *Journal of Antimicrobial Chemotherapy* **2003,** *52*, 538-542.

42. Zhanel, G. G.; Wiebe, R.; Dilay, L.; Thomson, K.; Rubinstein, E.; Hoban, D. J.; Noreddin, A. M.; Karlowsky, J. A., Comparative review of the carbapenems. *Drugs* **2007,** *67*, 1027-1052.

43. Nicolau, D. P., Carbapenems: A potent class of antibiotics. *Expert Opinion on Pharmacotherapy* **2008,** *9*, 23-37.

44. Papp-Wallace, K. M.; Endimiani, A.; Taracila, M. A.; Bonomo, R. A., Carbapenems: Past, present, and future. *Antimicrobial Agents and Chemotherapy* **2011,** *55*, 4943-4960.

45. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23*, 187-199.

46. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., The martini force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B* **2007,** *111*, 7812-7824.

47. Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J., Martini coarse-grained force field: Extension to carbohydrates. *Journal of Chemical Theory and Computation* **2009,** *5*, 3195-3210.

48. Periole, X.; Cavalli, M.; Marrink, S. J.; Ceruso, M. A., Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition. *Journal of Chemical Theory and Computation* **2009,** *5*, 2531-2543.

49. de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J., Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation* **2013,** *9*, 687-697.

50. Marrink, S. J.; Tieleman, D. P., Perspective on the martini model. *Chemical Society Reviews* **2013,** *42*, 6801-6822.

51. Wassenaar, T. A.; Pluhackova, K.; Bockmann, R. A.; Marrink, S. J.; Tieleman, D. P., Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation* **2014,** *10*, 676-690.

52. Ma, H. L.; Irudayanathan, F. J.; Jiang, W. J.; Nangia, S., Simulating gram-negative bacterial outer membrane: A coarse grain model. *Journal of Physical Chemistry B* **2015,** *119*, 14668-14682.

53. Ma, H. L.; Cummins, D. D.; Edelstein, N. B.; Gomez, J.; Khan, A.; Llewellyn, M. D.; Picudella, T.; Willsey, S. R.; Nangia, S., Modeling diversity in structures of bacterial outer membrane lipids. *Journal of Chemical Theory and Computation* **2017,** *13*, 811-824.

54. Graham, J. A.; Essex, J. W.; Khalid, S., Pycgtool: Automated generation of coarse-grained molecular dynamics models from atomistic trajectories. *Journal of Chemical Information and Modeling* **2017,** *57*, 650-656.

55. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

56. Krieger, E.; Vriend, G., New ways to boost molecular dynamics simulations. *Journal of computational chemistry* **2015,** *36*, 996-1007.

57. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., Swiss-model: Homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018,** *46*, W296-W303.

58. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., Charmm-gui: A web-based graphical user interface for charmm. *Journal of computational chemistry* **2008,** *29*, 1859-1865.

59. Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J., Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. *Journal of Chemical Theory and Computation* **2015,** *11*, 2144-2155.

60. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell Jr, A. D., Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry* **2010,** *31*, 671-690.

61. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984,** *81*, 3684-3690.

62. Pavelka, A.; Sebestova, E.; Kozlikova, B.; Brezovsky, J.; Sochor, J.; Damborsky, J., Caver: Algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM transactions on computational biology and bioinformatics* **2016,** *13*, 505-17.

63. Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14*, 33-38.

64. The pymol molecular graphics system, version 2.0 Schrödinger, LLC.

CHAPTER 3.

FINGERPRINTING PLASMA MEMBRANE LIPIDOME OF HUMAN RED BLOOD CELLS USING COMPUTATIONAL MODELING TECHNIQUES

3.1 Abstract

Although asymmetry in plasma membrane leaflets is well-established, the comprehensive understanding of this asymmetry in the physicochemical properties and biological function of the membrane is lacking. Here we report a molecular-level description of the human red blood cell (RBC) plasma membrane that replicates the experimentally determined composition of the lipid families that constitute the exoplasmic and cytoplasmic leaflets. Cholesterol exhibits an affinity for sphingomyelin lipids resulting in a 5:3 distribution between leaflets in favor of the exoplasmic leaflet (EL) leading to the formation of nanodomains. On the other hand, the cytoplasmic leaflet (CL) lacks nanodomains due to the presence of fewer saturated lipids. With the facilitation of cholesterols, EL containing more saturated lipids is more densely packed comparing to CL. We further illustrated this observation with several characterization methods, including Voronoi tessellation, lipid order parameter, and leaflet density map. Furthermore, we performed thermal annealing simulations to study temperature-triggered membrane asymmetry properties. Interestingly, cholesterols flip between the leaflets at an increasing rate while the overall distribution maintains the same. We noticed that the membrane lipids can be capable of tolerating moderate temperature changes from the environment. In summary, fingerprinting the RBC plasma membrane lipidome elucidates the complex relationships among different lipid families in conferring asymmetric properties to the membrane.

3.2 Introduction

Advances in membrane lipidomics have revealed a rich chemical diversity and leaflet asymmetry in eukaryotic cells.[66-73] These observations contrast to the rudimentary fluid-mosaic model of cell membranes that depicted the membrane as a simple mixture of lipids that accommodated membrane proteins.[74-75] Besides being a cell's first line of defense, the plasma membrane is a highly functional interface where many signal transduction events occur mediated via the peripheral and membrane-embedded proteins.[76-80] The plasma membrane leaflet asymmetry is evident in various eukaryotic cells—red blood cells (RBCs), platelets, and neuronal cells.

The lipid diversity manifests in two ways: (i) chemical structure of the lipids—charge, and type of the headgroups, length of the acyl chains, and the number and location of double bonds in the acyl chain backbone; and (ii) compositional diversity—the proportion of lipids in membrane leaflets.[81] These variations confer each leaflet its distinct biophysical property to interface with the cell's external and internal environment.[69-70] Alteration in leaflet composition can lead to a loss in cell function and even apoptosis.[70,82]

A recent study reported human RBC plasma membrane's leaflet asymmetry and distinct compositions of the exoplasmic and cytosolic leaflets.[83] The results showed that compared to the cytosolic leaflet (CL), the exoplasmic leaflet (EL) is more densely packed and rich in lipids with a higher percentage of saturated lipids. In particular, the EL is composed of phosphatidylcholine (PC), sphingomyelins (SM), and a small percentage of phosphatidylserine (PS) and phosphatidylethanolamine plasmalogen (PEP). In contrast, the CL has higher lipid head group diversity, including phosphatidylethanolamine (PE), PEP, PS, PC, SM, and phosphatidylinositols (PI). Although cholesterol was reported in both leaflets, the experimental analysis did not directly

provide the cholesterol distribution.[83] The detailed lipid distribution within the major lipid families is reproduced in Figure 3-1.

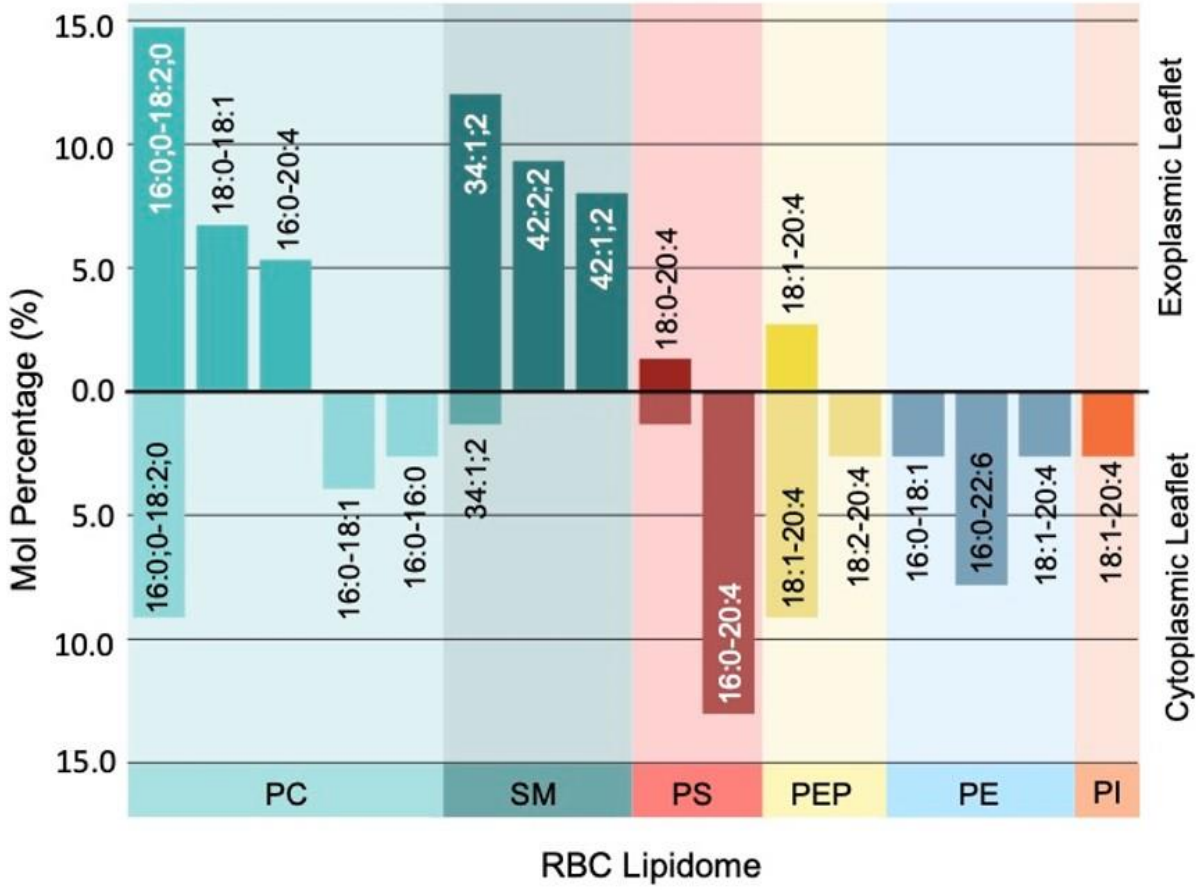


Figure 3-1. RBC plasma membrane composition distribution in exoplasmic leaflet and cytoplasmic leaflet. Lipids separated by category, phosphatidylcholine (PC, cyan), sphingomyelin (SM, dark cyan), phosphatidylserine (PS, red), phosphatidylethanolamine (PE, blue), phosphatidylethanolamine plasmalogen (PEP, yellow), and phosphatidylinositol (PI, orange). Lipid species annotated as (total number of carbon atoms in the acyl chain):(total number of double bonds in acyl chains);(total number of hydroxyl groups in the long acyl chain).

Despite the evidence of plasma membrane leaflet asymmetry, a comprehensive understanding of the RBC membrane leaflet's physicochemical properties and biological function is lacking. The unique lipid composition of each leaflet impacts the membrane's biophysical properties: cholesterol distribution (flip-flop rate and domain formation); membrane density (area per lipid); surface charge (lipid headgroup distribution); membrane protein distribution (protein

hydrophobic thickness and shape); and membrane protein's post-translational modifications (CL versus EL).

Several computational studies have focused on the structure and dynamics of model cell membranes with variable lipid complexities and compositions.[84-101] Ingólfsson et al. modeled an idealized plasma membrane consisting of 63 different lipid species that were asymmetrically distributed across the two leaflets.[91] Other computational studies have investigated asymmetric cholesterol distribution in the lipid bilayers.[102-108] A recent study correlated cholesterol distribution with the leaflet's lipid order; the more ordered the leaflet, the higher its cholesterol concentration—the type of lipids in the leaflet influence the cholesterol distribution and the flip-flop rate between the leaflets.[108] The model membranes have been foundational in deciphering molecular level lipid-lipid interactions.

However, the human RBC plasma membrane lipidome is more complex than an idealized model membrane previously investigated by computational techniques. The recently reported RBC membrane lipidome's unique feature is the highly saturated PC and SM lipids, which constitute ~56 mol% of the EL.[83] The abundance of these lipid acyl chains and their molecular interplay within the EL and lipids of the opposing CL remains unexplored.

Here, we investigate the structural and compositional diversities in RBC membrane leaflets using molecular dynamics at the coarse-grained (CG) resolution. We employed the Martini force field parameters,[47, 50, 113] which have been widely used to investigate biomembranes[19, 51, 53-54, 94, 108, 114] and proteins.[50-51, 108, 115-121] We modeled a subset of SM, PE, PEP, and PI lipids unavailable in the Martini force field. The results show the membrane asymmetry in terms of the lipid order parameters in each membrane leaflet, cholesterol distribution, leaflet densities, and membrane

thickness. The methods used and the results of the simulations are provided in sections 3.3 and 3.4, respectively.

3.3 Methods

Martini Force Field Parameters. The Martini CG parameters are based on a many-to-one mapping scheme, in which four neighboring non-hydrogen atoms are mapped into one CG bead.[47-48, 50, 52, 113] Martini has some flexibility to adopt a three-to-one or five-to-one mapping scheme, if required, to preserve some properties in the chemical structures. The RBC membrane constitutes of several lipids (Table S1) with the following lipid families: PC (PIPC, PAPC, POPC, and DPPC); SM (DPSM, PLSM, and PNSM); PEP (PEP1 and PEP2); PE (POPE, PUPE, and PEPE); PS (PAPS); and PI (PIP2).

The Martini force field parameters are available for most of the RBC membrane lipids (PIPC, PAPC, POPC, DPPC, DPSM, PNSM, POPE, PUPE, PAPS, and CHOL). For the remaining lipids, the parameters were constructed using the Martini headgroup and acyl chain bead types. Unlike the available PE lipids, the PEP lipids possess a vinyl-ether linkage at the sn-1 position instead of an ester linkage and an ester linkage at the sn-2 position,[122] for which Martini parameters had to be developed. The parameter sets for the newly developed lipids—PLSM, PEP1, PEP2, PEPE, and PIP2 (Table 3-2 – 3-6) are available for download from our research website.

Lipid Membrane. A 30×30 nm^2 patch membrane mimicking the human RBC plasma membrane composition was constructed in the CG presentation using a locally modified *insane.py* script.[60] The newly developed lipids are included in this modified *insane.py* script. The membrane EL and CL consisted of distinct lipid compositions. The membrane was

placed in a 30×30×15 nm^3 simulation box and solvated with standard Martini water (9:1 of W:WF) and 150 mM NaCl. The molecular composition of the simulation box is provided in Table 3-7.

CG MD Simulations. The GROMACS 5.1.2 package[56] was used for all MD simulations. The system was energy minimized until the maximum force on any bead was below the tolerance parameter of 10 $kJmol^{-1}nm^{-1}$. The minimization was followed by equilibration in the isothermal-isochoric (*NVT*) conditions for 25 ns, followed by a two-step isothermal-isobaric *NPT* (*NPT1* and *NPT2*) equilibration for 30 and 45 ns, respectively. Position restraints were used at lipid PO4 headgroup beads in *NVT*, *NPT1*, and *NPT2* with a force constant set at 100, 10, and 5 $kJmol^{-1}nm^{-2}$, respectively. A 20-fs time step was used for the equilibration and production runs. The temperature was maintained at 300 K for all systems using the v-rescale[123] thermostat with τ_t = 1 ps. A semi-isotropic pressure coupling of 1 bar was maintained using Parrinello-Rahman barostat[124] with τ_p = 8 ps. Both the nonbonded van der Waals and the electrostatic interaction cut-offs were set to 1.1 nm. The potential-shift-Verlet algorithm was applied to shift the van der Waals interactions beyond the cut-off. The Coulombic interactions were calculated using the reaction-field algorithm. Periodic boundary conditions were applied in all three dimensions. All position restraints were removed in the production runs. Flat-bottomed position restraints were applied to the phosphate beads of the lipid at 2.5 nm above and below the membrane with force constant of 1000 $kJmol^{-1}nm^{-2}$. The simulations were performed in triplicates, each for five microseconds, determined by the convergence of the mixing parameters (Figures 3-8 – 3-14). Thermal annealing simulations were performed from 280 K to 320 K with 5 K interval; each temperature step was equilibrated for 500 ns.

3.4 Results and Discussion

3.4.1 Plasma membrane lipid families confer distinct physicochemical properties to the EL and CL leaflets

The physicochemical properties of the RBC membrane were evaluated based on lipid-lipid interactions among the different lipid families (Figure 3-17). The lateral assembly of a lipid within the family (self-association) or with other lipid families (co-localization) was determined in two ways—local area density and lipid mixing parameters. The local area density was computed for each lipid family in a leaflet and is represented by the Voronoi plots.[125] If the area density of two or more lipid families coincided in a region, it suggested co-localization. In addition, if the area density was contiguous over an extended patch of the membrane, it indicated domain formation. On the other hand, the lipid mixing parameters provided a comparison of a lipid's affinity to establish self and cross interactions among various lipid types (Section 3.7.1). For example, in a well-equilibrated, homogenous lipid membrane, each lipid type will have an equal representation of all other lipids within its interacting distance. Conversely, a domain-forming lipid will exclude different lipid types from its interaction radius to the greatest extent possible.

The SM and PC families, key lipid families in the RBC membrane's EL, exhibit a heterogeneous lipid distribution (Figure 3-15). The SM lipids tend to self-associate and co-localize with cholesterol to form nanodomains. These nanodomains do not extend more than 3-4 nm and are separated by PC lipids (Figure 3-2a). The SM's tendency to form nanodomains is supported by the mixing parameters analysis. SM lipids have the highest proportion of self-association (44%) followed by a mixing affinity with cholesterol (30%) and PC (22%). The other lipid families

collectively represent <4% mixing. The SM concentration in the CL is negligible to form clusters (Figure 3-2a).

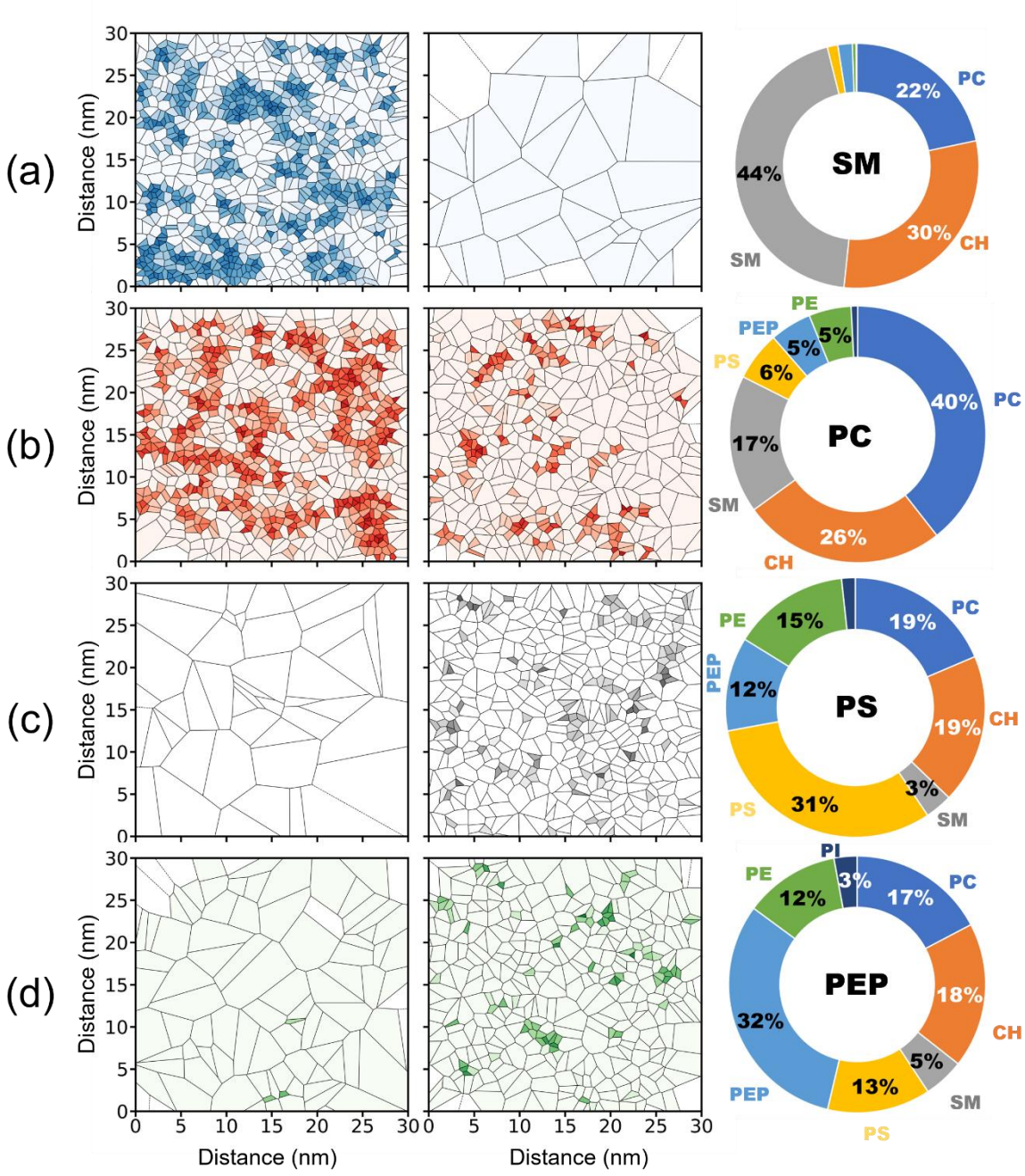


Figure 3-2. Voronoi tessellation of (a) SM, (b) PC, (c) PS, and (d) PEP lipids in the EL (left panel) and CL (center panel). The lipid density varies from high (darker shades) to low (lighter shades). The mixing parameters (right panel) of the lipid within its family and with other lipid families is shown in the donut charts. The name of each reference lipid is in the center of its respective donut. Lipid families that have low mixing (< 3%) with the reference lipid are omitted for clarity.

The PC lipids self-associate to a lesser extent than the SM. As a result, the PC lipids in the EL fill the area not occupied by the SM nanodomains leading to some local clustering (Figure 3-2b). The PC clustering in the CL is characteristically distinct from the EL; here the PC lipids are much more dispersed due to the lack of SM nanodomains in the CL and lower mole fraction. The remaining lipid families in the EL have a less significant contribution to the exterior surface of the RBC.

The negatively charged PS lipids have a self-associating affinity, but their clustering is not continuous to form domains (Figure 3-15). The PS lipids have almost equal mixing with PC, PE, and PEP lipids as indicated by their mixing parameters (Figure 3-2c).

The PE lipids are exclusively in CL and heterogeneously distributed with minimal self-association or clustering. The PE molecules do not exhibit affinity to any specific lipid family in the CL (Figure 3-3a). The PEP is a subclass of the PE lipid that shares the same headgroup but has a vinyl ether linkage for one of its carbon chains. The area density and mixing parameters mimic the PE lipid properties in the CL (Figure 3-2d). PEP constitutes less than 3% of the EL lipids and similar to PS, lacking the nanodomain formation in the EL (Figure 3-2d).

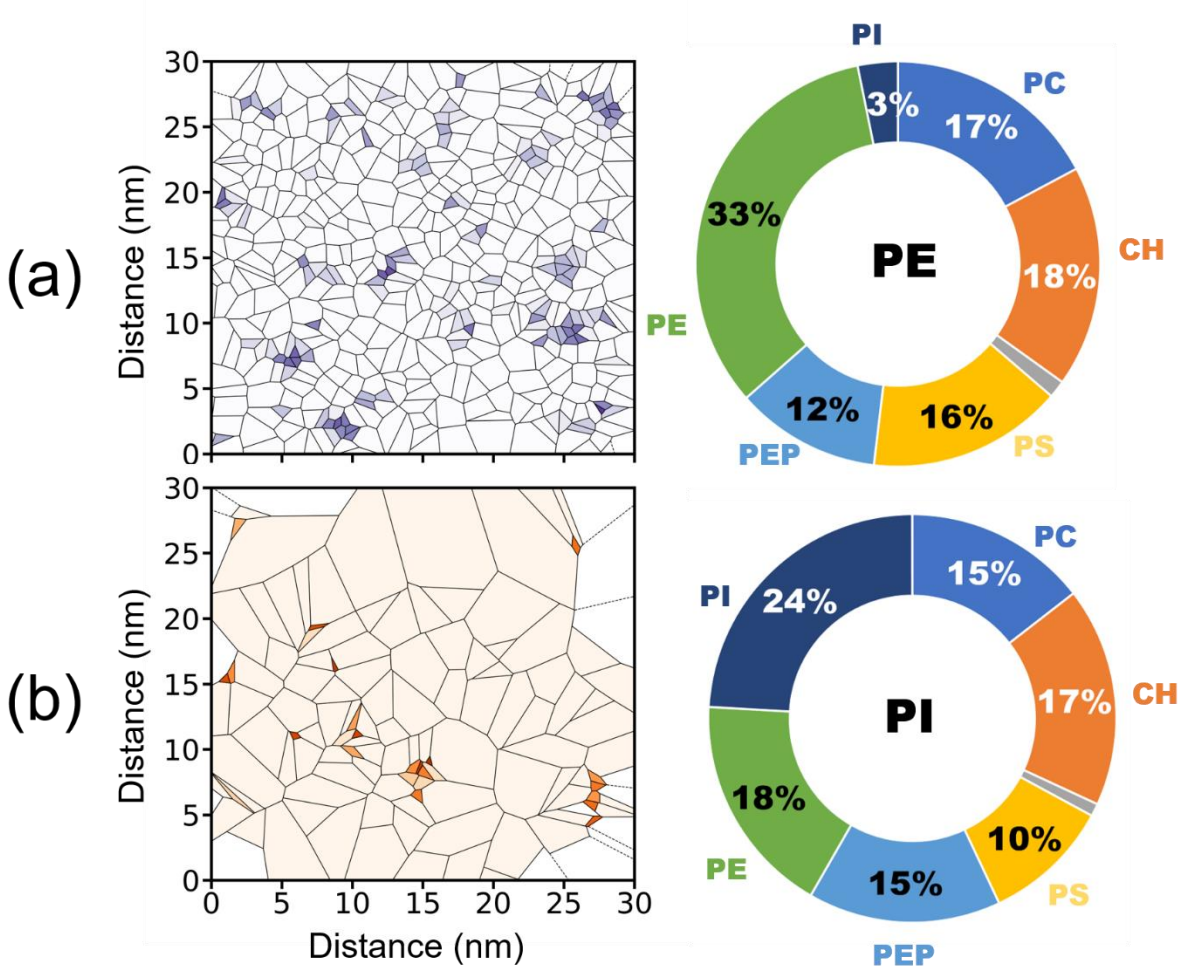


Figure 3-3 Voronoi tessellation of (a) PE and (b) PI lipids in the CL. The lipid density varies from high (darker shades) to low (lighter shades). The mixing parameters of the lipid within its family and with other lipid families is shown in the donut charts. The name of each reference lipid is in the center of its respective donut. Lipid families that have low mixing (< 3%) with the reference lipid are omitted for clarity.

Besides PE lipids, PI is also exclusive in the CL. The highly charged PI lipids have been the focus of several computational studies.[126-127] The PI lipids are in tiny mole fractions in the CL and do not exhibit colocalization with any lipid family. A few examples of PI clusters in the membrane protein interface have been reported.[128-129] The PI family forms non-specific interactions with the membrane lipids, which is evident in the Voronoi plot and the mixing parameters (Figure 3-3b).

The results display that human RBC plasma membrane leaflets have very distinct surface properties. The EL has distinct SM nanodomains while the CL has a near random distribution of PS, PE, and PEP lipid families.

3.4.2 Membrane cytoplasmic leaflet has lower lipid order than the exoplasmic leaflet

Membrane asymmetry is reflected in terms of membrane composition. Comparing lipid orders in EL and CL is another approach to show lipid packing in this asymmetric RBC plasma membrane. The order parameter, S_{CD}, was generated for the acyl chain beads for each lipid. Within each leaflet, the cumulative S_{CD} is calculated from the same set of carbon beads (Figure 3-4). The S_{CD} decreases as the acyl chain beads of interest move from headgroups towards lipid tails, meaning the farther the carbon bead is located away from the headgroups, the less ordered it gets. This trend holds true for both leaflets. Overall, EL has a larger S_{CD} and contains more ordered lipids than the CL, which is consistent with the latest reports.[83,108]

After analyzing order parameters from the perspective of position of the carbon beads, we computed the order parameter of each acyl chain in the lipid molecules (Table 3-8). It is evident from the order parameters that saturated and unsaturated acyl chains have remarkably different contribution to a lipids shape and packing in the membrane (Figure 3-16). The box plots of lipids in EL and CL are shown in Figure 3-5a and Figure 3-5b, respectively. For example, PAPC's 16:0 acyl chain is much more ordered (0.52±0.258) than the 20:4 (0.13±0.214) chain. There is remarkable disparity in the median values of the PUPE lipid's order parameters; the shorter saturated (16:0) chain is ordered (0.41±0.281), while the longer, unsaturated (22:6) is highly disordered (0.04±0.196). The PEP1 and PEP2 acyl chains also show differences in the vinyl ester

linked chain versus the ether linked acyl chain. Interestingly, the saturated DPPC lipid (16:0-16:0), sequestered in the CL in the plasma membrane, has high order parameters (0.49±0.272 and 0.46±0.274) for both acyl chains. Overall, the boxes in the EL are located higher than those in the CL, which can be used to reflect that the order parameters of lipids in the EL in general are higher than those in the CL. This observation is consistent with the profile from Figure 3-4, which again confirms that lipids in the EL are more ordered.

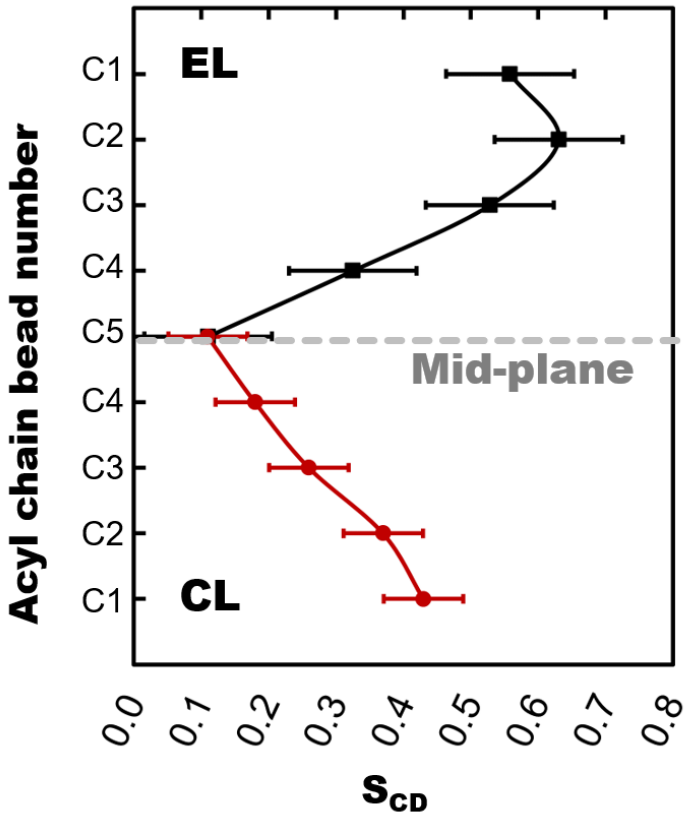


Figure 3-4 Order parameters of all lipids in EL (black) and CL (red), excluding cholesterols. The dashed line (gray) indicates the membrane's mid-plane.

The statistical analysis of the lipid order parameters highlight three important aspects: (a) the lipids adopt ordered-disordered-ordered layering of the membrane in EL, mid-plane, CL direction; (b) on average lipids in EL are more ordered than those in CL; (c) between the two acyl

chains of a lipid, the shorter and saturated acyl chain is more ordered than the longer, unsaturated acyl chain.

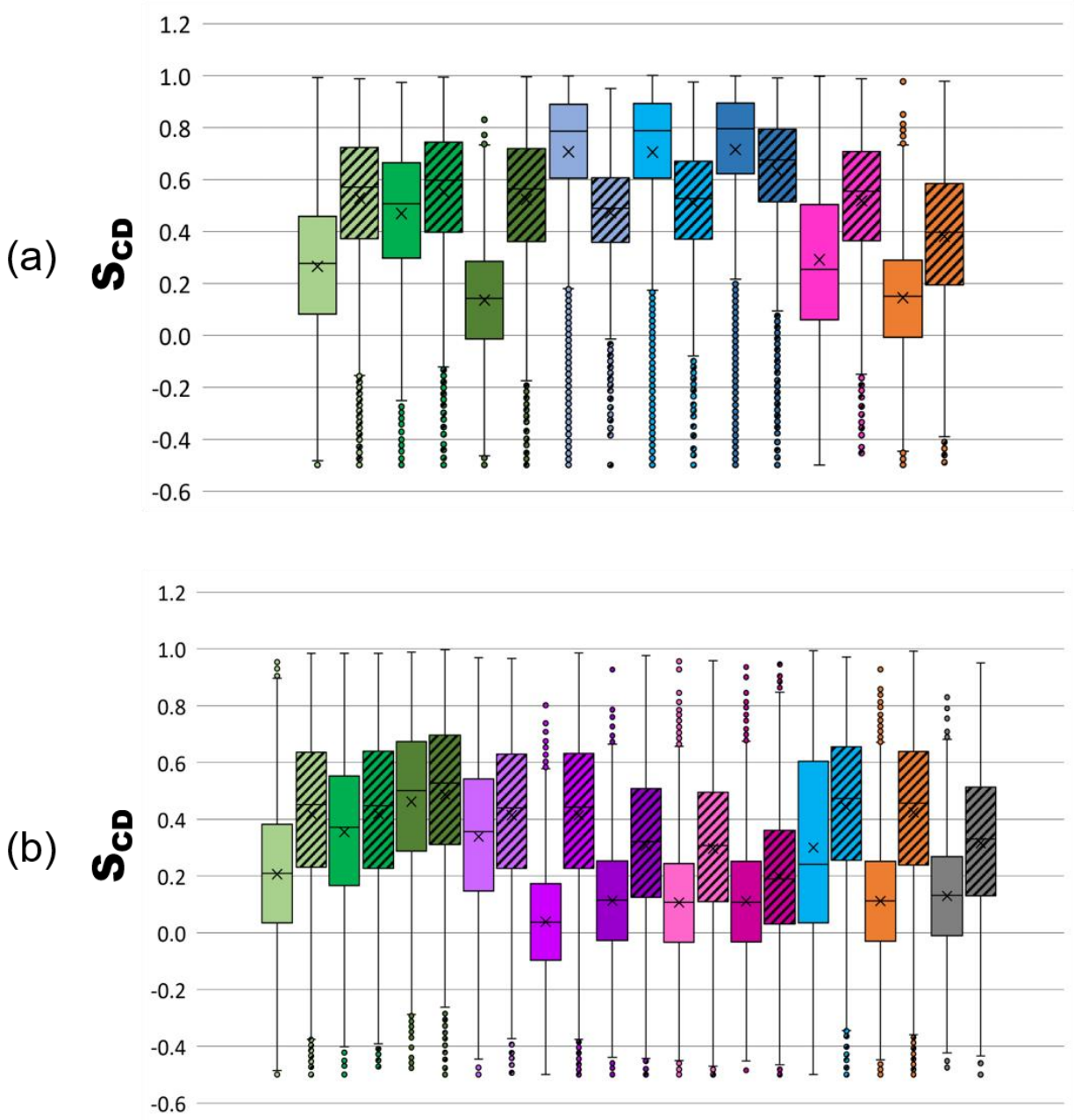


Figure 3-5 Order parameters of all lipids in EL (a) and CL (b) are shown as a box-whisker plot. Lipid species in (a) from left to right are PIPC, POPC, PAPC, PNSM, PLSM, DPSM, PAPS, and PEP1. Lipid species in (b) from left to right include PIPC, POPC, DPPC, POPE, PUPE, PEPE, PEP1, PEP2, DPSM, PAPS, and PIP2. The two acyl chains of a lipid are distinguished by solid and striped boxes.

The pairing of short/long, saturated/unsaturated, and charged/zwitterionic lipids along with asymmetric lipid compositions of the leaflets is critical in unique biophysical properties of the RBC plasma membrane.

3.4.3 Cholesterol flips between leaflets with preference for the exoplasmic leaflet

Several simulation studies have focused on evaluating the interaction of cholesterol with the lipids and the flip-flop dynamics between the membrane leaflets. Analyzing this dynamic behavior of cholesterol is experimentally challenging due to multiple confounding factors.[105, 130-131] The use of computational tools can capture the cholesterol flip-flop behavior.[102-104, 106-108]

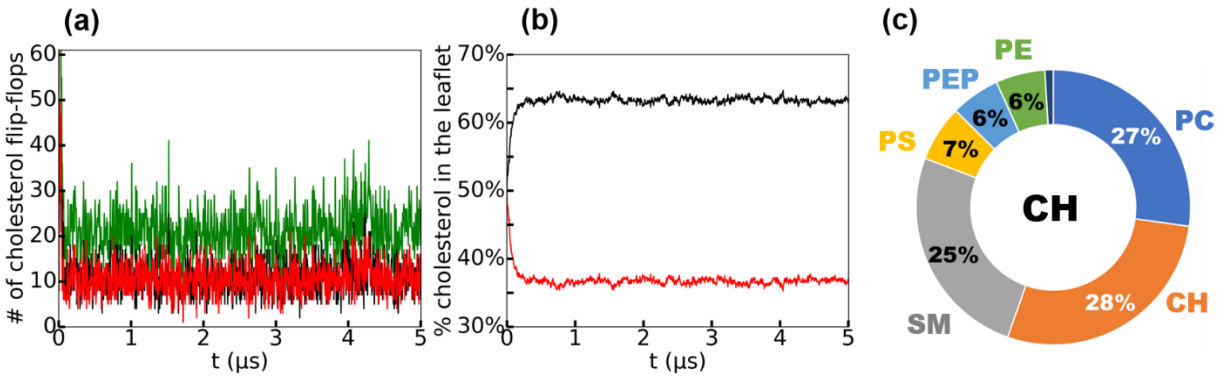


Figure 3-6 Quantitative analysis of cholesterol partitioning in the RBC membrane. (a) Number of cholesterol flips from CL to EL (back), from EL to CL (red), and the total number of flip-flops (green) as a function of simulation time. (b) Percent cholesterols in EL (black) and in CL (red) over the five microseconds. (c) Mixing parameters of cholesterol with other lipid families.

In the initial system set-up, cholesterol was distributed equally in each leaflet. In the pre-equilibrium NVT and NPT stages of the simulation, more molecules of cholesterol flipped from the CL to the EL than vice-versa. At equilibrium, the rate of cholesterol flip-flop between the two leaflets reached a stable distribution (Figure 3-6a and 3-6b); the EL and CL cholesterol distribution is 63±3% and 37±3%, respectively.

Previous experimental and computational studies have demonstrated the preferred interaction of cholesterol with the saturated lipids.[133] It has been suggested that the interaction between cholesterol molecules and SM lipids lead to the formation of lipid domains.[134-135] The lipid mixing parameters revealed the lateral association of cholesterol with membrane lipids: SM (27%), PC (25%), cholesterol (28%), PS (7%), PE (6%), PEP (6%), and PI (1%), as shown in Figure 3-6c. The preferential association of cholesterol with the SM and PC lipids is established within the first few microseconds of the simulations and remains unchanged after equilibration. The triplicate simulation runs show the same cholesterol lipid mixing trends. Thus, cholesterol localization occurs disproportionately in the SM-enriched EL.

3.4.4 Exoplasmic and cytosolic leaflets differ as temperature changes

We performed thermal annealing simulations to further characterize membrane asymmetry properties. First, EL and CL have different area per lipid values regardless of temperature. Due to variations in shapes and surface areas of different lipids, the EL has a smaller area per lipid (0.45± 0.04 lipids nm^{-2}) relative to the CL (0.55± 0.03 lipids nm^{-2}), which is consistent with earlier reports. The change in the area per lipid versus temperature in 280-320 K was evaluated. The area per lipid values increased linearly for both leaflets (Figure 3-7a); in the meantime, the rate of increase of area per lipid for EL is nearly identical to that of CL. On the other hand, as the temperature increases, lipids become more fluid and dynamic, which leads to the lateral expansion of the membrane patch and the transverse reduction in membrane bilayer thickness (Figure 3-7b).

The observations match well with each other, which show that the entire membrane stretch and become thinner in response to increasing temperature.

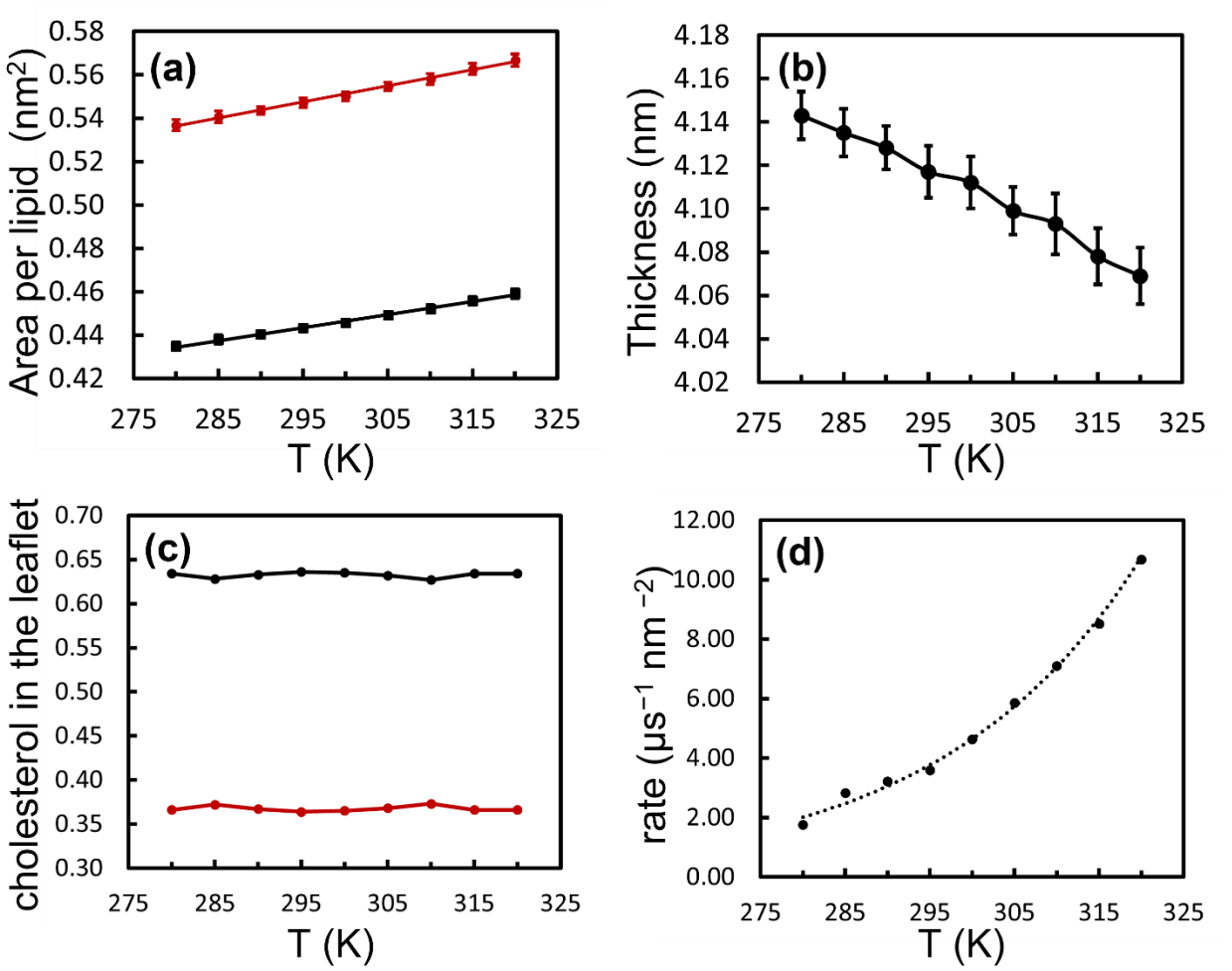


Figure 3-7 Variations in the area per lipid of EL (black) and CL (red), and thickness as a function of temperature in (a) and (b), respectively. (c) Cholesterol distribution between EL (black) and CL (red) as temperature increases from 280K to 320K. (d) Cholesterol flip-flop rate over the same range of temperatures.

Furthermore, we characterized the cholesterol dynamic behavior in terms of temperature. We noticed that the cholesterol distribution between leaflets remains the same (63±3% in EL and 37±3% in CL) as temperature spans from 280K to 320K (Figure 3-7c), which can be attributed to the unchanged lipid composition in each leaflet. This characterization can be used to prove that the major determinant of cholesterol distribution between asymmetric membrane leaflets is the respective composition, especially considering several key lipid families, such as SM and PC

lipids containing saturated acyl chains with higher affinity to cholesterols. However, we did observe that despite the constant cholesterol distribution in EL and CL, there are more cholesterols flipping between leaflets as the temperature increases (Figure 3-7d). It is obvious that the rate of cholesterol flip-flop is increasing exponentially from 2 $\mu s^{-1}nm^{-2}$ to 11 $\mu s^{-1}nm^{-2}$ at 280K to 320K, respectively. Therefore, increasing the temperature has a large impact on cholesterols flipping more rapidly even though the overall distribution remains unchanged between asymmetric leaflets.

3.5 Conclusions

We report a detailed molecular description of the human RBC plasma membrane that replicates the experimentally determined lipid composition of each of the two asymmetric membrane leaflets. Interestingly, cholesterol has a high binding affinity for SMs that results in an asymmetric 5:3 distribution of cholesterol in favor of the exoplasmic leaflet. The SM and PC lipids , and cholesterol lateral associate to form nanodomains in the exoplasmic leaflet. In contrast, the cytoplasmic leaflet has a near-random distribution of the PC, PS, PE, PEP, and PI families. All lipid families have the highest mixing parameter with the lipids of their species. As temperature increases, membrane bilayer becomes thinner and the area of the membrane patch stretches, resulting in larger area per lipid in both EL and CL. In addition to the increasing temperature, the rate of cholesterol flip-flop inreases while maintaining a constant distribution between the asymmetric leaflets. Fingerprinting the plasma membrane lipidome is pivotal in elucidating how cells exchange metabolites, transfer signals, and facilitate protein assembly.

3.6 Supporting Information

3.6.1 Methods and Analysis

Analyses were performed using multiple, in-house developed python scripts. Molecular visualization and graphics were generated using VMD and PyMOL.

Mixing parameter. The mixing parameter of membrane lipids is calculated as $p_{ij} = \frac{C_{ij}}{\sum C_{ik}}$, where p_{ij} is the mixing parameter between lipid i and lipid j, C_{ij} is the number of contacts made by lipid i with lipid j, and $\sum C_{ik}$ is the total contacts made by lipid i with all surrounding lipids, including itself. A contact is defined to have formed between two lipid molecules if the distance between a lipid species and the reference lipid is within a 1.1 nm cut-off.

Order parameter. The order parameter of lipid acyl chain is calculated from ^{2}H NMR as $S_{\mathrm{CD}} = \langle\frac{3\cos^2\theta - 1}{2}\rangle$, where S_{CD} is the order parameter between the carbon-hydrogen (C–D) bond, θ is the angle between the C–D bond vector and a reference axis, which is the membrane normal $x, y, z = 0, 0, 1$. The angular brackets determine that order parameters are weighted over time. Two types of order parameters were evaluated in this work, in viewpoint of carbon beads, order parameter was calculated between every two adjacent carbon beads and averaged over lipids of the exoplasmic leaflet (EL) and the cytoplasmic leaflet (CL), respectively. For the other purpose of analysis, each entire lipid acyl chain was calculated with one order parameter generated, averaging over lipids of the same type in EL and CL, respectively.

Density analysis. The number density of CG beads in EL and CL were calculated using the MDAnalysis library. The density profile is approximated by discretizing the simulation box into areas of $1nm \times 1nm$ and calculating the local number density in each region.

Voronoi tessellation. For each lipid type (PSM, PC, PI, PS, PE, PEP), the (x, y) coordinates of their C1, D1 and T1 beads on either of the leaflets were obtained using MDAnalysis package. These coordinates were used as input points to compute Voronoi diagram using scipy python package. The vertices of each Voronoi region were acquired from the scipy Voronoi object, and these vertices were used to calculate the polygon area using the "shoelace formula". The individual regions of the Voronoi diagrams were then colored based on their respective areas.

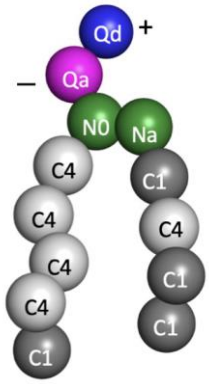
3.6.2 Supporting Tables

Table 3-1 Lipid names, corresponding head groups and tails, and assigned four letter names

Lipid names	Corresponding head groups and tails	Four letter names
1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine	PC 16:0-18:2	PIPC
1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphocholine	PC 16:0-20:4	PAPC
1-stearoyl-2-oleoyl-sn-glycero-3-phosphocholine	PC18:0-18:1	POPC
1-palmitoyl-2-palmitoyl-d31-sn-glycero-3-phosphocholine	PC 16:0-16:0	DPPC
N-Palmitoylsphingomyelin	SM 34:1;2	DPSM
Lysosphingomyelin	SM 42:1;2	PLSM
Neutral sphingomyelinase	SM 42:2;2	PNSM
1-stearoyl-2-arachidonoyl-sn-glycero-3-phospho-L-serine	PS 18:0-20:4	PAPS
1-(1-Enyl-stearoyl)-2-adrenoyl-sn-glycero-3-phosphoethanolamine	PE-O 18:1-20:4	PEP1
1-(1Z-octadecenyl)-2-arachidonoyl-sn-glycero-3-phosphoethanolamine	PE-O 18:2-20:4	PEP2
1-palmitoyl-2-docosahexaenoyl-sn-glycero-3-phosphoethanolamine	PE 16:0-22:6	PUPE
1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-arachidonoyl	PE 18:1-20:4	PEPE
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine	PE 16:0-18:1	POPE
1-stearoyl-2-adrenoyl-sn-glycero-3-phosphoinositol	PIP2 18:1-20:4	PIP2

Table 3-2 MARTINI CG beads and parameters for PEP1

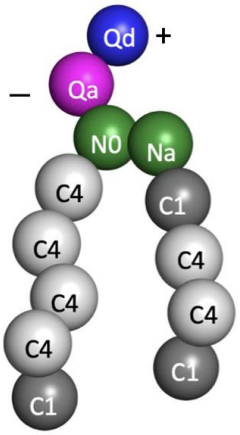
Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.47	1250
2-3	0.47	1250
3-4	0.37	1250
3-5	0.47	1250
5-6	0.47	1250
6-7	0.47	1250
7-8	0.47	1250
8-9	0.47	1250
4-10	0.47	1250
10-11	0.47	1250
11-12	0.47	1250
12-13	0.47	1250



Angle	θ (deg)	K_{angle} (kJ mol^{-1})
2-3-4	120	25
2-3-5	180	25
3-5-6	100	10
5-6-7	100	10
6-7-8	100	10
7-8-9	120	45
4-10-11	180	25
10-11-12	120	45
11-12-13	180	25

Table 3-3 MARTINI CG beads and parameters for PEP2

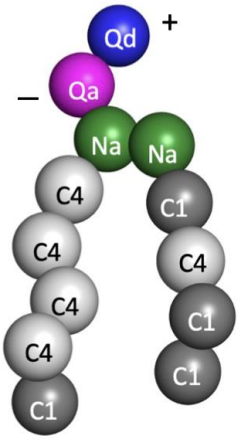
Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.47	1250
2-3	0.47	1250
3-4	0.37	1250
3-5	0.47	1250
5-6	0.47	1250
6-7	0.47	1250
7-8	0.47	1250
8-9	0.47	1250
4-10	0.47	1250
10-11	0.47	1250
11-12	0.47	1250
12-13	0.47	1250



Angle	θ (deg)	K_{angle} (kJ mol^{-1})
2-3-4	120	25
2-3-5	180	25
3-5-6	100	10
5-6-7	100	10
6-7-8	100	10
7-8-9	120	45
4-10-11	180	25
10-11-12	100	10
11-12-13	120	45

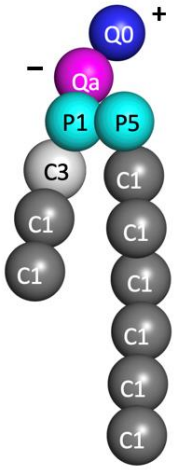
Table 3-4 MARTINI CG beads and parameters for PEPE

Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.47	1250
2-3	0.47	1250
3-4	0.37	1250
3-5	0.47	1250
5-6	0.47	1250
6-7	0.47	1250
7-8	0.47	1250
8-9	0.47	1250
4-10	0.47	1250
10-11	0.47	1250
11-12	0.47	1250
12-13	0.47	1250



Angle	θ (deg)	K_{angle} (kJ mol^{-1})
2-3-4	120	25
2-3-5	180	25
3-5-6	100	10
5-6-7	100	10
6-7-8	100	10
7-8-9	120	45
4-10-11	180	25
10-11-12	120	45
11-12-13	180	25

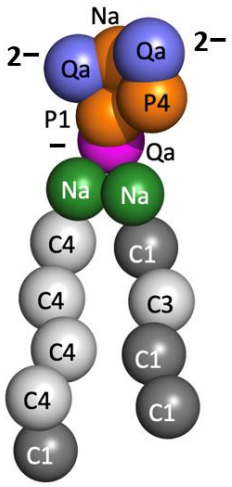
Table 3-5. MARTINI CG beads and parameters for PLSM



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.47	1250
2-3	0.47	1250
3-4	0.37	1250
3-5	0.47	1250
5-6	0.47	1250
6-7	0.47	1250
4-8	0.47	1250
8-9	0.47	1250
9-10	0.47	1250
10-11	0.47	1250
11-12	0.47	1250
12-13	0.47	1250

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
2-3-4	120	25
2-3-5	180	25
3-5-6	180	45
5-6-7	180	25
4-8-9	180	25
8-9-10	180	25
9-10-11	180	25
10-11-12	180	25
11-12-13	180	25

Table 3-6. MARTINI CG beads and parameters for PIP2



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.40	30000
1-3	0.40	30000
2-3	0.40	30000
2-5	0.30	25000
2-6	0.35	30000
1-5	0.40	25000
3-6	0.31	30000
5-6	0.60	25000
1-4	0.35	1250
4-7	0.47	1250
7-8	0.37	1250
7-9	0.47	1250
9-10	0.47	1250
10-11	0.47	1250
11-12	0.47	1250
12-13	0.47	1250
8-14	0.47	1250
14-15	0.47	1250
15-16	0.47	1250
16-17	0.47	1250

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-4-7	140	25
7-9-10	100	10
9-10-11	100	10
10-11-12	100	10
11-12-13	120	45
8-14-15	180	25
14-15-16	120	45
15-16-17	180	25

Table 3-7 Composition of RBC plasma membrane simulation system

Molecule/Ion	EL	CL	System	Total
PIPC	250	160		
PNSM	164	—		
PLSM	141	—		
PAPC	93	—		
POPC	118	68		
DPSM	211	22		
PAPS	22	251		
PEP1	47	160		
PIP2	—	45		
PUPE	—	137		
PEPE	—	68		
POPE	—	45		
PEP2	—	45		
DPPC	—	45		
CHOL	705	710		
W			65703	
WF			7300	
Na^+			1053	
Cl^-			555	
Total lipids excluding CHOL	1046	1046		78118

Table 3-8 Lipid order parameters (SCD) separated by exoplasmic and cytoplasmic leaflets and fatty acid tail

Lipid	Acyl chain	Exoplasmic Leaflet		Cytoplasmic Leaflet	
		Tail 1 (Left) *Mean* S_{CD}	*Tail 2 (Right)* *Mean* S_{CD}	*Tail 1 (Left)* *Mean* S_{CD}	*Tail 2 (Right)* *Mean* S_{CD}
PIPC	16:0-18:2	0.53±0.255	0.26±0.267	0.42±0.280	0.21±0.249
DPSM	34:1;2	0.71±0.256	0.63±0.221	0.44±0.277	0.30±0.348
PLSM	42:1;2	0.70±0.265	0.51±0.215	—	—
PNSM	42:2;2	0.71±0.259	0.47±0.187	—	—
PAPC	16:0-20:4	0.52±0.258	0.13±0.214	—	—
POPC	18:0-18:1	0.55±0.252	0.47±0.257	0.42±0.280	0.35±0.262
PAPS	18:0-20:4	0.52±0.250	0.29±0.326	0.40±0.280	0.10±0.201
PIP2	18:0-22:4	—	—	0.32±0.259	0.13±0.199
PEP1	18:1-22:4	0.38±0.265	0.14±0.218	0.30±0.265	0.11±0.202
PUPE	16:0-22:6	—	—	0.41±0.281	0.04±0.196
PEPE	18:1-20:4	—	—	0.31±0.264	0.11±0.203
POPE	16:0-18:1	—	—	0.41±0.280	0.34±0.264
PEP2	18:2-20:4	—	—	0.20±0.244	0.11±0.207
DPPC	16:0-16:0	—	—	0.49±0.272	0.46±0.274

3.6.3 Supporting Figures

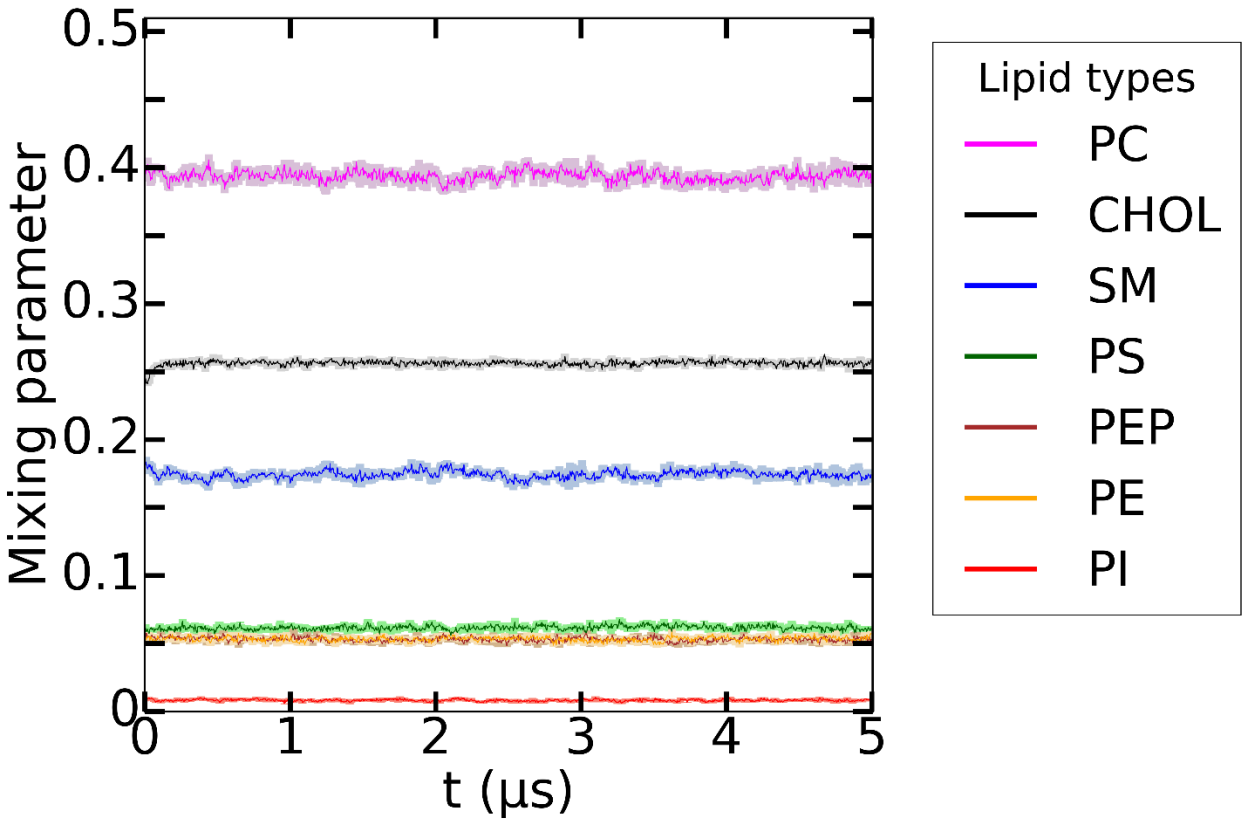


Figure 3-8 Mixing parameter profile of PC

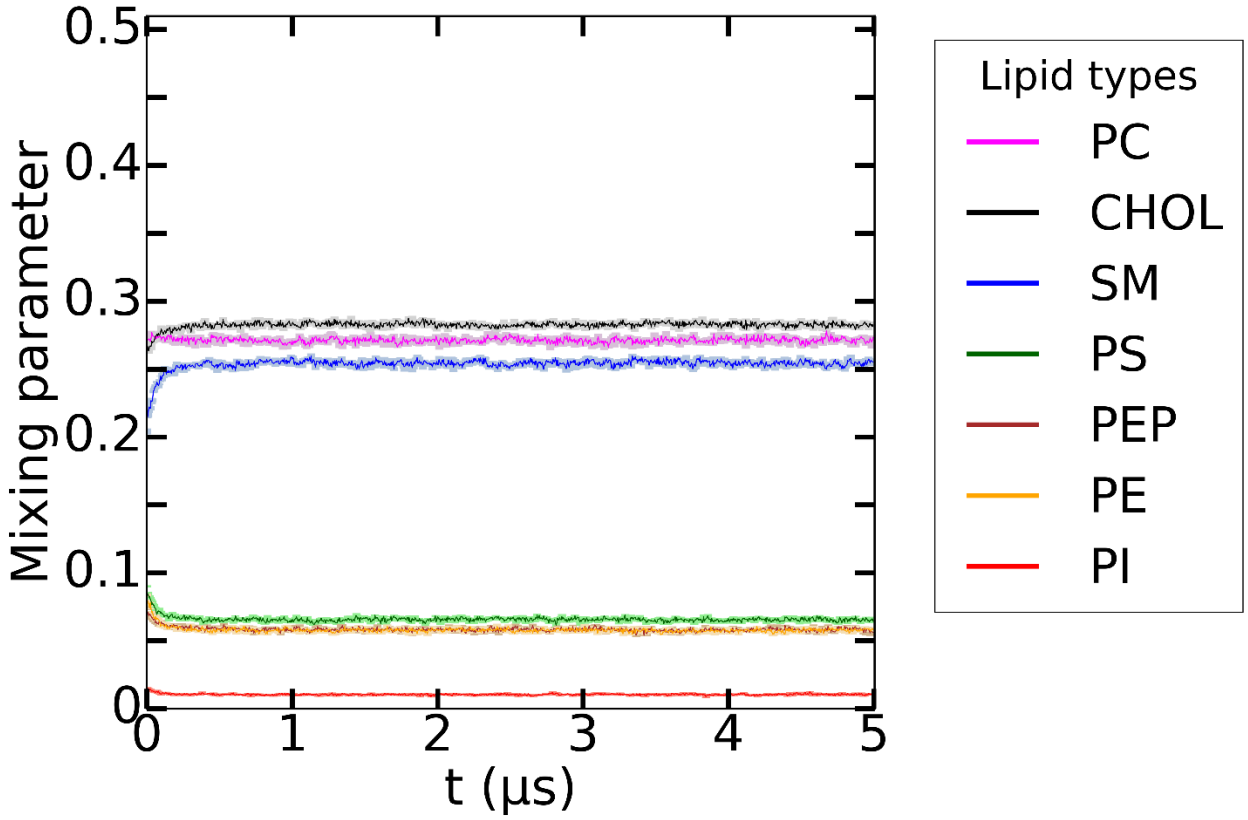


Figure 3-9 Mixing parameter profile of CHOL

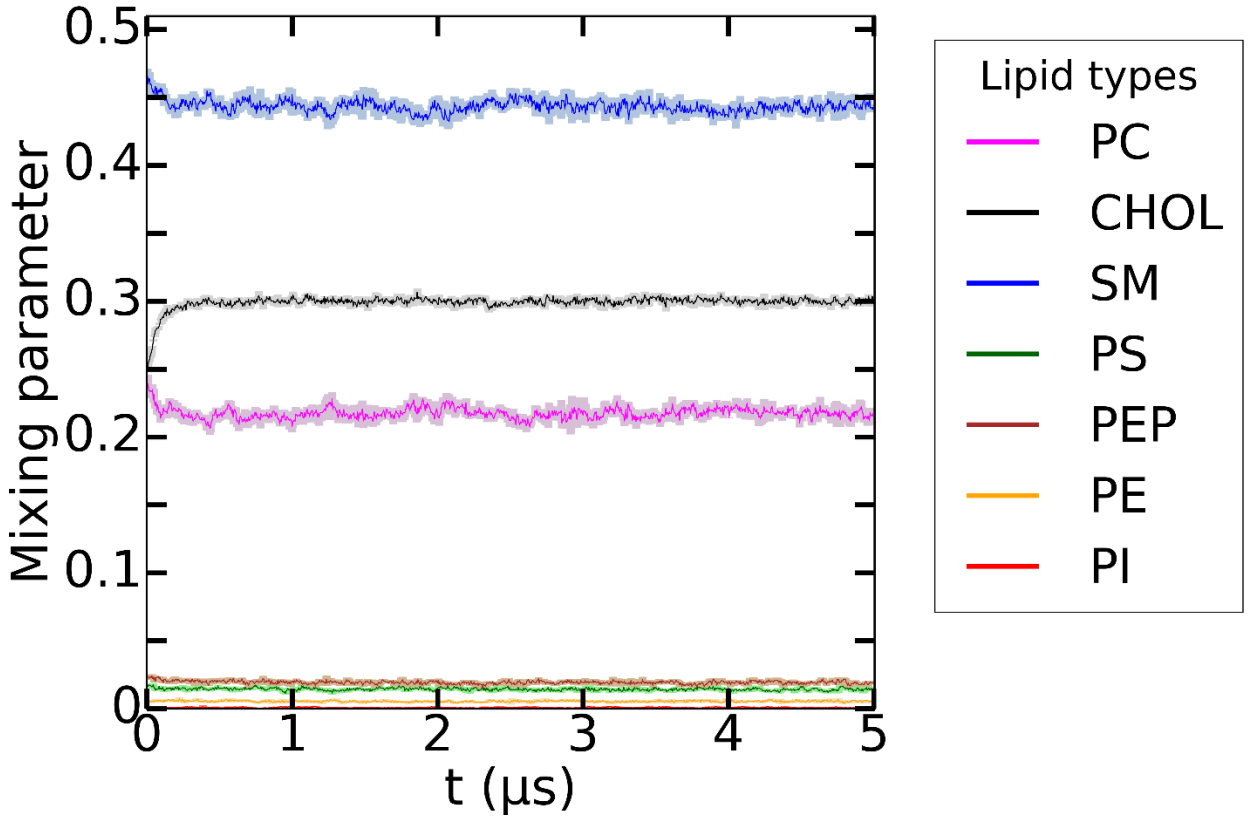


Figure 3-10 Mixing parameter profile of SM

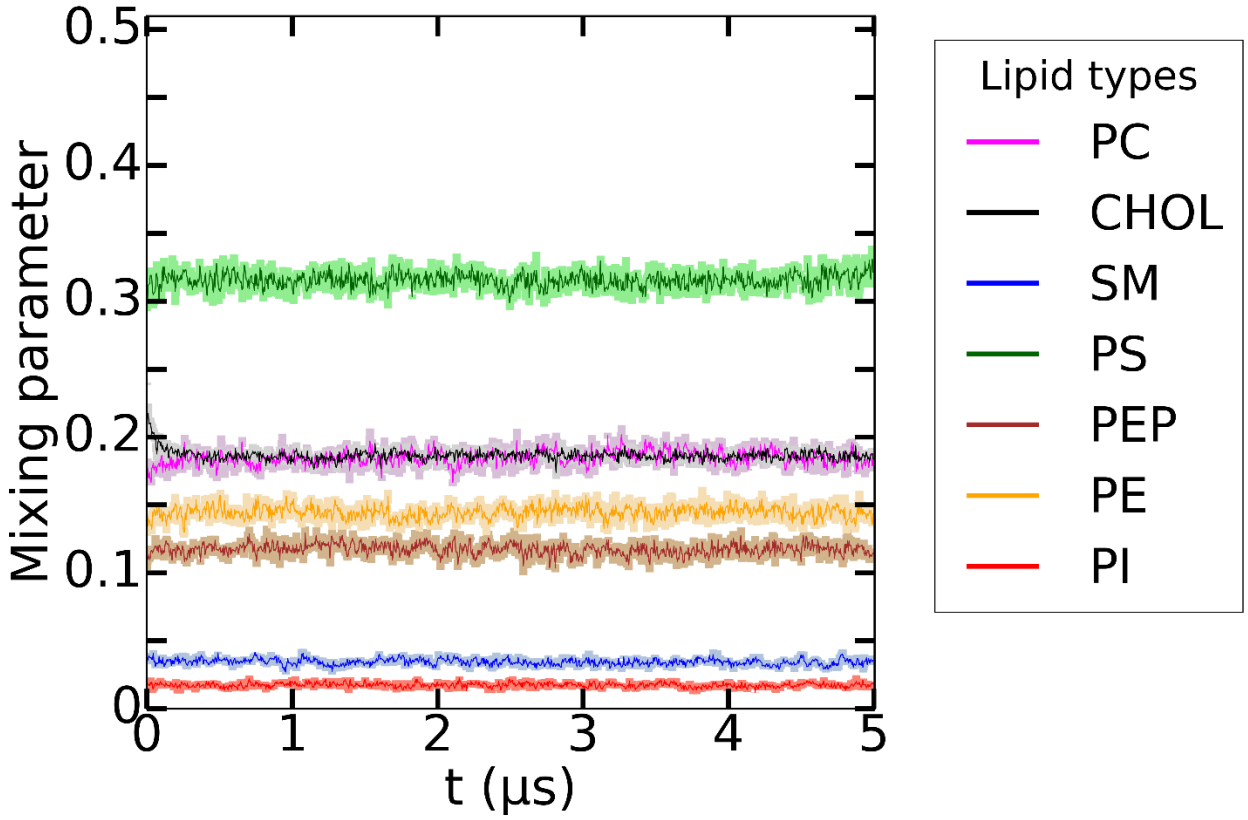


Figure 3-11 Mixing parameter profile of PS

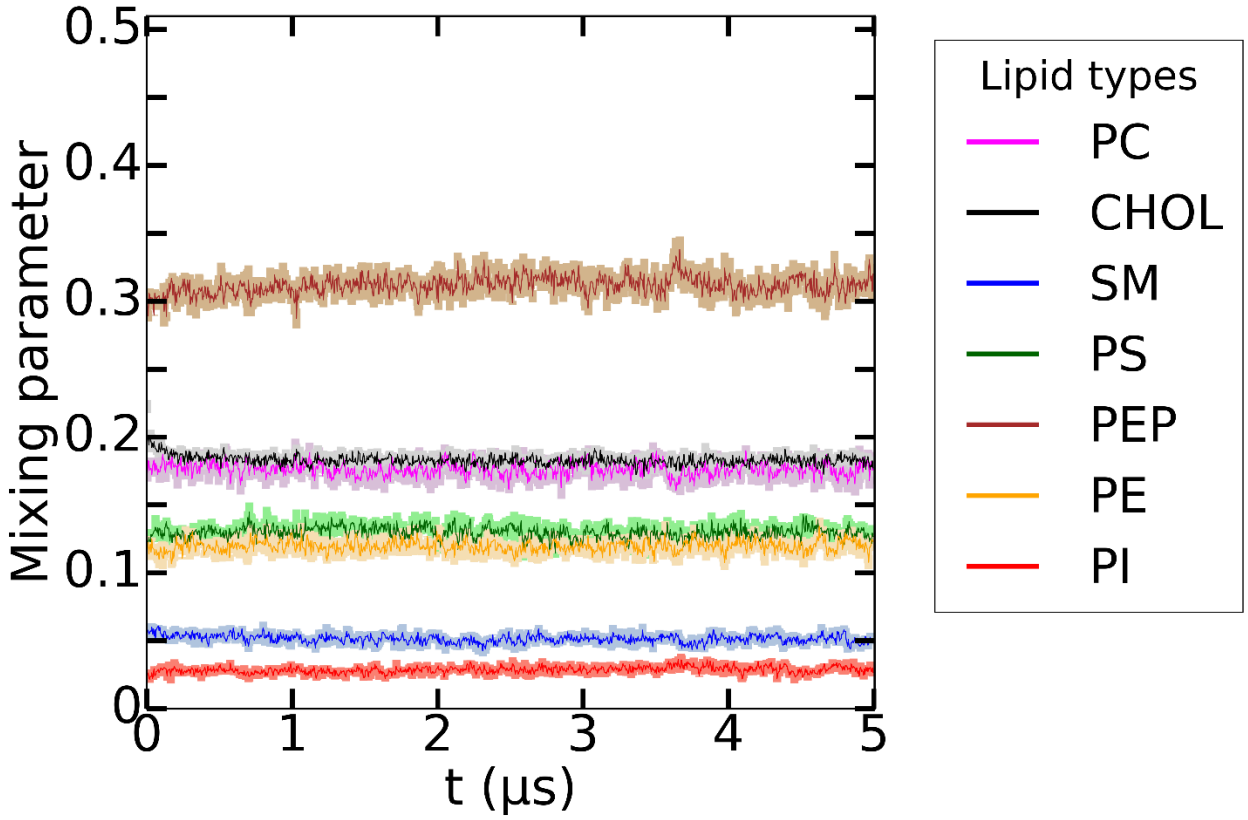


Figure 3-8 Mixing parameter profile of PEP

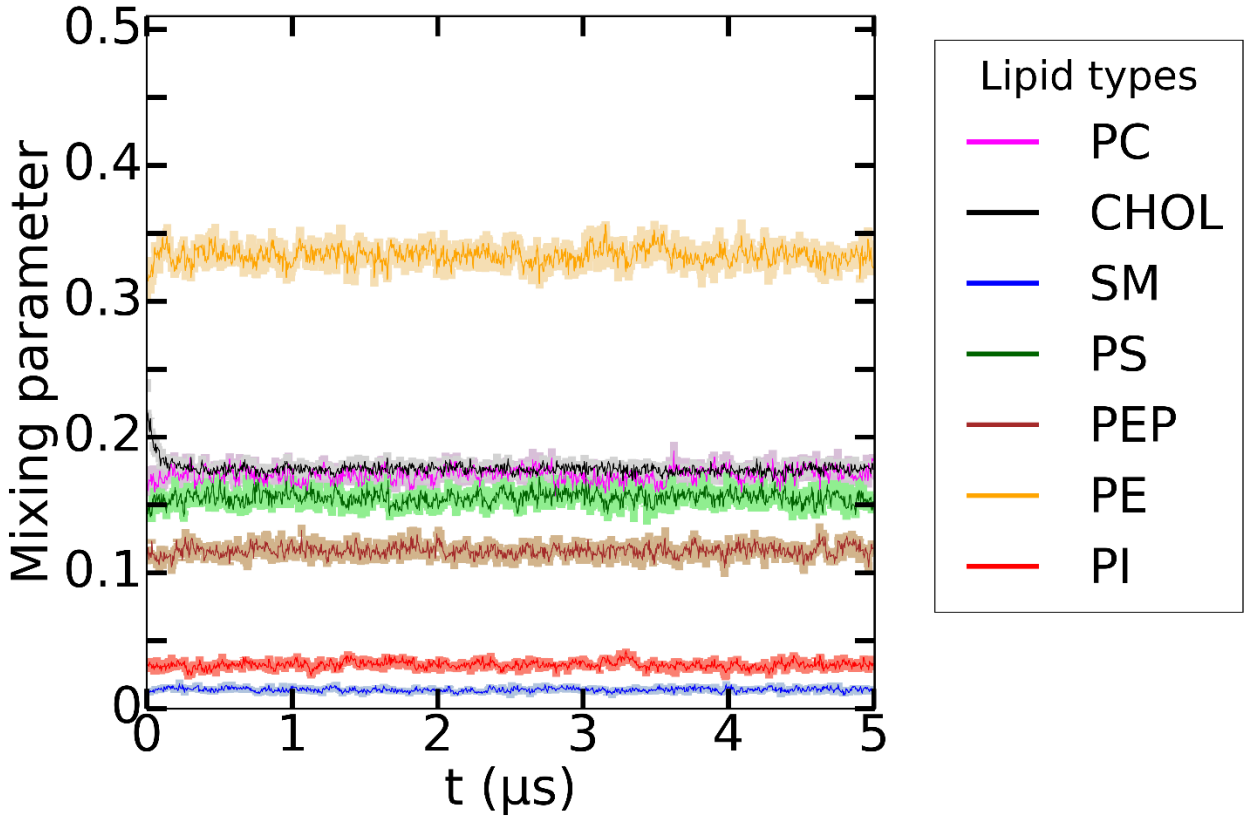


Figure 3-9 Mixing parameter profile of PE

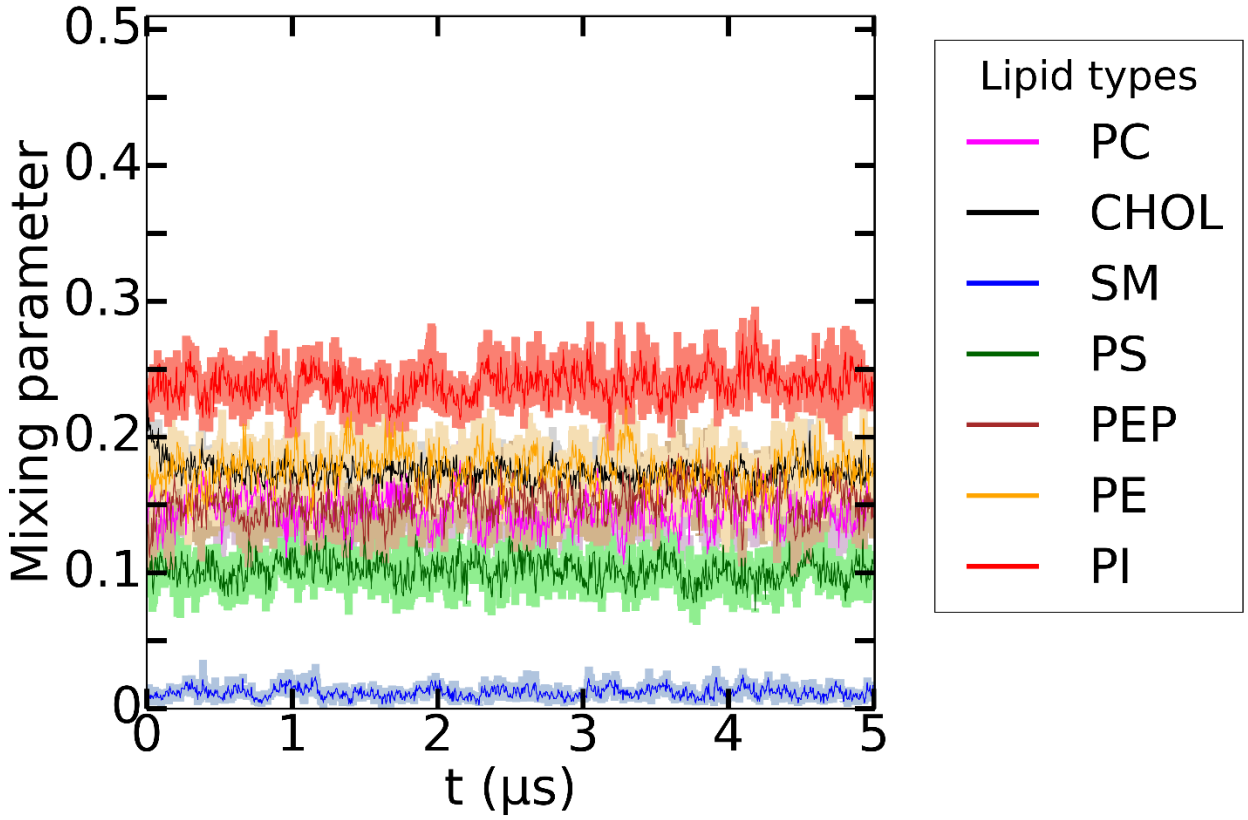


Figure 3-10 Mixing parameter profile of PI

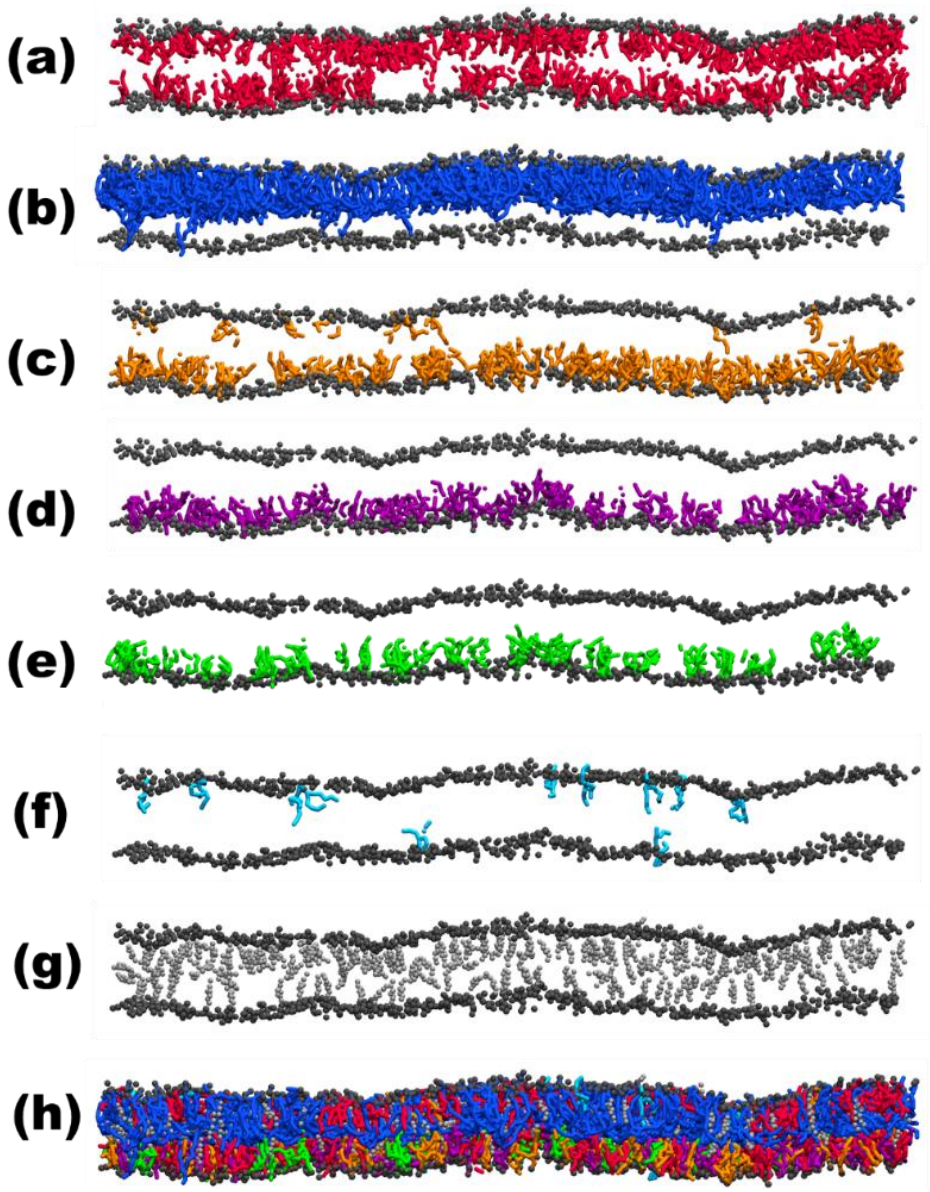


Figure 3-11 The cross-sectional profiles of equilibrated RBC plasma membrane. (a) PC, red; (b) SM, blue; (c) PS, orange; (d) PE, purple; (e) PEP, green; (f) PI, cyan;(g) cholesterol, gray; and (h) all lipid types. For clarity, all lipid headgroups (black) are shown

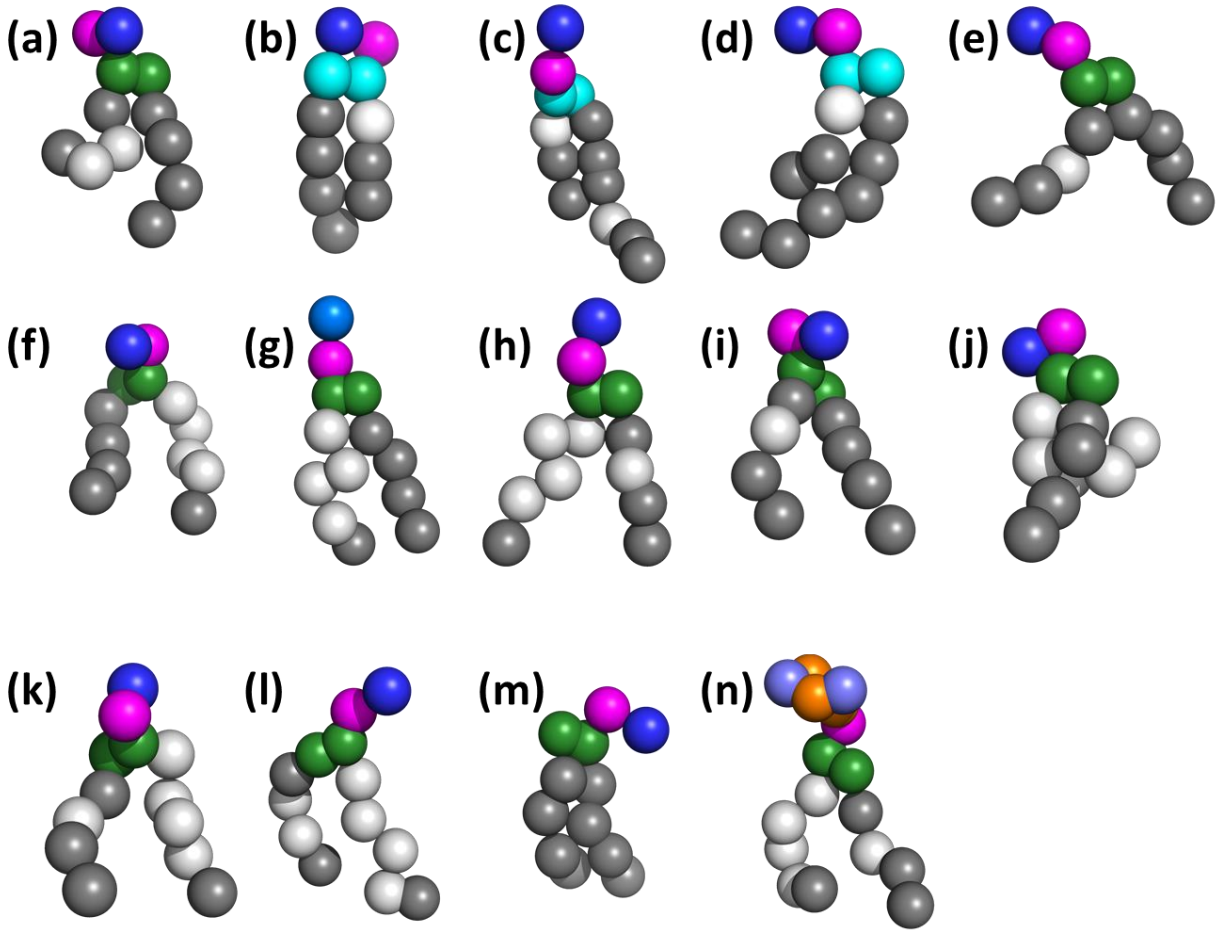


Figure 3-12 Representative lipid image after simulation (a) PIPC, (b) DPSM, (c) PNSM, (d) PLSM, (e) POPC, (f) PAPC, (g) PAPS, (h) PEP1, (i) POPE, (j) PUPE, (k) PEPE, (l) PEP2, (m) DPPC, and (n) PIP2. Coarse-grained bead types are denoted as saturated carbon (dark gray), unsaturated carbon (light gray), NC3 (dark blue), PO4 (pink), GL1 and GL2 (green), AM1 and AM2 (cyan), CNO (light blue), C1, C2, and C3 (orange), and P1 and P2 (purple).

	PC	CHOL	SM	PS	PEP	PE	PI
PC	39.4%	25.6%	17.4%	6.2%	5.2%	5.3%	0.8%
CHOL	27.2%	28.2%	25.4%	6.6%	5.8%	5.8%	1.0%
SM	21.7%	30.0%	44.4%	1.4%	1.9%	0.5%	0.1%
PS	18.6%	18.6%	3.4%	31.5%	11.7%	14.5%	1.7%
PEP	17.3%	18.3%	5.1%	13.0%	31.4%	12.0%	2.9%
PE	17.2%	17.7%	1.4%	15.6%	11.6%	33.3%	3.2%
PI	14.5%	17.4%	1.0%	10.1%	15.3%	17.6%	24.1%

Figure 3-17 Mixing parameter (%) of each lipid with other lipid families in the membrane. The color in each row displays lowest (blue) to highest (red) lipid contacts.

3.7 References

1. McIntyre, J. C.; Sleight, R. G., Fluorescence assay for phospholipid membrane asymmetry. *Biochemistry* **1991,** *30*, 11819-11827.

2. van Meer, G., Cellular lipidomics. *Embo Journal* **2005,** *24*, 3159-3165.

3. Ikeda, M.; Kihara, A.; Igarashi, Y., Lipid asymmetry of the eukaryotic plasma membrane: Functions and related enzymes. *Biological & Pharmaceutical Bulletin* **2006,** *29*, 1542-1546.

4. Gurtovenko, A. A.; Vattulainen, I., Lipid transmembrane asymmetry and intrinsic membrane potential: Two sides of the same coin. *Journal of the American Chemical Society* **2007,** *129*, 5358-5359.

5. Fadeel, B.; Xue, D., The ins and outs of phospholipid asymmetry in the plasma membrane: Roles in health and disease. *Critical Reviews in Biochemistry and Molecular Biology* **2009,** *44*, 264-277.

6. Leventis, P. A.; Grinstein, S., The distribution and function of phosphatidylserine in cellular membranes. In *Annual review of biophysics, vol 39*, Rees, D. C.; Dill, K. A.; Williamson, J. R., Eds. 2010; Vol. 39, pp 407-427.

7. Clarke, R. J.; Hossain, K. R.; Cao, K., Physiological roles of transverse lipid asymmetry of animal membranes. *Biochimica Et Biophysica Acta-Biomembranes* **2020,** *1862*.

8. Gupta, A.; Korte, T.; Herrmann, A.; Wohland, T., Plasma membrane asymmetry of lipid organization: Fluorescence lifetime microscopy and correlation spectroscopy analysis. *Journal of Lipid Research* **2020,** *61*, 252-266.

9. Singer, S. J.; Nicolson, G. L., The fluid mosaic model of the structure of cell membranes. *Science* **1972,** *175*, 720-731.

10. Kusumi, A.; Nakada, C.; Ritchie, K.; Murase, K.; Suzuki, K.; Murakoshi, H.; Kasai, R. S.; Kondo, J.; Fujiwara, T., Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: High-speed single-molecule tracking of membrane molecules. In *Annual review of biophysics and biomolecular structure*, 2005; Vol. 34, pp 351-378.

11. Ilangumaran, S.; Borisch, B.; Hoessli, D. C., Signal transduction via cd44: Role of plasma membrane microdomains. *Leukemia & Lymphoma* **1999,** *35*, 455-469.

12. Simons, K.; Toomre, D., Lipid rafts and signal transduction. *Nature Reviews Molecular Cell Biology* **2000,** *1*, 31-39.

13. Conner, S. D.; Schmid, S. L., Regulated portals of entry into the cell. *Nature* **2003,** *422*, 37-44.

14. Tian, T. H.; Harding, A.; Inder, K.; Plowman, S.; Parton, R. G.; Hancock, J. F., Plasma membrane nanoswitches generate high-fidelity ras signal transduction. *Nature Cell Biology* **2007,** *9*, 905-914.

15. Kusumi, A.; Fujiwara, T. K.; Chadda, R.; Xie, M.; Tsunoyama, T. A.; Kalay, Z.; Kasai, R. S.; Suzuki, K. G. N., Dynamic organizing principles of the plasma membrane that regulate signal transduction: Commemorating the fortieth anniversary of singer and nicolson's fluid-mosaic model. In *Annual review of cell and developmental biology, vol 28*, Schekman, R., Ed. 2012; Vol. 28, pp 215-250.

16. Harayama, T.; Riezman, H., Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology* **2018,** *19*, 281-296.

17. Fadok, V. A.; de Cathelineau, A.; Daleke, D. L.; Henson, P. M.; Bratton, D. L., Loss of phospholipid asymmetry and surface exposure of phosphatidylserine is required for phagocytosis of apoptotic cells by macrophages and fibroblasts. *Journal of Biological Chemistry* **2001,** *276*, 1071-1077.

18. Lorent, J. H.; Levental, K. R.; Ganesan, L.; Rivera-Longsworth, G.; Sezgin, E.; Doktorova, M. D.; Lyman, E.; Levental, I., Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nature Chemical Biology* **2020,** *16*, 644-652.

19. Chiu, S. W.; Vasudevan, S.; Jakobsson, E.; Mashl, R. J.; Scott, H. L., Structure of sphingomyelin bilayers: A simulation study. *Biophysical Journal* **2003,** *85*, 3624-3635.

20. Pandit, S. A.; Jakobsson, E.; Scott, H. L., Simulation of the early stages of nano-domain formation in mixed bilayers of sphingomyelin, cholesterol, and dioleylphosphatidylcholine. *Biophysical Journal* **2004,** *87*, 3312-3322.

21. Niemela, P. S.; Hyvonen, M. T.; Vattulainen, I., Influence of chain length and unsaturation on sphingomyelin bilayers. *Biophysical Journal* **2006,** *90*, 851-863.

22. Vacha, R.; Berkowitz, M. L.; Jungwirth, P., Molecular model of a cell plasma membrane with an asymmetric multicomponent composition: Water permeation and ion effects. *Biophysical Journal* **2009,** *96*, 4493-4501.

23. Perlmutter, J. D.; Sachs, J. N., Interleaflet interaction and asymmetry in phase separated lipid bilayers: Molecular dynamics simulations. *Journal of the American Chemical Society* **2011,** *133,* 6563-6577.

24. Wennberg, C. L.; van der Spoel, D.; Hub, J. S., Large influence of cholesterol on solute partitioning into lipid membranes. *Journal of the American Chemical Society* **2012,** *134,* 5351-5361.

25. Choubey, A.; Kalia, R. K.; Malmstadt, N.; Nakano, A.; Vashishta, P., Cholesterol translocation in a phospholipid membrane. *Biophysical Journal* **2013,** *104,* 2429-2436.

26. Ingolfsson, H. I.; Melo, M. N.; van Eerden, F. J.; Arnarez, C.; Lopez, C. A.; Wassenaar, T. A.; Periole, X.; de Vries, A. H.; Tieleman, D. P.; Marrink, S. J., Lipid organization of the plasma membrane. *Journal of the American Chemical Society* **2014,** *136,* 14554-14559.

27. Koldso, H.; Shorthouse, D.; Helie, J.; Sansom, M. S. P., Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers. *Plos Computational Biology* **2014,** *10*.

28. Sodt, A. J.; Sandar, M. L.; Gawrisch, K.; Pastor, R. W.; Lyman, E., The molecular structure of the liquid-ordered phase of lipid bilayers. *Journal of the American Chemical Society* **2014,** *136,* 725-732.

29. Sharma, S.; Kim, B. N.; Stansfeld, P. J.; Sansom, M. S. P.; Lindau, M., A coarse grained model for a lipid membrane with physiological composition and leaflet asymmetry. *Plos One* **2015,** *10*.

30. Metzler, R.; Jeon, J. H.; Cherstvy, A. G., Non-brownian diffusion in lipid membranes: Experiments and simulations. *Biochimica Et Biophysica Acta-Biomembranes* **2016,** *1858*, 2451-2467.

31. Gu, R. X.; Ingolfsson, H. I.; de Vries, A. H.; Marrink, S. J.; Tieleman, D. P., Ganglioside-lipid and ganglioside-protein interactions revealed by coarse-grained and atomistic molecular dynamics simulations. *Journal of Physical Chemistry B* **2017,** *121*, 3262-3275.

32. Ingólfsson, H. I.; Carpenter, T. S.; Bhatia, H.; Bremer, P.-T.; Marrink, S. J.; Lightstone, F. C., Computational lipidomics of the neuronal plasma membrane. *Biophysical Journal* **2017,** *113*, 2271-2280.

33. Carter, J. W.; Gonzalez, M. A.; Brooks, N. J.; Seddon, J. M.; Bresme, F., Flip-flop asymmetry of cholesterol in model membranes induced by thermal gradients. *Soft Matter* **2020,** *16*, 5925-5932.

34. Seo, S.; Murata, M.; Shinoda, W., Pivotal role of interdigitation in interleaflet interactions: Implications from molecular dynamics simulations. *Journal of Physical Chemistry Letters* **2020,** *11*, 5171-5176.

35. Wilson, K. A.; MacDermott-Opeskin, H. I.; Riley, E.; Lin, Y. C.; O'Mara, M. L., Understanding the link between lipid diversity and the biophysical properties of the neuronal plasma membrane. *Biochemistry* **2020,** *59*, 3010-3018.

36. Leonard, A. N.; Wang, E.; Monje-Galvan, V.; Klauda, J. B., Developing and testing of lipid force fields with applications to modeling cellular membranes. *Chemical Reviews* **2019,** *119*, 6227-6269.

37. Wood, W. G.; Igbavboa, U.; Muller, W. E.; Eckert, G. P., Cholesterol asymmetry in synaptic plasma membranes. *Journal of Neurochemistry* **2011,** *116*, 684-689.

38. Marquardt, D.; Heberle, F. A.; Greathouse, D. V.; Koeppe, R. E.; Standaert, R. F.; Van Oosten, B. J.; Harroun, T. A.; Kinnun, J. J.; Williams, J. A.; Wassall, S. R.; Katsaras, J., Lipid bilayer thickness determines cholesterol's location in model membranes. *Soft Matter* **2016,** *12*, 9417-9428.

39. Marquardt, D.; Kucerka, N.; Wassall, S. R.; Harroun, T. A.; Katsaras, J., Cholesterol's location in lipid bilayers. *Chemistry and Physics of Lipids* **2016,** *199*, 17-25.

40. Steck, T. L.; Lange, Y., Transverse distribution of plasma membrane bilayer cholesterol: Picking sides. *Traffic* **2018,** *19*, 750-760.

41. Thallmair, S.; Ingolfsson, H. I.; Marrink, S. J., Cholesterol flip-flop impacts domain registration in plasma membrane models. *Journal of Physical Chemistry Letters* **2018,** *9*, 5527-5533.

42. Gu, R. X.; Baoukina, S.; Tieleman, D. P., Cholesterol flip-flop in heterogeneous membranes. *Journal of Chemical Theory and Computation* **2019,** *15*, 2064-2070.

43. Aghaaminiha, M.; Farnoud, A. M.; Sharma, S., Quantitative relationship between cholesterol distribution and ordering of lipids in asymmetric lipid bilayers. *Soft Matter* **2021,** *17*, 2742-2752.

44. Niemelä, P. S.; Hyvönen, M. T.; Vattulainen, I., Influence of chain length and unsaturation on sphingomyelin bilayers. *Biophysical journal* **2006,** *90*, 851-863.

45. Fujimoto, T.; Parmryd, I., Interleaflet coupling, pinning, and leaflet asymmetry-major players in plasma membrane nanodomain formation. *Frontiers in cell and developmental biology* **2017,** *4*, 155-155.

46. Róg, T.; Orłowski, A.; Llorente, A.; Skotland, T.; Sylvänne, T.; Kauhanen, D.; Ekroos, K.; Sandvig, K.; Vattulainen, I., Interdigitation of long-chain sphingomyelin induces coupling of membrane leaflets in a cholesterol dependent manner. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016,** *1858*, 281-288.

47. Chiantia, S.; London, E., Acyl chain length and saturation modulate interleaflet coupling in asymmetric bilayers: Effects on dynamics and structural order. *Biophys J* **2012,** *103*, 2311-9.

48. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., The martini force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B* **2007,** *111*, 7812-7824.

49. Yesylevskyy, S. O.; Schafer, L. V.; Sengupta, D.; Marrink, S. J., Polarizable water model for the coarse-grained martini force field. *Plos Computational Biology* **2010,** *6*.

50. de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J., Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation* **2013,** *9*, 687-697.

51. Marrink, S. J.; Tieleman, D. P., Perspective on the martini model. *Chemical Society Reviews* **2013,** *42*, 6801-6822.

52. Ma, H. L.; Irudayanathan, F. J.; Jiang, W. J.; Nangia, S., Simulating gram-negative bacterial outer membrane: A coarse grain model. *Journal of Physical Chemistry B* **2015,** *119*, 14668-14682.

53. Ma, H. L.; Cummins, D. D.; Edelstein, N. B.; Gomez, J.; Khan, A.; Llewellyn, M. D.; Picudella, T.; Willsey, S. R.; Nangia, S., Modeling diversity in structures of bacterial outer membrane lipids. *Journal of Chemical Theory and Computation* **2017,** *13*, 811-824.

54. Ibsen, K. N.; Ma, H. L.; Banerjee, A.; Tanner, E. E. L.; Nangia, S.; Mitragotri, S., Mechanism of antibacterial activity of choline-based ionic liquids (cage). *Acs Biomaterials Science & Engineering* **2018,** *4*, 2370-2379.

55. Ma, H. L.; Khan, A.; Nangia, S., Dynamics of ompf trimer formation in the bacterial outer membrane of escherichia coli. *Langmuir* **2018,** *34*, 5623-5634.

56. Irudayanathan, F. J.; Trasatti, J. P.; Karande, P.; Nangia, S., Molecular architecture of the blood brain barrier tight junction proteins-a synergistic computational and in vitro approach. *Journal of Physical Chemistry B* **2016,** *120*, 77-88.

57. Irudayanathan, F. J.; Wang, X. Y.; Wang, N.; Willsey, S. R.; Seddon, I. A.; Nangia, S., Self-assembly simulations of classic claudins-insights into the pore structure, selectivity, and higher order complexes. *Journal of Physical Chemistry B* **2018,** *122*, 7463-7474.

58. Rajagopal, N.; Irudayanathan, F. J.; Nangia, S., Computational nanoscopy of tight junctions at the blood-brain barrier interface. *International Journal of Molecular Sciences* **2019,** *20*.

59. Rajagopal, N.; Irudayanathan, F. J.; Nangia, S., Palmitoylation of claudin-5 proteins influences their lipid domain affinity and tight junction assembly at the blood-brain barrier interface. *Journal of Physical Chemistry B* **2019,** *123*, 983-993.

60. Rajagopal, N.; Nangia, S., Obtaining protein association energy landscape for integral membrane proteins. *Journal of Chemical Theory and Computation* **2019,** *15*, 6444-6455.

61. Irudayanathan, F. J.; Nangia, S., Paracellular gatekeeping: What does it take for an ion to pass through a tight junction pore? *Langmuir* **2020,** *36*, 6757-6764.

62. Rajagopal, N.; Durand, A. J.; Nangia, S., Predicting selectivity of paracellular pores for biomimetic applications. *Molecular Systems Design & Engineering* **2020,** *5*, 686-696.

63. Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J., Martini coarse-grained force field: Extension to carbohydrates. *Journal of Chemical Theory and Computation* **2009,** *5*, 3195-3210.

64. Wassenaar, T. A.; Pluhackova, K.; Bockmann, R. A.; Marrink, S. J.; Tieleman, D. P., Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation* **2014,** *10*, 676-690.

65. Rog, T.; Koivuniemi, A., The biophysical properties of ethanolamine plasmalogens revealed by atomistic molecular dynamics simulations. *Biochimica et biophysica acta* **2016,** *1858*, 97-103.

66. Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J., Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. *Journal of Chemical Theory and Computation* **2015,** *11*, 2144-2155.

67. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

68. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126*, 014101.

69. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981,** *52*, 7182-7190.

70. Aurenhammer, F., Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.* **1991,** *23*, 345–405.

71. Fuller, J. C.; Martinez, M.; Wade, R. C., On calculation of the electrostatic potential of a phosphatidylinositol phosphate-containing phosphatidylcholine lipid membrane accounting for membrane dynamics. *Plos One* **2014,** *9*.

72. Bradley, R. P.; Slochower, D. R.; Janmey, P. A.; Radhakrishnan, R., Divalent cations bind to phosphoinositides to induce ion and isomer specific propensities for nano-cluster initiation in bilayer membranes. *Royal Society Open Science* **2020,** *7*, 192208.

73. McLaughlin, S.; Murray, D., Plasma membrane phosphoinositide organization by protein electrostatics. *Nature* **2005,** *438*, 605-611.

74. Sun, F.; Schroer, C. F. E.; Palacios, C. R.; Xu, L.; Luo, S.-Z.; Marrink, S. J., Molecular mechanism for bidirectional regulation of cd44 for lipid raft affiliation by palmitoylations and pip2. *PLOS Computational Biology* **2020,** *16*, e1007777.

75. Garg, S.; Porcar, L.; Woodka, A. C.; Butler, P. D.; Perez-Salas, U., Noninvasive neutron scattering measurements reveal slower cholesterol transport in model lipid membranes. *Biophysical journal* **2011,** *101*, 370-377.

76. Steck, T. L.; Lange, Y., How slow is the transbilayer diffusion (flip-flop) of cholesterol? *Biophysical journal* **2012,** *102*, 945-949.

77. Elkins, M. R.; Bandara, A.; Pantelopulos, G. A.; Straub, J. E.; Hong, M., Direct observation of cholesterol dimers and tetramers in lipid bilayers. *The Journal of Physical Chemistry B* **2021,** *125*, 1825-1837.

78. Bennett, W. F. D.; Shea, J. E.; Tieleman, D. P., Phospholipid chain interactions with cholesterol drive domain formation in lipid membranes. *Biophys J* **2018,** *114*, 2595-2605.

79. Samsonov, A. V.; Mihalyov, I.; Cohen, F. S., Characterization of cholesterol-sphingomyelin domains and their dynamics in bilayer membranes. *Biophysical Journal* **2001,** *81*, 1486-1500.

80. Wang, E.; Klauda, J. B., Examination of mixtures containing sphingomyelin and cholesterol by molecular dynamics simulations. *The Journal of Physical Chemistry B* **2017,** *121*, 4833-4844.

81. Gu, R. X.; Baoukina, S.; Tieleman, D. P., Phase separation in atomistic simulations of model membranes. *Journal of the American Chemical Society* **2020,** *142*, 2844-2856.

CHAPTER 4.

NON-CANONICAL LIPOPROTEINS WITH PROGRAMMABLE ASSEMBLY AND ARCHITECTURE

Reprint with permission from "Hossain, M. S.; Maller, C.; Dai, Y.; Nangia, S.; Mozhdehi, D. Non-canonical lipoproteins with programmable assembly and architecture. Chem. Commun., 2020, 56, 10281.
DOI: 10.1039/D0CC03271A" Copyright The Royal Society of Chemistry 2020.

4.1 Abstract

The substrate promiscuity of an acyltransferase is leveraged to synthesize artificial lipoproteins bearing a non-canonical PTM (ncPTM). The non-canonical functionality of these lipoproteins results in a distinctive hysteretic assembly—absent from the canonical lipoproteins—and is used to prepare hybrid multiblock materials with precise and programmable patterns of amphiphilicity. This study demonstrates the promise of expanding the repertoire of PTMs for the development of nanomaterials with a unique assembly and function.

4.2 Introduction

The disordered regions on polypeptide chains contribute significantly to the properties of natural biomaterials and inspire the design of intrinsically disordered peptide-polymers (IDPPs) for mimicking natural biomaterials. Contrary to common strategies of IDPPs assembly, post-translational modifications (PTMs) are conducted to control the configuration of proteins.

The application of PTMs in material science is a rare attempt. In this project, non-canonical PTM (ncPTM) is used by Mozhdehi lab at Syracuse University to develop recombinant nanomaterials with controlled hierarchical assembly. In comparison with the traditional canonical PTM modified IDPP with myristic acid (M), ncPTM uses the artificial lipid, 12-azidododecanoic acid (ADA), which shows distinctive temperature-triggered emergent assembly that is absent in M-IDPP. We developed an *in-silico* model from the molecular-scale perspective to explain the differences in the macromolecular assembly of ADA-IDPP compared to M-IDPP, which can provide insights in designing novel lipoproteins.

4.3 Methods

The molecular structure of M-peptide and ADA-peptide molecules were built in two-steps. First, the three-dimensional structure of the recognition peptide (GLYASKLFSNL) was determined using the I-TASSER[1] webserver. In the second step, the M and ADA lipid tails were added to the glycine residue of the peptide using the CHARMM-GUI[2] webserver. The individual M-peptide and ADA-peptide structures were equilibrated at 300 K using the CHARMM36 all-atom force field available within the CHARMM-GUI workspace.

The equilibrated structures were used to construct two explicitly solvated systems containing: (i) 15 molecules of M-peptide and (ii) 15 molecules of ADA-peptide using CHARMM-GUI Builder.[2] Each system contained 150 mM NaCl solution and TIP3P water molecules in a cubic box with a box length of 10 nm. The systems were energy minimized and equilibrated at T = 300 K in the isothermal-isochoric (NVT) and at P = 1 bar in isothermal-isobaric (NPT) ensemble constraints in the CHARMM-GUI workflow. The equilibrated output from these runs was used to study self-assembly behavior.

The self-assembly molecular dynamics simulations were performed using the GROMACS 2019.4 simulation package.[3] The M-peptide and ADA-peptide molecules and ions were modeled using the CHARMM36 all-atom force field.[4] Water was modeled using TIP3P.[5] The NPT production runs were run for 400 ns using 2 fs timestep. The temperature was maintained at 303.15 K using the Nose-Hoover thermostat[6] with $\tau t = 1.0$ ps. The pressure was maintained at 1 bar using isotropic coupling with Parrinello-Rahman barostat[7] with $\tau P = 5$ ps and compressibility of 4.5 x 10−5 bar-1. Periodic boundary conditions were applied in all three dimensions. The nonbonded van der Waals interactions were calculated using a cutoff of 1.2 nm. The long-range electrostatic interactions were studied with particle mesh Ewald (PME) algorithm at a 1.2 nm cutoff.[8] Further,

thermal annealing simulations were performed for both systems over 280−360 K temperature range (280, 290, 300, 310, 320, 330, 340, 350, and 360 K) over a total simulation time of 400 ns. Analyses of the results from thermal annealing simulations, including radius of gyration (Rg) and solvent accessible surface area (SASA), were performed using GROMACS built-in utilities. Molecular visualization was performed using VMD software.[9]

4.4 Results and Discussion

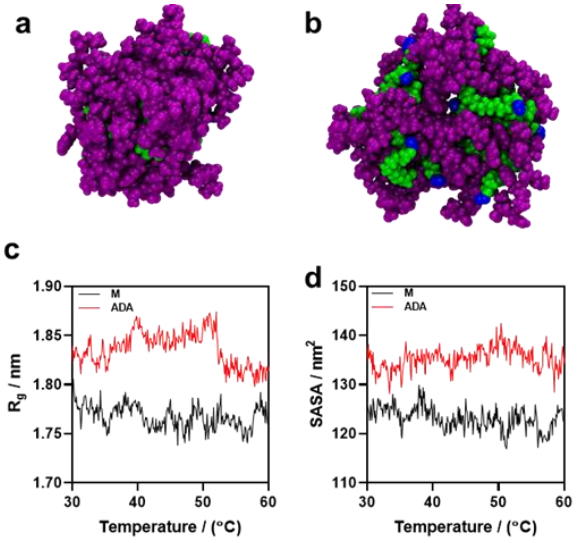


Figure 4-1 Result of in silico self-assembly of M–peptide and ADA–peptide molecules using all-atom molecular dynamics simulations. a) Snapshot of 15 M–peptide molecules showing core-shell structure with myristoyl chains (green) in the core and peptides (purple) forming the shell at 40 °C. b) Snapshot of 15 ADA–peptide aggregate with ADA chains (green) and terminal azide (blue) on the surface of aggregate along with the peptides (purple) at 40 °C. c) The variation in the radius of gyration (Rg) of the M–peptide (black) and ADA–peptide (red) aggregates over 30–60 °C. d) The variation in the solvent accessible surface area (SASA) of the M–peptide (black) ADA–peptide (red) aggregates over 30–60 °C.

We developed an *in-silico* model to explain the differences in the self-assembly of ADA-IDPP compared to canonical lipoprotein (M-IDPP)—hysteresis and non-equilibrium assembly into fibers— and to bridge our understanding from a molecular structure level to macromolecular assembly. Given the 99.5% similarity between the two lipoproteins, we confined our atomistic simulations to the N-terminal amphiphilic region (lipid–recognition sequence peptide); the single

letter amino acid sequence of the recognition peptide used is GLYASKLFSNL. The aggregation of the M–peptides occurs via the hydrophobic interactions of the myristoyl chains (shown in green) that form a micellar core while the peptides (purple) form the shell (Figure 4-1a). In contrast, in ADA–peptide assembly, the charged azide ($-N_3$) groups (blue beads in Figure 4-1b) are hydrophilic and remain solvent-exposed, preventing efficient packing and formation of a lipid core. Further, the ADA–peptide aggregate has a consistently higher radius of gyration (R_g) and solvent accessible surface area (SASA) compared to M–peptide aggregate (Figure 4-1c,d) over the 30–60 °C temperature range. The Combination of computational and experimental results show for the first time that despite the increased hydrophilicity of the ADA (compared to the canonical lipid), it can drive the self-assembly of the recombinant lipoprotein into micelles.

4.5 Conclusions

In summary, the well-documented substrate-promiscuity of lipidation machinery has been extensively leveraged in the field of chemical biology.[10-12] We applied this strategy to design novel lipoproteins with emergent material properties such as a stimuli-responsive shape-shifting nanomorphology. We foresee several opportunities for the design of dynamic nano-biomaterials in this untapped chemical design space. For example, a programmable morphological change from nanoparticles to fibers can be used to simultaneously release encapsulated cargo and provide a scaffold for cell-adhesion and growth. Non-canonical lipids can also be used as chemical handles for structural elaboration and synthesis of hybrid materials with a unique and precise amphiphilic pattern. These hybrid systems can be programmed to assemble into complex 2D and 3D morphologies to form materials with unique optical and mechanical properties.

4.6 Acknowledgements

This research was supported by a start-up grant from the Syracuse University (to D. M.) and received partial support from the National Science Foundation under Grant No. CHE-1659775 (REU) and CBET-1453312 (to S. N.).

4.7 References

1. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y., The I-TASSER Suite: protein structure and function prediction. *Nature Methods* **2015,** *12* (1), 7-8.
2. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008,** *29* (11), 1859-1865.
3. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2,* 19-25.
4. Huang, J.; MacKerell Jr, A. D., CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* **2013,** *34* (25), 2135-2145.
5. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
6. Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* **1992,** *97* (4), 2635-2643.
7. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981,** *52* (12), 7182-7190.

8. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.
9. Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", J. Molec. Graphics, 1996, vol. 14, pp. 33-38.
10. D. A. Ciulla, A. G. Wagner, X. Liu, C. L. Cooper, M. T. Jorgensen, C. Wang, P. Goyal, N. K. Banavali, J. L. Pezzullo, J.-L. Giner and B. P. Callahan, Chem. Commun., 2019, 55, 1829–1832.
11. K. N. Chuh, A. R. Batt and M. R. Pratt, Cell Chem. Biol., 2016, 23, 86–107.
12. X. Zhang, Z. Xu, D. S. Moumin, D. A. Ciulla, T. S. Owen, R. A. Mancusi, J.-L. Giner, C. Wang and B. P. Callahan, Bioconjugate Chem., 2019, 30, 2799–2804.

CHAPTER 5.

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

In our research, we have adopted Multiscale MD modeling, which combines the all-atom (AA) and the coarse-grained (CG) resolutions, along with other advanced MD techniques, to apply computational simulations to the biomolecular field of study and provide molecular insights that are undetectable via currently available experimental methods. In this work, we described the development and validation of a computational method to accelerate the antibiotic discovery pipeline, we performed comprehensive investigation of plasma membrane lipidome asymmetry of human red blood cells to unveil physicochemical perperties and biological functions, we collaborated with an experimental group to understand variations in self-assembly of canonical and non-canonical lipoproteins at a molecular-level for inspiring macro-scale design of novel biomaterials.

In Chapter 2, we report the development of a new computational screening platform (CLASP) that is designed to accelerate the antibiotic discovery process. The CLASP outputs comprehensive thermodynamic and kinetic data, which includes free energy profile, energy barrier, translocation rate constant, contact analysis of the molecule with the pore-lining residues, orientation analysis of the molecule within the porin, and the overall mechanism of the antibiotic uptake. In this first application of CLASP, we demonstrate the translocation profile of six well-established antibiotics from the carbapenem family through *P. Aeruginosa*‘*s* OccD3 channel. The results show excellent agreement with the barrier heights, the key residues, and the translocation rates of the drug molecules. We were able to identify critical pore-lining residues in the OccD channel that provide steric gating. Mutation of these residues to less bulky groups enhanced the predicted rates.

In Chapter 3, we report a detailed molecular description of the human RBC plasma membrane that replicates the experimentally determined lipid composition of each of the two asymmetric membrane leaflets. We demonstrated distinct physicochemical properties to the exoplasmic and cytoplasmic leaflets conferred by plasma membrane lipid families. We characterized the membrane asymmetry by studying lipid mixing parameters and order parameters, cholesterol flip-flop dynamics, membrane thickness and area per lipid. Fingerprinting the plasma membrane lipidome is pivotal in elucidating how cells exchange metabolites, transfer signals, and facilitate protein assembly.

In Chapter 4, we developed *in-silico* models of canonical and non-canonical post-translational modified intrinsically disordered peptide-polymers. We characterized temperature-triggered self-assembly of natural and artificial lipoproteins via thermal annealing atomistic simulations. Agreement between computational and experimental results reinforced the robustness of computational modeling and simulation, which bridges the understanding between micro- to macro-scale self-assembly phenomenon and provides insights in designing novel nano-biomaterials with unique physical and chemical properties.

In summary, the physics-driven computational simulation method in conjunction of enhanced sampling technique, multiscale modeling, and other supportive computational tools and high performance computing resources are becoming a robust engine and a necessity in understanding biological processes at molecular-level and solve scaled-up real life problems.

5.2 Future Work

For the ongoing CLASP project, we are focusing on the study of small molecule translocation through another major family of Outer membrane Carboxylate Channel (Occ) from *P. aeruginosa*, the OccK subfamily, which is distinct from OccD subfamily by phylogenetic analysis.[1] The porin size of OccK channels are substantially larger than those of OccD subfamily and are less selective for substrates.[2] We have some preliminary results at this stage. Figure 5-1 shows the visual representation of homology modeled and well equilibrated structures of OccK1-8. We plan to conduct a systematic comparison between different OccK porins, and with OccD subfamily porins. We expect to provide further knowledge to the free energy profile and interaction pattern between drug molecules and porin residues.

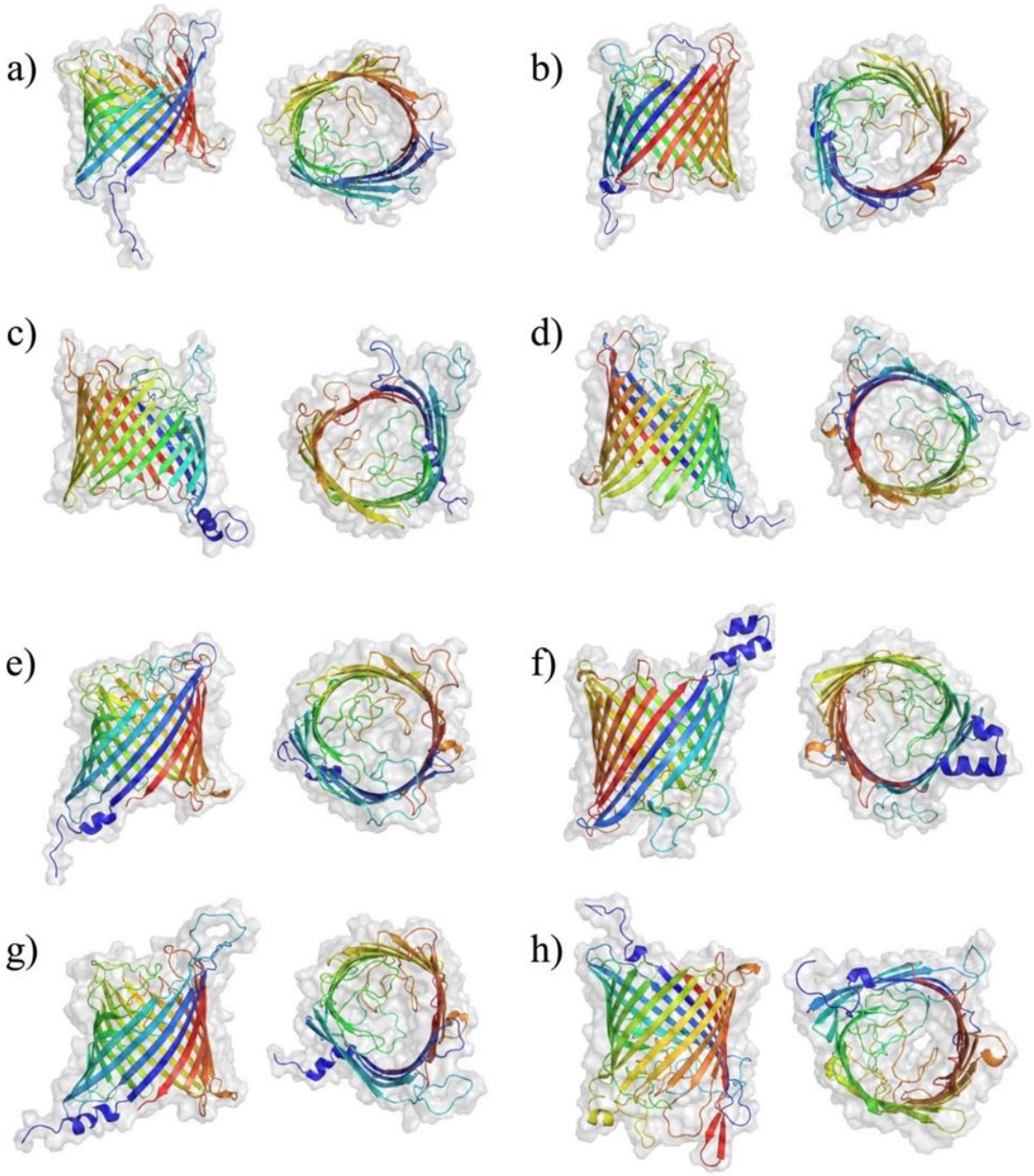


Figure 5-1 Homology modeled *P. aeruginosa* porin structures displayed in front and top view. (a) OccK1, (b) OccK2, (c) OccK3, (d) OccK4, (e) OccK5, (f) OccK6, (g) OccK7, and (h) OccK8. The porin surface is displayed as a transparent shadow.

On the other hand, the major limitation and the potential challenge for the application of CLASP is its heavy dependency on molecular database, which may either be unavailable or incompatible with the current CLASP algorithm. We envision to establish a database of experimentally solved and homology modeled membrane proteins embedded in biological membrane lipid systems. Here is a few features we would like to achieve as enhancements for the current CLASP platform:

- We would like to develop a graphical user interface for CLASP so that it can be freely distributed and serve as a robust engine in screening and identifying highly permeable antibiotic candidates.
- In the meantime, we would like to establish more well-equilibrated protein inserted membrane systems at both all-atom and coarse-grained models to ensure they are ready for use at different resolutions.
- Ultimately, we expect to collaborate with industry to publish CLASP as a crucial step to state-of-the-art high-throughput virtual screening method to identify drug candidates.

5.3 References

1. Tamber, S.; Ochs, M. M.; Hancock, R. E. W., Role of the Novel OprD Family of Porins in Nutrient Uptake in Pseudomonas aeruginosa. *Journal of Bacteriology* **2006,** *188* (1), 45.
2. Eren, E.; Vijayaraghavan, J.; Liu, J.; Cheneke, B. R.; Touw, D. S.; Lepore, B. W.; Indic, M.; Movileanu, L.; van den Berg, B., Substrate specificity within a family of outer membrane carboxylate channels. *PLoS Biol* **2012,** *10* (1), e1001242-e1001242.

VITA

SUMMARY OF QUALIFICATIONS

- Doctorate degree in Computational Biophysics with a focus on Molecular Dynamics simulation of biological systems
- 4 years of hands-on experience in multiscale modeling under GROMACS using various force fields
- 4 years of strong knowledge of computer-aided drug design software, databases, and webservers
- 4 years of experience in workflow automation via shell scripting and Python programming under Linux/Unix environment
- 4 years of extensive experience with a cloud-based computational environment such as a SLURM based cluster
- Strong verbal/written presentation skills, excellent interpersonal, collaborative, team-building, and communication skills

CORE COMPETENCIES

- **Simulation Method**: Molecular modeling, homology modeling, molecular docking, binding site prediction
- **Numerical Method:** Molecular Dynamics, umbrella sampling, metadynamics
- **Programming**: Python (NumPy, SciPy, pandas, MDAnalysis), Unix/Linux, Shell scripting, MATLAB
- **Scientific Software**: GROMACS, VMD, PyMOL, CAVER, I-TASSER, YASARA, Avogadro, CHARMM-GUI

EDUCATION

Doctor of Philosophy, Biomedical & Chemical Engineering, *Syracuse University* 8/2017 – 5/2021

Master of Science, Chemical & Biomolecular Engineering, *University of California, Berkeley* 8/2014 – 5/2015

Bachelor of Science, Chemical Engineering, *University of Rochester* 8/2010 – 5/2014

PROFESSIONAL EXPERIENCES

Graduate Researcher – *Syracuse University*, Syracuse, NY 8/2017 – 5/2021

- Developed high performance antibiotic drug screening platform using all-atom and coarse-grain molecular dynamics simulation
- Characterized structural, orientational, conformational, mutational, kinetic, and thermodynamic details of drug transport
- Automated workflow via shell scripting and Python for simulation initiation, deployment, and data analysis

Graduate Researcher – *Syracuse University*, Syracuse, NY 8/2019 – 5/2020

- Developed in-silico models of canonical and non-canonical post-translational modified intrinsically disordered peptide-polymers
- Characterized temperature-triggered self-assembly of artificial lipoproteins via thermal annealing all-atom simulation
- Collaborated with interdisciplinary teams to bridge understanding from molecular to macro-scale aggregation of nanobiomaterials

Research Consultant - *Advanced Biotech. Inc.*, Totowa, NJ 4/2016 – 4/2017

- Teamed with validation engineers to inspect bottlenecks and pitfalls in product manufacturing pipelines
- Researched and presented reaction pathways for bio-enzymatic fermentation of food fragrance and additives
- Delivered reports and presentations regarding research progress to principle scientists and senior managers

HONORS & EXTRACURRICULAR ACTIVITIES

- Stevenson Biomaterials Poster Award, Syracuse University 2020
- Undergraduate Summer Program Mentor (8 students), Syracuse University 2020
- Graduate Teaching Assistant (advanced fluid dynamics, process control, etc.), Syracuse University 2017 – 2019
- Syracuse University Women in Science and Engineering (WiSE) member 2018 – 2021
- UC, Berkeley Product Development scholarship ($20,000/yr.) 2015

PUBLICATIONS & PRESENTATIONS

- **Y Dai**, A Karmazyn, N Rajagopal, S Nangia, Fingerprinting plasma membrane lipidome of human red blood cells using computational modeling techniques, to be submitted.
- **Y Dai**, H Ma, M Wu, T Welsch, S Vora, D Ren and S Nangia, Development of the computational antibiotic screening platform (CLASP) to aid in the discovery of new antibiotics, *Soft Matter*, 2021.
- M Hossain, C Maller, **Y Dai**, S Nangia and D Mozhdehi, Non-canonical lipoproteins with programmable assembly and architecture, *Chemical Communications*, 2020.
- **Y Dai**, and S Nangia, High-throughput screening of small-molecule libraries in search of new antibiotics, *AIChE Annual Meeting*, 2020.