James Ludwig

Contributing to the Understanding of Cell Membranes and Uremic Toxins by Combining Molecular Dynamics with Machine Learning and Statistical Modeling

James Ludwig

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy in Chemistry

University of Washington

2021

Reading Committee:

Walter Pfaendtner, Chair

Gary Drobny

Stephanie Valleau

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Contributing to the understanding of cell membranes and uremic toxins by combining molecular dynamics with machine learning and statistical modeling

James Ludwig

Chair of the Supervisory Committee:
Walter Pfaendtner
Department of Chemistry

The calculation of thermodynamic and kinetic quantities related to biophysical phenomena is an integral part of understanding how biological systems function. Even the most basic interactions, such as ligand to protein residue, are critical pieces of information when interpreting complex biophysical processes. To this end, modeling of various systems and challenges, from detailed kinetics calculations to specific interactions in nature, can be accomplished at an atomistic level using molecular dynamics (MD) simulations. MD provides an ideal opportunity to implement statistical modeling and, moreover, machine learning (ML) methods for deep system design and results analysis due to the often large-scale high-dimensional data sets produced. In this dissertation, two distinct biophysical settings are investigated. The first section studies model membranes and the effects of small molecule partitioning, using Gaussian mixture models and enhanced sampling to identify potential changes in phase composition and solvation free energies. The second section begins by exploring uremic toxins bound to human serum albumin and provides long time-scale binding dynamics through Markov state models as well as detailed kinetic information regarding unbinding using deep learning to optimize unbinding coordinates. Finally, I model adsorbent polymer materials for the capture of uremic toxins and design methods for

screening a range of system configurations. This work highlights the wide range of biophysical challenges that can be rigorously tackled by the combination of MD and statistical modeling/ML.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENTS

My Ph.D. has certainly been an exciting one. In my first year alone, I gained two incredible mentors in Lutz Maibaum and Sarah Keller as well as planned and completed my wedding with Taylor Ludwig, the person who has been my partner for 10 years this summer. All of this to say, I could never have imagined that I would get the opportunity to join the incredible Pfaendtner research group halfway through my Ph.D. while having the support and love of Taylor alongside our many wonderful friends and family throughout this entire journey.

Both Taylor and I will never forget how enthusiastic Lutz Maibaum was when attending a visitor's weekend, dragging me over to meet the research group that would become my team for the first half of my Ph.D., specifically Shushan He, Nihit Pokhrel, and Raymond Jin. Thanks to these individuals who have all since gone onto amazing post graduate careers, I was able to learn so much about how to begin my first graduate research project in simulating model membranes and general biophysical computational practices. I also learned skills that would be crucial to my future career path, such as enhanced sampling techniques from Nihit and the application of machine learning to large data sets from Shushan. Also at the beginning stages of my graduate career was the incredible research group led by Sarah Keller. This group acted jointly alongside the Maibaum group, giving me an exciting way to work directly with experimental collaborators, another skill that has remained critical to my future endeavors. Specifically, my first paper, focused on a computational approach for studying model membrane interactions with n-alcohols, was inspired by the work of and discussions with Caitlin Cornell. This first half of my graduate career gave me exactly what I needed to succeed when I was required to find a new research group midway through my degree, and I couldn't be more thankful to all those mentioned and more.

When in need of a new research group, I was suggested to meet with Jim Pfaendtner. I became immediately excited at the thought due to Jim's great knowledge and history with the enhanced sampling techniques I had begun using in my own work. However, I could not have imagined that joining the Pfaendtner lab would provide me a way to continue my Ph.D. smoothly alongside an incredible mentor and team. Jim Pfaendtner has provided me a setting that not only feels like home, but a mentoring style that helped me excel. Whether support is thoughtful research advice or providing Oculus headsets for the group, Jim tries his hardest to keep everyone happy and healthy while moving forward through each of our unique career paths. This is all before mentioning how fantastic the Pfaendtner group's members are. Josh Smith was not only my first interaction with the group but is also the team member whose work I picked up. Josh taught me data science skills to use in analyzing molecular dynamics simulations while giving me a starting point for my new research studying uremic toxins. I am beyond grateful for the support Josh has given over the years, continuing on as I began to work toward a data science career path. Other past and current members such as Sarah Alamdari, Janani Sampath, Chris Nyambura, Orion Dollar, Luke Gibson, Xin Qi, Wesley Beckner, Arushi Prakash, and Sabiha Rustam have all done so much to help me in my own research and life. I can only hope that I have provided them a similar amount of happiness and support in their own lives. I also can't forget to mention Jesse Prelesnik, who has joined me on this roller coaster of a Ph.D. path from day one and has remained a true friend along the way. Finally, the Pfaendtner group allowed me to work with the Center for Dialysis Innovation. This project helped me understand my personal desire to use my work to contribute towards greater health missions such as the center's. I am so grateful to all the mentorship and discussion provided by those I had the pleasure of working with such as Buddy Ratner, Bruce Hinds, Johnathan Himmelfarb, Anna Galperin, and Runbang Tang. My experiences with and thanks to the Pfaendtner group will remain incredibly useful and memorable as I move forward with my post graduate career.

The final group of people I am forever indebted to are my friends and family. As I have already touched upon, Taylor Ludwig has provided me far too much love and kindness over our decade together while being the best friend anyone could ever ask for. Taylor always amazes me with what she is able to achieve. I have also been blessed with other wonderful people in my life that I hold dear and thank for their endless support. My brother and oldest friend, Mathew, has always been there to help me relive my nostalgia and remind me of the hobbies I enjoy so much. I couldn't be prouder of Mathew for the wonderful and caring person he has always been alongside his own personal accomplishments. Furthermore, both mine and Taylor's parents have strived to learn and understand us despite any differences between us. Specifically, my father Darrin, who has always made his presence and love for me known no matter the distance. Our grandparents as well have had a huge impact on our lives. Taylor's maternal grandmother, Mary J. Lewis has provided so much love and support to Taylor, myself, my brother, and all of our friends/family that I know each of us would not be where we are today without her. Finally, I want to thank our friends who we have relied on daily and are so grateful for; Cameron and Rachel Howard, Tristain Carter, Israel Price, Darcey Woodall, Nicklaus Palmer Del., Donna West, David Ulloa, Steven Holmes, and Johnny Bennet. Even as I wrote this section, they made me laugh, smile and cry with happiness. Each of them is far more incredible than they would admit, and I am so lucky to have met them all. Somehow, there are still many more family and friends not mentioned here that have helped me and I truly am forever grateful and love you all.

DEDICATION

for my partner, Taylor, who has given me both immeasurable happiness and a belief in myself that I never thought possible.

Chapter 1. **Introduction**

Kinetics and thermodynamics act as the backbone to numerous biochemical challenges which are key to novel drug/materials design and, more generally, the chemical understanding of such challenges. However, systems that would benefit from these types of information have remained difficult to unravel due to their chemical complexities or even restrictive requirements for study. In order to overcome these hurdles, molecular dynamics (MD) in combination with statistical modeling has proven to be both an effective and efficient tool, providing reliable results and cutting-edge discoveries. At the intrinsic level, MD allows for bridging between the provided atomistic resolution of MD and macroscopic observable properties of biophysical systems ranging from, but certainly not limited to, protein-ligand and cellular membrane dynamics. Beyond this, MD can be used for inherently hard to sample processes such as ligand unbinding or protein folding by using enhanced sampling techniques like metadynamics.[1–4] Combining MD with statistical modeling at the front and back ends provides opportunity to both improve simulation design and extend the reach of results. More than ever, the most useful statistical approaches in this domain fall under the label of machine learning (ML) due to the vast data sets being produced both computationally as well as experimentally. ML provides not only an expansive number of post-processing techniques, generally unsupervised learning methods like dimensionality reduction and clustering, but also a means for iterating and improving upon system design through deep learning. Thus, combining MD and machine learning provides one the ability to conduct a comprehensive investigation of biochemical challenges that are of current importance and can lead to advancements in the respective fields.

This dissertation presents my efforts towards tackling a set of relevant biophysical challenges through a computational lens whilst taking advantage of various ML methods. Each chapter presents either a unique system or mechanism and my accomplishments towards answering a relevant challenge to that topic. Chapter 2 delves into the coexisting liquid-ordered and liquid-disordered phases in multicomponent lipid bilayers that has received widespread attention due to its relevance for biological systems. I investigate the effects of alcohol additions into model membranes since their anesthetic properties may arise from the modulation of bilayer composition. This chapter also offers insight regarding the use of coarse-grained MD simulations to allow for the exploration of large model membrane systems in a computationally realistic and efficient time frame. In Chapters 3-5, a focus is placed on investigating the troublesome uremic toxins that are a leading contributor to negative chronic kidney disease patient outcomes. I begin in Chapter 3 by providing time-scale dynamics for 4 protein bound uremic toxins (PBUTs) within their primary binding pocket of human serum albumin (HSA), implementing a set of ML methods for post-processing and the creation of a Markov state model (MSM). Chapter 4 continues this look into PBUT bound to HSA by studying the unbinding kinetics and thermodynamics using enhanced sampling MD combined with deep learning for the a priori creation of an optimized sampling coordinate. Wrapping up my work on uremic toxins is Chapter 5 with an alternate backdrop in which I computationally model and seek to optimize a toxin capture strategy focused on highly specific adsorbent polymer materials. More specifically, I provide mechanistic information regarding an experimentally tested molecularly imprinted polymer (MIP) and potential improvements that can be made by altering the composition in a batch automated manner. I use Chapter 6 as a conclusion of my work, summarizing the impact I have made on the study of various biophysical fields and beyond.

Chapter 2. **Effect of alcohol on phase separation in model membranes**

2.1 *Introduction*

Many tertiary bilayers that consist of high melting point lipids, low melting point lipids, and cholesterol exhibit phase separation into liquid-ordered (L_o) and liquid-disordered (L_d) phases over a wide range of compositions.[5–7] This phenomenon is relevant to our understanding of biological systems because cellular membranes are believed to exhibit a similar, albeit highly dynamic and short ranged, heterogeneity in bilayer properties in the form of lipid rafts.[8,9] To further explore the nature of this phase separation and its connection to biological processes it is important to understand how the presence of additional molecules alters the properties of the coexisting phases. In this work, we study to what extent the addition of linear n-alcohols changes the composition and structure of the ordered and disordered phases in tertiary lipid membranes.

Our interest in n-alcohols stems from their anesthetic properties, which are believed to be related to their effects on membrane structure. Recent experiments on model membrane systems show that the addition of short chain alcohols to tertiary giant unilamellar vesicles (GUVs) stabilizes the separation into coexisting phases: the transition temperature increases, and so does the difference in order between the L_o and L_d phases.[10] It is not clear if this change in order is due to the alcohols directly or whether the alcohols alter the lipid compositions of the coexisting phases. Interestingly, adding such alcohols to giant plasma membrane vesicles (GPMVs) has the opposite effect: the transition temperature decreases, signaling a stabilization of the homogeneous phase.[11]These changes in transition temperature can be understood in terms of the partitioning of the additive into the different phases.[12]

To gain further insight into the interaction between n-alcohols and phase-separating lipid bilayers we perform Molecular Dynamics (MD) simulations using the coarse-grained (CG) MARTINI model, which is known to capture many relevant aspects of Lo-Ld phase separation in mixed lipid bilayers.[13–16] In particular, it has been shown to exhibit phase separation at low temperatures in tertiary mixtures of DPPC (1,2-dipalmitoyl-sn-glycero-3-phospocholine), DIPC (1,2-dilinoleoyl-sn-glycero-3-phosphocholine), and cholesterol (CHOL).[17] Using the MARTINI force field allows us to access the length and time scales necessary to reach equilibrium, which would be difficult to achieve with more detailed simulation models. The MARTINI model can be used to test various amphiphilic molecule's effects on biophysical properties and thermodynamic conditions in membranes using both a large library of molecule types and analysis tools.[18–22] It is worth noting that this approach is not without limitations as the MARTINI force field has been shown to struggle, for example, in correctly capturing realistic membrane protein aggregation, alongside other general pitfalls which have been discussed in the literature.[23,24] Despite such shortcomings, the membrane simulation community has made great strides using the MARTINI model in a variety of areas such as nanoparticle-membrane interactions, lipid fingerprinting, protein insertion, and phase separation in model membranes.[25–28] Finally, the MARTINI forcefield has also been used successfully to study n-alcohols over the past decade in regards to general bilayer properties, bacterial membranes, and bolalipid membranes.[29–32]

In this chapter we study the effect of alcohol on a ternary bilayer containing equal amounts DPPC, DIPC, and cholesterol. As others before us, we find that the latter spontaneously separates into coexisting liquid-ordered and liquid-disordered domains. We then isolate those domains and introduce alcohols, which allows us to directly observe structural, dynamical, and thermodynamic changes induced by the additive (Sections 2.3.1-2.3.4). Next, we study phase separation in presence of the alcohol to test whether the latter directly affects the coexisting compositions (Section 2.3.5). The latter is one possible explanation of the experimentally observed increase in ordering of the ordered phase if, for example, alcohol caused a repartitioning of cholesterol from the L_d to the L_o phase. Answering these questions will allow us to shed light onto several open questions surrounding these complicated systems.

In Section 2.2 we describe our simulation methods, details of free energy and analysis calculations, and the conditions of thermodynamic equilibrium in coexisting bilayer systems. We present our results in Section 2.3 for isolated L_d and L_o systems as well as a combined system that allows exchange of all molecule types between the phases. Finally, we present some concluding remarks regarding the simulation of model membranes and the MARTINI model in Section 2.4.

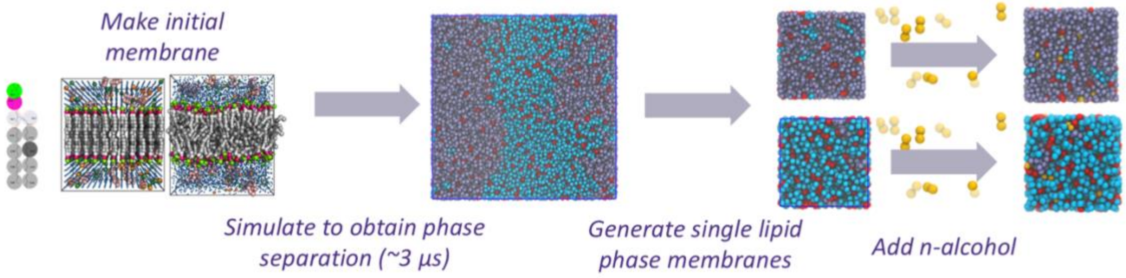


Figure 2.1 - Schematic for building MARTINI model membranes.

2.2 ***Methods***

2.2.1 *Simulation details*

Our starting point are simulations of spontaneously phase-separating bilayers of equal parts DPPC, DIPC, and CHOL. Simulations are run using the CG MARTINI 2.1 force field[33] in the GROMACS 5.1.4 suite.[34] The MARTINI model applies a 4-to-1 mapping of heavy atoms to so-called beads (radii ~0.26 nm).[35] Following this scheme, MARTINI n-alcohols are made up of at least one polar bead containing the hydroxyl group and additional beads for every four carbons beyond the four of the shortest single bead butanol. Membranes are constructed and remain perpendicular to the z-axis throughout all simulations. Initial configurations are created efficiently using the Insane (**INS**ert membr**ANE**) tool.[20] We have adapted Insane in order to handle n-alcohols/alkanes, allowing for quick construction of all systems.

All simulations followed the standard MARTINI simulation settings unless noted otherwise [36,37]. Simulations were performed using a time step of 30 fs and configurations were saved every 1000 steps (30 ps). Both the electrostatic and van der Waals interactions were truncated at 1.1 nm. A semi-isotropic Berendsen barostat was used with a reference point of 1 bar to control pressure and effectively maintain zero surface tension for the membrane.[16,38] A stochastic velocity rescaling thermostat was used to maintain a temperature of 300 K.[39] Each system is first energy minimized using steepest descent for 1000 steps before a brief relaxation run at 350 K for 30 ns.

2.2.2 *Formation of ternary membranes solvated with n-alcohol*

For the initial control system without alcohol, Insane places 33 molecules each of DPPC, DIPC, and CHOL each into each leaflet of a bilayer that is centered in our simulation box. This bilayer is then solvated. The system undergoes our equilibration procedure and is then replicated in both lateral directions 3 times to approximately 20 nm x 20 nm, followed by additional equilibration. The exact number of lipids and solvent of the control system is shown in **Table 2.1**. Equilibration of all systems was monitored both by tracking general membrane properties for consistency (box size, pressure, and interaction energies) alongside visualization of their density profiles. We observed spontaneous separation into liquid-ordered and liquid-disordered phases, consistent with previous simulation studies.[13,16,40,41] Simulations ran for 7.5 μs in the

NPT ensemble, with phase separation occurring between 2-3 μs. The last 4.5 μs of these trajectories were used for the calculation of reported quantities for the phase separated membranes, while 3 μs were used for the homogeneous single-phase systems. The phase separated membranes are approximately 4.5 nm thick with a ~2.25 nm thick solvent layer above and below the membrane (see **Table 2.1** for solvent amounts in the octanol system). The single-phase systems used for the free energy calculations have taller boxes (~17 nm) in the z direction to accommodate the removed cholesterol during enhanced sampling simulations. These result in solvent layers between 6.5-7.5 nm thick (corresponding to 4600-7000 MARTINI water beads). Otherwise, single-phase systems are created to be the same size as the phase separated system, but with lipid compositions that yield homogenous bilayers. All visual representations of simulations are created using the VMD package.[42] A snapshot of the final frame of such a simulation is seen in **Figure 2.2**.

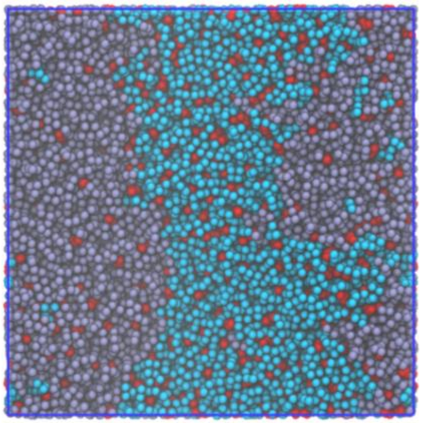


Figure 2.2 - Phase separated membrane of equal parts DPPC/DIPC/CHOL after 7.5 μs. DPPC is shown in blue, DIPC in purple, and CHOL in red. Striped phase domains form as expected for such a simulated square patch of membrane with periodic boundary conditions.

In order to test the effects of n-alcohols, the modified Insane tool is used to add approximately 1, 3, 5, 10, or 16 mole percent of n-alcohols (where n = 4, 8, 12, and 16) to the systems. These systems undergo the same minimization, equilibration and production scheme as the control system. **Table 2.1** shows the exact number of molecules in the phase-separated octanol membranes, and details the scheme used for the other alcohols and the following single-phase membranes. Biophysical properties are calculated using GROMACS and the MDAnalysis toolkits.[34,43] We use a Gaussian Mixture Model (GMM) analysis of the localized density distribution to obtain the compositions of the coexisting phases in these spontaneously phase-separating bilayers.[16] In this approach, we calculate the bilayer composition in a moderately sized (larger than a single molecule, but much smaller than the entire system) observation window. In a phase-separated system, this composition should reflect one of the two coexisting phases, unless the observation window happens to cover the interface between the two domains in which case an intermediate composition is observed. Fitting this data to both a single Gaussian and a triple Gaussian model allows us to identify whether phase separation has occurred, and if so what the compositions of the coexisting phases are. We find the equilibrium compositions to have a 6:71:23 (L_d) and 57:3:40 (L_o) DPPC/DIPC/CHOL ratio for the control (alcohol-free) system.

Table 2.1 - System outline of components for phase separated membranes solvated with octanol.

	Number of Molecules in system			
Component	**0 – Octanol (Control)**	**1 - Octanol**	**5 - Octanol**	**10 - Octanol**
DPPC	297	297	297	297
DIPC	297	297	297	297
CHOL	297	297	297	297
Octanol	0	9	45	90
Water	18000	17982	17910	17820

These compositions found from the phase separated control system are then used to construct initial configurations for homogeneous systems of only L_o or L_d phase bilayers. We create single-phase systems with alcohols of the same lengths and concentrations as before, again using the modified Insane program. Membrane thickness is calculated by measuring the distance between the phosphate head group beads of the upper and lower leaflets. The acyl-chain ordering parameter is calculated between two MARTINI lipid tail beads A and B as

$$S_{AB} = < \frac{3cos^2\theta - 1}{2} >$$

where θ is the angle between the AB bond and the z direction.[17] $S_{AB} = 1$ indicates perfect alignment, $S_{AB} = 0$ is random orientation, and $S_{AB} = -0.5$ is anti-alignment. S_{AB} is not necessarily equivalent to the experimental measure of order (for instance the deuterium order parameter), but it has proven a useful tool in detailing the flexibility of lipid tails.[44] The mole fraction partition coefficient of the alcohols is calculated as

$$K_X = X_{a,b}/X_{a,w}$$

where $X_{a,b}$ is the mole fraction of alcohol within the bilayer and $X_{a,w}$ is the mole fraction of alcohol in solution.[45]

2.2.3 *Calculation of Transfer Free Energies*

If two phases are in thermodynamic coexistence, the chemical potential of each component must be the same in both phases. For example, if μ_{L_o} and μ_{L_d} denote the chemical potential of cholesterol in coexisting ordered and disordered phases, then

$$\mu_{L_o} = \mu_{L_d}$$

must hold. In general, the difference $\Delta G_{trans} = \mu_{L_o} - \mu_{L_d}$ denotes the free energy cost of moving a cholesterol molecule from the ordered to the disordered phase, which is zero at coexistence. Because we are interested in how the addition of alcohol might affect the partitioning of cholesterol into the two phases, it would be useful to calculate the chemical potential in both ordered and disordered phases as a function of alcohol content. Calculating the chemical potential directly, however, is a challenging task using computer simulations. We therefore choose to instead calculate the free energy cost of moving a single cholesterol molecule from a single-phase lipid bilayer into the aqueous solvent. Denoting these energies ΔG_{L_o} and ΔG_{L_o} for the ordered and the disordered phase, respectively, we can reconstruct the transfer free energy using

$$\Delta G_{trans} = \Delta G_{L_o} - \Delta G_{L_d}$$

(see **Figure 2.3**). The use of a common reference state, a single cholesterol molecule in aqueous solution, thereby allows us to calculate the transfer free energies between bilayers of different composition without the complications of computing the chemical potential directly or performing a separate free energy calculation for every possible pair of membranes.

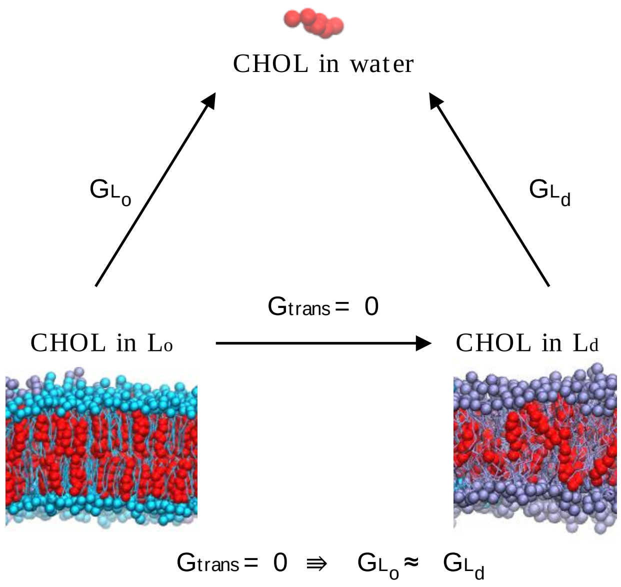


Figure 2.3 - Calculation of cholesterol transfer free energies. For each type of homogeneous bilayer, for example a liquid-ordered (bottom left) or liquid-disordered (bottom-right) one, we compute the free energy required to move one cholesterol molecule from the bilayer into aqueous solution (top). Because the latter state is the same in all calculations, we can use this data to reconstruct transfer free energies of cholesterol from one bilayer to another. If two bilayers are at coexistence, this transfer free energy is zero.

2.2.4 *Cholesterol removal*

Removing cholesterol from a membrane is an unfavorable process that will not occur in a simulation over accessible time scales. Rather, simulations will mostly sample regions of configuration space that are near potential energy minima. Events with an energy barrier above k_BT, where T is the absolute temperature and k_B is the Boltzmann's constant, will rarely be sampled. This is the case for cholesterol removal. However, such rare events can be observed using accelerated sampling techniques to overcoming the large energy barrier of removal.[46] This work uses the umbrella sampling (US) method.[47,48] After defining a reaction coordinate s, a continuous parameter that distinguishes between the two thermodynamic states of interest, US restrains this coordinate around a particular value by imposing a bias potential. A harmonic potential is commonly used as the bias due to its simplicity,

$$V_i(s) = \frac{K}{2}(s - s_i)^2,$$

and we choose K, the bias strength, to be 1000 (kJ/mol/nm^2). Our reaction coordinate is the distance between the membrane center of mass (COM) and the COM of a target cholesterol. 40 simulations are created varying the reaction coordinate across a range of interest (0 nm - 4.0 nm). Each simulation is run for 0.4 μs, for a total of 16 μs simulation time for a given system. We combine these simulations using the weighted histogram analysis method (WHAM)[49,50], producing the unbiased free energy profile. Statistical uncertainties were estimated using bootstrapping in g_wham, a GROMACS suite tool.[34,51] To improve sampling we used replica exchange, allowing neighboring simulation windows to swap configurations at a predetermined frequency based on the Metropolis criterion.[52,53]

2.2.5 *Calculation of alcohol and cholesterol trans-bilayer motion*

Another interesting biophysical property we measure is the rate of trans-bilayer motion, or flip-flop, of components within a membrane. Lipid flip-flop is used biologically in order to maintain an asymmetric distribution of lipids which plays a role in processes like blood coagulation and cell apoptosis.[54,55] Cholesterol flip-flop has been studied in heterogeneous membranes, yielding flip-flop rates faster than that of phospholipids.[56,57] It has also been shown that the cholesterol flip-flop rate depends upon membrane order[58], giving another reason to investigate this property in our systems. The flipping rate is calculated for both CHOL and n-alcohols in the single-phase membranes. To measure the flip-flop rate of cholesterol and alcohols in our simulations, we assign a molecule to one of multiple possible states based on the z-position of its polar group relative to the membrane: a molecule can be either part of the top leaflet, the bottom leaflet, the region in-between the two leaflets at the center of the bilayer, or in solution. These regions were chosen based on the density profile of the bilayer in the normal direction. Using an in-house python script, the position and widths of each region as well as the membrane thickness were determined from Gaussian fits to each leaflet's density profile (this code is available at the repository: https://github.com/jamesprg/n_alchohol_MARTINI). To avoid measuring an artificially high transition rate due to molecules fluctuating back and forth across the boundaries, these regions are not connected; instead, they are separated by several Angstroms. If a molecule falls into one of those gaps its state is not determined by its current position, but instead by the last and the next positively identified state using the transition-based assignment method.[59] We count each successful flip-flop event from one leaflet to the other that proceeded through the interior of the bilayer (and not through the periodically replicated solvent). The flip rate of component x, F_x, is then calculated as follows

$$F_x = \frac{f_x}{t \times N_x}$$

where f_x is the number of flip-flop events, t is the simulation time used for analysis, and N_x is the number of x molecules in the system.

2.3 ***Results and discussion***

2.3.1 *Alcohol partitioning between solvent and membrane*

In order to study the partitioning of alcohols between MARTINI model membranes and water we compute the partitioning coefficient. Our results are shown in **Figure 2.4 (a)/(b)**. As expected, we find that the longer n-alcohols result in a larger percentage of membrane partitioning, likely due to the long hydrophobic tails having a greater preference for the hydrophobic region of the membranes. When comparing the two phases, both hexadecanol and dodecanol partition more strongly into the ordered phase than into the disordered phase. However, in the case of octanol, partitioning into either phase becomes more similar in magnitude. Finally, butanol partitions approximately twice as strongly into the disordered phase over the ordered phase for all compositions. We see that for the MARTINI model, in which the smallest n-alcohols are single bead molecules, the alcohol shorter than octanol studied here (butanol) partitioned and remained more stably in the disordered L_d phase. When moving to longer alcohols such as dodecanol, the trend appears to reverse. This could be due to the alcohols becoming more similar in size to the lipids, comparable to cholesterol. This fact, alongside their larger hydrophobic region, could create a preference for the L_o phase due to a strong driving force to keep these longer hydrophobic tails well shielded from the solvent by the more stable and thicker ordered membrane.

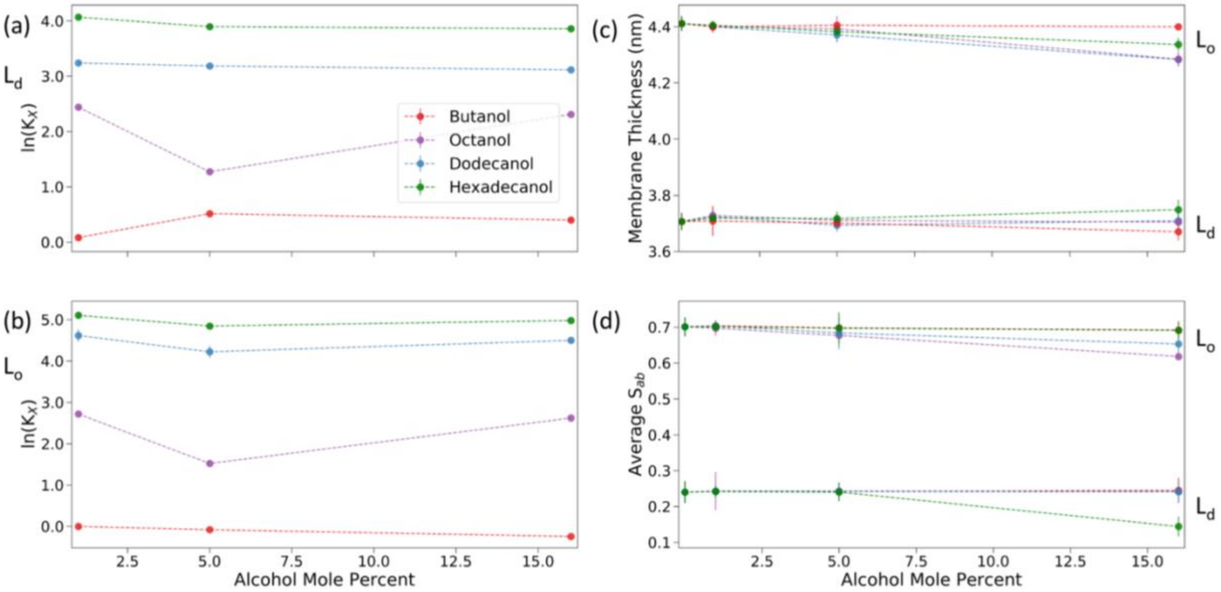


Figure 2.4 - Alcohol partitioning and induced structural changes in homogeneous ordered and disordered membranes. Panels (a) and (b) show the partition coefficient as a function of alcohol mole percent for the L_d and L_o phase, respectively. Alcohol concentrations vary between 0, 1, 5, and 16 mole percent. Panels (c) and (d) show the membrane thickness and average order parameter as a function of alcohol mole percent for the L_d (lower sets of data) and L_o (upper sets of data) phase, respectively. Statistical uncertainties were calculated from observations in 4 independent trials for each data point.

2.3.2 *Alcohol effects on the structure of L_o and L_d membranes*

We now compare the structural properties of the single-phase systems. **Figure 2.4(c)** shows the average distance between the head group beads of lipids in the upper and lower leaflets as a function of alcohol mole percent for various n-alcohols. All four alcohols tested do not affect the L_o or L_d membrane thickness by more than 0.2 nm for any mole percent. Neither the increase in alcohol length nor mole percent led to a monotonic increase or decrease in membrane thickness for any case except for dodecanol in the L_o phase. This indicates that the addition of the n-alcohols tested does not significantly alter membrane thickness of either phase regardless of mole percent.

We reach similar conclusions by analyzing the average acyl-chain order parameters shown **Figure 2.4(d)**. An increase in octanol and dodecanol tends to slightly decrease the average order parameter. We note that this observation is the opposite of that made in experiments, which showed an increase in order of the Lo phase in the presence of alcohol.[10] However, the two systems are not directly comparable: unlike the experiment, we here add the alcohols to single-phase systems. In addition, membrane order was measured in the experiment using laurdan polarization, whereas here we use the tail order parameter. Our simulation data show that increasing the alcohol length in the 16 mole percent systems from butanol to octanol leads to a larger decrease in order parameter (0.074) than compared to the decrease found when going to dodecanol instead (0.035). However, hexadecanol is within the calculated uncertainty of butanol, indicating that the minimal disordering effect of the L_o phase observed is only occurring for alcohols in between the shortest and longest tested. The pure L_d phase shows no difference in order parameter between all alcohols/compositions tested except for hexadecanol. Furthermore, the notable decrease in order only occurs in the 16% hexadecanol L_d system, similar to the trend in the L_o phase as well as the general increase in non-lipid component. We suspect hexadecanol's difference in this regard to be related to its similar length to the lipids in the system. Overall, these observations imply that the addition of alcohols does not significantly increase the order of the pure L_o or L_d membranes. However, it seems that alcohols which begin to reach the length of the lipids in the system could have greater effects, and in this case, lead to a decrease in order in either phase, more prominently so in the L_d phase. Changes in order parameter might be difficult to detect using coarse-grained methods due to the general loss of information compared to all-atom models. To our knowledge, a systematic comparison of the ordering of small molecules such as n-alcohols in model membranes in different forcefields does not yet exist. However, there has been work done studying general lipids in this regard, which has found coarse graining to lead to more flexible hydrocarbons[60] as well as confirming order parameters between all-atom and MARTINI forcefields.[44]

2.3.3 *Alcohol effects on cholesterol chemical potential in L_o and L_d membranes*

As described in Sections 2.2.3/2.2.4, we calculate the free energy of moving a cholesterol molecule out of the bilayer and into solution. We begin by studying homogeneous L_o and L_d bilayers whose compositions were obtained from the phase-separating control simulation without alcohol (**Figure 2.5(a)**). We find similar free energies of that $\Delta G_{L_o} = 90.9 \pm 2.8\ kJ$ for the ordered and $\Delta G_{L_d} = 82.1 \pm 3.4\ kJ$ for the disordered phase, as expected for two phases that are in equilibrium. That they are not exactly equal is likely due to uncertainties in the exact composition and limited flexibility of setting the composition in such small systems.

In panels (b) and (c) we show the same free energy profiles in the presence of various mole fractions of octanol. Neither system deviates outside of the uncertainty range for any of the four compositions studied. This suggests that the alcohol does not alter the favorability for cholesterol to remain in either pure phase membrane. It also affirms the accuracy of our simulations. We emphasize that these results are produced from single-phase membrane systems, and we will later compare them to changes caused by alcohol additions in systems containing both L_o and L_d phases.

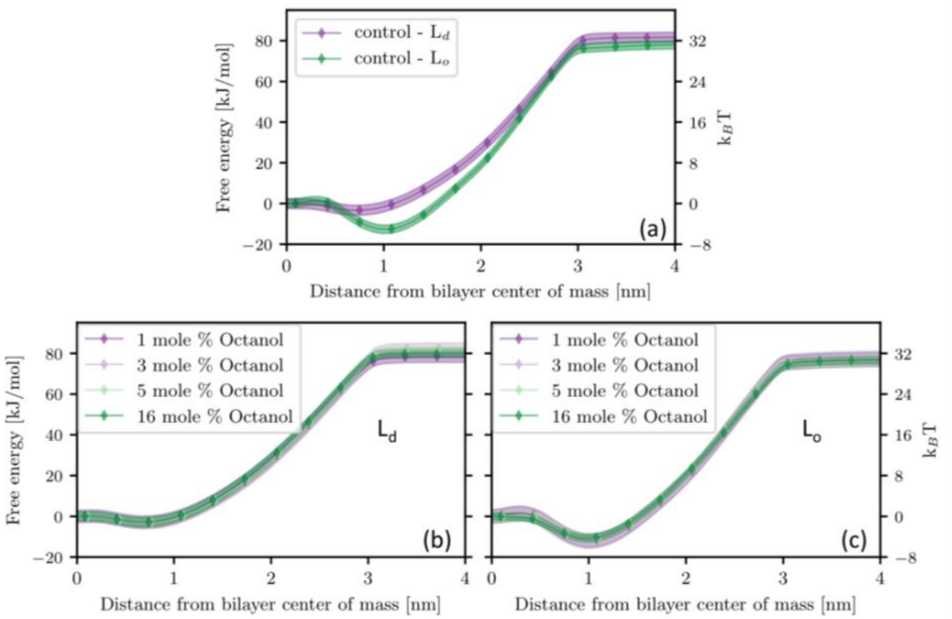


Figure 2.5 - Free energy profiles for moving cholesterol out of single-phase bilayers. (a) The free energy profile of removing a cholesterol molecule from a liquid-ordered membrane shows a more pronounced minimum than for a disordered membrane in the control simulations. Adding alcohols does not alter the free energy profiles for disordered (b) or ordered (c) bilayers.

2.3.4 Trans-bilayer motion of n-alcohols and cholesterol in L_o and L_d membranes

Our results for the flip-flop rate of alcohols and cholesterol are summarized in **Table 2.2/2.3** for single-phase disordered and ordered bilayers, respectively. Flip-flop rates for cholesterol range from 2.1 x 10^7 to 3.6 x 10^7 s^{-1} in the L_d phase and from 9.31 x 10^4 to 6.36 x 10^5 s^{-1} in the L_o phase. In agreement with previous studies, the disordered phase results in a more frequent amount of flipping events, likely due to the decreased rigidity of the lipid acyl-chain tails and increased fluidity.[58] In comparison, the n-alcohol flip-flop rates range from 5.0 x 10^6 to 6.2 x 10^7 s^{-1} in the L_d phase and from 7.5 x 10^4 to 3.0 x 10^5 s^{-1} in the L_o phase. All n-alcohols also exhibit flip rates at least an order of magnitude faster in the L_d phase compared to the L_o phase. To compare the rates between alcohol and cholesterol, we calculate the flip rate ratio (alcohol rate divided by cholesterol rate); results are shown in the final column of each table. We find that only butanol (for both L_o and L_d phases) shows smaller flipping rates than cholesterol.

In all other systems, the alcohol flip rate is found to be faster, an effect most prominent in the L_o phase. This result shows that n-alcohol flip-flop rates are less reliant on the order of a particular membrane. We also note that octanol observes a uniquely higher flip-flop rate in the L_o phase compared both to its L_d counterpart and all other n-alcohols studied, which might relate to the partition coefficient cutoff behavior discovered in section 2.3.1. All n-alcohols observe a monotonic increase in flip rate with total alcohol concentration, aside from butanol which remains constant within calculated uncertainties. Also notable is the result that the cholesterol flip rate increases monotonically with total alcohol concentration for all systems (again aside from butanol for the L_o phase only), which one might interpret as a sign of increased disorder for both the L_d and L_o phases. However, this conclusion is not supported by our measurements of the order parameters discussed in section 2.3.2. Thus, it seems that the flip-rate is acting less like a measure of the tail order parameter, but instead correlates more strongly with other measures of fluidity. Furthermore, the length of the alcohol is also correlated to the monotonic increase in cholesterol flip rate with alcohol concentration in the L_d systems. In general, the liquid-ordered system shows a maximum in the flip rate ratios for octanol; for longer alcohols the flip rate ratio decreases but remains greater than unity. This again is indicative of a potential cut-off in terms of ordering effects once alcohols reach lengths similar to that of the lipid components.

Table 2.2 - Average flip-flop of CHOL and n-alcohols in L_d phase membranes.

		Alcohol		Cholesterol		
n-alcohol	**Alcohol %**	**Total Flips**	**Flip Rate (s^{-1})**	**Total Flips**	**Flip Rate (s^{-1})**	**Flip Rate Ratio**
Butanol	**1**	7.6×10^{1}	5.3×10^{6}	6.9×10^{3}	2.1×10^{7}	0.25
	5	3.0×10^{2}	5.0×10^{6}	5.9×10^{3}	2.2×10^{7}	0.22
	16	9.7×10^{2}	5.1×10^{6}	5.2×10^{3}	2.3×10^{7}	0.22
Octanol	**1**	3.1×10^{2}	2.6×10^{7}	6.3×10^{3}	2.3×10^{7}	1.13
	5	2.0×10^{3}	2.8×10^{7}	7.8×10^{3}	2.5×10^{7}	1.14
	16	6.5×10^{3}	3.4×10^{7}	6.3×10^{3}	2.8×10^{7}	1.23
Dodecanol	**1**	5.8×10^{2}	4.8×10^{7}	7.5×10^{3}	2.7×10^{7}	1.76
	5	3.1×10^{3}	5.2×10^{7}	8.0×10^{3}	3.0×10^{7}	1.72
	16	1.2×10^{4}	6.2×10^{7}	8.0×10^{3}	3.5×10^{7}	1.77
Hexadecanol	**1**	5.5×10^{2}	4.6×10^{7}	8.0×10^{3}	2.9×10^{7}	1.60
	5	3.1×10^{3}	5.2×10^{7}	8.1×10^{3}	3.1×10^{7}	1.69
	16	1.1×10^{4}	5.9×10^{7}	8.2×10^{3}	3.6×10^{7}	1.66

Table 2.3 - Average flip-flop of CHOL and n-alcohols in L_o phase membranes.

		Alcohol		Cholesterol		
n-alcohol	**Alcohol %**	**Total Flips**	**Flip Rate (s^{-1})**	**Total Flips**	**Flip Rate (s^{-1})**	**Flip Rate Ratio**
Butanol	**1**	1.3	8.7×10^{4}	6.6×10^{1}	1.14×10^{5}	0.76
	5	4.5	7.5×10^{4}	4.5×10^{1}	9.76×10^{4}	0.77
	16	1.9×10^{1}	9.8×10^{4}	3.8×10^{1}	9.31×10^{4}	1.05
Octanol	**1**	9.3	7.7×10^{5}	4.9×10^{1}	1.02×10^{5}	7.55
	5	9.6×10^{1}	1.3×10^{6}	1.0×10^{2}	1.85×10^{5}	7.15
	16	5.8×10^{2}	3.0×10^{6}	2.6×10^{2}	6.36×10^{5}	4.74
Dodecanol	**1**	5.0	4.2×10^{5}	5.2×10^{1}	1.07×10^{5}	3.88
	5	4.7×10^{1}	7.9×10^{5}	7.1×10^{1}	1.55×10^{5}	5.08
	16	3.1×10^{2}	1.6×10^{6}	1.4×10^{2}	3.46×10^{5}	4.74
Hexadecanol	**1**	3.8	3.1×10^{5}	4.7×10^{1}	9.74×10^{4}	3.21
	5	2.1×10^{1}	3.5×10^{5}	5.7×10^{1}	1.25×10^{5}	2.77
	16	8.9×10^{1}	4.6×10^{5}	5.9×10^{1}	1.45×10^{5}	3.19

2.3.5 *Alcohol effects on equilibrium compositions of membranes exhibiting phase coexistence*

In this section we switch our focus to bilayer systems in which liquid-ordered and liquid-disordered regions coexist. We determine the composition of the L_o and L_d phase in relatively large systems of about 20 nm per lateral dimension. Number density profiles were generated, from which we calculate the mole fraction of each component. Three separate simulations were averaged to obtain the density profiles shown in **Figures 2.6 (a)/(b)**. These profiles account only for the total number of lipids, as the alcohol is able to leave the membrane and is therefore considered part of the solvent. The smoothness of these profiles is indicative of the small variance between replicate simulations. In **Table 2.4**, each component's mole fraction from the density profiles is given as an average over a particular membrane strip in the x direction that was found to be of a single phase for the simulation time analyzed.

Table 2.4 - Lipid mole fractions determined from number density profiles.

	L_o			L_d		
Lipid type	**0% Octanol**	**5% Octanol**	**10% Octanol**	**0% Octanol**	**5% Octanol**	**10% Octanol**
DPPC	0.542 ± 0.001	0.492 ± 0.001	0.531 ± 0.001	0.039 ± 0.001	0.064 ± 0.001	0.043 ± 0.000
DIPC	0.020 ± 0.001	0.074 ± 0.001	0.023 ± 0.001	0.802 ± 0.001	0.778 ± 0.002	0.814 ± 0.001
CHOL	0.438 ± 0.001	0.434 ± 0.001	0.446 ± 0.001	0.159 ± 0.002	0.158 ± 0.001	0.143 ± 0.001

In order to analyze the composition changes between systems, we average the portions of the curves within the bulk regions of either phase. We see that in the 5 mole percent octanol system, DIPC had a slight decrease in the L_d phase by 2.4%. We find the opposite trend when increasing to 10 mole percent octanol in the L_d phase as the DIPC composition increases by 3.6%. The DIPC fraction in the L_o phase (or DIPC poor phase) of the 5 mole percent octanol system increases by 5.4%. Interestingly, this shift is reversed when increasing the octanol mole percent to 10, where we also observe a rise in the L_d phase's DIPC fraction. We also determine shifts in both DPPC and cholesterol between different systems. The largest change to the cholesterol profiles is observed in the L_d phase with 10 mole percent octanol, differing from both the 0 and 5 mole percent systems by at least 1.5%. Furthermore, the DPPC and cholesterol together appear to make up for the shifts previously observed in the DIPC profiles: in L_o phase, when jumping from the 0 to 5 octanol mole percent systems where DIPC increases by 5.4%, DPPC decreases by 5.0%. This becomes equal if we factor in the 0.4% decrease also observed for cholesterol. When moving from 5 to 10 octanol mole percent, there is an increase in DPPC by 3.9%. Again, this matches the DIPC decrease (5.1%) if we account for the cholesterol increase by 1.2%. Finally, the 3.6% increase in DIPC mole fraction for the L_d phase from 5 to 10 octanol mole percent is accounted for by the 2.1% and 1.5% decreases in DPPC and CHOL mole percent respectively.

We also calculate the number density profile for octanol in **Figure 2.6 (b)**. This profile illustrates directly octanol's preference for the disordered over the ordered phase: octanol prefers the L_d region over the L_o by 3.6:1 and 2.9:1 for the 5 and 10 mole percent systems, respectively. The magnitude of this preference is larger than what one might expect from the calculations of partition coefficient into single-phase bilayers discussion in section 2.3.1. However, the two calculations are not equivalent: while in the single-phase systems the composition of both the ordered and disordered phases are fixed, here they can vary by lipids and cholesterol moving from one phase to the other in response to the addition of alcohols. An exact agreement between these two calculations should therefore not be expected.

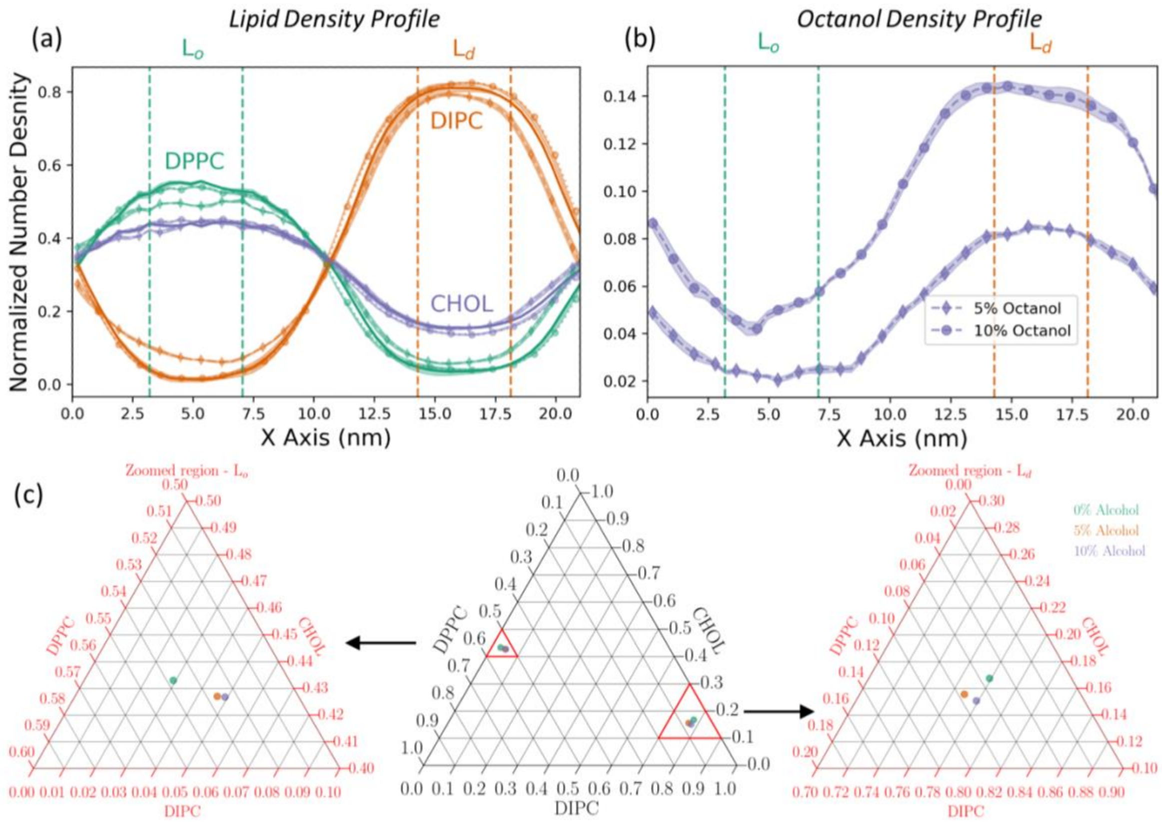


Figure 2.6 - Membrane component density profiles and ternary composition diagrams. Upper row: Number density profiles are plotted corresponding to the three octanol mole fractions: 0% (solid), 5% (diamonds), and 10% (circles). The center of the plot represents the L_d phase while the right and left sides are L_o phase. Dashed lines denote the two regions used to average over for analysis. Panel (a) is calculated using lipid components only while (b) shows the profiles for octanol. The number density is normalized based on all components. (c) Ternary phase triangles with compositions of various octanol concentrations. Each point is an average over three simulation runs, where compositions were calculated using the GMM method. Each axis on a triangle shows the mole fraction of the given component. The central triangle shows the full composition space while the left and right triangles show magnified regions highlighting L_o and L_d phases.

Determination of the compositions of coexisting phases from density profiles is hampered by fluctuations of the interface between the phases as well as lacking information about the correlations between the compositions of different species. We therefore tested the accuracy of our results by using a Gaussian Mixture Model analysis of our simulation data in the three-dimensional composition space.[16] **Figure 2.6 (c)** shows the ternary composition diagram with the coexisting equilibrium compositions obtained for the 0, 5, and 10 octanol mole percent systems, again averaging over three separate simulations. In this composition triangle, L_d and L_o

phases are easily distinguished. As with the number density profiles, we look to determine the shift in composition at various octanol concentrations. Similar to our results from the number density profiles, DPPC and CHOL mole fractions in the L_o phase both decrease by approximately 0.01, while that of DIPC increases by approximately 0.02 mole fraction. Analysis of the L_d phase ternary triangle also reinforces the density profiles. We find that the DIPC fraction initially decreases to 0.768 in the 5 octanol mole percent system before increasing back to 0.779 in the 10 octanol mole percent system, following the trend observed in the density profiles. Each component's mole fraction is given in **Table 2.5**.

Table 2.5 - Lipid mole fractions determined from Gaussian Mixture Models.

	L_o			**L_d**		
Lipid type	**0% Octanol**	**5% Octanol**	**10% Octanol**	**0% Octanol**	**5% Octanol**	**10% Octanol**
DPPC	0.538 ± 0.006	0.527 ± 0.003	0.524 ± 0.004	0.054 ± 0.005	0.076 ± 0.006	0.071 ± 0.005
DIPC	0.029 ± 0.005	0.046 ± 0.004	0.049 ± 0.006	0.780 ± 0.010	0.768 ± 0.005	0.780 ± 0.010
CHOL	0.433 ± 0.002	0.427 ± 0.008	0.427 ± 0.002	0.170 ± 0.010	0.156 ± 0.002	0.150 ± 0.010

2.4 *Conclusion*

Our results on single-phase model membranes suggest that butanol, octanol, and dodecanol additions cause little to no effect in membrane thickness or order parameter at concentrations at or below 16 mole percent. Neither the single-phase nor the full model membrane systems displayed any monotonic shifts in order with increasing alcohol mole percent or length as anticipated by the ordering effects seen in experimental GUVs. However, flip-flop rates suggest a decrease in order, represented by the increase in cholesterol flip-flop, for all systems aside from butanol in the L_o single-phase membranes. We show that MARTINI model membranes do not exhibit changes in lipid composition greater than 5.4 mole percent between systems with 0, 5, or 10 mole percent of octanol. Our GMM analysis supports very small composition changes, finding the maximum compositional change to be a decrease in DPPC mole fraction by 2.23 in the L_o phase after adding 5 mole percent octanol to the membrane. The smaller maximum change found by the GMM supports the choice of this method, avoiding the larger fluctuations found in the density profiles alone while still observing similar behavior. Finally, we used umbrella sampling to calculate the free energy of cholesterol removal from each single-phase membrane and found that increasing the octanol content caused no significant change to cholesterol's preference for either phase.

Despite the minute composition changes observed when adding particular alcohols to our systems, the addition of alcohols led to a decrease in the distinction between the two phases. This occurs when either phase loses a portion of its natural lipid type or gains lipids of the alternate type. Also, of note is the preference octanol exhibits for the L_d phase over the L_o phase by at least 2.9:1 in the phase separated membranes. This preference has been discussed in recent literature but is still not fully understood.[61] A potential explanation for this effect is that the disordered tails of L_d phase lipids lead to an easier partitioning for shorter alcohols, which our data suggests are those of octanol's length and below. As indicated by our partition coefficient calculations in single-phase membranes, longer alcohols such as dodecanol show stronger partitioning into the L_o phase. This effect could be due to their similarly long hydrophobic tails which would incur larger free energy penalties compared to shorter alcohols if not properly shielded by ordered head groups.

In contrast to our results in homogeneous single-phase membranes, the experimental results by Cornell and coworkers display a clear monotonic increase in the order of the L_o phase of model membranes in GUVs that contain both liquid phases. There are multiple possible reasons for this inconsistency. First, the generalized polarization of laurdan that was used in the experiments is only an indirect probe of membrane order and might yield information that cannot be directly compared to the membrane thickness and tail order parameter discussed here. Second, the lipids used in our simulation study are not the same as in the experiments: due to a deficiency in the Martini model, phase separation does not occur spontaneously in DPPC:DOPC:CHOL membranes, and we used the more highly unsaturated DIPC lipid instead. The difference in acyl chain chemistry may be the origin for the difference in observed behavior. Third, it is possible that the MARTINI force field in its current form does capture the intricate interactions in these systems with sufficient accuracy. While its coarse-grained descriptions of phospholipids and cholesterol have been well-tested, its combination of alcohols has received much less scrutiny so far. Previous simulation studies have found that the effect of alcohol on bilayer properties can be rather subtle.[32]

Our simulation results suggest several potential experiments that could shed further light on these phenomena. It would be useful to perform experiments on butanol in addition to the longer chain alcohol to fully appreciate the variety of effects that can be induced by alcohols. It would also be beneficial to obtain NMR or other experimental data on the tail ordering in ternary lipid bilayers in the presence of alcohol, as this would allow a direct comparison with simulations. Finally, it would be interesting to repeat the laurdan experiments on single-phase GUVs that are either completely liquid-ordered or completely liquid-disordered, to test the extent to which the currently observed changes are due to changes in lipid composition upon the addition of alcohol.

Chapter 3. **Analyzing the long time-scale dynamics of uremic toxins bound to Sudlow site II in human serum albumin**

3.1 *Introduction*

Despite increased attention received in recent years, the efficient removal of protein bound uremic toxins (PBUTs) remains a challenge towards the improvement of outcomes for chronic kidney disease (CKD) patients.[62–67] PBUTs are uremic retention solutes, generally < 500 Da, which have a high affinity for carrier proteins like human serum albumin (HSA).[64] The difficulties surrounding PBUT removal are twofold: 1) uremic toxins binding transport proteins have complex binding pathways that are not well understood and 2) uremic toxins have low free fraction in the bloodstream making them resistant to traditional dialysis methods. To overcome these difficulties related to PBUT capture, two major strategies have been suggested. The first strategy involves the use of competitive binders in protein binding pockets to increase toxin free fraction in the bloodstream. Competitive binders such as ibuprofen have been shown to increase the amount of unbound PBUTs that bind the HSA Sudlow Site II such as indoxyl sulfate (ISO), p-cresyl sulfate (PCS), indole-3-acetic acid (IAA), and hippuric acid (HA).[68] The second strategy focuses on improving the ability to capture free toxin in solution. One example being the design of highly specific nanoporous adsorbents, developed with internal surface areas suitable for non-specific adsorption of hydrophobic molecules, which has been shown to improve the capture of free toxin.[69,70] Such approaches take steps towards improving the management of PBUTs in CKD patients. However, gaps in our molecular level understanding of these mechanisms prevent widespread use or rational design of more efficient toxin capture materials. Gaining atomic-scale insight into the dominant interactions that drive PBUT binding within a native binding pocket could greatly assist in this goal.

We have previously used molecular dynamics to provide general mechanisms of the PBUTs binding in Sudlow site II alongside insight into toxin-specific interactions.[71] This was done using experimental structures of ISO bound to Sudlow site II of HSA to create initial binding poses of 3 PBUTs (PCS, IAA, and HA).[72] These toxins were chosen due to the body of literature supporting their link to negative CKD patient outcomes and chemical similarities.[62,63,67,68,73] These four toxins span a chemically diverse space as highlighted in **Figure 3.1** which details the substitution of either the anionic or hydrophobic group (or both) between systems. Sudlow site II is an appropriate binding site to study since it is already the primary binding site for two of the four toxins ISO and PCS.[74] Our earlier work provided equilibrium characterization of the predicted binding pockets on the time scale of 100s of ns. Herein, we expand this by studying the long-timescale dynamics within the binding pocket of each protein-toxin system. In order to study these binding mode kinetics, we build Markov state models (MSM), which has the ability to estimate timescales for events that are much longer than typically accessible in biomolecular MD simulations.[75,76] In recent years, using MSMs to interpret and expand the results of classical MD simulations has found great use in a variety of biomolecular challenges.[77,78] For example, MSMs have been used to great effect in areas such as protein folding,[79,80] protein-ligand binding,[81,82] intrinsically disordered proteins,[75,83,84] and native state conformational changes.[85,86] Thus, we provide an improved understanding of interactions within the binding pocket and the possible pathways PBUTs may travel before unbinding.

Using an MD approach while desiring to obtain meaningful kinetic information from protein-ligand interactions forces us to employ a method for handling the high dimensional data produced. However, a natural advantage of MSMs is that they require a dimensional reduction of the input features in order to extract reliable kinetic information from such large data sets.[75] In spite of the great potential of MSMs, one obstacle is the selection of key degrees of freedom that quickly decorrelate. Schwantes and Pande have addressed this through the use of time-structure independent component analysis (tICA) as a means of picking out the quickly decorrelating degrees of freedom.[87] Analysis using tICA involves finding the linear combinations of a set of features that maximizes the autocorrelation function of that projection while requiring each linear combination be uncorrelated to the previous ones. By implementing these methods, we can project high-dimensional MD data onto the slowest degrees of freedom. Thus, we will use tICA for constructing an MSM in order to describe the kinetic transitions of our protein-ligand binding pocket interactions.

In this work, we combine MD simulations with tICA-informed MSMs in order to quantify the binding mode kinetics of four previously investigated PBUT-HSA complexes. We detail the major binding modes for each system as well as estimate the timescales for the slowest transitions within the binding pocket. Furthermore, we use these models to help outline potential pathways for toxin unbinding. This work increases understanding of the HSA binding pocket on a molecular level by analyzing four important PBUTs relevant to CKD. Furthermore, analysis of the long timescale dynamics within the binding pocket should help in the development of ideal reaction coordinates for future studies using enhanced sampling techniques to study PBUT unbinding. This direction would allow for estimation of the kinetic properties of uremic toxin unbinding which has been shown to be the leading marker in determining the clinical efficacy of drugs.[88,89] However, these properties remain very challenging to measure experimentally and calculate using simulations. Thus, taking steps to improve the design and understanding of the HSA-PBUT system's dynamics in a binding pocket of interest could prove critical to the future improvement of drugs designed treat CKD.

The remainder of this chapter is organized as follows. In section 3.2, we highlight our simulation setup, methods for selecting the relevant features of each protein-toxin system, and construction of the MSMs. To start section 3.3, we verify the validity of our MSMs before detailing the discovered binding modes and key interactions in each system. Analysis of the MSM eigenvectors is then used to report and discuss kinetic information surrounding toxin transitions within the binding site. Concluding remarks highlight areas of future work for mechanistic studies of PBUTs in section 3.4.

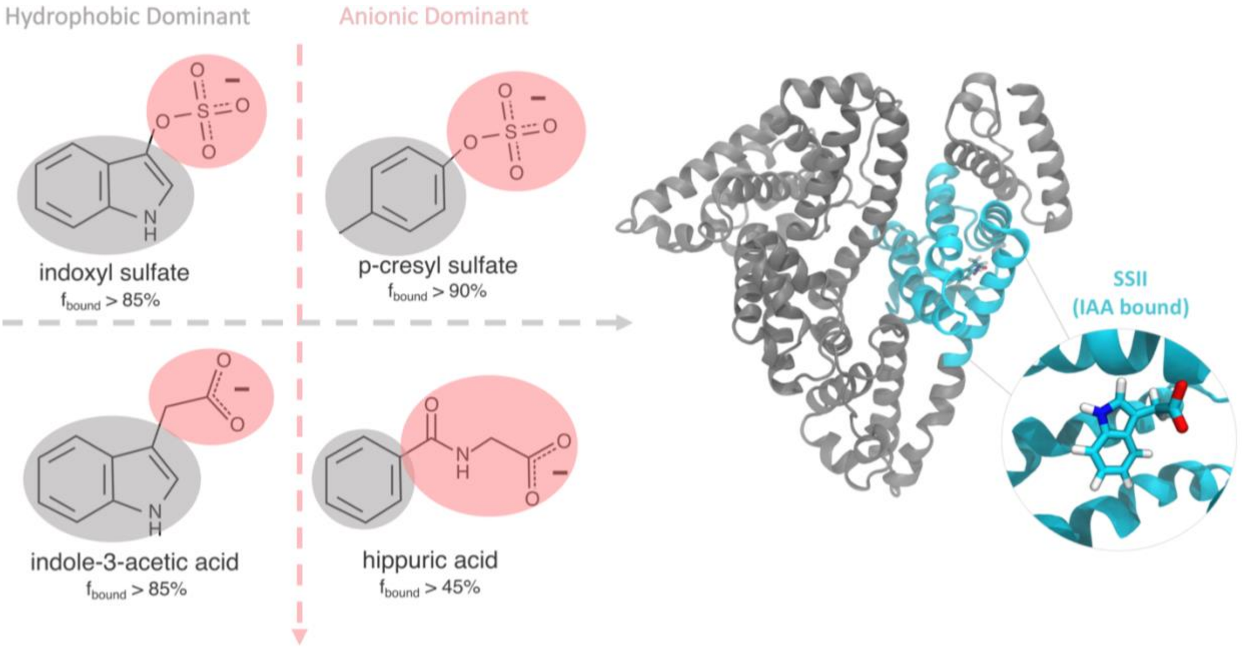


Figure 3.1 - Target toxin design space and HSA binding site. The left panel shows our toxin design space, where the vertical axis tracks a shift in the hydrophobic group and the horizontal tracks the anionic group. The right panel is a snapshot of a molecular dynamics run, displaying the HSA protein bound by IAA at Sudlow site II.

3.2 *Methods*

3.2.1 *Initial structures and details of MD simulations*

The four starting structures for toxin bound to Sudlow site II were adapted from Smith and Pfaendtner.[71] These individual toxins were combined with the HSA protein from an experimentally resolved structure (RCSB PDB ID: 2BXH) using the Amber 14 forcefield parameters.[72,90] The systems were solvated in a cubic box (L=10.8 nm) with TIP3P water and 14 sodium counter ions.[91] Following this system setup, MD simulation was conducted using Gromacs 2016.3.[34] System equilibration was done in a three-step process as follows; energy minimization, annealing, and a short NPT simulation. For energy minimization, each system was treated with a steepest descent algorithm for 10000 steps. All simulations beyond the point use an MD timestep of 2 fs, using the LINCS algorithm to constraint all X-H bonds.[92] Temperature was ramped over 250 ps to 298K, followed by a short NPT simulation of 500 ps for fast mechanical equilibration with the Berendsen barostat at 1 bar.[93]

Based on guidelines from prior studies recommending at least 12 production runs[94], we performed 15 NPT simulations for each protein-toxin complex with a randomly generated initial velocities. Systems were held at 298 K and 1 bar using the Bussi-Donadio-Parrinello thermostat and Parrinello-Rahman barostat.[39,95] Each simulation was carried out for 250 ns of which the first 50 ns are not analyzed (used for the purpose of extra relaxation). Furthermore, frames are saved every 2 ps for analysis. Any error bars obtained are calculated from an aggregation of the set of simulations for a particular system with standard deviations calculated from each mean. The Plumed 2.4 plugin was then used to post process the raw simulation data as described in the following section.[96]

3.2.2 *Key residue determination in binding pocket-toxin interactions*

In order to obtain the key interactions taking place in the complex protein-ligand binding pocket, an initial set of residues for study needed to be determined. These residues were selected from the static starting structures, prior to analysis of any MD results. Each list of residues was started by identifying any protein residue within 0.6 nm of any atom of a toxin. The list was then curated by calculating the number of atomic and hydrophilic contacts using the following switching function:

$$N_{contacts}(\boldsymbol{r}) = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{r_{ij} - d}{r}\right)^{6}}{1 - \left(\frac{r_{ij} - d}{r}\right)^{12}},$$

where A and B are the toxin and residue's heavy atoms (non-hydrogen), respectively, *d* (diameter) and *r* (radius) control the shell formed around each atom, and r_{ij} is the distance between atoms *i* and *j*. Contacts were calculated every 10 fs (5 MD steps) counting atomic contacts within a coordination sphere of radius 0.4 nm (r = 0.4 nm, d = 0 nm) while hydrophilic contacts were counted using a coordination annulus set to 0.27 nm in diameter and 0.03 nm thick (r = 0.03 nm, d = 0.27 nm) as used in our previous work.[71] We then cut the initial list of residues using the resulting highest number of contacts (given error bars produced from the standard deviation of the 15 individual data sets) and knowledge of which residues would play an interesting role in determining binding modes.[72] Each toxin had the following final number of key residues based on the previous criteria; ISO - 9, PCS - 11 , IAA - 14, and HIP - 13. The center of mass distances between toxin and each (1) key residue and (2) pair of helices at the mouth of Sudlow site II (H1-H2, H1-H3, H2-H3) were then calculated every 2 ps per simulation and results were analyzed using the Python programming language. The full list of key residues and calculated descriptors is available in Appendix II (**Figure I.1** and **Table I.1**).

3.2.3 *Time independent component analysis as a means of dimensionality reduction*

We used time-structure independent component analysis (tICA) to project the 9+ structural descriptors onto a 2-dimensional space in order to aid in identifying the binding modes from otherwise uninterpretable data. The tICA space is built by selecting a lag time, also known as the correlation lag time, which is a parameter of the autocorrelation function used to assess the slowest degrees of freedom in a data set. While the role of this parameter is well documented in the literature[87], we pay particular attention to its selection in the construction of a MSM as it is one of the few hyperparameters we can tune and control the trade-off between interpretability and accuracy in our model. We will further present our choice of parameter selection in section 3.3.1. The resulting tICA space is discretized into microstates (>50) using an automated pattern recognition algorithm. We use the k-means clustering method which has been shown as an effective approach to handling the dense, low dimensional data produced by tICA.[75,87,97,98] Furthermore, the k-means clustering algorithm has been used to good effect for building MSMs due to its ability to highlight more densely sampled regions while assigning few clusters with lower counts, helping to improve transition probabilities.[75,99,100]

3.2.4 *Construction, scoring, and analysis of Markov State Models*

The details of using MSMs for biomolecular simulations has already been discussed in depth.[77,78,101,102] We closely follow previous examples in our work here. Of most importance to our work is the construction of timescales, t, that are decomposed into independent processes through our MSM. In short, an MSM divides the phase space of our complex MD simulations into discrete microstates. We use the MSMbuilder python package, which starts by building a count matrix, $\boldsymbol{C}$, where the C_{ij} elements are the number of transitions from state i to j at time $t + \tau$, summed over all of t. By computing a transition matrix, $\boldsymbol{T}$, with elements that measures the probability of our system moving from microstate i to j in time τ as

$$T_{ij} = P(s_{t+\tau} = j | s_t = i)$$

where $S = (s_t)$ is a trajectory in the discretized micro-state space of length N and $s_t \in \{1, \ldots, \text{k}\}$ is the microstate-index at time t. Our combination of tICA and k-means clustering discretizes our data into the micro-state space while highlighting the slowest degrees of freedom, allowing the MSM to model long-time dynamics that would otherwise be otherwise unseen. MSMbuilder performs a maximum likelihood estimation in order to maximize the probability that a given $\boldsymbol{T}$ will generate some observed trajectory

$$(T) = P(S|\boldsymbol{T}) = \prod_{t=0}^{N-\tau} T_{s_t, s_{t+\tau}} = \prod_{i,j}^{k} T_{ij}^{C_{ij}}.$$

The timescales describing the various binding mode transitions are then determined from spectral analysis on the i eigenvectors of $\boldsymbol{T}$ which encompass the various processes of the system. Specifically, $\boldsymbol{T}$ has right eigenvectors, $\widetilde{\psi_\iota}$, and left eigenvectors, $\widetilde{\phi_\iota}$. Furthermore, if $\boldsymbol{T}$ is irreducible, a stationary distribution, π, can be found by normalizing the eigenvector corresponding to the unique eigenvalue of norm 1. Since $\boldsymbol{T}$ fulfills detailed balance (by its Markovianity), the left and right eigenvectors are related by

$$\widetilde{\phi_\iota} = diag(\pi)\widetilde{\psi_\iota}.$$

Finally, implied timescales of these processes are obtained from the matrix's eigenvalues, λ, as

$$t_i = -\frac{\tau}{\ln|\lambda_i|}.$$

With a set of discretized data in hand, MSMs are able to be built and optimized using the general matrix Rayleigh quotient (GRMQ) as a scoring metric for a 5-fold cross validation procedure which has been shown to further validate MSMs.[103] This method performs a grid search over the hyperparameters; tICA lag time (0.002-0.2 ns) and k-means clusters (50-300). After obtaining the best set of parameters for each system, we use the PCCA+ spectral clustering algorithm to coarse-grain the MSM for improved interpretability, available through the msmtools library. Analysis and construction of the MSM was done using the MSMBuilder software package.[104] The model cross validation method was conducted with the scikit-learn library.[105]

3.3 *Results and discussion*

3.3.1 *Construction of optimized MSMs by GMRQ scoring and ITS verification*

In order to build an MSM that estimates the slowest dynamics of each PBUT-ligand system, an optimization of hyperparameters must be done. We performed a grid search over tICA lag times and k-means clusters for each system with results for the IAA timescales being shown in **Figure 3.2(a)**. Both tICA and MSM will always underestimate the timescales of the true process by way of the variational principle of conformational dynamics.[76] However, the MSM uses step functions rather than tICA's linear combination of order parameters. As such, the MSMs estimate is a more reliable estimate of the dominant nonlinear eigenfunction which describe the timescales.[106] The final IAA MSM uses a lag time of 1.00 ns and 4 timescales while the GMRQ chosen parameters for the all systems are reported in **Table I.2**. **Figure 3.2(b)** shows the implied timescales plot for the chosen parameters. A proper lag time leads to a convergence in timescales after the selected time while also remaining as small as possible, which has been shown to increase the accuracy of approximations with fineness of discretization.[101] For IAA, the implied timescales after the chosen lag time of 1 ns which is a factor of 36 smaller than the slowest process, highlighting the strength of discretization in this parameter set.

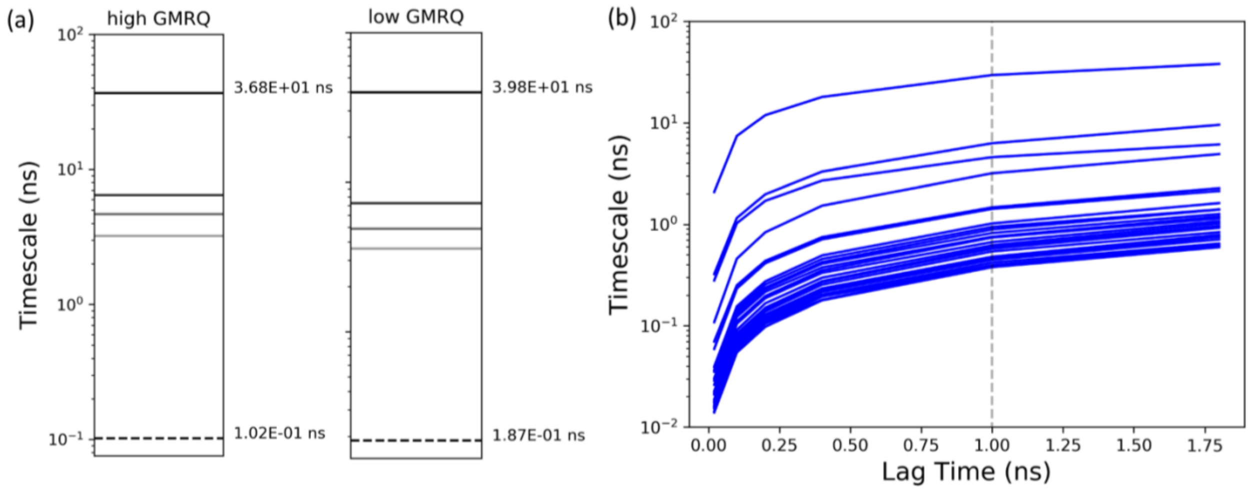


Figure 3.2 - Implied timescales calculated from hyperparameters handed to tICA and MSM (a) In both plots, solid lines represent the implied timescale of the best and a worse MSM rated by GMRQ score. The dashed line shows the slowest timescale estimated by tICA alone, which is markedly faster than the more reliable MSMs estimate of the slowest process. It should also be noted that increasing tICA lag time will inherently slow down the slowest timescale of the model and is a good sanity check as to why one should not expect the highest GMRQ scored model to have the slowest 1st process. (b) Implied timescales (ITS) plotted vs MSM lag time. The dashed line represents the final choice of lag time and highlights where the timescales begin to converge. Beyond the 4th timescale, overlap of ITS indicates the lack of unique transitions in the system and a strong choice for the final MSM.

3.3.2 *MSMs identify PBUT-HSA binding modes and key residue interactions*

Obtaining accurate kinetic information from our data required breaking the phase space into microstates that are too small to be interpretable. We perform spectral clustering with Perron cluster analysis (PCCA+) to group these microstates into macrostates that can be associated with conformational changes of the slow relaxation modes obtained from our MSM.[107] We chose the number of clusters by determining the smallest number of macrostates while retaining all information key to the molecular interpretation of the MSM. Using the 2D Euclidean distance, the MD frame with tICs closest to the center of each binding mode is determined. Pose View, a type 2D interaction diagram[108], and VMD[42] are then used to generate images which visualize the key residue-toxin interactions as depicted in **Figure 3.3/3.4** respectively. While the SI will contain additional figures for each toxin, the following sections will focus on comparing both a bulkier hydrophobic toxin (IAA) and a less bulky toxin with a larger anionic group (PCS). As such, **Figure 3.3** is detailed with the transition probability matrix of the IAA system, built by summing inter-macrostate transition probabilities of the microstates assigned to each coarse-grained cluster. The three determined binding modes makeup 0.07 (State 2), 0.12 (State 1), and 0.81 (State 0) of the total MD sampling.

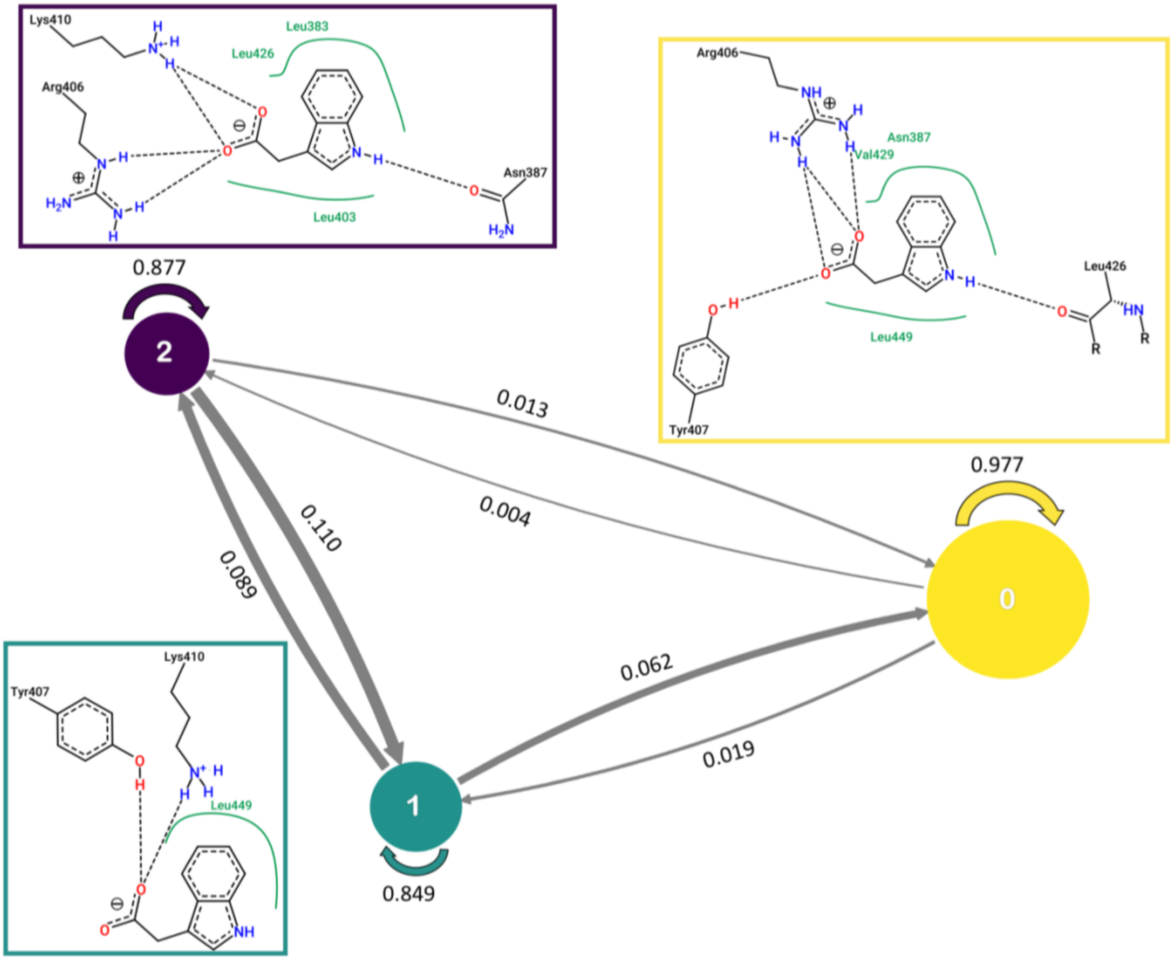


Figure 3.3 - Pose View images of frame representative of each IAA-HSA binding mode accompanied by transition probabilities. Each Pose View image is boxed by the binding modes particular color. Hydrogen bonds are represented by dashed black lines, hydrophilic interactions are in green, and polar atoms are either red (oxygen) or blue (nitrogen). The transition probabilities are shown for transitions between modes with proportionally sized arrows. The size of each colored sphere is proportional to the binding modes total fraction of MD frames. It’s worth reminding the reader that the images are not necessarily characteristic of the averaged structure, but rather a single snapshot ideally estimating certain interactions.

Analysis of the binding modes represented by the clusters shows some clear defining characteristics. The primary binding mode (the yellow mode 1) features hydrophobic interactions (residues V433, N391, and L453) which pin the toxin deep in the binding pocket. The toxin is also able to form a dual hydrogen bond with R410 that is not seen in the other modes. As the toxin transitions to the secondary binding modes, it loses some of those hydrophobic interactions and particularly gains the capability to hydrogen bond with K414. These secondary modes are thus closer to the mouth of the binding pocket and it appears that binding mode 1 is described by half as many residue toxin interactions than 2. By examining the left panel of **Figure 3.4**, we see that the toxin moves up toward the mouth of the protein as previously indicated by the pose view images. Moreover, the ability of the toxin to form a hydrogen bond between either of the carboxyl oxygens with K414 is understood by the flipping of the toxin's functional group up further towards the mouth.

In stark comparison, the right panel of **Figure 3.4** again highlights the most probable frame for each binding mode in PCS (a less bulky toxin). In the three states determined by the MSM, we see much slighter shifts in the position of PCS compared to IAA. Specifically, PCS does not observe a transition from a deeper hydrophobic dominated position towards the mouth of the pocket. Visualization of the residues deeper in the binding pocket, such as G434, show less fluctuation in the PCS system compared to the IAA system. Hydrophobic interactions deep within Sudlow site II are found to be unique to the bulky toxins, IAA and ISO, only. These bulkier toxins also end up being more interpretable (fewer binding modes required and/or better correlation with tICs) when compared to HIP and PCS (**Figures II.4-II.6**). Thus, the hydrophobic group is critical in understanding these binding site interactions even before studying the timescale dynamics.

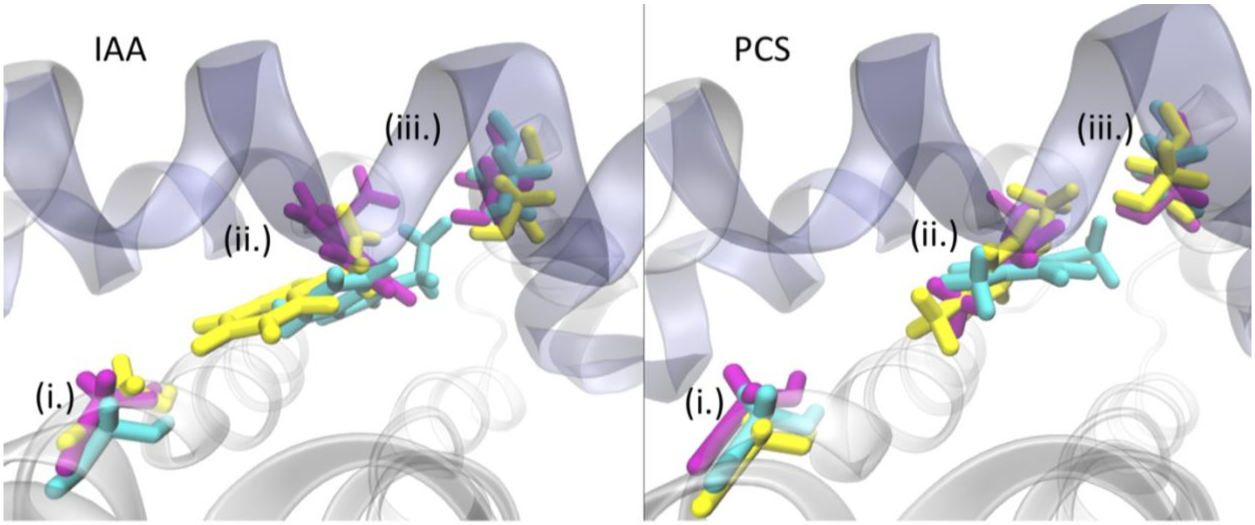


Figure 3.4 - VMD images for frame representative of each binding mode. A set of three MD frames are visualized using VMD and color coded according to the particular binding mode for two toxin systems; IAA and PCS in the left and right panels respectively. Each frame contains a set of molecules to help orient the binding pocket: (i) Residue G434 deep in the binding pocket (ii) toxin position (iii) S489 at the mouth of the binding pocket. The three helices that make up the mouth of Sudlow site II are colored in light blue to further orient the pocket.

3.3.3 *Long timescale dynamics and eigenvector analysis of MSM transition matrix*

We now analyze features of the MSMs to interpret the two types of interactions, based upon hydrophobic or anionic moieties, of dominant cluster images discovered in section 3.2.2. Specifically, the relaxation timescales of our MSM are used to highlight transition mechanisms of toxins in the binding pocket. We can interpret timescales of particular molecular change by taking the Pearson's correlation coefficient, r, between each eigenvector mode projection of an MSM and its basis function. From **Table 3.1**, we see that the slowest transition (the 1st eigenvector) has the strongest anticorrelation with the IAA-G434 center of mass distance alongside a strongest key residue correlation with IAA-S489 center of mass distance. This transition is depicted in the left panel of **Figure 3.4**, in which the IAA of state 0 (yellow) moves away from the deep G434 residue and jumps toward the S489 residue at the mouth of the pocket, as seen in the IAA of state 2 (purple). The more subtle, second slowest transition (the 2nd eigenvector column) is defined by its largest anti-correlation with residue R410. This relationship is indicative of the breaking of a hydrogen bond during this process and can be seen in the shift between dominant cluster images of the left panel of **Figure 3.4** (transitioning between states 1 and 2). These first two processes make up the unique transitions of the indole-containing compounds. The mechanism described above is further detailed in **Figure 3.5**, incorporating the correlation coefficients of **Table 3.1** and highlighting the core to mouth transition of process 1 versus the quicker hydrophilic based transition of process 2. The third process contains high correlations with R410 alongside new interactions with a couple key residues particularly the bulky N391, again indicating the nature of a hydrogen bond formation during this process as well as unique positioning within the binding pocket. The 4th set of correlations begins to highlight the possibility to obtain degenerate processes as we expand our number of projections, noted by the strong correlation to deep hydrophobic residues alongside anticorrelation to both the 1st eigenvector and hydrogen bonding partners in the mouth. This relationship appears to capture a subset of the transition seen fully in the slowest 1st process.

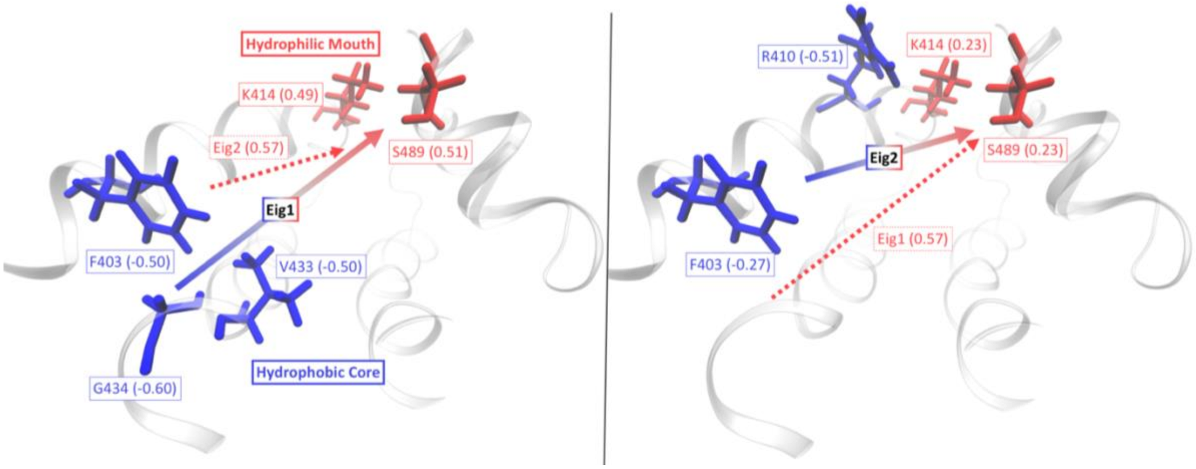


Figure 3.5 - Mechanism of IAA systems slowest processes expressed by eigenvector correlations to order parameters. The transitions captured by the first and second eigenvectors (Eig1 or Eig2) are shown in the left and right panels respectively with that process's eigenvector described by the line transitioning from blue to red. Each process is also detailed with the highest correlated order parameters/eigenvectors to itself. Residues in blue describe the starting position of the process (largest negative correlation coefficient) while those in red are the ending position (largest positive correlation coefficient). The dashed lines show the correlation between eigenvectors (following a similar coloring scheme) where positive correlations between eigenvectors can be understood by the process moving in a similar manner to some extent (and vice-versa). Both processes share the hydrophilic mouth as the ending position as well a strong correlation and anti-correlation to S489 and F403 respectively. These details explain the strong correlation between the first two eigenvectors as well as the overall toxin transition mechanism.

In the case of the less bulky anionic PCS toxin system, **Table 3.2** highlights the first three sets of correlations between order parameters and MSM eigenvectors. When compared to IAA, the 1st eigenvector no longer contains a pair of residues which indicate large shifts in the binding pocket. Instead, this transition is more similar to IAA's 2nd and 3rd eigenvectors. This 1st process observes strongest positive correlation to a hydrogen bonding partner (Y411) while being stabilized by a higher positioned hydrophobic residue in L430. This process also captures an anti-correlation with the common binding partner R410 and a potential aromatic π-stacking partner, F488, all of which keep PCS further from the hydrophobic core then IAA. The 2nd column captures the strongest correlation of the system in an R410 salt bridge formation with anticorrelation to L430 which provided stabilization in the 1st transition. As in **Table 3.1**, the final column shows signs that the 3rd process is degenerate with the first, that being 1) the opposite correlation behavior for previously dominant order parameters and 2) the relatively weak correlations overall that do not contribute to a unique process.

Table 3.1 - Top six correlations between order parameters and projection on the eigenvector modes of the IAA MSM.

1st Eigenvector		**2nd Eigenvector**		**3rd Eigenvector**		**4th Eigenvector**	
Residue (ranked)	r	Residue (ranked)	r	Residue (ranked)	r	Residue (ranked)	r
G434	(-)0.60	Eig1	0.57	R410	0.72	Eig1	(-)0.48
Eig2	0.57	R410	(-)0.51	N391	(-)0.37	G434	0.31
S489	0.51	Eig3	(-)0.31	L407	0.34	S489	(-)0.27
V433	(-)0.50	F403	(-)0.27	Eig2	(-)0.31	R485	(-)0.25
F403	(-)0.50	K414	0.23	R485	0.29	F403	0.19
K414	0.49	S489	0.23	Y411	0.27	V433	0.18

Table 3.2 - Top six correlations between order parameters and projection on the eigenvector modes of the PCS MSM.

1st Eigenvector		**2nd Eigenvector**		**3rd Eigenvector**	
Residue (ranked)	r	Residue (ranked)	r	Residue (ranked)	R
Y411	0.34	R410	0.78	Y411	(-)0.13
F488	(-)0.21	Y411	0.49	L387	(-)0.09
R410	(-)0.18	L430	(-)0.39	L430	0.08
L430	0.18	L453	(-)0.36	L453	(-)0.07
R485	0.12	H2-H3	0.35	H2-H3	(-)0.07
H2-H3	0.12	L407	0.29	R485	(-)0.05

Furthermore, **Figure 3.6** shows the projection of two left eigenvector modes onto the tIC plot in order to connect the slowest timescales to a particular progression between binding modes for IAA and PCS (the other two toxins can be found in **Figure I.7**). For the IAA system, the slowest process found by the model is approximately 6 times as long as the second slowest process found. Process 1 is confirmed to be the transition from deep within the binding pocket of state 0 to the states sitting closer to the mouth of the pocket, 1 and 2, noted in **Figure 3.4** as well as starting and ending positions shown in the upper panel images of **Figure 3.6**. The much faster second process captures the transition from state 2 to 1 where we see the carboxyl group retracting into the binding pocket. As noted from the correlations of **Table 3.1**, the key change here is the breaking of the

IAA-R410 salt bridge that forms in State 2, a much quicker event than the traversal of the binding pocket seen in Process 1. We see this quicker transition again in process 4, but near the mouth where IAA forms hydrogen bonds with R485 or S489. Thus, we see that the tIC 2 direction is highly correlated with the breaking/formation of the R410 salt bridge while the tIC 1 direction corresponds to the IAA-S489 distance (effectively how deep the toxin is within the pocket).

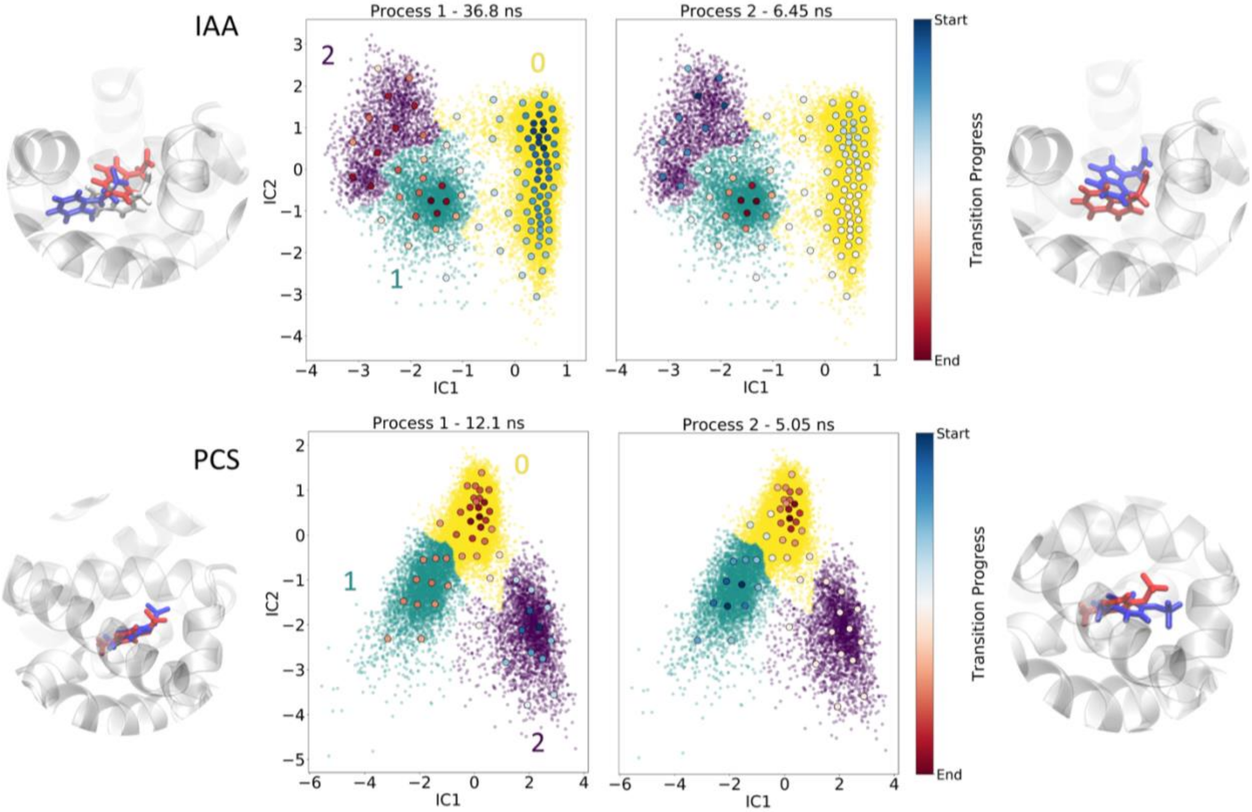


Figure 3.6 - tIC plots of the coarse grained MSM overlaid by the projection of each eigenvector. Each of the 4 tIC plots details the results of the tICA of our 15 MD simulations after being first treated with k-means clustering and then coarse grained using PCCA+ clustering. Toxins are varied between IAA and PCS in upper and lower respectively. Each pair of plots for toxin differ in the MSM eigenvector projection overlaid atop which details the progression of the particular timescale discovered by our MSM. Each plot also has a corresponding snapshot highlighting the toxin's shift between starting (blue) and ending (red) positions.

The lower panel of **Figure 3.6** reveals that the first two processes of the more anionic PCS system are less than a third the speed of the IAA system's processes. The two systems do have similarities in that 3 macrostates are used to fully describe the binding modes alongside observing two distinct transitions. However, in the case of PCS, the two processes each start in one of the two less dominant modes and transition into the primary binding mode. This is different from the IAA system in which only one of the two transitions involved a jump to/from the primary binding mode deeper in the hydrophobic core of the pocket. As noted, the slowest process of the PCS system is a fraction of the timescale of the slowest IAA process which jumps away from the core to the mouth of the binding site. This discrepancy in times further supports the idea that bulkier toxin transitions require longer timescales to occur. Alongside the correlations of **Table 3.2**, the 1st PCS process is identified as a transition between hydrogen bonding partners R485/R410 as well as aromatic stabilizing partners, potentially providing π-π stacking, Y411/F88. The 2nd PCS process is akin to the previously found R410 salt bridge formation of the IAA system (also process 2). In fact, the R410 salt bridge formation process occurs across our entire toxin design space, always on the timescale of <10 ns. The lower right snapshot of **Figure 3.6** shows how this formation causes a reorientation of the toxin toward the mouth of the binding pocket and R410. This reorientation is seen similarly in IAA's 2nd process but is instead breaking this bond and thus shifting slightly lower.

In the case of indole-containing toxins, like IAA, each tIC axis is well correlated to the transitions discovered by our MSM since the movement along each axis correlates to a particular kinetic movement (such as the R410 salt bridge formation along IAA's tIC 2 axis). On the other hand, PCS and the other less bulky toxin HIP (in **Figure I.8**), are not as clear due to the discovered processes containing a mix of each axis' components. Thus, in order to better interpret each systems kinetic space, **Figure 3.7** shows kinetic maps with coordinates provided by the toxin's two slowest left eigenvectors (i.e. processes 1 and 2).[75] Left eigenvectors were chosen over right eigenvectors as they are weighted by the stationary distribution leading to a mapping of higher stationary probability around the profile's edge. Plotting these eigenvector spaces provides a sense of transition speed as nearer points can more quickly reach each other compared to those further apart. In all, we find highly populated microstates at the borders with transition pathways connecting these states through the less populated microstates. The upper left panel of **Figure 3.7** shows the kinetic map for IAA and further confirms the two transitions (from state 0 to 1 and 2 to 1) from our eigenvector projection into tICA space. The lower left panel of **Figure 3.7** depicts PCS and of note is how evenly spaced the microstates are mapped. IAA runs into a bottleneck in the kinetic map when attempting to leave the deeper yellow binding mode by route of its slowest transition along the first left eigenvector. The second bulky toxin, ISO, also observes a tighter bottleneck, requiring a transition through a select few low populated states in order to leave either of the two poses in the hydrophobic core of the binding pocket (yellow and teal binding modes).

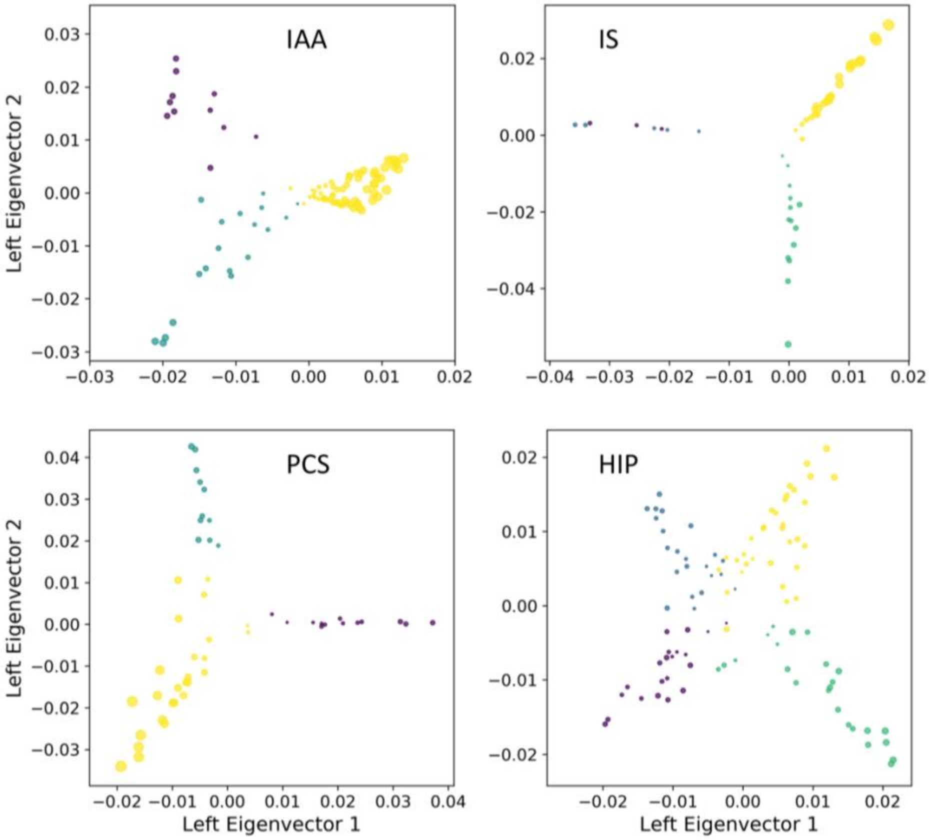


Figure 3.7 - Kinetic maps of the two slowest processes for each toxin in the design space bound to Sudlow site II of HSA. Each point in the eigenvector space represents 1 of the microstates determined by tICA. The slowest relaxation processes follow along the left eigenvector 1/horizontal axis (as this is process 1 for each MSM) and the second slowest along the 2nd/vertical axis. Points have size weighted by the microstates population as well as a color coding indicating the particular macrostate/binding mode of the point. Distances in the eigenvector space relate to kinetic separation for each system. The bulkier toxins lie in the top row (IAA and ISO) while the less bulky and more anionic toxins lie in the lower row (PCS and HIP). The less bulky toxins observe much less of a bottleneck-like effect when transitioning in/out of primary (or any) binding modes.

3.4 *Conclusion*

We have shown an effective method for estimating MSMs in order to describe the transition timescales of a uremic toxin design space bound to a predominant plasma protein, HSA. We found that the indole containing toxins of the design space undergo a bottleneck-like effect in terms of their transitions out of a primary binding mode stabilized by hydrophobic interactions deep in the binding pocket. As we transition to the more anionic toxins, this bottleneck disappears and so too does the increased timescale in which the slowest transition occurs. The knowledge of these processes' timescales will provide further intuition about designing synthetic materials for targeted toxin capture. In terms of future work the binding poses and transition timescales determined by our tICA and MSMs respectively could serve as critical tools in the design of enhanced sampling based unbinding studies. For example, in recent years, many methods of discovering effective order parameters (discussed in the great overview done by Wang et al.[109]) have been developed and combined with enhanced sampling, such as infrequent metadynamics, to allow for the calculation of thermodynamic and kinetic quantities, such as k_{off}, k_{on}, unbinding times, K_d, and ΔG.[94,110–112] However, the complexity of the Sudlow II binding site and unbinding process has frustrated the application of these methods to the study of HSA and PBUTs. The knowledge of each pose and transition timescale will be a valuable addition to mechanistic studies that develop enhanced sampling schemes to study unbinding kinetics and thermodynamics in HSA like those displayed in the following chapter.

Chapter 4. **Machine learning and metadynamics to study uremic toxin unbinding**

4.1 *Introduction*

As long as Chronic Kidney Disease (CKD) remains a regular health concern, efficient removal of the troublesome protein bound uremic toxins (PBUTs) underlying this condition will be of upmost importance. PBUTs are small uremic retention solutes (<500 Da) that have a high affinity for carrier proteins, a particular issue for the predominant blood plasma protein in humans, Human Serum Albumin (HSA). PBUTs, such as indoxyl sulfate (ISO) and p-cresyl sulfate (PCS), have been directly linked to negative CKD patient outcomes, requiring the need for efficient removal strategies to be developed.[67,73,113] In general, CKD patients rely on forms of hemodialysis in order to adequately remove small water-soluble molecules which are not adequate for the removal of PBUTs due to their binding strength and thus low free-fraction in the bloodstream. Thus, a bevy of research has been undertaken in recent years with the goal of developing new strategies for PBUT removal, for example the use of competitive binding drugs as inhibitors or highly specific nanoporous adsorbents for adsorption of hydrophobic molecules.[69,70] However, the mechanistic details regarding the binding of uremic toxins to transport proteins are complex and not well understood. Moreover, these materials designed to aid in the capture of PBUTs are in competition with the high affinity between uremic toxins and HSA. To this end, there is a continued need for increasing the atomic-scale insight surrounding the interactions of PBUTs and the kinetics and thermodynamics of their binding with HSA.

One method of efficiently obtaining atomic-level information regarding PBUTs and other protein-ligand systems is by way of molecular dynamics (MD). This approach gives the opportunity to not only obtain key interaction information of bound ligands[71,114], but also the kinetic and thermodynamic information that can be difficult to obtain and gain insight from without MD. To do so, enhanced sampling methods have shown wide success in supporting these endeavors by providing MD with the ability to sample rare events (time scales of milliseconds or more) which are otherwise unreasonable when limited to timesteps of a few femtoseconds.[82,115–117] Furthermore, it has been shown that kinetic quantities, such as the rate of dissociation for ligands (k_{off}), are crucial in the development and/or improvement of drugs.[118,119] One way to better the current understanding and strategies for PBUT removal would be to take advantage of these types of studies. However, for systems with particularly complex binding pathways (PBUTs, G protein-coupled receptors, etc.) both experimental and computational results are difficult to obtain accurately.[120] Thankfully, the metadynamics based enhanced sampling schemes have achieved accurate calculations of k_{off} by using the rare event nature of these biochemical challenges.[111,116,121–124]

Enhanced sampling methods require the selection of a collective variable (CV), some slow degree of freedom, that can adequately describe the biological rare event of interest. The search for a proper CV is no simple matter and can require a combination of factors such as chemical intuition, MD experience, and an understanding of the underlying free energy space in order to find success. Even in those circumstances, complex biochemical pathways can still lead to a difficult time in manually choosing a proper RC. Recently, the use of a machine learned RC, optimized on a set of computationally determined order parameters, obtained kinetic results validated experimentally by using infrequent metadynamics (InMetaD), a scheme of metadynamics that applies bias slowly so as not to perturb the transition state allowing for the recovery of kinetic quantities.[119,125] In particular, combining Automatic Mutual Information Noise Omission (AMINO)[126] and Reweighted Autoencoded Variational Bayes for Enhanced sampling

(RAVE)[127] allows for the selection of potentially useful order parameters from a larger set and the learning of an optimized CV respectively. By combining these methods, one can speed up the process of determining the least redundant set of inputs for a CV while accessing kinetic results on a higher level of efficiency and accuracy than previous InMetaD based results.

In this work, we use an optimized machine learned CV alongside an alternate version of InMetaD, known as frequency adaptive metadynamics (FaMetaD), in order to obtain the first set of computational based unbinding kinetic results for two uremic toxins, ISO and PCS, from Sudlow site II of HSA. This work provides results with a high level of accuracy while continuing to shed light on methods which lower the computational cost of large biochemical ligand unbinding simulations. Furthermore, we also provide both thermodynamic results and analysis of the mechanics behind the unbinding paths discovered through our work. From here, such an approach could be used on an assortment of other challenging toxins and/or competitive binders/techniques to gain insight crucial to improving the knowledge of these conditions and, potentially, even the outcomes of CKD patients overall.

The remainder of this manuscript is organized as follows. We begin in section 4.2 by providing methodological details of our simulation setup, machine learning (ML) based methods for creating an optimum CV, styles of metadynamics used (frequency adaptive and funnel), approach for dimensionality reduction, as well as calculation of kinetic and thermodynamics quantities. In section 4.3, we then delve into analysis of our learned CV before providing the kinetic results discovered from their usage. From here, the funnel Metadynamics based thermodynamic unbinding results are presented and compared to the available experimental results found in literature. We end section 4.3 by presenting a ML method for the analysis of unbinding pathways for ISO and PCS, detailing the mechanics of each. The chapter is then completed with section 4.4, a discussion on the reliability and overall applicability of these ML based metadynamics methods for the study of complex biochemical challenges such as those presented in the study of PBUTs within Sudlow site II of HSA.

4.2 *Methods*

4.2.1 *Molecular dynamic simulation setup and design*

The starting structures for ISO and PCS bound to Sudlow site II were adapted from our previous works focused on binding pocket mechanics.[71] Meanwhile, the HSA protein originates from an experimentally resolved structure (RCSB PDB ID: 2BXH) using the Amber 14 forcefield parameters.[72,90] The systems were solvated in a cubic box (L=10.8 nm) with TIP3P water and 14 sodium counter ions.[91] Following this system setup, MD simulation was conducted using Gromacs 2020.5[34] and hydrogen mass repartitioning to use a time step of 4 fs. System equilibration was then done in a three-step process. First, the system was energy minimized with a steepest descent algorithm for 10000 steps. Next, the temperature was ramped over 250 ps to 298K during an annealing step. Finally, a short NPT simulation of 1 ns was used for fast mechanical equilibration with the Berendsen barostat at 1 bar.[93] All simulations use an MD timestep of 4 fs, using the LINCS algorithm to constraint all *X*-H bonds.[92]

As a means to obtain an adequate number of unbinding events for the calculation of k_{off}, we performed 15 NVT simulations. The same initial toxin pose was used for each of the 15 simulations in a given k_{off} calculation. Based upon our previous experience and studies of uremic toxins in Sudlow site II[71,114], we investigated two particular poses for ISO; (1) ISO_{S489}: near the mouth of the pocket dominated by hydrogen bond interactions with residues such as S489 and R410 and (2) ISO_{Y411}: a rarely observed position (likely along an unbinding path itself) that is at a similar pocket height compared to the first pose, but with its indole group flipped in the opposite direction. A PCS system was also studied that had an initial pose similar to the ISO_{S489}pose described. Systems were held at 300 K using the Bussi-Donadio-Parrinello thermostat.[39] Each FaMetaD simulation was carried out until the toxin left the binding pocket, indicated by the center of mass (COM) distance between the toxin and binding pocket residues greater than 1.5 nm. Furthermore, frames were saved every 2 ps for analysis. The open-source, community-developed PLUMED library version 2.8 (with the funnel Metadynamics module enabled) was used to implement all metadynamics procedures described in the following sections.[96,128]

4.2.2 *Machine learned collective variable for rare events in MD*

In order to take advantage of FaMetaD for the calculation of kinetics, we must first produce an ideal CV. This task is not at all trivial, but we follow the recent approach used by Ribeiro et al. in combining two powerful machine learning methods, AMINO and RAVE, improving the ability to obtain a useful coordinate and optimizing it.[119] While the details of these methods have been well documented in the literature, we aim to highlight our general process as it will hopefully shed light on how to properly prepare a CV for the biasing of complex unbinding pathways.

First, we use AMINO in order to take a large set of potentially useful input parameters, such as the COM distance between a ligand and each C-α's within a set distance of the ligand, and screen for redundancy by way of a mutual information based distance function which measures the dissimilarity between parameters.[126] After extensive testing and conversations with the Tiwary group (creators of AMINO), we settled on the following HSA features: (1) the C-α's of residues 340-489 within 2.0 nm of ISO and (2) all charged or hydrophobic side chains within 1.5 nm of ISO (choosing a heavy atom as the feature). In order to better account for the duality of the functional groups within ISO, we chose two features instead of a simple COM, the indole group's nitrogen and the sulfate group's sulfur atoms. Plumed was used in order to track the distances between each HSA to ISO feature along a 40 ns trajectory with a 1 ps resolution. This results in a total of $40{,}000 \times 300$ features as inputs to AMINO for each pose. A python 3 script using the serial, kernel density estimation version of AMINO, available on the Tiwary group's github, was used with the default parameters for maximum number of clusters (output features) and bins, being 20. The resulting parameters chosen by AMINO are provided in **Table II.1**.

An important note, we found the use of residues 340-489 to be a deciding factor in our systems performance. This region was used as it contains the majority of Subdomain IIIA in which Sudlow site II is located. However, the inclusion of residues beyond 489 leads to a prominent issue later due to biasing this region of the protein. The structure of the residues between 490 and 500 constitute a loose coil unlike what is seen in the dense structured binding pocket. AMINO produces a set of parameters with the least amount of redundancy based upon parameter dissimilarity. Thus, when submitting parameters which are representative of a single domain (like a binding pocket), one will generally find at least a single output feature in any distinctly different region included as input, in our case a coil feature between residues 490-500. Such a distinct feature on its own may be perfectly acceptable and potentially useful for certain studies using AMINO, for example identifying unique structural domains in a system. However, when building a ligand unbinding CV, the inclusion of such unstructured features can lead to unnatural movement in the affected regions around the binding pocket. For example, once the ISO unbinds from Sudlow site II, the toxin-coil interaction will extend further into solution, both influencing the unbinding path as well as destabilizing the protein. Due to these issues, we found AMINO to be most successful when keeping the input order parameters limited to features either found within the pocket or without distinct structural differences to the pocket.

After choosing a set of order parameters with the help of AMINO, the selected set is used as input for the next machine learning method, RAVE. We will now review the basics of the method as well as any pertinent details regarding the implementation for our system specifically as RAVE has been explained thoroughly in a variety of literature.[127,129,130] Generally speaking, RAVE uses a variational autoencoder (VAE) in order to obtain a detailed mapping of the trajectory into its probability distribution in latent space.[127] This dimensionally reduced representation of the input is known as a predictive information bottleneck (PIB), which can be interpreted as an optimal CV that enables maximal prediction of a trajectory's future based upon its past.[119,130,131] Another key aspect to RAVE lies in the ability to then use the constructed PIB as a bias in order to better explore a region and, in turn, produced an optimal estimate of the probability distribution with future iterations. We will highlight our use of this approach below after discussing the general framework of the neural network (NN) architecture employed by RAVE.

When taking advantage of a VAE, we must pay attention to the suitability of the neural network architecture for a given challenge. Our toxin unbinding systems observe similar complex unbinding mechanics to recent works using RAVE for the calculation of kinetics and thus we follow a similar neural network acrchitecture.[119] A basic encoder first maps the n order parameters chosen by AMINO into a linear bottleneck using

$$Z = \boldsymbol{c} \cdot \boldsymbol{x}$$

where c is the n-dimensional coefficients vector of the NN, x is the n-dimensional vector of input order parameters, and Z is the resulting latent space bottleneck value. Our decoder then acts in a two-step process, first mapping the latent variable sampled from a Gaussian distribution into a set of non-linear $N = 128$ dimensional layers before transforming the N-dimensional layer back into the n-dimensional space of the input trajectory without any non-linearity. Transforming from the latent variable to the N-dimensional space was performed using the following:

$$\boldsymbol{d} = \psi(C_1\psi(C_2Z + b_1) + b_2)$$

where C_1 and C_2 are coefficient matrices with shapes $N \times 1$ and $N \times N$ respectively, b_i are vectors of coefficients, and ψ is the exponential linear unit function which applies non-linearity to each layer.[132] Transforming back into the n-dimensional space of the input trajectory without non-linearity was completed using:

$$\boldsymbol{x}_{\Delta t} = C_3\boldsymbol{d} + b_3$$

where C_3 is again a matrix of coefficients now with shape $n \times N$, b_3 is now capturing the n-dimensional bias, and $\boldsymbol{x}_{\Delta t}$ describes our resulting NN trained using a small forward propagation in time.

This NN architecture was handled using the Keras python library for deep learning.[133] All training was handled by the RMSprop optimizer for 500 epochs with a learning rate of 0.0002. The bias introduced by the NN architecture can be sufficiently accounted for when using small lag times of $\Delta t = 2\ ps$.[129] The loss for each system converged to an acceptable value below 0.1 and convergence profiles can be found in **Figure II.1**. Finally, RAVE was used in a two-step process beginning with the initial 40-ns unbiased trajectory as input for the creation of a trial CV. This CV was then used for a 200 ns biased simulation to allow for improved exploration of the binding pocket. This trajectory is then provided as the second RAVE input, producing the optimized RAVE CV used for all productions runs.

4.2.3 *Funnel and Frequency adaptive Metadynamics*

Two styles of metadynamics were used throughout the production runs of this work. Frequency adaptive metadynamics (FaMetaD) is an alternate take on the infrequent metadynamics

(InMetaD) methodology which has been used to calculate unbiased kinetic properties thanks to its ability to leave the transition state of rare events unperturbed.[134,135] This delicate task is accomplished by adding a bias, $V(s,t)$, where s is a collection of one or more CVs that distinguish transition states and t is the simulation time, in such a way that only the free energy basins, and not the barriers, feel the bias potential. Furthermore, the bias applied can be accounted for by way of the following acceleration factor,

$$\alpha(\tau) = \langle e^{\beta V(s,t)} \rangle,$$

where β is $(k_b T)^{-1}$ and the angle brackets denote an average of metadynamics simulation time until τ. FaMetaD cleverly builds upon InMetaD by using an adaptive bias deposition scheme as follows

$$\tau_{dep}(t) = min[\tau_0 \cdot max\left[\frac{\alpha(t)}{\theta}, 1\right], \tau_c],$$

where τ_0 is the initial pace of bias deposition, τ_c is the maximum pace of bias deposition allowed, $\alpha(t)$ is the instantaneous acceleration factor at time t, and θ acts a threshold for which $\alpha(t)$ must meet before changing to a slower pace. This scheme provides for more frequent bias deposition while in free energy basins before slowing down over time as the trajectory approaches a barrier/transition state. Not only has this bias deposition scheme been shown to improve the computational cost of often arduous kinetic calculations, but also the accuracy of said results when compared to InMetaD.[135]

The second MetaD variant used was Funnel MetaD (FM), which applies a funnel-shaped potential used to improve the sampling of unbinding events. Such a potential allows for proper exploration of the binding pocket while also limiting the region of space explored outside of the pocket.[136–138] For our systems, the bulk of the conic section was positioned within the binding pocket to ensure proper exploration as well as rebinding while the transition from cone to cylinder began just above the mouth of the pocket, as shown in **Figure II.2**. Each funnel was chosen to have a cylindrical radius of 0.55 nm, a cone height (distance from end of cone until the cylinder) of 0.4 nm, and a cone amplitude of 1.10 rad. All other parameters are the default funnel parameters, aside from the maximum values of the funnel which are determined by careful consideration of the distances each systems ligand should be exploring. All specifics can be found in the plumed files in this projects github repository.

For each of our FaMetaD simulations (used for the calculation of kinetic quantities described in section 4.2.4), we chose an initial bias deposition pace (τ_0) of 1 ps, max bias deposition pace (τ_c) of 100 ps, an acceleration threshold (θ) of 10^4, and a biasfactor of 20. On the other hand, when calculating thermodynamic properties using a funnel shaped bias procedure, we set the bias deposition pace to 1 ps and the biasfactor to 15. All metadynamics simulations used an initial hill height of 0.5 kJ/mol and hill widths are calculated as half of the standard deviation from an unbiased monitoring of each CV. The CVs used for both approaches using metadynamics (kinetic or thermodynamic calculations) use the RAVE optimized CV and either a COM distance between toxin and binding pocket or the position along a funnel bias (these are specified and discussed in the results section).

4.2.4 *Kinetic and thermodynamic property calculations*

Recent work using the InMetaD/FaMetaD methods have shown success in calculating kinetics of unbinding events.[116,119,121–124,139–141] This can be achieved by running sets of independent infrequently biased trajectories starting with the ligand in a bound pose and stopping once an unbinding event occurs. From these runs, we can extract the unbiased unbinding times using

$$t_{unb} = \sum_{t_i < t_D} \Delta t e^{\beta V(s,t_i)},$$

where $V(s, t_i)$ is the instantaneous bias at time t_i, Δt is the timestep (0.004 ps), and the sum is over all steps until the unbinding event occurs, as described in section 4.2.3. Due to the rare-event nature of ligand unbinding, a set of such times is expected to be exponentially distributed around the actual residence time (τ) with the following probability density function (PDF)

$$f(t_{unb}) = \tau^{-1} e^{-t_{unb}\tau}.$$

We can estimate τ by fitting the cumulative distribution function (CDF) to a constructed empirical cumulative distribution function (ECDF) of our unbinding times.[139] Assessing the reliability of this fit, and thus our estimation of τ, is done by performing a Kolmogorov-Smirnov (KS) test.[142,143] Specifically, we test the null hypothesis that both the set of metadynamics based unbinding times and a larger set of times drawn randomly from $f(t_{unb})$ are derived from the same distribution. The null-hypothesis would be rejected if the associated $p - value < 0.05$. For each calculation, 15 FaMetaD simulations were completed following the setup and procedures described in the previous sections. Our implementation of this procedure follows directly from the seminal work of Salvalaglio et al.[139] and can be found in the associated files and python notebook of this work's github repository for reproducibility.

For the calculation of unbinding free energies, we used a Boltzmann averaged difference of the free energy between the bound and unbound states by defining two regions to represent each state. The regions are able to be defined based upon the distance, r, between the COM of the toxin's heavy atoms and the residues of the binding pocket, where the bound and unbound regions are represented by r_0 to r_1 and r_1 to r_2 respectively.[144,145] The equations describing both the free energies of each state, G_{bound} or $G_{unbound}$, and the overall binding free energy, ΔG are as follows,

$$G_{bound} = \frac{1}{r_1 - r_0} \int_{r_0}^{r_1} e^{-\beta G(r)}\, dr \quad , \quad G_{unbound} = \frac{1}{r_1 - r_0} \int_{r_0}^{r_1} e^{-\beta G(r)}\, dr \quad ,$$

$$\Delta G = -\frac{1}{\beta}\ln\left(\frac{A_{bound}}{A_{unbound}}\right)$$

where $A(r)$ is the free energy of the system at a particular r. In order to calculate the absolute binding free energy, ΔG_b^0, a correction must be applied to the binding free energy from a funnel metadynamics run with sufficient sampling of binding events, ΔG, as follows,

$$\Delta G_b^0 = \Delta G - \frac{1}{\beta} ln\left(\pi R_{cyl}^2 C^0\right)$$

where R_{cyl} is the radius of the cylindrical portion of the funnel and C^0 is 1/1,660 Å^{-3} the standard concentration.[136] All comparisons with experiment will use the absolute binding free energy, ΔG_b^0, while error bars are calculated as averages over the last converged region of the trajectory (final 30%).

4.2.5 *Dimensionality reduction and clustering of unbinding trajectories*

In order to handle the unpredictably large number of frames produced from groups of FaMetaD simulations in combination with the high dimensionality of feature sets produced by AMINO, a method for dimensionally reducing the data becomes an important tool. We used a principal component analysis (PCA) to project each system's set of AMINO chosen features, a COM distance between binding pocket and toxin, as well as the RAVE CV at every 0.2 ns of simulation time onto a 2-dimensional space. PCA has been used to obtain useful kinetic results in a variety of protein-ligand works and more, particularly useful with the high dimensional data used for the creation of our optimized unbinding CV.[71,87,119] Only the first 2 principal components are displayed and used for discussing the unbinding mechanics of our system as we found that increasing this number did not change the proceeding cleaning or analysis of the data. As in Chapter 3, we used k-means clustering for handling the dense, low dimensional data produced by PCA in order to identify unique interactions or steps along the unbinding route.

4.3 *Results and Discussion*

4.3.1 *Feature selection for optimized collective variable*

We first determine an optimal set of parameters from a 40 ns unbiased trajectory by way of AMINO for use in biased MD to determine unbinding kinetics. In each of the three systems studied, $\mathrm{ISO}_{\mathrm{S489}}$, $\mathrm{ISO}_{\mathrm{Y411}}$, and PCS, AMINO selected 13, 10, and 14 total OPs respectively to describe the toxin-binding pocket interactions. From here, a RAVE NN was trained on the unbiased trajectories for each set of OPs to create 3 CVs for exploring the binding pocket effectively. After using the CVs to explore each system in biased MD, a RAVE NN is again used, this time to improve the obtained CVs in preparation for unbinding simulations. The following three linear combination CVs were produced for the $\mathrm{ISO}_{\mathrm{S489}}$, $\mathrm{ISO}_{\mathrm{Y411}}$, and PCS systems respectively,

$$\boldsymbol{CV_{ISO_{S489}}} = 0.21\, C_{L347-s} + 0.12\, C_{L347-r} + 0.68\, C_{E342-r} + 0.13\, C_{K402-r} + 0.42\, C_{F403-s} + 0.03\, C_{C437-r} - 0.03\, C_{V462-s} + 0.06\, C_{V462-r} + 0.15\, C_{C487-r} + 0.21\, S_{P384-s} - 0.21\, S_{P384-r} + 0.42\, S_{P416-s} + 0.03\, S_{E442-s},$$

$$\boldsymbol{CV_{ISO_{Y411}}} = -0.13\, C_{V343-s} + 0.02\, C_{L346-r} + 0.44\, C_{L347-r} + 0.12\, C_{K402-r} + 0.18\, C_{V426-r} + 0.55\, C_{R445-r} + 0.60\, C_{V462-r} + 0.07\, C_{V473-r} + 0.13\, S_{P384-r} - 0.24\, S_{F488-r},$$

$$\boldsymbol{CV_{PCS}} = -0.11\, C_{E396-r} + 0.15\, C_{F403-s} - 0.22\, C_{F403-r} + 0.11\, C_{T412-r} - 0.29\, C_{E425-s} + 0.79\, C_{E425-r} - 0.15\, C_{S434-s} + 0.16\, C_{L453-s} + 0.08\, C_{V462-s} + 0.19\, C_{V473-r} - 0.19\, S_{Y407-s} + 0.13\, S_{K413-r} + 0.14\, S_{P416-s} - 0.17\, S_{P441-r},$$

where C_{x-y} is the distance between x, a C-α atom, and y, an atom in either the ring (r) or sulfate (s) region of the toxin, while S_{x-y} follows the same scheme, but x is a charged or hydrophobic sidechain atom instead. Regarding the AMINO OPs, the three systems have a similar total number of OPs, which likely arises from the consistent input parameters, trajectory sizes, and general structural similarity between systems. Furthermore, AMINO selected more C-α residue atoms over charged/hydrophobic side chains in all systems. Given that there were fewer side chain targets, however, these OPs and the differences between systems may be even more crucial to the later results. For example, AMINO selected 3 sidechain residues for the ISO$_{S489}$ system for targeting the sulfate group of ISO, opposite the 2 total sidechain residues of the ISO$_{Y411.}$system which target the indole group of ISO, which may contribute to differences in unbinding pathways from each pose.

Also, we can explore the RAVE NN determined coefficients for our final unbinding CV. The weights of each AMINO selected OP help in highlighting which features may be contributing most to the resulting unbinding pathways. Overall, the PCS system obtains the largest weighted OP of 0.79 withC_{E421-r} with the next largest appearing in ISO$_{S489}$ at 0.68 with a similar OP C_{E389-r} detailing the importance placed on GLU residues for these unbinding CVs/pathways. In fact, the PCS system's next most important OP is the same residue targeting the sulfate group, C_{E389-s}, and only weighted to -0.29, further showing the importance of GLU and the lack of weight being given to the other 12 features. None of the OPs observe entirely unimportant weights, with only 4, 2, and 1 OPs weighted below a value of 0.1 for the ISO$_{S489}$, ISO$_{Y411}$, and PCS systems respectively. The information obtainable by the selection of OPs and weights is novel on its own and could offer support of future toxin unbinding simulations beyond the following results.

4.3.2 *Calculating unbinding kinetics for ISO and PCS*

For each of the three systems explored, 15 FaMetaD simulations were run, ending once the specific toxin left the binding pocket (signified by a COM distance between toxin and pocket of 1.5 nm). We aimed to efficiently obtain accurate kinetic results for two difficult systems of study (PCS and ISO bound in HSA Sudlow site ii), reporting dissociation constants and unbinding/residence times. Following the methods of Tiwary et al.[134], we implement a KS test to verify whether or not our resulting set of unbinding times' CDF comes from a Poisson distribution, as expected for such rare events in MD. In each figure shown in this section, 3 variable features and 1 constant will be highlighted. In terms of system dependent quantities, we (1) plot the ECDF of 15 unbinding times, (2) plot the CDF of a best fit Poisson distribution, and (3) display the p-value of a KS test that measure how well our unbinding times fit that Poisson distribution. The single constant feature is the grey vertical line (and approximated error) of what is currently, to our knowledge, the only known comparable experimental kinetic result. Specifically, the plotted experimental value ($6.7 \times 10^{-4} s$) reports the unbinding time of IS bound to bovine serum albumin (BSA) measured using ^{1}H relaxation dispersion spectroscopy.[146] Again, experimental results for our exact system are not currently reported, but this comparison is thought to be adequate due to the similarity of the proteins and, more specifically, Sudlow site II. **Table 4.1** reports the results for all 4 sets of FaMetaD trajectories studied: the two ISO poses and PCS using a RAVE and COM distance CV alongside an ISO$_{S489}$ system using a RAVE and funnel CV.

Table 4.1 - Kinetics results for FaMetaD trajectories and statics of Poisson fit.

System	$\boldsymbol{\tau\ (s)}^{a}$	$\boldsymbol{k_{off}\ (s^{-1})}^{b}$	$\boldsymbol{p_{KS}}$	$\boldsymbol{\mu/\sigma}^{c}$
ISO$_{S489}$	$(3.2 \pm 0.42) \times 10^{-2}$	31.4	0.17	0.33
ISO$_{Y41}$	$(5.23 \pm 0.14) \times 10^{-3}$	191	0.36	0.54
PCS	$(7.51 \pm 1.24) \times 10^{-1}$	1.32	0.20	0.56
ISO$_{S489}$ (funnel CVs)	$(3.71 \pm 0.14) \times 10^{-2}$	30.0	0.54	0.32

[a] Error bars for kinetics are calculated as the 90% confidence intervals of 1000 bootstrap samples pulling 15 data points per sample with replacement.
[b] k_{off} is calculated as the inverse of unbinding time (τ).
[c] A goodness of fit to the Poisson distribution where $\mu/\sigma = 1$.

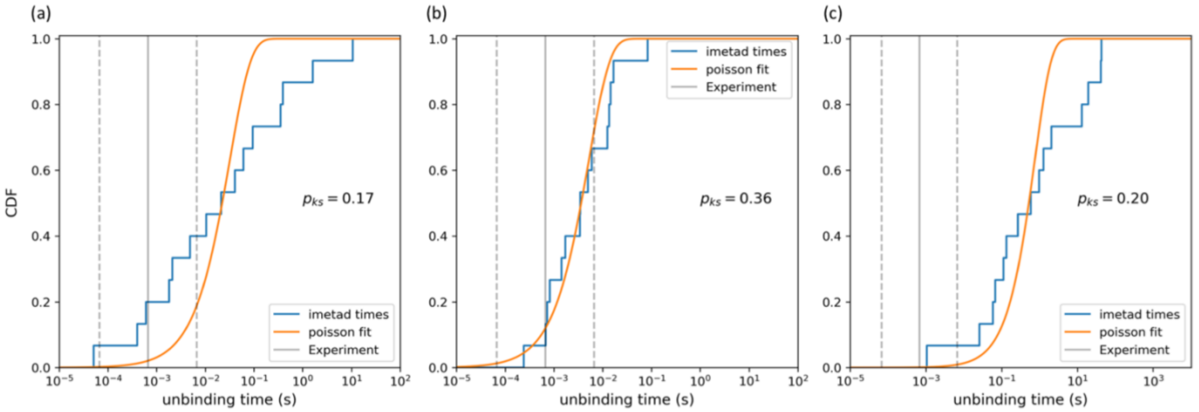


Figure 4.1 - Plotting ECDF of unbinding times with Poisson fit and experimental result for three FaMetaD systems. Panels (a), (b), and (c) show the results for the ISO$_{S489}$, ISO$_{Y411}$, and PCS systems respectively. Approximate error of the experimental value is plotted with dashed vertical lines. The p-value from a KS test of the ECDF and best fit Poisson line is provided to the right of each set of curves.

We first show results for the three systems where bias was applied to both the RAVE produced CV alongside a COM distance between toxin and binding pocket CV in **Figure 4.1**. The resulting range of times for both ISO poses includes the experimental BSA unbinding time, a decent comparison to experiment. However, the resulting unbinding times found by fitting to the nearest Poisson distribution for ISO$_{S489}$ and ISO$_{Y411}$ are two and one order of magnitude larger respectively. We point out now that these differences are still rather small (especially when considering unbinding times encompass the experimental result) and, more importantly, are in line with what we might expect for a dependency upon binding pose. The ISO$_{Y411}$ pose observes a faster τ than the more common ISO$_{S489}$ pose which supports the idea that ISO$_{Y411}$, as a rarely visited pose and may represent a step further along the unbinding path. PCS observes the slowest τ, another magnitude of order slower than ISO$_{S489}$. This result could imply that the anionic toxins are more stable in such mouth poses than the indole containing compounds, which have been shown to find their most stable configuration deeper near the hydrophobic core of the pocket.[114] At present, it seems likely that the experimental ISO result may be indicative of a step further along the unbinding path than those used as starting positions in this work. Further work could seek to study such a pose which we highlight later in this chapter.

Figure 4.2 shows the ECDF and fit result of our fourth and final set of FaMetaD results, ISO$_{S489}$ with an alternate second CV, ISO position along a funnel bias. The resulting τ of $(3.71 \pm 0.10) \times 10^{-2}$ s is remarkably close to the previous measure using the COM distance CV for ISO$_{S489}$, highlighting the strength of the results and method. Furthermore, the p-value for the Funnel CV method is 3.17 times greater than that of the COM distance CV of the same system. In fact, this is the highest p-value achieved across all systems, potentially indicting the reliability of a funnel-based approach when choosing FaMetaD CVs for protein-ligand unbinding, especially when dealing with complex binding pocket interactions. Further computational work could investigate a second binding pose of PCS, a third deeper pose of ISO and experimental work could seek to confirm of this approach's reliability.

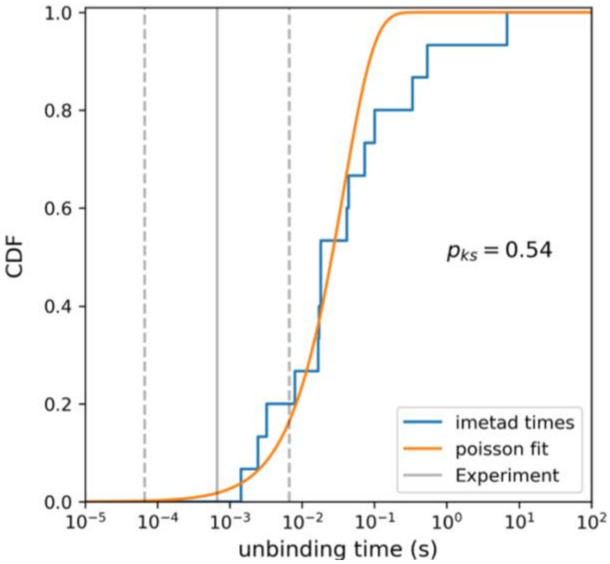


Figure 4.2 Plotting ECDF of unbinding times with Poisson fit and experimental result for ISO$_{S489}$ system biasing RAVE and Funnel MetaD parameters. Approximate error of the experimental value is plotted with dashed vertical lines. The p-value from a KS test of whether the ECDF and best fit Poisson line is provided to the right of each set of curves.

4.3.3 *Funnel Metadynamics for free energy profiles*

Binding energies were calculated for PCS and ISO using funnel metadynamics for comparison with experimental results. **Figure 4.3** shows both the convergence of free energy profiles and raw results in the form of CV data/deposited Gaussian hill height. Correspondingly, **Table 4.2** reports the free energy between bound and unbound regions after the funnel correction is applied as well as dissociation constants with comparison to experimental results. The simulated free energies and errors are calculated by taking the mean and standard deviation over the last 30% of the trajectory due to the region remaining within thermal fluctuations. The final corrected unbinding free energies are each within one order of magnitude of the experimentally determined values in healthy serum, a decent measure of agreement for the funnel metadynamics calculations. We find that PCS has a higher unbinding energy than ISO, following both the experimental trend as well as the slower unbinding time comparison in section 4.3.2. The stronger unbinding energy of PCS further supports the hypothesis that the anionic toxins have a higher affinity for the mouth region of Sudlow site II than the bulkier indole containing toxins like ISO. Future work could investigate additional toxins, HSA conditions, and/or CVs. For instance, the RAVE optimized CV used for biasing with FaMetaD could be implemented for free energy calculations with careful attention paid to the RAVE value in relation to unbinding steps. Experimentally, various HSA conditions are studied, and it would be useful to model alternate conditions or the effects of multiple toxins on unbinding.[147]

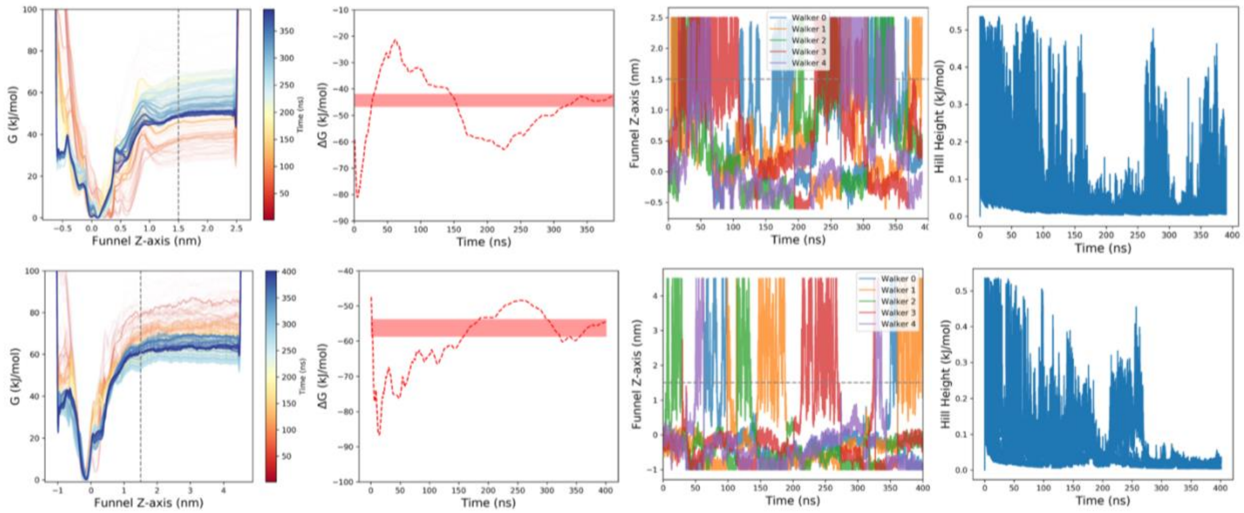


Figure 4.3 - Unbinding free energy determination for ISO and PCS. The first and second rows correspond to ISO and PCS systems respectively while each column display a particular criterion or check for system convergence. Column 1 reveals the free energy profiles as it evolves over time while column 2 plots the free energy difference between bound and unbound regions (denoted by the dashed line of figures in column 1). Columns 3 and 4 plot the distance of the particular toxin along the funnel's z-axis with a dashed line to indicate the boundary between regions and the Gaussian hill height versus time respectively.

Table 4.2 - Unbinding free energy results and comparison to experiment.

System	$^{a}\boldsymbol{\Delta G_b^0}$ **(kJ/mol)**	$K_d\ (\mu M)$	$^{b}K_{d,exp}\ (\mu M)$
ISO	-41.34 ± 5.03	$(6.16 \pm 0.75) \times 10^{-2}$	1.11×10^{-3}
PCS	-53.29 ± 2.69	$(5.07 \pm 0.26) \times 10^{-4}$	3.07×10^{-4}

[a]Unbinding free energies after funnel correction (described in 4.2) was applied.
[b]Dissociation constants from Deltombe et. al. using healthy serum.[147]

4.3.4 *Unbinding path analysis using PCA and K-means clustering*

The unbinding paths of uremic toxins could further improve understandings of the general mechanisms key to competitive binding or other toxin removal approaches. To this end, we implemented principal component analysis (PCA) for the dimensionality reduction of our sets of 15 trajectories, ideal for such dense sets of data containing over 1×10^6 frames and at least 13 order parameters/dimensions. Specifically, in the ISO$_{S489}$ and PCS systems, PCA was applied on the set of AMINO chosen order parameters along with the COM distance between ISO and the binding pocket as well as the RAVE CV to obtain a low dimensional space of the first 2 principal components. We then used K-means clustering to label each input frame of our trajectory with one of four clusters using the principal components. We find that each cluster represent a unique binding pose in Sudlow site II and/or a step along the toxin's unbinding route. **Figure 4.4/4.5** plot the PC space of 15 trajectories at a 0.2 ns interval colored by the K-means determined cluster with representative MD frames selected using the nearest 2D Euclidean distance to the cluster center.

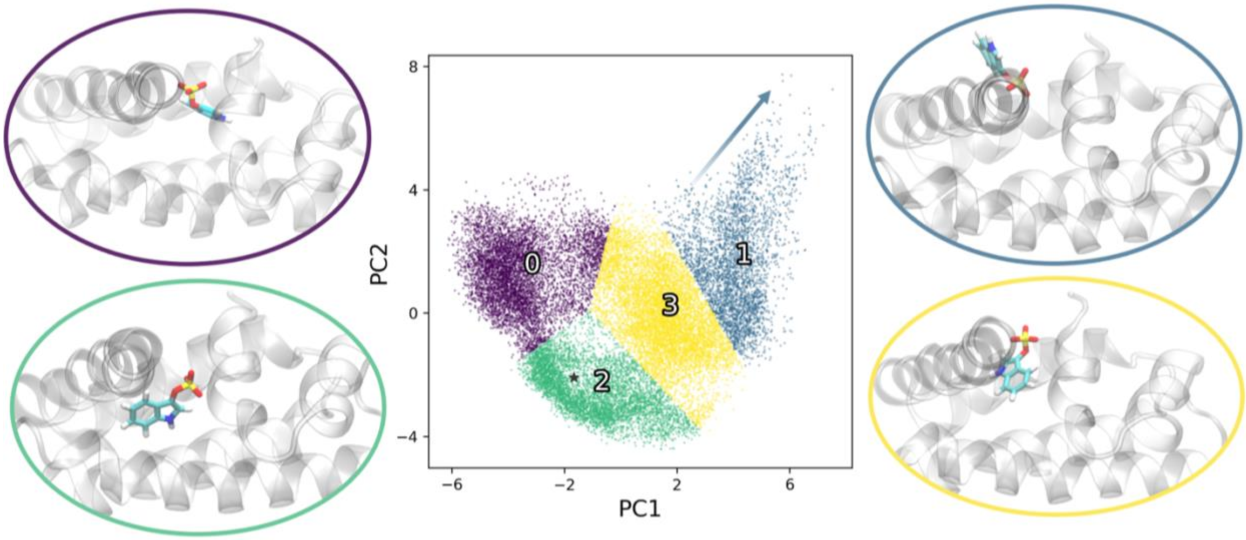


Figure 4.4 - PCA with K-mean clustering applied for ISO_{S489} trajectories. The center scatter plot shows the first 2 PC values for every 0.2 ns of simulation time colored and labeled by cluster with the label at the location of cluster center. A star in cluster 2 represents the initial ISO_{S489} pose and an arrow shows the unbinding direction in PC space. The 4 surrounding VMD snapshots are representations of each cluster chosen as the frame closest to each clusters center with outlines of the particular node's color.

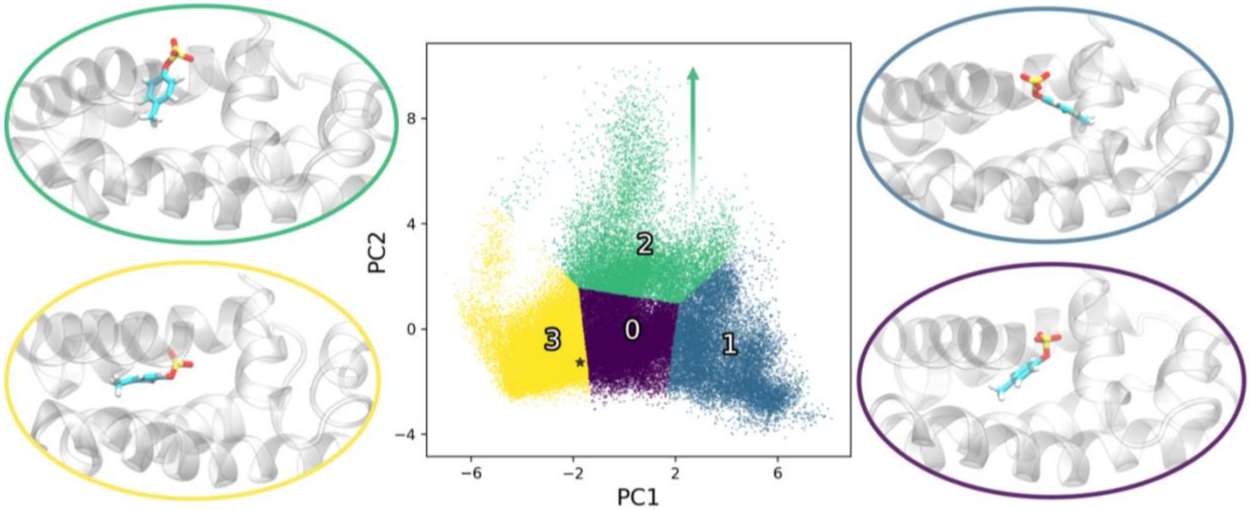


Figure 4.5 - PCA with K-mean clustering applied for PCS trajectories. The center scatter plot shows the first 2 PC values for every 0.2 ns of simulation time colored and labeled by cluster with the label at the location of cluster center. A star in cluster 2 represents the initial PCS pose and an arrow shows the unbinding direction in PC space. The 4 surrounding VMD snapshots are representations of each cluster chosen as the frame closest to each clusters center with outlines of the particular node's color.

Both the ISO and PCS systems were split into 4 clusters as any more or fewer led to a similar amount of mechanistic information overall or a loss respectively. Clusters 1 and 2 represent positions closest to the point of unbinding for the ISO_{S489} and PCS system's respectively. These unbinding clusters were determined both by the associated MD snapshots and the final unbinding frames being spent in the cluster (shown by the time-colored data points in **Figure II.3/II.4**). Longer trajectories tend to transition across the 4 clusters before unbinding. We also see that the unbinding cluster contains the lowest fraction of data compared to all others, 0.15 and 0.18 for ISO_{S489} and PCS respectively, indicating the nature of both systems to unbind relatively quick once the unbinding cluster is found. ISO_{S489}'s unbinding cluster is more visually clear than that of PCS, having the indole group flipped up outside of the pocket with just the sulfate group remaining. Our work in chapter 3 highlights the speed at which such interactions with the sulfate group break (<10 ns), supporting the short time spent in the cluster. Opposite ISO, PCS's unbinding cluster finds the sulfate group outside of the pocket with the bulkier section still bound. The swapping of these bound functional groups between toxins further reveal the differences between the two unbinding paths. These poses may also support why ISO unbinds quicker than PCS, since ISO is bound by weak hydrogen bonds while the PCS system is predominantly bound by interactions with hydrophobic sidechains that provide more steric hindrance. In fact, the cluster nearest the unbinding cluster for ISO (3/yellow) is bound by similar interactions as the discovered unbinding cluster for PCS. This similarity indicates that a fraction of the discussed PCS unbinding pose contains an even more transient step as seen in ISO which would explain the larger fraction of data in the labeled unbinding cluster of PCS compared to ISO. This is confirmed in **Figure II.5** which splits the PCS PC space into an increased 6 clusters and finds a further transient pose which makes up only 0.03 fraction of the total data, highlighting the transient nature of the pose. Interestingly, the respective MD snapshot confirms the unbinding path differences between ISO and PCS suggested earlier. While PCS first breaks the sulfate group interaction and then works its ring group out of the pocket slowly, ISO first flips the indole group out of the pocket and then observes a more transient step of simply breaking the sulfate interactions.

Additionally, the 3 clusters aside from the unbinding pose found for each system provide kinetic understanding of the unbinding paths. We find that the additional 3 clusters of PCS capture: (0/purple) toxin centered in the binding pocket with sulfate group oriented toward the mouth, (1/blue) toxin positioned opposite the unbinding point with the sulfate group oriented toward the mouth, (3/yellow) toxin in deepest pose with sulfate group pushed down (toxin parallel to mouth) below the unbinding point. The PC space does well in capturing the distinct position of PCS within the pocket since the x-axis/PC1 correlates with movement across the binding pocket while the y-axis/PC2 correlates with movements up and along the unbinding path. As mentioned earlier, the extra cluster for PCS in a 6-cluster configuration provides degenerate information captured in the

4-cluster configuration, only complicating the PC space. For the ISO system, the 3 additional clusters capture: (0/purple) the toxin positioned opposite the unbinding point with the sulfate group oriented toward the mouth, (2/green) both sulfate and indole groups are held just below the unbinding point, (3/yellow) similar to the PCS unbinding cluster, the sulfate group reaches out of the pocket while the indole group is held at the mouth. We see that the 2 PCs again capture the kinetic space well, finding the deepest positions to have more negative PC 1 and 2 values which increase as the toxin moves along the unbinding path, passing through cluster 3 and, finally, 1.

4.4 *Conclusion*

This chapter acts as both a direct follow-up to the previous regarding uremic toxins in Sudlow site II and a comprehensive application of using a machine learned CV for protein-ligand unbinding studies and successive analysis. Our work provides both kinetic and thermodynamic metrics for the unbinding of two troublesome uremic toxins in the field of CKD, ISO and PCS. We discovered that two distinct ISO binding poses in Sudlow site II unbind faster than PCS, supporting a link between unbinding speed and toxin chemistry being that the bulkier toxins which favor the deeper regions of the pocket unbind more quickly from the mouth. Furthermore, alternate funnel-based CVs were tested which led to improved statistics and computational time, again providing routes for optimization of these approaches for the study of uremic toxin binding. The thermodynamic results showed decent agreement with experiment while matching the overall trend between toxins, observing a stronger binding energy for PCS than ISO when bound near the mouth of Sudlow site II. Finally, PCA and K-means clustering were used to analyze the FaMetaD trajectories, providing mechanistic insight surrounding the unbinding paths of each toxin. For ISO unbinding, the indole group flips out of the pocket while transient interactions with the sulfate group attempt to hold onto the toxin before unbinding occurs. Alternatively, PCS adopts a more brute force method of breaking the transient sulfate group interactions before working the bulkier indole group out of the pocket more slowly. Although these results have room to grow by the study of additional toxins or poses, these results act as a strong basis for both PCS and ISO unbinding mechanics.

Subsequent work has a variety of directions in which to move forward. Experimental kinetic results, although difficult to obtain, would be valuable in the confirmation of these methods. A majority of the challenges met in this work came due to the complexity of this particular protein-ligand system and it certainly adds to why there are so few explorations of this systems kinetics. However, with so few experimental comparisons, validation of these important studies is left in an unclear position. Outside of experimental endeavors, simulation of alternative HSA conditions would be an important next step. The conditions studied in this section do not completely reflect physiological conditions. Moving toward more physiological settings could provide more accurate results and an improved understanding of the impact components such as fatty acids or varied oxidation have on toxin binding properties. Finally, the application of ML for CV selection is still evolving and has room to improve. Future advances in AMINO, RAVE, or alternative methods could lead to improved speeds and accuracy of results. Any of the discussed future directions will continue the push toward understanding and improving strategies for the handling of PBUTs, leading to potential improvements for CKD patient outcomes.

Chapter 5. **Modeling molecularly imprinted polymers for uremic toxin capture**

5.1 *Introduction*

As mentioned in Chapters 3 and 4, one goal of studying uremic toxins is to determine effective strategies which allow for management of PBUT concentrations in CKD patients. This approach may help in the development of portable devices which alleviate or completely remove the need for dialysate. In particular, the use of highly specific adsorbent materials has found some success in improving free toxin capture.[69,70] These strategies have improved by focusing on a materials porosity and surface chemistry, without which, adsorbent material could induce adverse effects in the target.[148] To this end, continued efforts have gone toward designing these adsorbent and/or porous materials as they could lead to the direct removal of PBUTs if designed properly.[149]

For example, the use of molecular imprinted polymers (MIPs) offers great potential to the field of PBUT management due to their ability to bind target molecules with high specificity, low cost, and portability. These polymer materials have found great success in the past as both sensors for cancer biomarkers and adsorbents.[150,151] A MIP is built by polymerizing a set of functional monomers (FM) by way of a cross-linker (CL) around a target template molecule, for example a toxin. The interactions between FM/CL and template are of reversible covalent or non-covalent type and, once the template is removed, forms a binding pocket optimized for the template.[152,153] MIPs have already shown their usefulness in this type of capture for particular uremic toxins, for example creatine.[151,154] One difficulty with a MIP based approach for toxin binding, however, is the amount of work required to design a proper MIP. Each piece of the MIP (FM, CL, template, and overall ratio) requires careful consideration and optimization in order to obtain the desired high specificity. Such a requirement, alongside the fact that the capacity for a general MIP is low compared to other adsorbents, poses as a barrier moving forward.

As a solution, MD simulations can act as an efficient tool to both speed up and reduce the costs of MIP testing. Currently, MD has already shown success in helping to optimize the preparation of MIPs by offering atomistic level detail regarding interactions between FM, CL, and template.[155–160] Such studies focus on investigating the pre-polymerization phase due to the ease of translation in MD as well as the importance of this phase in the production of high affinity chemistries.[161–163] Thus, there is ample motivation for exploring the mechanics of a MIP's components in the pre-polymerization phase using MD especially since this offers the opportunity for increased scaling and high-throughput results. To this end, we investigate a set of relevant FMs, a toxin (ISO) as a template, and a common cross-linker, ethylene glycol dimethacrylate (EGM). The FMs chosen at the onset of this study were 2-(dimethylamino)ethyl methacrylate (DEM), 2-hydroxyethyl methacrylate (HEM), and styrene (STY). In all, these MIP components were chosen for three reasons, (1) each has found extensive use as successful MIP components throughout both computational and experimental works[158,160,164–168] (2) they were tested by our experimental collaborators and (3) they offer potential for both hydrogen bonding and pi interactions with ISO which we have shown to be relevant in the competing binding poses of HSA. As such, our in-depth study of these components in the pre-polymerization phase helps to guide the development of MIPs aimed to capture uremic toxin in physiologically relevant conditions as well as the design of computational MIP experiments by offering novel directions for studying pre-polymerization phase MIPs.

The following sections will highlight our methods, results, and conclusions regarding MIP studies through an all-atom MD lens. We begin by discussing the basic tools and setup used when configuring and running our MIP systems in section 5.2. From there, section 5.3 is split into two categories: 5.3.1-5.3.3 focuses on the pre-polymerization phase while 5.3.4 investigates a novel approach for modeling interactions of the imprinted material with a template. Finally, the chapter concludes by offering insight as to what this work may offer to future efforts in MIP design both experimentally and computationally, which, for the later, is given close attention due to its potential for future growth and usefulness.

5.2 *Methods*

5.2.1 *Simulation design and setup*

In order to study the MIPs components of potential interest to our experimental collaborators and ISO binding specifically, partial charges were needed for the DEM, HEM, STY, EGM, and extra solvent component methanol (MeOH). These charges were assigned using the restrained electrostatic potential (RESP)[169] method performed in Gaussian 09[170] using a B3LYP hybrid functional with the 6-31G(d) basis set.[171,172] All water was modeled using the TIP3P forcefield.[91] Each MIP pre-polymerization system was initially configured using the packmol software, placing all MIP components and solvent within a 6 nm box randomly (unless described otherwise). These structure files were then prepared, equilibrated, and used for production runs with the GROMACS 2018.3 software. All systems have at least the same basic 3-step equilibration process: (1) 10,000 steps undergoing steepest descent energy minimization, (2) 0.25 ns annealing to heat the system from an initial 5 K to the target 300 K, (3) and a 0.5 ns NPT simulation using the v-rescale thermostat alongside the Berendsen barostat.[39,93] Finally, 50 ns long production runs were completed in the NPT ensemble with 2 fs timestep. Post processing of trajectories was done using various GROMACS suite tools and Plumed 2.4.[96] In particular, COM minimum distances were calculated through Plumed using the following equation,

$$s = \frac{\beta}{\log \sum_i e^{\beta/s_i}},$$

where β is a user defined parameter chosen to be the largest value that does not wash out the distances (100) and s_i is our distance metric (COM distance between components). The following subsections describe the exact compositions of the systems studied as well as any particular details regarding further setup and equilibration where required.

5.2.2 *Styrene and methanol systems*

We began by modeling an experimentally tested MIP ratio between the FMs, CL, and toxin of 1:8:16:8:40 for the ISO:DEM:HEM:STY:EGM respectively. The number of molecules used for the system exactly matches the ratio and is accompanied by a solvent of water and MeOH at a 1:3 ratio selected to match the experimental concentration. The exact number of molecules used for all components can be found in **Table III.1**. Potassium ions are added to neutralize the system's charge (provided by ISO). Equilibration and production were completed as described in the previous sub-section, with three trials trajectories used for averaging and calculation of error bars. These systems acted as a baseline for our initial investigations of pre-polymerization MIPs.

In order to monitor concentration effects on FM-toxin interactions, we tested varying the solvent concentration at intervals of 20% the total solvent for both water and MeOH. Each system contained a single STY and ISO, followed the equilibration scheme, and underwent a 40 ns NVT production run. We used the well-tempered metadynamics (WTMetaD) method in order to improve sampling of the FM-toxin interaction and achieve smooth convergence across multiple walkers (using 40 identical systems).[1] Bias was deposited to the COM distance between STY and ISO every 500 steps (1 ps) with an initial height of 1 kJ/mol and biasfactor of 6. The well-tempered metadynamics scheme was implemented using the Plumed 2.4 software package.[96,173] The files to reproduce both the unbiased baseline and biased solvent tests are available on GitHub at https://github.com/jamesprg/Batch-Simulation-With-Varied-Composition.

5.2.3 *Batch generation of improved systems with varied composition*

After the initial baseline trials testing the 1:8:16:8:40 ratio in a water methanol solvent were completed, improvements were found in order to further streamline the surface chemistry of the MIPs. Specifically, STY was removed (for reasons we will highlight in the results section) and thus MeOH was no longer necessary to help dissolve STY in the pre-polymerization phase. As such, the focus of this chapter is dedicated to studying the ISO:DEM:HEM:EGM system and how varying the ratio of components effects various physical properties. The goal of this process was to help discover a method that would both provide the ideal MIP for ISO while also detailing a method for efficiently deciding upon a ratio for MIP components. In order to complete this task, we required a large amount of MIP systems to be generated and prepared quickly. By way of python script to batch automate the process, a set of systems were generated which vary one component (DEM, HEM, or EGM) using a multiplier from 2-10 of the baseline ratio, 8:16:40 for DEM:HEM:EGM respectively. Each component that was not varied was held constant at the baseline value. This resulted in a set of $r * n + 1 = 28$ systems, where r is the number of ratio multipliers (9), n is the number of FMs and CLs (3), and 1 accounting for the baseline control system. All batch generated systems follow the previously described equilibration and production schemes, resulting in a 50 ns NVT simulation for each ratio tested.

By using the baseline ratio that was determined experimentally as a positive control, we then moved to choosing negative controls. This approach provided an effective method for determining how appropriate particular analysis methods were. As such, two more batch simulation sets were made as negative controls: (1) a set of systems with a single FM (DEM) or (2) a set of systems replacing either FM with an alternative that showed poor results experimentally, methacrylic acid (MAA). These negative controls were decided upon after discussion with our experimental collaborators. Each system was generated in the same manner as the batch automated systems based on the baseline ratio described previously.

In order to best model conditions of the experimental pre-polymerization phase, we implemented a dialysate solvent for all batch automated simulations. Specifically, we matched the experimentally selected 0.9% NaCl solution. This decision further allows for our baseline 1:8:16:40 system to meet the conditions a of a positive experimental result as well as remain consistent with our negative control systems. Furthermore, these conditions allow our systems to be more physiologically representative of the conditions expected for the MIPs. All forcefield files, initial structures, topologies, and scripts to replicate our methods can be found on the associated github.

5.2.4 *Using constraints to monitor bias of varied compositions*

In order to investigate the potential interactions of a binding pocket formed with the FMs and CL of interest, a new set of systems were created which used restraints to study, what we label, the pseudo-binding pocket behavior possible in a polymerized MIP. WTMetaD was used to determine an ideal distance for each FM/CL to be held, following the same 40 walker process as before with a single FM/CL and ISO for each possible interaction. For the three components, DEM, HEM, and EGM, the following maximum distances were chosen respectively: 0.52 nm, 0.48 nm, and 0.54 nm. We then batch simulate a new set of systems which contains all combinations of 1-3 molecules for each FM and CL (27 systems) with each component bound within the determined maximum distance of a single ISO. Each system followed the 3 step equilibration process described in section 5.2.1 with two additional steps: (4) a steered MD simulation which led each FM/CL (minimum 3 and maximum 9) to be 0.3 nm below the optimal COM distance found from WTMetaD followed by (5) an additional NPT equilibration (same simulation scheme as before) with an upper wall potential implemented with Plumed 2.5[96] as follows,

$$V_{wall}(x) = \begin{cases} k(x-a)^2, & x > a \\ 0, & x \leq a \end{cases}$$

where k is the restraining force constant (50 kJ/mol), x is the COM distance between the particular FM residue and ISO, and a is the optimal distance found by WTMetaD for each FM/CL at which the restraint is turned on. Afterwards, each system's production run followed the same simulation protocol except for the addition of the upper wall potential, $V_{wall}(x)$. Using this restraint, we were able to investigate the potential favorability of different binding pocket confirmations. As before, all files used to setup the restraint based pseudo-binding pocket systems are available on GitHub.

5.3 ***Results and discussion***

5.3.1 *Initial 1:8:16:8:40 baseline pre-polymerization phase MIP*

In order to initially determine the solvent averaged spatial distribution across the experimentally determined MIP ratio for ISO capture, we calculated the average radial distribution function (RDF) across three trials. Specifically, we monitored various FM/CL-ISO interactions, mainly consisting of hydrogen bonds (HB) between donor and acceptor of either the FM or ISO. The more unique interactions are the aromatics possible between ISO-STY and STY-STY due to the various configurations possible for pi-pi stacking.[174,175] For this purpose, RDFs were calculated between each symmetrically unique aromatic carbon in STY and ISO. The choice of carbon did not change the peak position/shape in the RDF and thus all following STY-ISO RDF profiles will use the same carbon to carbon atom selection. Each pair of atoms used for RDF calculations can be found in **Table 5.1**. The resulting RDF profiles from the average across all 3 trajectories is shown in **Figure 5.1**.

Table 5.1 - Pre-polymerization component interactions used for RDF.

Interaction Pair (X-Y)	**X atom[a]**	**Y atom[a]**
ISO-DEM	H6	O1
ISO-STY	C1	C1
STY-STY	C1	C1
ISO-EGM	H6	O2
ISO-HEM	O3	H8
ISO-MeOH	O3	H1

[a] Atoms are labeled by their atom name in the provided amber forcefield files.

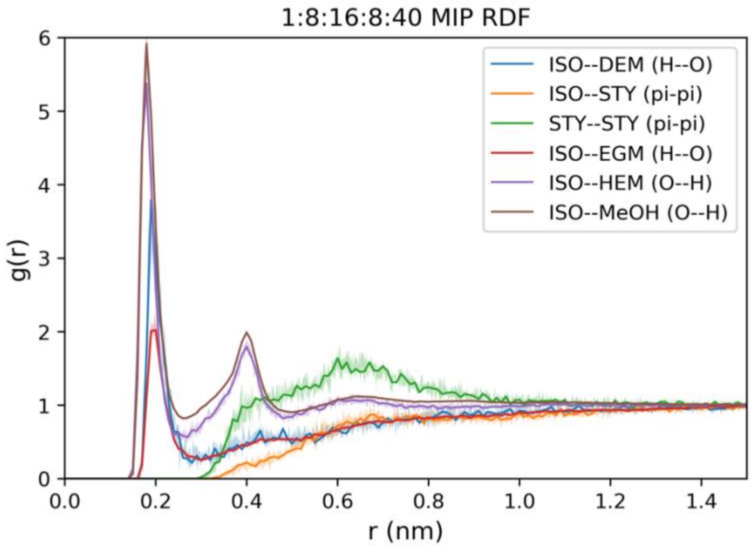


Figure 5.1 - RDF profiles between each FM/CL and ISO. Profiles are generated using the Gromacs rdf tool with interactions chosen as described in Table 5.1 Error bars are standard deviations across the 3 trials. Each pre-polymerization component can be found in the legend. We highlight the fact that each HB interaction sits at the expected position around 0.2 nm meanwhile the aromatic interaction between ISO and STY shows no particular peak.

Analysis of **Figure 5.1** reveals both the most and least prominent interactions of the pre-polymerization components. Additionally, we can calculate the area under each profile up to some cutoff in order to compare the strength of each peak. The strongest HB peak comes from the ISO-MeOH HB (0.31 ± 0.005) followed by the ISO-HEM HB (0.26 ± 0.017). The ISO-HEM HB makes up the bulk of the HB interaction which use ISO as the acceptor, highlighting the comparative usefulness of HEM for interacting with ISO. On the other hand, DEM and EGM makeup the bulk of the HB interactions with ISO as HB acceptors, with an area under the first peak of (0.16 ± 0.03) and (0.11 ± 0.01) respectively. The final FM, STY, did not show any detectable structure through the RDF profile. The STY-STY profile did in fact observe a broad peak around 0.5-0.7 nm, tentatively matching an aromatic T-shaped carbon-carbon interaction.[176] This may indicate that a STY is prefers to interact with other STY rather than ISO.

Further insight regarding the interactions between pre-polymerization components can be gained by analyzing particular distance metrics between ISO and FM/CL. We chose to monitor the COM minimum distance between each FM/CL-ISO pair versus time in **Figure 5.2**. This metric provides a general idea as to each component's interactions with ISO along a trajectory. MeOH observes the bulk of the lowest minimum distance. This is likely due to a majority in solvent and lack of steric hinderance with ISO. The second lowest ranking distances are found with HEM, as expected based upon the RDF peaks. However, we can now see that the slight differences in RDF peak may not fully capture the differences in dominant interactions of the pre-polymerization phase. The RDF profiles are able to better parse the exact types of interactions, since components such as HEM and EGM appear in reverse priority here, despite the pure difference in frequency of potential HB partners noted by the RDF. Another useful comparison between metrics is the lack of specificity in STY. Again, no particular structure is found between STY-ISO, since the minimum distance is constantly fluctuating across a wide range of values. This effect is much less prominent for HEM/EGM, where a lower average COM minimum distance is achieved, even with an expected level of intermittency between frames.

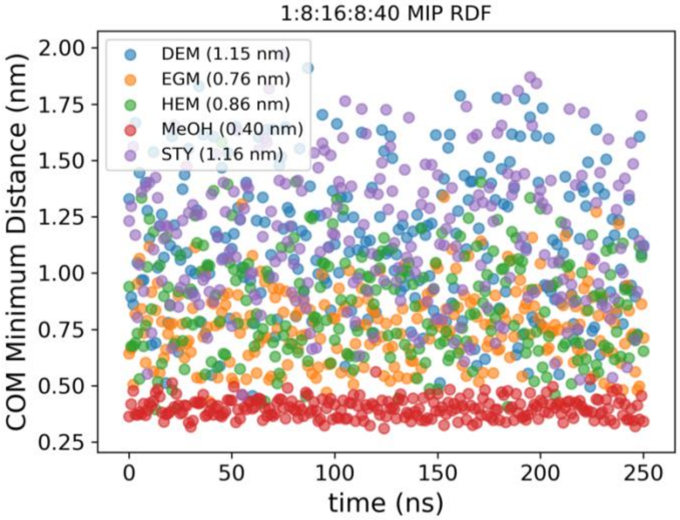


Figure 5.2. - COM min distances calculated by Plumed. Average COM minimum distance is labeled in parenthesis of the legend. Keep in mind, these average values are not representative of the average interacting minimum distance since data points are not screened for whether or not the point represents a potential interaction with ISO nor weighted for the total amount of each component.

5.3.2 *Solvent effect on FM-ISO interaction in pre-polymerization phase*

Due to the level of ISO-MeOH interaction as well as lack of ISO-STY interaction shown in the baseline system, we investigated the potential solvent effects on ISO-STY interactions. In this way, we first produced a set of trajectories varying the amount of each solvent component (MeOH and water) in 20 mole percent intervals of the total solvent present in the baseline system. **Figure 5.3** indicates an important trend when combining these particular MIP components in the pre-polymerization phase. That is, in decreasing the amount of water and accordingly increasing MeOH in our solvent, the STY-ISO interaction washes away. We see that as the MeOH concentration increases from 20 to 30 percent, the free energy barrier of the ISO-STY interaction decreases dramatically by approximately 61%. This seems to be indicative of the nature of STY-

MeOH reactions that we know to exist and be necessary to dissolve STY in our system experimentally. However, this lack of interaction may be reason to seek out alternative FM when optimizing a MIP as the end goal is to obtain an ideal surface chemistry for the target toxin.

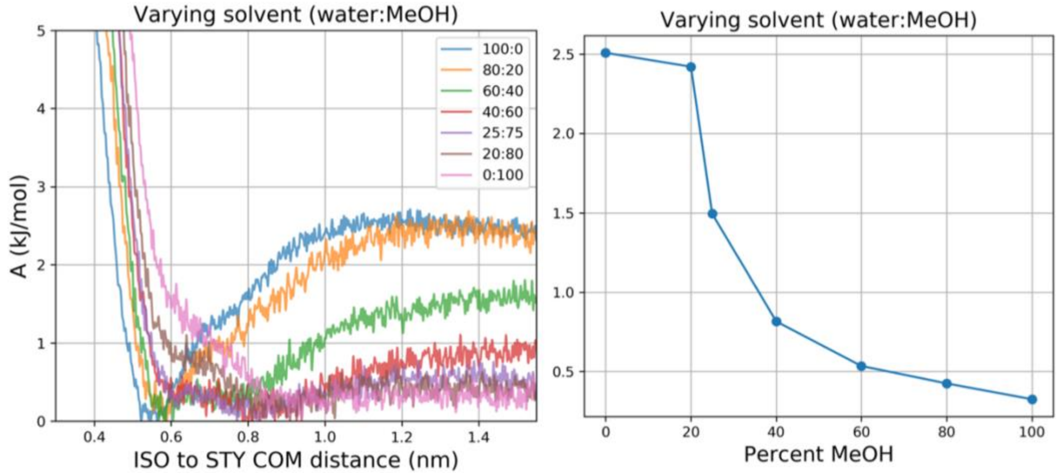


Figure 5.3 - Free energy profiles with varied solvent concentrations. The left panel shows each free energy profile generated from 40 walkers of WTMetaD run for 50 ns a piece while biasing the COM distance between the single ISO and STY in the system. The accompanying right panel provides the calculated free energy barrier versus decreasing water concentration, highlighting the loss of interaction with MeOH addition.

To investigate the solvent effects further, we also compared some of the other interactions present in pure solvent systems. Alongside the pure MeOH ISO-STY system, we calculated the STY-MeOH and ISO-MeOH interactions, as shown in **Figure 5.4 (a)**. Examination of the two profiles which include MeOH helps explain the lack of ISO-STY interaction (also shown in the green profile), as MeOH and ISO find a favorable interaction in the region which is otherwise unsampled in the ISO-STY profile. The slight differences in the left side of the MeOH profiles can also be explained by expected HB interaction between ISO-MeOH that would not be possible for STY-MeOH. We also generated another set of single FM-ISO systems using both DEM and HEM. **Figure 5.4 (b)** shows the resulting free energy profiles of each ISO-FM COM distance in either pure water or pure MeOH. Interestingly, the washing out of ISO-FM interactions caused by MeOH seen in the ISO-STY system is still present. Even with the free energy barrier of both the DEM and HEM profiles being twice as large as the STY system, the barrier is entirely lost. Thanks to these findings, we continued forward without MeOH as a solvent due to its high interaction energy with the pre-polymerization components. In addition, this led to the removal of STY as well since MeOH was a necessary component for homogeneity experimentally. This approach proved successful, finding the same baseline ratio 1:8:16:40, now without STY, to observe a high removal capacity as before in experiment, indicating that both STY and MeOH were not necessary for an effective MIP.

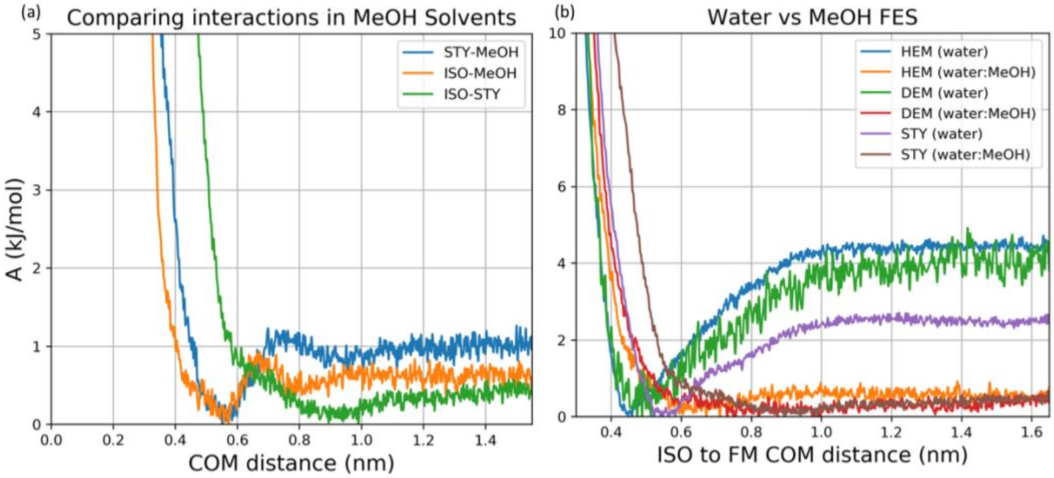


Figure 5.4 - Free energy profiles in pure solvents between ISO-FM and ISO-MeOH. (a) Free energy profiles in a system containing a single ISO and STY in MeOH solvent, highlighting the preference of both ISO and STY to interact with MeOH over each other. (b) Free energy profiles between ISO and the 3 FMs in 2 different solvents each (6 systems total), showing the increased interaction with the removal of MeOH. All profiles are generated using WTMetaD from 40 walkers with 40 ns per walker.

5.3.3 *Effect of varied composition on ISO:DEM:HEM:EGM MIP*

With the updated baseline ratio in hand, we performed a set of batch automated simulations which varied the amount of each MIP component across a range while holding all other components at the baseline. To support the set of trajectories based upon the baseline ratio, we also generated two more sets of systems which have been shown to act as negative controls experimentally. Specifically, we used the following two systems: (1) replacing DEM with MAA and (2) using only a single FM (DEM). From the three sets of batch generated trajectories various average structure features were calculated for the purpose of determining the optimal ratio of MIP components efficiently. To this end, we first calculated the average COM minimum distance between each component and ISO as shown for ISO to EGM in **Figure 5.5** (with additional results provided in **Figure III.I**). Following from our previous use of this metric, a cutoff distance is now implemented, only accounting for distances within 0.5 nm which we treat as the range of interaction. The profiles reveal a distinct lack of differentiation between systems despite a change in FM/CL amount, with no distinct trends correlating to these variations. Furthermore, if we compare across the three differing system setups, (i) baseline, (ii) with MAA, (iii) and single FM, there is again a lack of differentiation between the systems. When examining the ISO-EGM average COM distance, most if not all values fall within error of any other system setup and composition. We note specifically that the baseline dialysate systems (DEM composition 8, HEM composition 16, and EGM composition 40) do not observe higher average interaction distances when compared to either negative control. Thus, these metrics alone proved unsuitable in the determination of an optimal MIP ratio.

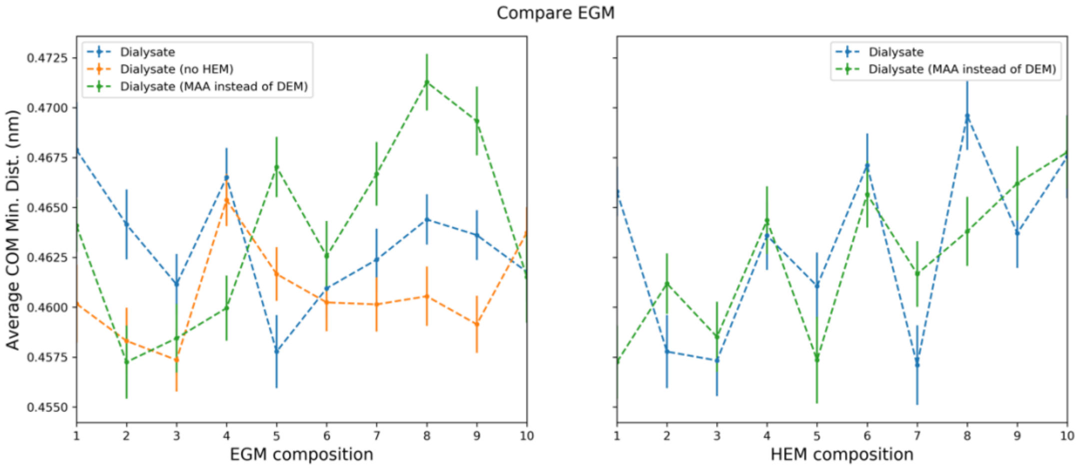


Figure 5.5 - Average COM minimum distances between EGM and ISO within a 0.5 nm cutoff. The left panel displays the results between 3 cases: dialysate, dialysate without HEM, and dialysate with DEM replaced by MAA, for increasing EGM ratio. The right panel shows the results for increasing HEM ratio

of dialysate and dialysate with DEM replaced by MAA only since HEM is removed from the other test case.

As a follow up to the average structure calculations, we investigated the use of a potentially more suitable metric in the form of HB lifetime distributions. The use of HB lifetime analysis for pre-polymerization phase MIP simulations has shown the ability to determine between FMs in the past.[155,177] However, it has yet to be shown if this approach could be used for the determination of an optimal FM ratio across a finely discretized region. Although the basic HB lifetime has the potential to forgo a similar result as the average structure calculations, HB interactions themselves can also be extracted from the trajectories on-hand and then transformed into a distribution of lifetimes. **Figure 5.6** shows the distributions for each baseline component when varying DEM composition. The HEM-ISO distributions (**Figure 5.6 (a)**) observe the most similar set of distributions across the 10 systems. This supports the hypothesis that HEM is an effective HB partner for ISO and a FM in general and may be less impacted by increasing the amount of DEM. In **Figure 5.6 (b)** and **Figure 5.6 (c)**, we see that the peaks of the EGM-ISO and DEM-ISO distributions, respectively, cover a range of probabilities twice as large as the HEM-ISO distributions. Again, this could help support the interaction between ISO and either DEM or EGM being less consistent with ISO as the donator, opposite the acceptor role with HEM-ISO. These roles are further supported by the average HB lifetimes themselves, with HEM achieving the highest average lifetime of 13.89 ps. Furthermore, DEM’s range of lifetimes/distribution shapes may be indicative of the effect changing FM composition has on DEM-ISO interactions. That is, the DEM-ISO average HB lifetime increases by a factor of 1.92 when moving from 8 to 16 DEM, an increase unseen for any HEM or EGM transitions. In fact, **Figure III.2** shows that DEM observes drastic jumps in both the set of HEM and EGM varied HB distributions as well. Further development of an HB lifetime distribution-based approach for the analysis of pre-polymerization phase MIPs could support a method for determining optimal component ratios.

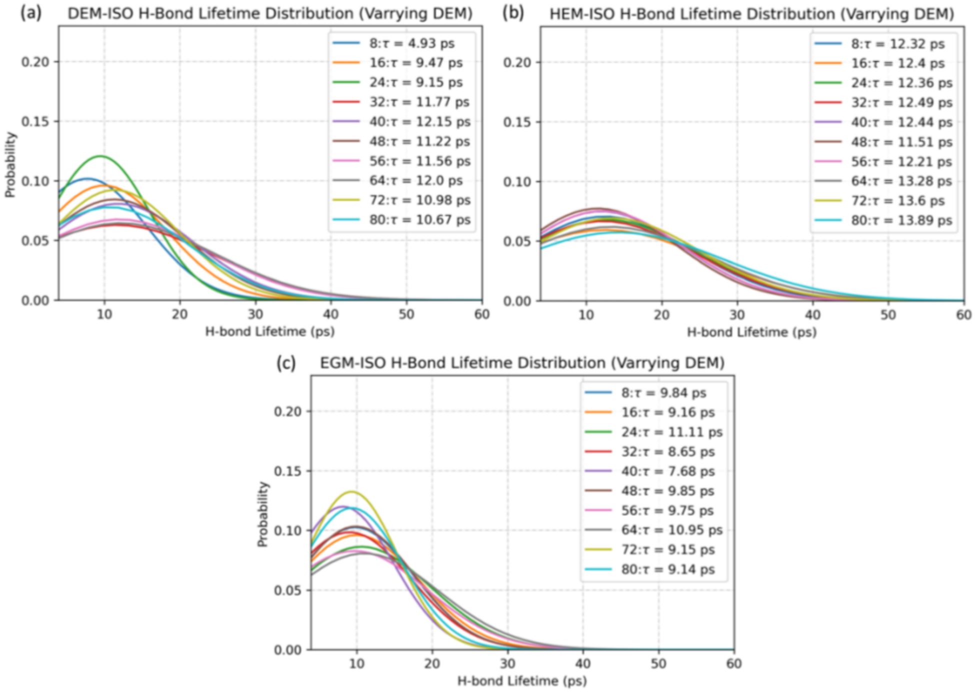


Figure 5.6 - HB lifetime distributions with increasing DEM in systems. The (a), (b), and (c) panels show the HEM-ISO, EGM-ISO, and DEM-ISO HB lifetime distributions respectively. Each lifetime reported in the legend is an average calculated across the full trajectory.

5.3.4 *Pseudo-binding pocket interactions of ISO:DEM:HEM:EGM MIPs*

In order to investigate an alternative approach to computationally screening potential MIP components for effective toxin removal capacity, we simulated what we label a pseudo-binding pocket framework. To this end, we generate a new set of batch automated systems covering a range of 1-3 molecules for each FM/CL restrained around a single ISO. We then analyze the resulting applied bias across each trajectory to identify which systems were the most stable. **Figure 5.7** shows an ordering of the systems based upon each component's average applied bias. The resulting orders have one clear similarity between each other, being that the highest amount of bias occurs in the largest (3:3:3) system as the pseudo-binding pocket verges on becoming overfilled. As noted in our previous studies, the pre-polymerization phase does not form any complex structures for extended periods of time, thus we expect large increases in applied bias with system growth. More interestingly, however, are the differences in each component's associated orders. For example, all three components find different lowest average bias systems: DEM – 2:1:1, HEM – 1:2:1, and EGM – 1:1:3. This fact may help reveal the optimal number of each components which will regularly interact with ISO in particular conditions (i.e. ratios). Another notable trend is DEM's preference for 2 and then 3 constituents before 1, which is seen best by observing ranks 1-7. Rank 3 for the DEM sort is the only rank prior to 8 which contains a DEM:1 ratio, with DEM instead finding lower average bias with 2 or 3 DEM molecules. In fact, the first 7 DEM ranks makeup all possible combinations of ratios with DEM:2 or DEM:3 that do not include HEM or EGM at a ratio of 3 aside from the 2:2:2 or 3:2:2 system. This trend could be explained by DEM having a preference for higher DEM counts over other components as well as DEM ratios above 1. Similarly, the HEM sort's trend matches DEM exactly for ranks 1-9, however rank 10 for DEM (2:2:2) falls to rank 11 for HEM which interestingly prefers a 1:3:3 configuration slightly more, potentially indicating HEM's preference for HEM/EGM presence over DEM. Finally, the EGM sort does not follow the exact same rank 1-9 systems as HEM and DEM. Instead of preferring a self-ratio of 2 before 3 for the lowest average bias, EGM places priority on 3 specifically. Furthermore, EGM pushes any system with a self-ratio of 1 to rank 7 and below, standing apart from both DEM and HEM which has the smallest system (1:1:1) at rank 3. This could also be a sign of EGM ratio preferences, showing an even greater ability than DEM to include 3 constituents in a pseudo-binding pocket model.

Average interaction energies between ISO and each of the three MIP components were also calculated as show in **Figure 5.8**, sorted by coulombic energies (E_C) . Additionally, each set of interaction energies sorted by Lennard-jones energies (E_{LJ}) can be found in **Figure III.3**. The HEM interaction energies support the strength of HEM-ISO interaction (HB), observing higher

E_C than EGM or DEM E_C across all systems. Furthermore, the HEM $-$ E_C trends toward smaller magnitudes with particular combinations of smaller system sizes, most notably 1:1:1 observes the lowest HEM $-$ E_C of all systems. Not only does HEM observe the most negative E_C, it also shows the most variation across the range of systems, achieving a standard deviation of 6.3 in E_C. Interestingly, HEM observes a crossing point in both E_C and E_{LJ} rankings that is unseen for all other FMs. As seen in both **Figure 5.8**/**Figure III.3**, DEM and EGM observe little changes to either E_C or E_{LJ}, with the largest standard deviation across all system energies being $2.2\ kJ/mol$ for EGM $-$ E_{LJ}. EGM and DEM profiles also offer similar results in terms of average interaction energies, with an $\overline{E_C}$ of $-4.6\ \pm 0.8\ kJ/mol$ and $-4.9\ \pm 0.9\ kJ/mol$ and $\overline{E_{LJ}}$ of $-17.9\ \pm 2.2\ kJ/mol$ and $-17.9\ \pm 1.9\ kJ/mol$ respectively. The most notable difference is the jump between the final two EGM $-$ E_{LJ} ranked systems (-16.6 to $-8.7\ kJ/mol$), signaling a distinct disfavor with EGM interactions in the bulkiest system compared to all others. However, each interaction does have a similarly low result for the bulkiest system as we would assume based upon the amount of component packing. Results such as these, especially the HEM crossing point, could prove critical in determining the minimum conditions necessary for more favorable FM-toxin interactions.

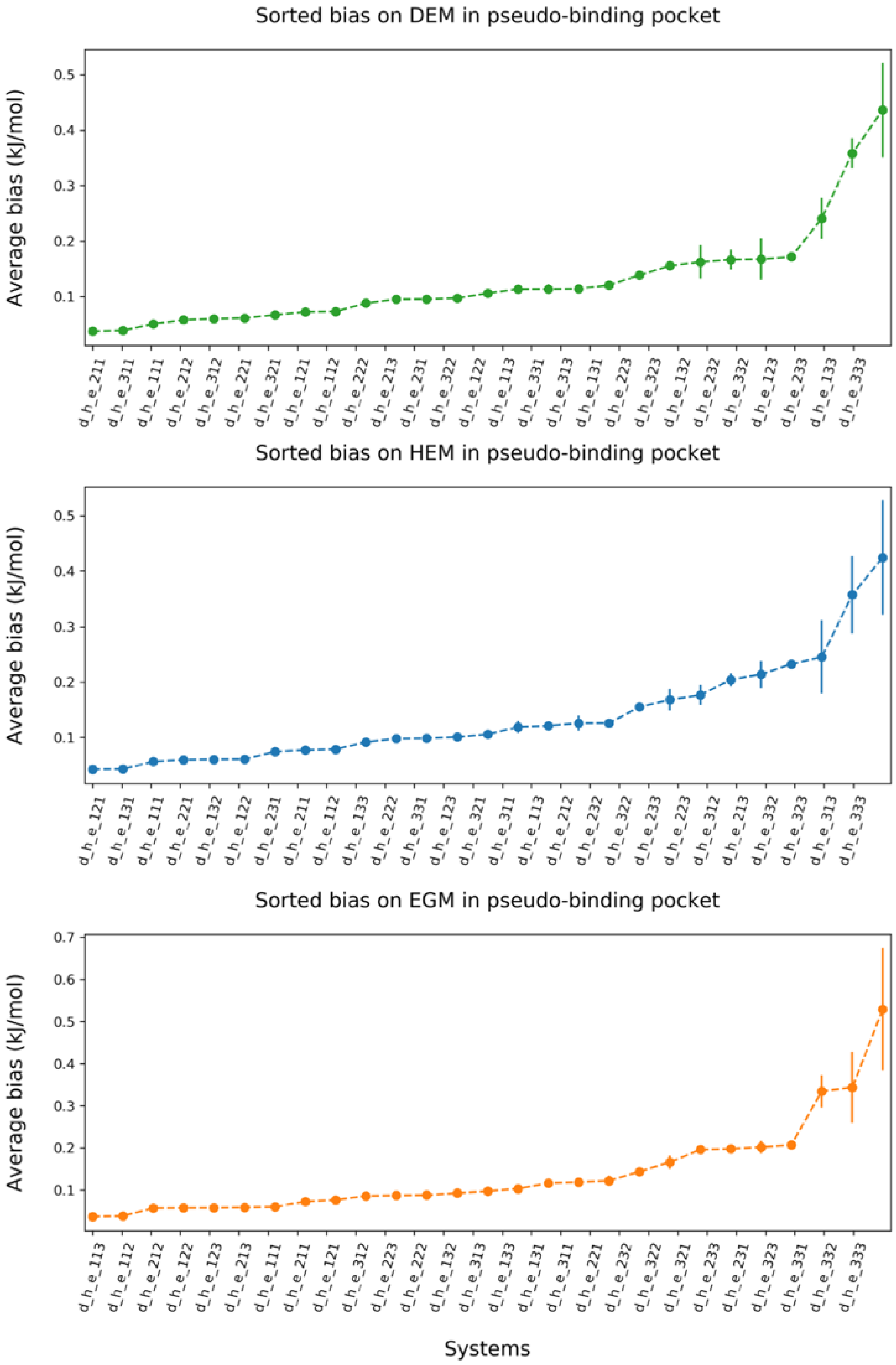
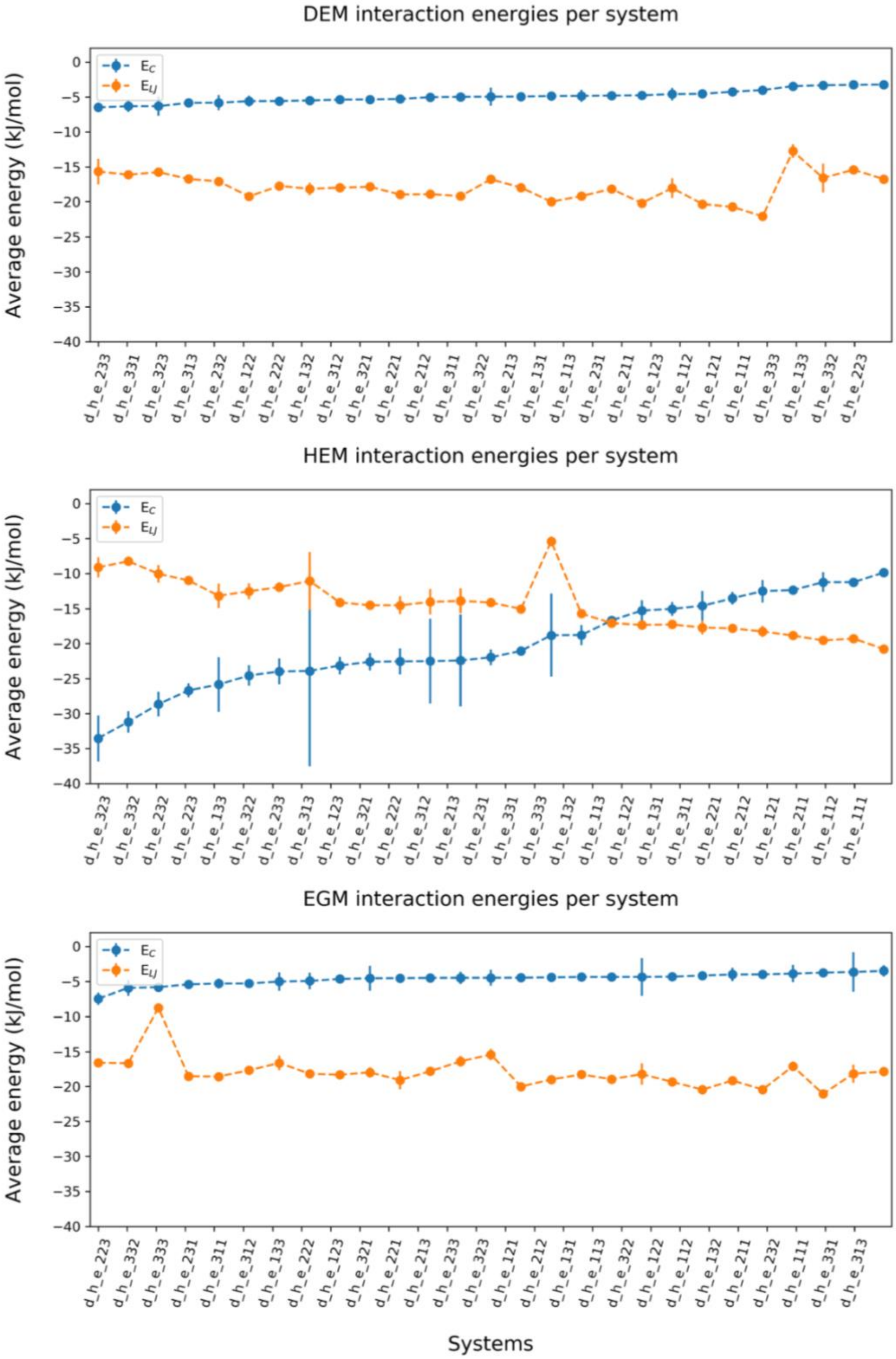


Figure 5.7 - Average bias per MIP component applied across pseudo-binding pocket trajectory. Averages and error bars are calculated by splitting trajectories into thirds. Each average is normalized to account for the total number of components present.



DEM interaction energies per system

HEM interaction energies per system

EGM interaction energies per system

Figure 5.8 - The interaction energies per MIP component in pseudo-binding pocket ranked by E_C. Average Coulombic (E_C) and Lennard-Jones (E_{LJ}) energies for each system are reported (sorted by E_C) per component. Error bars are calculated using 3 equal partitions of the trajectory. Each average is normalized to account for the total number of components present.

5.4 *Conclusion*

While pre-polymerization phase MIPs have proved a difficult system to use for the determination of an optimal component ratio, there was still plenty to be gained from these efforts. First of all, our work has supported the usefulness of a computational approach for the design of MIPs. Despite average structure properties providing little in the way of an optimal ratio on their own, alternative metrics begin to provide a clearer understanding of component interactions. We have shown that the monitoring of HB lifetime distributions can provide a deeper level of comparison between MIP configurations than average properties. Our initial findings in the pre-polymerization phase were bookended by HEM's consistent ability to observe both stable and longer lasting interactions with ISO. Meanwhile, DEM was found to observe the most variation in those same respects. However, when investigating an alternative model for MIP interactions, a pseudo-binding pocket configuration, we found that HEM is the most likely of all components to show interaction energy variation, albeit with a consistently strong interaction. Even these findings may prove useful in MIP synthesis, providing insight into the intricacies of interactions encountered during MIP synthesis and use.

Despite the handful of results successfully comparing the baseline effectiveness between FMs/CLs in the literature, determining optimal combinations of ideal components is a still unmet challenge. Forthcoming work should be aware of the warnings found in this chapter regarding average solvent properties and their struggle to meet this goal. Although our work has taken steps towards determining the optimal conditions for one particular MIP, the strategies employed need further development to provide the type of efficient determination necessary for experimental progress and clinical use. Attempting to model steps of the MIP process outside of simply the pre-polymerization phase has high potential, as shown in our investigation of pseudo-binding pocket structures. This has also been shown, for example, in Zink et. al. with the simulations of Virtually imprinted polymers (VIPs) by way of a unique virtual chromatography simulation.[178] Alternatively, computational approaches have also found success outside of any particular step of the MIPs building scheme using local density of state calculations for the optimization of FM suitability.[179] These efforts pave the way not only toward understanding the mechanics and theoretical basis of MIPs, but also optimal MIPs design, leading to future improvements in CKD patient outcomes.

Chapter 6. **Conclusions**

6.1 ***Impact***

My journey as a Ph.D. student has provided me the opportunity to work in a variety of settings. It is not necessarily standard that one makes a transition between research groups yet continues growing one’s original interests while branching into new ones. This, however, perfectly describes my experience when shifting from my original cell membrane inspired work to the world of uremic toxins as described in the previous chapters. I have had the great opportunity of working in two very distinct biophysical domains with various personal and collaborative laboratories, all of which my contributions have affected in unique ways.

Work in both the Keller and Maibaum Labs has brought clarity to phase behavior in cell membranes and the potential of MD for modeling these long-time scale dynamics in large systems respectively.[10,16,52,180] In order to support the lacking clarity regarding n-alcohol anesthesia, my work offered a unique bridge between the use of MD and small molecule partitioning by calculating phase composition and free energy changes as well as flip-flop rates of bilayer components using statistical modeling. My efforts also shed light on the benefits and pitfalls of using a coarse-grained MD approach for biophysical investigations, especially when looking at both macro and microscopic properties.

After joining the Center for Dialysis Innovations team, my work brought various continued and novel results that supported their goal in renal replacement research. My initial work with PBUTs acted as a continuation of the first MD based molecular-level study of these toxins and provided novel long time-scale dynamical insight while identifying sets of previously unknown binding poses for indole-3-acetic acid and hippuric acid. Following this, I implemented a ML approach for calculating both kinetic and thermodynamic unbinding properties by taking advantage of deep learning for optimizing an unbinding coordinate as well as enhanced sampling. Finally, I offered atomic scale resolution to the interactions and properties surrounding molecularly imprinted polymers, creating methods to begin searching for optimal component configurations as well as simulating post-template removal interactions that, to our knowledge, have yet to be investigated computationally. My studies of uremic toxins have both proved useful for the design of certain toxin removal strategies, such as competitive binders and adsorbent materials, as well as continued an important effort in highlighting the effectiveness of MD and statistical modeling for these complex issues.

Overall, my Ph.D. has culminated in emphasizing the potential for combining MD and statistical modeling to ascertain kinetic, thermodynamic, or mechanistic information in biophysical systems. In doing so, I have contributed to relevant biomedical domains like renal replacement therapy and anesthesiologic understanding as well as continued a push toward MD and ML based approaches for future research. Finally, my analysis codes have been made accessible through software sharing resources in order to strive towards improving both general and computational scientific reproducibility.

6.2 ***Future work***

Recent findings in the literature could prove useful in furthering the endeavors of my first project regarding alcohol partitioning into coarse grained model membranes. On the experimental side, the past year has provided many important discoveries in the study of anesthesiology, further revealing the importance of lipid phase formation in these processes.[181] Computationally, a new version of the Martini model (Martini 3) has been published that provides improvements for a variety of issues, including difficulties that were mentioned in Chapter 2 regarding bead/system sizes in the calculation of macroscopic properties.[182] My hope is that new endeavors in this field will take advantage of these resources for the continued modeling of more complex systems as we seek to understand the contributing factors in phase formation as well as general anesthesia. Furthermore, my implementation of deep learning for the optimization of unbinding coordinates could prove useful for the calculation of various membrane thermodynamic properties, aiding in the determination of small molecule partitioning effects.

Despite my many advances forward in the study of PBUTs, there is certainly a bevy of work still to be done. While my efforts have continued to provide pertinent information as well as methods for the investigation of uremic toxin removal strategies, continuing upon this work is essential. The modeling of more physiological and alternative conditions is an important next step in both producing more comparable results while also pushing towards the testing of evolving strategies. I have prepared my work to be continued forward with investigating the effects of fatty acids bound to HSA due to the potential to both act as a competitive binder and/or induce allosteric alterations to these systems that could otherwise go undetected. After this, studies can be done regarding the effect of oxidation changes which have also shown potential in this area. Both of these categories are simple next steps, that, when combined with the ML approaches applied in Chapters 3 and 4, could provide as suitable beginnings to the development of new materials or strategies for the purpose of improving CKD patient outcomes.

My efforts with MIPs and uremic toxins is certainly the least covered domain in the literature of all my works and could benefit from future endeavors. Most simply, a continued study of the pre-polymerization phase testing the alternate uremic toxins that have been generated by myself and previous Pfaendtner lab members may be useful. This work could benefit by implementing larger trial sizes in combination with unsupervised learning to better determine the minute differences between system/toxin results. More critically, my work was only able to begin diving into steps beyond the pre-polymerization phase of MIP production, modeling the potential interactions of a template within a structured MIP. However, this direction has plenty of potential and, again, little has been done computationally in this domain. Moving in this direction provides the opportunity to obtain a great deal of insight yet to be discovered regarding MIP construction, benefitting those working with such uremic toxin capture projects as well as other synthetic receptors.

Bibliography

(1) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603. https://doi.org/10.1103/PhysRevLett.100.020603.

(2) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nat. Rev. Phys.* **2020**, *2* (4), 200–212. https://doi.org/10.1038/s42254-020-0153-0.

(3) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1* (5), 826–843. https://doi.org/10.1002/wcms.31.

(4) Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (11), 5062–5067. https://doi.org/10.1021/acs.jctc.5b00846.

(5) Baumgart, T.; Hess, S. T.; Webb, W. W. Imaging Coexisting Fluid Domains in Biomembrane Models Coupling Curvature and Line Tension. *Nature* **2003**, *425* (6960), 821–824. https://doi.org/10.1038/nature02013.

(6) Veatch, S. L.; Keller, S. L. Separation of Liquid Phases in Giant Vesicles of Ternary Mixtures of Phospholipids and Cholesterol. *Biophys. J.* **2003**, *85* (5), 3074–3083. https://doi.org/10.1016/S0006-3495(03)74726-2.

(7) Zhao, J.; Wu, J.; Heberle, F. A.; Mills, T. T.; Klawitter, P.; Huang, G.; Costanza, G.; Feigenson, G. W. Phase Studies of Model Biomembranes: Complex Behavior of DSPC/DOPC/Cholesterol. *Biochim. Biophys. Acta - Biomembr.* **2007**, *1768* (11), 2764–2776. https://doi.org/https://doi.org/10.1016/j.bbamem.2007.07.008.

(8) Simons, K.; Ikonen, E. Functional Rafts in Cell Membranes. *Nature* **1997**, *387* (6633), 569–572. https://doi.org/10.1038/42408.

(9) Levental, I.; Veatch, S. The Continuing Mystery of Lipid Rafts. *J. Mol. Biol.* **2016**, *428* (24 Pt A), 4749–4764. https://doi.org/10.1016/j.jmb.2016.08.022.

(10) Cornell, C. E.; McCarthy, N. L. C.; Levental, K. R.; Levental, I.; Brooks, N. J.; Keller, S. L. N-Alcohol Length Governs Shift in Lo-Ld Mixing Temperatures in Synthetic and Cell-Derived Membranes. *Biophys. J.* **2017**, *113* (6), 1200–1211. https://doi.org/10.1016/j.bpj.2017.06.066.

(11) Machta, B. B.; Gray, E.; Nouri, M.; McCarthy, N. L. C.; Gray, E. M.; Miller, A. L.; Brooks, N. J.; Veatch, S. L. Conditions That Stabilize Membrane Domains Also Antagonize N-Alcohol Anesthesia. *Biophys. J.* **2016**, *111* (3), 537–545. https://doi.org/10.1016/j.bpj.2016.06.039.

(12) Schick, M. Shift in Membrane Miscibility Transition Temperature upon Addition of Short-Chain Alcohols. *Phys. Rev. E* **2016**, *94* (6), 1–6. https://doi.org/10.1103/PhysRevE.94.062114.

(13) Risselada, H. J.; Marrink, S. J. The Molecular Face of Lipid Rafts in Model Membranes. *Proc. Natl. Acad. Sci.* **2008**, *105* (45), 17367 LP – 17372. https://doi.org/10.1073/pnas.0807527105.

(14) Baoukina, S.; Tieleman, D. P. Simulations of Lipid Monolayers. *Methods Mol. Biol.* **2013**, *924*, 431–444. https://doi.org/10.1007/978-1-62703-017-5_16.

(15) Ackerman, D. G.; Feigenson, G. W. Lipid Bilayers: Clusters, Domains and Phases. *Essays Biochem.* **2015**, *57*, 33–42. https://doi.org/10.1042/bse0570033.

(16) He, S.; Maibaum, L. Identifying the Onset of Phase Separation in Quaternary Lipid

Bilayer Systems from Coarse-Grained Simulations. *J. Phys. Chem. B* **2018**, *122* (14), 3961–3973. https://doi.org/10.1021/acs.jpcb.8b00364.

(17) Hakobyan, D.; Heuer, A. Key Molecular Requirements for Raft Formation in Lipid/Cholesterol Membranes. *PLoS One* **2014**, *9* (2). https://doi.org/10.1371/journal.pone.0087369.

(18) Stansfeld, P. J.; Goose, J. E.; Caffrey, M.; Carpenter, E. P.; Parker, J. L.; Newstead, S.; Sansom, M. S. P. MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **2015**, *23* (7), 1350–1361. https://doi.org/10.1016/j.str.2015.05.006.

(19) Qi, Y.; Ingólfsson, H. I.; Cheng, X.; Lee, J.; Marrink, S. J.; Im, W. CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.* **2015**, *11* (9), 4486–4494. https://doi.org/10.1021/acs.jctc.5b00513.

(20) 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. *J. Chem. Theory Comput.* **2015**, *11* (5), 2144–2155. https://doi.org/10.1021/acs.jctc.5b00209.

(21) Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Böckmann, R. A. High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *J. Chem. Theory Comput.* **2015**, *11* (5), 2278–2291. https://doi.org/10.1021/ct5010092.

(22) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014**, *10* (2), 676–690. https://doi.org/10.1021/ct400617g.

(23) Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; De Vries, A. H.; Marrink, S. J. Pitfalls of the Martini Model. *J. Chem. Theory Comput.* **2019**, *15* (10), 5448–5460. https://doi.org/10.1021/acs.jctc.9b00473.

(24) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive Aggregation of Membrane Proteins in the Martini Model. *PLoS One* **2017**, *12* (11), e0187936.

(25) Kumar Basak, U.; Roobala, C.; Basu, J. K.; Maiti, P. K. Size-Dependent Interaction of Hydrophilic/Hydrophobic Ligand Functionalized Cationic and Anionic Nanoparticles with Lipid Bilayers. *J. Phys. Condens. Matter* **2020**, *32* (10). https://doi.org/10.1088/1361-648X/ab5770.

(26) Löpez, C. A.; Vesselinov, V. V.; Gnanakaran, S.; Alexandrov, B. S. Unsupervised Machine Learning for Analysis of Phase Separation in Ternary Lipid Mixture. *J. Chem. Theory Comput.* **2019**, *15* (11), 6343–6357. https://doi.org/10.1021/acs.jctc.9b00074.

(27) Khan, H. M.; Souza, P. C. T.; Thallmair, S.; Barnoud, J.; de Vries, A. H.; Marrink, S. J.; Reuter, N. Capturing Choline–Aromatics Cation−π Interactions in the MARTINI Force Field. *J. Chem. Theory Comput.* **2020**, *16* (4), 2550–2560. https://doi.org/10.1021/acs.jctc.9b01194.

(28) Thallmair, S.; Vainikka, P. A.; Marrink, S. J. Lipid Fingerprints and Cofactor Dynamics of Light-Harvesting Complex II in Different Membranes. *Biophys. J.* **2019**, *116* (8), 1446–1455. https://doi.org/10.1016/j.bpj.2019.03.009.

(29) Kondela, T.; Gallová, J.; Hauß, T.; Barnoud, J.; Marrink, S.-J.; Kučerka, N. Alcohol Interactions with Lipid Bilayers. *Molecules* **2017**, *22* (12), 2078. https://doi.org/10.3390/molecules22122078.

(30) Klacsová, M.; Bulacu, M.; Kučerka, N.; Uhríková, D.; Teixeira, J.; Marrink, S. J.; Balgavý, P. The Effect of Aliphatic Alcohols on Fluid Bilayers in Unilamellar DOPC Vesicles - A Small-Angle Neutron Scattering and Molecular Dynamics Study. *Biochim. Biophys. Acta - Biomembr.* **2011**, *1808* (9), 2136–2146. https://doi.org/10.1016/j.bbamem.2011.04.010.

(31) Bulacu, M.; Périole, X.; Marrink, S. J. In Silico Design of Robust Bolalipid Membranes. *Biomacromolecules* **2012**, *13* (1), 196–205. https://doi.org/10.1021/bm201454j.

(32) Melo, M. N.; Arnarez, C.; Sikkema, H.; Kumar, N.; Walko, M.; Berendsen, H. J. C.; Kocer, A.; Marrink, S. J.; Ingólfsson, H. I. High-Throughput Simulations Reveal Membrane-Mediated Effects of Alcohols on MscL Gating. *J. Am. Chem. Soc.* **2017**, *139* (7), 2664–2671. https://doi.org/10.1021/jacs.6b11091.

(33) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42* (16), 6801–6822. https://doi.org/10.1039/c3cs60093a.

(34) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(35) De Jong, D. H.; Singh, G.; Drew Bennett, W. F.; Arnarez, C.; Wassenaar, T. A.; Schä, L. V; Periole, X.; Peter, D.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. **2012**. https://doi.org/10.1021/ct300646g.

(36) 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. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824. https://doi.org/10.1021/jp071097f.

(37) Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108* (2), 750–760. https://doi.org/10.1021/jp036508g.

(38) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1), 43–56. https://doi.org/https://doi.org/10.1016/0010-4655(95)00042-E.

(39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1). https://doi.org/10.1063/1.2408420.

(40) Ackerman, D. G.; Feigenson, G. W. Multiscale Modeling of Four-Component Lipid Mixtures: Domain Composition, Size, Alignment, and Properties of the Phase Interface. *J. Phys. Chem. B* **2015**, *119* (11), 4240–4250. https://doi.org/10.1021/jp511083z.

(41) Rosetti, C.; Pastorino, C. Comparison of Ternary Bilayer Mixtures with Asymmetric or Symmetric Unsaturated Phosphatidylcholine Lipids by Coarse Grained Molecular Dynamics Simulations. *J. Phys. Chem. B* **2012**, *116* (11), 3525–3537. https://doi.org/10.1021/jp212406u.

(42) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(43) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

(44) Saeedimasine, M.; Montanino, A.; Kleiven, S.; Villa, A. Role of Lipid Composition on the Structural and Mechanical Features of Axonal Membranes: A Molecular Simulation Study. *Sci. Rep.* **2019**, *9* (1), 1–12. https://doi.org/10.1038/s41598-019-44318-9.

(45) Rowe, E. S.; Zhang, F.; Leung, T. W.; Parr, J. S.; Guy, P. T. Thermodynamics of Membrane Partitioning for a Series of N-Alcohols Determined by Titration Calorimetry: Role of Hydrophobic Effects. *Biochemistry* **1998**, *37* (8), 2430–2440. https://doi.org/10.1021/bi9721602.

(46) Kästner, J. Umbrella Sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (6), 932–942. https://doi.org/10.1002/wcms.66.

(47) Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28* (4), 578–581. https://doi.org/10.1016/0009-2614(74)80109-0.

(48) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199. https://doi.org/https://doi.org/10.1016/0021-9991(77)90121-8.

(49) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. https://doi.org/10.1002/jcc.540130812.

(50) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135* (1), 40–57. https://doi.org/10.1016/S0010-4655(00)00215-0.

(51) Hub, J. S.; de Groot, B. L.; van der Spoel, D. G_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6* (12), 3713–3720. https://doi.org/10.1021/ct100494z.

(52) Pokhrel, N.; Maibaum, L. Free Energy Calculations of Membrane Permeation: Challenges Due to Strong Headgroup-Solute Interactions. *J. Chem. Theory Comput.* **2018**, *14* (3), 1762–1771. https://doi.org/10.1021/acs.jctc.7b01159.

(53) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (39), 13749–13754. https://doi.org/10.1073/pnas.0506346102.

(54) Kobayashi, T.; Menon, A. K. Transbilayer Lipid Asymmetry. *Curr. Biol.* **2018**, *28* (8), R386–R391. https://doi.org/10.1016/j.cub.2018.01.007.

(55) Fadeel, B.; Xue, D. The Ins and Outs of Phospholipid Asymmetry in the Plasma Membrane: Roles in Health and Disease. *Crit. Rev. Biochem. Mol. Biol.* **2009**, *44* (5), 264–277. https://doi.org/10.1080/10409230903193307.

(56) Wimley, W. C.; Thompson, T. E. Exchange and Flip-Flop of Dimyristoylphosphatidylcholine in Liquid-Crystalline, Gel, and Two-Component, Two-Phase Large Unilamellar Vesicles. *Biochemistry* **1990**, *29* (5), 1296–1303. https://doi.org/10.1021/bi00457a027.

(57) Bennett, W. F. D.; MacCallum, J. L.; Hinner, M. J.; Marrink, S. J.; Tieleman, D. P. Molecular View of Cholesterol Flip-Flop and Chemical Potential in Different Membrane Environments. *J. Am. Chem. Soc.* **2009**, *131* (35), 12714–12720. https://doi.org/10.1021/ja903529f.

(58) Gu, R. X.; Baoukina, S.; Tieleman, D. P. Cholesterol Flip-Flop in Heterogeneous Membranes. *J. Chem. Theory Comput.* **2019**, *15* (3), 2064–2070. https://doi.org/10.1021/acs.jctc.8b00933.

(59) Buchete, N.-V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112* (19), 6057–6069. https://doi.org/10.1021/jp0761665.

(60) Buslaev, P.; Gushchin, I. Effects of Coarse Graining and Saturation of Hydrocarbon Chains on Structure and Dynamics of Simulated Lipid Molecules. *Sci. Rep.* **2017**, *7* (1), 1–15. https://doi.org/10.1038/s41598-017-11761-5.

(61) Polley, A. Partition of Common Anesthetic Molecules in the Liquid Disordered Phase Domain of a Composite Multicomponent Membrane. *Phys. Rev. E* **2018**, *98* (1), 12409. https://doi.org/10.1103/PhysRevE.98.012409.

(62) Shi, Y.; Zhang, Y.; Tian, H.; Wang, Y.; Shen, Y.; Zhu, Q.; Ding, F. Improved Dialytic Removal of Protein-Bound Uremic Toxins by Intravenous Lipid Emulsion in Chronic Kidney Disease Rats. *Nephrol. Dial. Transplant.* **2019**, *34* (11), 1842–1852. https://doi.org/10.1093/ndt/gfz079.

(63) Gryp, T.; Vanholder, R.; Vaneechoutte, M.; Glorieux, G. P-Cresyl Sulfate. *Toxins (Basel).* **2017**, *9* (2), 1–24. https://doi.org/10.3390/toxins9020052.

(64) Ackley, W.; Soiefer, L.; Etinger, A.; Lowenstein, J. Uremic Retention Solutes. *Asp. Dial.* **2018**.

(65) Leong, S. C.; Sirich, T. L. Indoxyl Sulfate-Review of Toxicity and Therapeutic Strategies. *Toxins (Basel).* **2016**, *8* (12). https://doi.org/10.3390/toxins8120358.

(66) Rocchetti, M. T.; Cosola, C.; di Bari, I.; Magnani, S.; Galleggiante, V.; Scandiffio, L.; Dalfino, G.; Netti, G. S.; Atti, M.; Corciulo, R.; et al. Efficacy of Divinylbenzenic Resin in Removing Indoxyl Sulfate and P-Cresol Sulfate in Hemodialysis Patients: Results from an in Vitro Study and an in Vivo Pilot Trial (Xuanro4-Nature 3.2). *Toxins (Basel).* **2020**, *12* (3). https://doi.org/10.3390/toxins12030170.

(67) Chiu, C. A.; Lu, L. F.; Yu, T. H.; Hung, W. C.; Chung, F. M.; Tsai, I. T.; Yang, C. Y.; Hsu, C. C.; Lu, Y. C.; Wang, C. P.; et al. Increased Levels of Total P-Cresylsulphate and Indoxyl Sulphate Are Associated with Coronary Artery Disease in Patients with Diabetic Nephropathy. *Rev. Diabet. Stud.* **2010**, *7* (4), 275–284. https://doi.org/10.1900/RDS.2010.7.275.

(68) Tao, X.; Thijssen, S.; Kotanko, P.; Ho, C. H.; Henrie, M.; Stroup, E.; Handelman, G. Improved Dialytic Removal of Protein-Bound Uraemic Toxins with Use of Albumin Binding Competitors: An in Vitro Human Whole Blood Study. *Sci. Rep.* **2016**, *6* (December 2015), 2–10. https://doi.org/10.1038/srep23389.

(69) Sandeman, S. R.; Howell, C. A.; Phillips, G. J.; Zheng, Y.; Standen, G.; Pletzenauer, R.; Davenport, A.; Basnayake, K.; Boyd, O.; Holt, S.; et al. An Adsorbent Monolith Device to Augment the Removal of Uraemic Toxins during Haemodialysis. *J. Mater. Sci. Mater. Med.* **2014**, *25* (6), 1589–1597. https://doi.org/10.1007/s10856-014-5173-9.

(70) Sandeman, S. R.; Zheng, Y.; Ingavle, G. C.; Howell, C. A.; Mikhalovsky, S. V.; Basnayake, K.; Boyd, O.; Davenport, A.; Beaton, N.; Davies, N. A Haemocompatible and Scalable Nanoporous Adsorbent Monolith Synthesised Using a Novel Lignin Binder Route to Augment the Adsorption of Poorly Removed Uraemic Toxins in Haemodialysis. *Biomed. Mater.* **2017**, *12* (3). https://doi.org/10.1088/1748-605X/aa6546.

(71) Smith, J.; Pfaendtner, J. Elucidating the Molecular Interactions between Uremic Toxins and the Sudlow II Binding Site of Human Serum Albumin. *J. Phys. Chem. B* **2020**, *124* (19), 3922–3930. https://doi.org/10.1021/acs.jpcb.0c02015.

(72) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S. Structural Basis of the Drug-Binding Specificity of Human Serum Albumin. *J. Mol. Biol.* **2005**, *353* (1), 38–52. https://doi.org/10.1016/j.jmb.2005.07.075.

(73) Sirich, T. L.; Meyer, T. W.; Gondouin, B.; Brunet, P.; Niwa, T. Protein-Bound Molecules:

A Large Family With a Bad Character. *Semin. Nephrol.* **2014**, *34* (2), 106–117. https://doi.org/10.1016/j.semnephrol.2014.02.004.

(74) Florens, N.; Yi, D.; Juillard, L.; Soulage, C. O. Using Binding Competitors of Albumin to Promote the Removal of Protein-Bound Uremic Toxins in Hemodialysis: Hope or Pipe Dream? *Biochimie* **2018**, *144*, 1–8. https://doi.org/10.1016/j.biochi.2017.09.018.

(75) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139* (1). https://doi.org/10.1063/1.4811489.

(76) Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014**, *10* (4), 1739–1752. https://doi.org/10.1021/ct4009156.

(77) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010**, *52* (1), 99–105. https://doi.org/10.1016/j.ymeth.2010.06.002.

(78) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144. https://doi.org/10.1016/j.sbi.2014.04.002.

(79) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (45), 19011–19016. https://doi.org/10.1073/pnas.0905466106.

(80) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-Level Description of Ubiquitin Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (15), 5915–5920. https://doi.org/10.1073/pnas.1218321110.

(81) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133* (24), 9181–9183. https://doi.org/10.1021/ja202726y.

(82) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (25), 10184–10189. https://doi.org/10.1073/pnas.1103547108.

(83) Lin, Y. S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S. Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys. J.* **2012**, *102* (2), 315–324. https://doi.org/10.1016/j.bpj.2011.12.002.

(84) Qiao, Q.; Bowman, G. R.; Huang, X. Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations That Potentially Seed Aggregation. *J. Am. Chem. Soc.* **2013**, *135* (43), 16092–16101. https://doi.org/10.1021/ja403147m.

(85) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation Pathway of Src Kinase Reveals Intermediate States as Targets for Drug Design. *Nat. Commun.* **2014**, *5* (1), 1–11. https://doi.org/10.1038/ncomms4397.

(86) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-Based Simulations on Google Exacycle Reveal Ligand Modulation of GPCR Activation Pathways. *Nat. Chem.* **2014**, *6* (1), 15–21. https://doi.org/10.1038/nchem.1821.

(87) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9* (4), 2000–2009. https://doi.org/10.1021/ct300878a.

(88) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-Target Residence Time and Its

Implications for Lead Optimization. *Nat. Rev. Drug Discov.* **2006**, *5* (9), 730–739. https://doi.org/10.1038/nrd2082.

(89) Copeland, R. A. The Drug-Target Residence Time Model: A 10-Year Retrospective. *Nat. Rev. Drug Discov.* **2016**, *15* (2), 87–95. https://doi.org/10.1038/nrd.2015.18.

(90) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(91) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79*, 926–935. https://doi.org/10.1063/1.445869.

(92) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(93) 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. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(94) Marques, S. M.; Bednar, D.; Damborsky, J. Computational Study of Protein-Ligand Unbinding for Enzyme Engineering. *Front. Chem.* **2019**, *7* (JAN), 1–15. https://doi.org/10.3389/fchem.2018.00650.

(95) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.

(96) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613. https://doi.org/10.1016/j.cpc.2013.09.018.

(97) Molgedey, L.; Schuster, H. G. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Lett.* **1994**, *72* (23), 3634–3637. https://doi.org/10.1103/PhysRevLett.72.3634.

(98) Kriegel, H. P.; Kröger, P.; Zimek, A. Clustering High-Dimensional Data: A Survey on Subspace Clustering, Pattern-Based Clustering, and Correlation Clustering. *ACM Trans. Knowl. Discov. Data* **2009**, *3* (1). https://doi.org/10.1145/1497577.1497578.

(99) Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. Complete Protein-Protein Association Kinetics in Atomic Detail Revealed by Molecular Dynamics Simulations and Markov Modelling. *Nat. Chem.* **2017**, *9* (10), 1005–1011. https://doi.org/10.1038/nchem.2785.

(100) Pinamonti, G.; Zhao, J.; Condon, D. E.; Paul, F.; Turner, H.; Bussi, G.; States, U.; Science, C. 40054843348_20180727_IIB.Pdf. **2018**, *13* (2), 926–934. https://doi.org/10.1021/acs.jctc.6b00982.Predicting.

(101) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schtte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17). https://doi.org/10.1063/1.3565032.

(102) Noé, F.; Fischer, S. Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 154–162. https://doi.org/10.1016/j.sbi.2008.01.008.

(103) McGibbon, R. T.; Pande, V. S. Variational Cross-Validation of Slow Dynamical Modes in

Molecular Kinetics. *J. Chem. Phys.* **2015**, *142* (12), 1–12. https://doi.org/10.1063/1.4916292.
(104) Harrigan, M. P.; Sultan, M. M.; Hernández, C. X.; Husic, B. E.; Eastman, P.; Schwantes, C. R.; Beauchamp, K. A.; McGibbon, R. T.; Pande, V. S. MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys. J.* **2017**, *112* (1), 10–15. https://doi.org/10.1016/j.bpj.2016.10.042.
(105) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
(106) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139* (1), 015102. https://doi.org/10.1063/1.4811489.
(107) Deuflhard, P.; Weber, M. Robust Perron Cluster Analysis in Conformation Dynamics. *Linear Algebra Appl.* **2005**, *398* (1–3), 161–184. https://doi.org/10.1016/j.laa.2004.10.026.
(108) Fährrolfes, R.; Bietz, S.; Flachsenberg, F.; Meyder, A.; Nittinger, E.; Otto, T.; Volkamer, A.; Rarey, M. ProteinsPlus: A Web Portal for Structure Analysis of Macromolecules. *Nucleic Acids Res.* **2017**, *45* (W1), W337–W343. https://doi.org/10.1093/nar/gkx333.
(109) Wang, Y.; Lamim Ribeiro, J. M.; Tiwary, P. Machine Learning Approaches for Analyzing and Enhancing Molecular Dynamics Simulations. *Current Opinion in Structural Biology*. Elsevier Ltd April 1, 2020, pp 139–145. https://doi.org/10.1016/j.sbi.2019.12.016.
(110) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K. Frequency Adaptive Metadynamics for the Calculation of Rare-Event Kinetics. *J. Chem. Phys.* **2018**, *149* (7), 1–15. https://doi.org/10.1063/1.5024679.
(111) Brotzakis, Z. F.; Limongelli, V.; Parrinello, M. Accelerating the Calculation of Protein-Ligand Binding Free Energy and Residence Times Using Dynamically Optimized Collective Variables. *J. Chem. Theory Comput.* **2019**, *15* (1), 743–750. https://doi.org/10.1021/acs.jctc.8b00934.
(112) Wang, Y.; Martins, J. M.; Lindorff-Larsen, K. Biomolecular Conformational Changes and Ligand Binding: From Kinetics to Thermodynamics. *Chem. Sci.* **2017**, *8* (9), 6466–6473. https://doi.org/10.1039/c7sc01627a.
(113) Vanholder, R.; Schepers, E.; Pletinck, A.; Nagler, E. V.; Glorieux, G. The Uremic Toxicity of Indoxyl Sulfate and P-Cresyl Sulfate: A Systematic Review. *J. Am. Soc. Nephrol.* **2014**, *25* (9), 1897–1907. https://doi.org/10.1681/ASN.2013101062.
(114) Ludwig, J.; Smith, J.; Pfaendtner, J. Analyzing the Long Time-Scale Dynamics of Uremic Toxins Bound to Sudlow Site II in Human Serum Albumin. *J. Phys. Chem. B* **2021**, acs.jpcb.1c00221. https://doi.org/10.1021/acs.jpcb.1c00221.
(115) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science (80-.).* **2011**, *334* (6055), 517–520. https://doi.org/10.1126/science.1208351.
(116) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein–Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci.* **2015**, *112* (5), E386–E391. https://doi.org/10.1073/pnas.1424461112.
(117) Paul, F.; Wehmeyer, C.; Abualrous, E. T.; Wu, H.; Crabtree, M. D.; Schöneberg, J.; Clarke, J.; Freund, C.; Weikl, T. R.; Noé, F. Protein-Peptide Association Kinetics beyond the Seconds Timescale from Atomistic Simulations. *Nat. Commun.* **2017**, *8* (1), 1–10. https://doi.org/10.1038/s41467-017-01163-6.
(118) Swinney, D. C. Biochemical Mechanisms of Drug Action: What Does It Take for

Success? *Nat. Rev. Drug Discov.* **2004**, *3* (9), 801–808. https://doi.org/10.1038/nrd1500.

(119) Lamim Ribeiro, J. M.; Provasi, D.; Filizola, M. A Combination of Machine Learning and Infrequent Metadynamics to Efficiently Predict Kinetic Rates, Transition States, and Molecular Determinants of Drug Dissociation from G Protein-Coupled Receptors. *J. Chem. Phys.* **2020**, *153* (12), 124105. https://doi.org/10.1063/5.0019100.

(120) Pollard, T. D. A Guide to Simple and Informative Binding Assays. *Molecular Biology of the Cell*. The American Society for Cell Biology December 1, 2010, pp 4061–4067. https://doi.org/10.1091/mbc.E10-08-0683.

(121) Tiwary, P.; Parinello, M. From Metadynamics to Dynamics.

(122) Smith, Z.; Pramanik, D.; Tsai, S.-T.; Tiwary, P. Multi-Dimensional Spectral Gap Optimization of Order Parameters (SGOOP) through Conditional Probability Factorization. *J. Chem. Phys.* **2018**, *149* (23), 234105. https://doi.org/10.1063/1.5064856.

(123) Tiwary, P. Molecular Determinants and Bottlenecks in the Dissociation Dynamics of Biotin–Streptavidin. *J. Phys. Chem. B* **2017**, *121* (48), 10841–10849. https://doi.org/10.1021/acs.jpcb.7b09510.

(124) Tiwary, P.; Berne, B. J. How Wet Should Be the Reaction Coordinate for Ligand Unbinding? *J. Chem. Phys.* **2016**, *145* (5), 054113. https://doi.org/10.1063/1.4959969.

(125) Smith, Z.; Ravindra, P.; Wang, Y.; Cooley, R.; Tiwary, P. Discovering Protein Conformational Flexibility through Artificial-Intelligence-Aided Molecular Dynamics. *J. Phys. Chem. B* **2020**, *124* (38), 8221–8229. https://doi.org/10.1021/acs.jpcb.0c03985.

(126) Ravindra, P.; Smith, Z.; Tiwary, P. Automatic Mutual Information Noise Omission (AMINO): Generating Order Parameters for Molecular Systems. **2019**.

(127) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted Autoencoded Variational Bayes for Enhanced Sampling (RAVE). *J. Chem. Phys.* **2018**, *149* (7), 072301. https://doi.org/10.1063/1.5025487.

(128) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; et al. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nature Methods*. Nature Publishing Group August 1, 2019, pp 670–673. https://doi.org/10.1038/s41592-019-0506-8.

(129) Wang, Y.; Tiwary, P. Understanding the Role of Predictive Time Delay and Biased Propagator in RAVE. *J. Chem. Phys.* **2020**, *152* (14). https://doi.org/10.1063/5.0004838.

(130) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Past–Future Information Bottleneck for Sampling Molecular Reaction Coordinate Simultaneously with Thermodynamics and Kinetics. *Nat. Commun.* **2019**, *10* (1), 1–8. https://doi.org/10.1038/s41467-019-11405-4.

(131) Still, S. Information Bottleneck Approach to Predictive Inference. *Entropy* **2014**, *16* (2), 968–989. https://doi.org/10.3390/e16020968.

(132) Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.* **2015**.

(133) Chollet, F.; et al. Keras https://github.com/keras-team/keras (2015).

(134) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111* (23), 1–5. https://doi.org/10.1103/PhysRevLett.111.230602.

(135) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K. Frequency Adaptive Metadynamics for the Calculation of Rare-Event Kinetics. *J. Chem. Phys.* **2018**, *149* (7), 072309. https://doi.org/10.1063/1.5024679.

(136) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding

Free-Energy Method. *Proc. Natl. Acad. Sci.* **2013**, *110* (16), 6358–6363. https://doi.org/10.1073/pnas.1303186110.
(137) Troussicot, L.; Guillière, F.; Limongelli, V.; Walker, O.; Lancelin, J. M. Funnel-Metadynamics and Solution NMR to Estimate Protein-Ligand Affinities. *J. Am. Chem. Soc.* **2015**, *137* (3), 1273–1281. https://doi.org/10.1021/ja511336z.
(138) Capelli, R.; Bochicchio, A.; Piccini, G.; Casasnovas, R.; Carloni, P.; Parrinello, M. Chasing the Full Free Energy Landscape of Neuroreceptor/Ligand Unbinding by Metadynamics Simulations. *J. Chem. Theory Comput.* **2019**, *15* (5), 3354–3361. https://doi.org/10.1021/acs.jctc.9b00118.
(139) Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the Reliability of the Dynamics Reconstructed from Metadynamics. *J. Chem. Theory Comput.* **2014**, *10* (4), 1420–1425. https://doi.org/10.1021/ct500040r.
(140) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a P38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139* (13), 4780–4788. https://doi.org/10.1021/jacs.6b12950.
(141) Pramanik, D.; Smith, Z.; Kells, A.; Tiwary, P. Can One Trust Kinetic and Thermodynamic Observables from Biased Metadynamics Simulations: Detailed Quantitative Benchmarks on Millimolar Drug Fragment Dissociation. *bioRxiv*. bioRxiv February 25, 2019, p 558601. https://doi.org/10.1101/558601.
(142) Miller, L. H. Table of Percentage Points of Kolmogorov Statistics. *J. Am. Stat. Assoc.* **1956**, *51* (273), 111–121. https://doi.org/10.1080/01621459.1956.10501314.
(143) Massey, F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **1951**, *46* (253), 68. https://doi.org/10.2307/2280095.
(144) Sultan, A. M.; Westcott, Z. C.; Hughes, Z. E.; Palafox-Hernandez, J. P.; Giesa, T.; Puddu, V.; Buehler, M. J.; Perry, C. C.; Walsh, T. R. Aqueous Peptide-TiO2 Interfaces: Isoenergetic Binding via Either Entropically or Enthalpically Driven Mechanisms. *ACS Appl. Mater. Interfaces* **2016**, *8* (28), 18620–18630. https://doi.org/10.1021/acsami.6b05200.
(145) Alamdari, S.; Pfaendtner, J. Impact of Glutamate Carboxylation in the Adsorption of the α-1 Domain of Osteocalcin to Hydroxyapatite and Titania. *Mol. Syst. Des. Eng.* **2020**, *5* (3), 620–631. https://doi.org/10.1039/c9me00158a.
(146) Moschen, T.; Grutsch, S.; Juen, M. A.; Wunderlich, C. H.; Kreutz, C.; Tollinger, M. Measurement of Ligand-Target Residence Times by 1H Relaxation Dispersion NMR Spectroscopy. *J. Med. Chem.* **2016**, *59* (23), 10788–10793. https://doi.org/10.1021/acs.jmedchem.6b01110.
(147) Deltombe, O.; de Loor, H.; Glorieux, G.; Dhondt, A.; Van Biesen, W.; Meijers, B.; Eloot, S. Exploring Binding Characteristics and the Related Competition of Different Protein-Bound Uremic Toxins. *Biochimie* **2017**, *139*, 20–26. https://doi.org/10.1016/j.biochi.2017.05.010.
(148) Cheung, A. K. Biocompatibility of Hemodialysis Membranes. *J. Am. Soc. Nephrol.* **1990**, *1* (2), 150–161.
(149) Ma, Y.; Li, S.; Tonelli, M.; Unsworth, L. D. Adsorption-Based Strategies for Removing Uremic Toxins from Blood. *Microporous Mesoporous Mater.* **2021**, *319*, 111035. https://doi.org/10.1016/j.micromeso.2021.111035.
(150) Selvolini, G.; Marrazza, G. MIP-Based Sensors: Promising New Tools for Cancer Biomarker Determination. *Sensors (Switzerland)*. MDPI AG April 1, 2017.

https://doi.org/10.3390/s17040718.
(151) Tsai, H. A.; Syu, M. J. Synthesis of Creatinine-Imprinted Poly(β-Cyclodextrin) for the Specific Binding of Creatinine. *Biomaterials* **2005**, *26* (15), 2759–2766. https://doi.org/10.1016/j.biomaterials.2004.07.037.
(152) Baggiani, C.; Anfossi, L.; Giovannoli, C. Molecular Imprinted Polymers as Synthetic Receptors for the Analysis of Myco- and Phyco-Toxins. *Analyst*. Royal Society of Chemistry 2008, pp 719–730. https://doi.org/10.1039/b711352h.
(153) Baggiani, C.; Giovannoli, C.; Anfossi, L. Man-Made Synthetic Receptors for Capture and Analysis of Ochratoxin A. *Toxins (Basel).* **2015**, *7* (10), 4083–4098. https://doi.org/10.3390/toxins7104083.
(154) Tsai, H. A.; Syu, M. J. Preparation of Imprinted Poly(Tetraethoxysilanol) Sol-Gel for the Specific Uptake of Creatinine. *Chem. Eng. J.* **2011**, *168* (3), 1369–1376. https://doi.org/10.1016/j.cej.2011.02.047.
(155) Golker, K.; Karlsson, B. C. G.; Olsson, G. D.; Rosengren, A. M.; Nicholls, I. A. Influence of Composition and Morphology on Template Recognition in Molecularly Imprinted Polymers. *Macromolecules* **2013**, *46* (4), 1408–1414. https://doi.org/10.1021/ma3024238.
(156) Olsson, G. D.; Karlsson, B. C. G.; Shoravi, S.; Wiklander, J. G.; Nicholls, I. A. Mechanisms Underlying Molecularly Imprinted Polymer Molecular Memory and the Role of Crosslinker: Resolving Debate on the Nature of Template Recognition in Phenylalanine Anilide Imprinted Polymers. *J. Mol. Recognit.* **2012**, *25* (2), 69–73. https://doi.org/10.1002/jmr.2147.
(157) Kong, Y.; Wang, N.; Ni, X.; Yu, Q.; Liu, H.; Huang, W.; Xu, W. Molecular Dynamics Simulations of Molecularly Imprinted Polymer Approaches to the Preparation of Selective Materials to Remove Norfloxacin. *J. Appl. Polym. Sci.* **2016**, *133* (1), 1–11. https://doi.org/10.1002/app.42817.
(158) Cleland, D.; Olsson, G. D.; Karlsson, B. C. G.; Nicholls, I. A.; McCluskey, A. Molecular Dynamics Approaches to the Design and Synthesis of PCB Targeting Molecularly Imprinted Polymers: Interference to Monomer-Template Interactions in Imprinting of 1,2,3-Trichlorobenzene. *Org. Biomol. Chem.* **2014**, *12* (5), 844–853. https://doi.org/10.1039/c3ob42399a.
(159) Golker, K.; Nicholls, I. A. The Effect of Crosslinking Density on Molecularly Imprinted Polymer Morphology and Recognition. *Eur. Polym. J.* **2016**, *75*, 423–430. https://doi.org/10.1016/j.eurpolymj.2016.01.008.
(160) Karlsson, B. C. G.; O'Mahony, J.; Karlsson, J. G.; Bengtsson, H.; Eriksson, L. A.; Nicholls, I. A. Structure and Dynamics of Monomer-Template Complexation: An Explanation for Molecularly Imprinted Polymer Recognition Site Heterogeneity. *J. Am. Chem. Soc.* **2009**, *131* (37), 13297–13304. https://doi.org/10.1021/ja902087t.
(161) Andersson, H. S.; Karlsson, J. G.; Piletsky, S. A.; Koch-Schmidt, A. C.; Mosbach, K.; Nicholls, I. A. Study of the Nature of Recognition in Molecularly Imprinted Polymers, II [1]: Influence of Monomer-Template Ratio and Sample Load on Retention and Selectivity. *J. Chromatogr. A* **1999**, *848* (1–2), 39–49. https://doi.org/10.1016/S0021-9673(99)00483-5.
(162) Nicholls, I. A.; Andersson, H. S.; Golker, K.; Henschel, H.; Karlsson, B. C. G.; Olsson, G. D.; Rosengren, A. M.; Shoravi, S.; Suriyanarayanan, S.; Wiklander, J. G.; et al. Rational Design of Biomimetic Molecularly Imprinted Materials: Theoretical and Computational Strategies for Guiding Nanoscale Structured Polymer Development. *Analytical and*

Bioanalytical Chemistry. Springer Verlag April 9, 2011, pp 1771–1786. https://doi.org/10.1007/s00216-011-4935-1.

(163) Baggiani, C.; Giovannoli, C.; Anfossi, L.; Passini, C.; Baravalle, P.; Giraudi, G. A Connection between the Binding Properties of Imprinted and Nonimprinted Polymers: A Change of Perspective in Molecular Imprinting. *J. Am. Chem. Soc.* **2012**, *134* (3), 1513–1518. https://doi.org/10.1021/ja205632t.

(164) Viveiros, R.; Rebocho, S.; Casimiro, T. Green Strategies for Molecularly Imprinted Polymer Development. *Polymers (Basel).* **2018**, *10* (3), 306. https://doi.org/10.3390/polym10030306.

(165) Busato, M.; Distefano, R.; Bates, F.; Karim, K.; Bossi, A. M.; López Vilariño, J. M.; Piletsky, S.; Bombieri, N.; Giorgetti, A. MIRATE: MIps RATional DEsign Science Gateway. *J. Integr. Bioinform.* **2018**, *15* (4), 1–6. https://doi.org/10.1515/jib-2017-0075.

(166) Da Silva, M. S.; Viveiros, R.; Morgado, P. I.; Aguiar-Ricardo, A.; Correia, I. J.; Casimiro, T. Development of 2-(Dimethylamino)Ethyl Methacrylate-Based Molecular Recognition Devices for Controlled Drug Delivery Using Supercritical Fluid Technology. *Int. J. Pharm.* **2011**, *416* (1), 61–68. https://doi.org/10.1016/j.ijpharm.2011.06.004.

(167) Oral, E.; Peppas, N. A. Hydrophilic Molecularly Imprinted Poly(Hydroxyethyl-Methacrylate) Polymers. *Journal of Biomedical Materials Research - Part A*. J Biomed Mater Res A July 2006, pp 205–210. https://doi.org/10.1002/jbm.a.30725.

(168) Hsu, C. Y.; Lin, H. Y.; Thomas, J. L.; Wu, B. T.; Chou, T. C. Incorporation of Styrene Enhances Recognition of Ribonuclease A by Molecularly Imprinted Polymers. *Biosens. Bioelectron.* **2006**, *22* (3), 355–363. https://doi.org/10.1016/j.bios.2006.05.008.

(169) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280. https://doi.org/10.1021/j100142a004.

(170) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian 09. **2010**.

(171) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913.

(172) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A Complete Basis Set Model Chemistry. I. The Total Energies of Closed-Shell Atoms and Hydrides of the First-Row Elements. *J. Chem. Phys.* **1988**, *89* (4), 2193–2218. https://doi.org/10.1063/1.455064.

(173) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; et al. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nature Methods*. Nature Publishing Group August 1, 2019, pp 670–673. https://doi.org/10.1038/s41592-019-0506-8.

(174) Ran, J.; Hobza, P. On the Nature of Bonding in Lone Pair···π-Electron Complexes: CCSD(T)/Complete Basis Set Limit Calculations. *J. Chem. Theory Comput.* **2009**, *5* (4), 1180–1185. https://doi.org/10.1021/ct900036y.

(175) Lee, E. C.; Hong, B. H.; Lee, J. Y.; Kim, J. C.; Kim, D.; Kim, Y.; Tarakeshwar, P.; Kim, K. S. Substituent Effects on the Edge-to-Face Aromatic Interactions. *J. Am. Chem. Soc.* **2005**, *127* (12), 4530–4537. https://doi.org/10.1021/ja037454r.

(176) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. Estimates of the Ab Initio Limit for π-π Interactions: The Benzene Dimer. *J. Am. Chem. Soc.* **2002**, *124* (36), 10887–10893.

https://doi.org/10.1021/ja025896h.

(177) Golker, K.; Karlsson, B. C. G.; Wiklander, J. G.; Rosengren, A. M.; Nicholls, I. A. Hydrogen Bond Diversity in the Pre-Polymerization Stage Contributes to Morphology and MIP-Template Recognition - MAA versus MMA. *Eur. Polym. J.* **2015**, *66*, 558–568. https://doi.org/10.1016/j.eurpolymj.2015.03.018.

(178) Zink, S.; Moura, F. A.; Autreto, P. A. D. S.; Galvão, D. S.; Mizaikoff, B. Virtually Imprinted Polymers (VIPs): Understanding Molecularly Templated Materials: Via Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2018**, *20* (19), 13145–13152. https://doi.org/10.1039/c7cp08284c.

(179) Zink, S.; Moura, F. A.; Autreto, P. A. D. S.; Galvão, D. S.; Mizaikoff, B. Efficient Prediction of Suitable Functional Monomers for Molecular Imprinting: Via Local Density of States Calculations. *Phys. Chem. Chem. Phys.* **2018**, *20* (19), 13153–13158. https://doi.org/10.1039/c7cp08283e.

(180) Rayermann, S. P.; Rayermann, G. E.; Cornell, C. E.; Merz, A. J.; Keller, S. L. Hallmarks of Reversible Separation of Living, Unperturbed Cell Membranes into Two Liquid Phases. *Biophys. J.* **2017**, *113* (11), 2425–2432. https://doi.org/10.1016/j.bpj.2017.09.029.

(181) Pavel, M. A.; Petersen, E. N.; Wang, H.; Lerner, R. A.; Hansen, S. B. Studies on the Mechanism of General Anesthesia. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (24), 13757–13766. https://doi.org/10.1073/pnas.2004259117.

(182) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; et al. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* **2021**, *18* (4), 382–388. https://doi.org/10.1038/s41592-021-01098-3.

Appendix I – Supplemental materials for Chapter 3

Figures

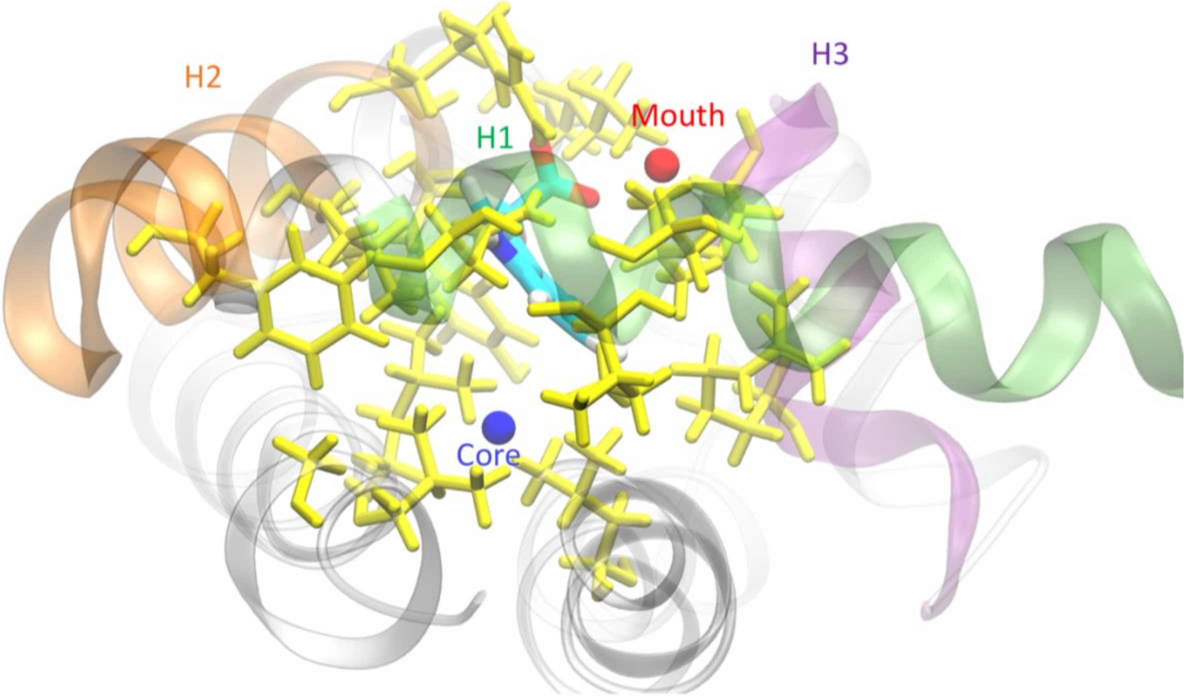


Figure I.1 - IAA-HSA system with the 14 key residues considered after filtering by atomic contacts. Key residues are colored in yellow. Indole-3-acetic acid is colored by atom type with hydrogen, carbon, oxygen, and nitrogen atoms colored white, cyan, red, and blue, respectively. The hydrophobic core and the more hydrophilic mouth of the pocket are labeled with blue and red spheres respectively. Finally, the three helices used in our set of collective variables are labeled and color coated.

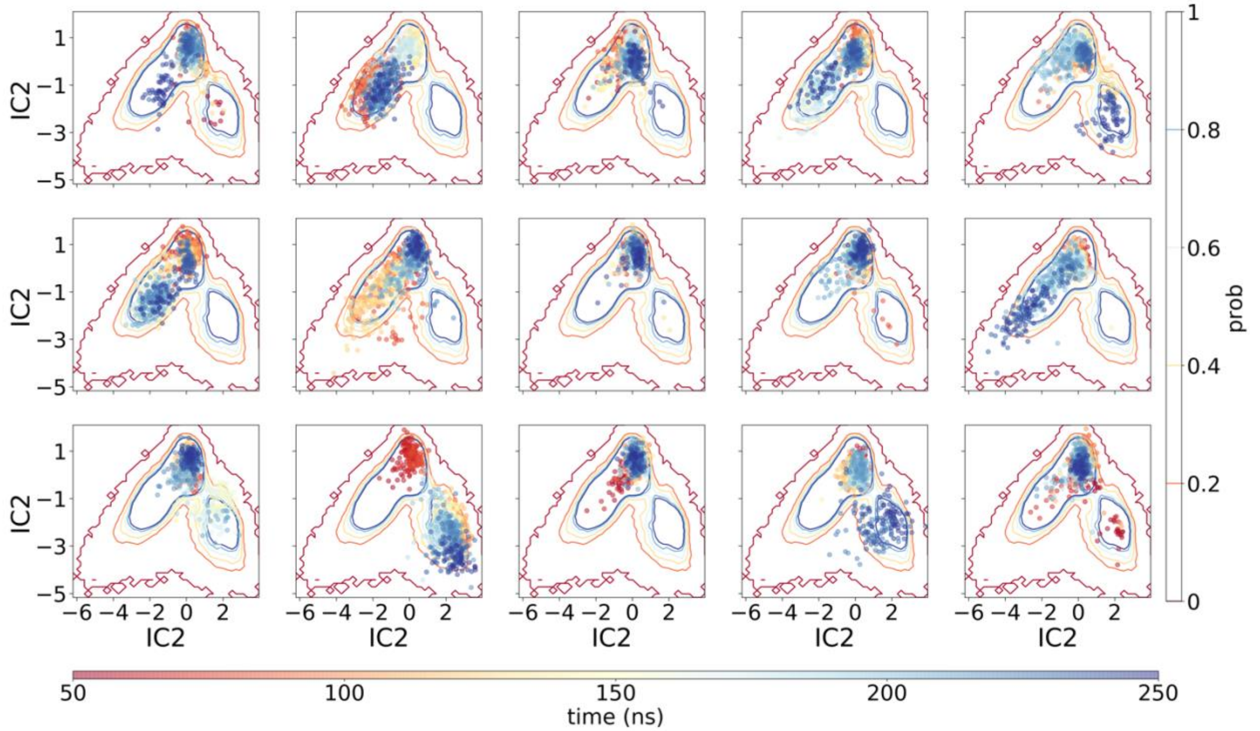


Figure I.2 - tICA data for a single toxin (PCS-HSA) system. The tICA subspace data is colored by time over each 250 ns simulation. The underlying contour plots represent the probability of the system being found in a given region of the tICA space.

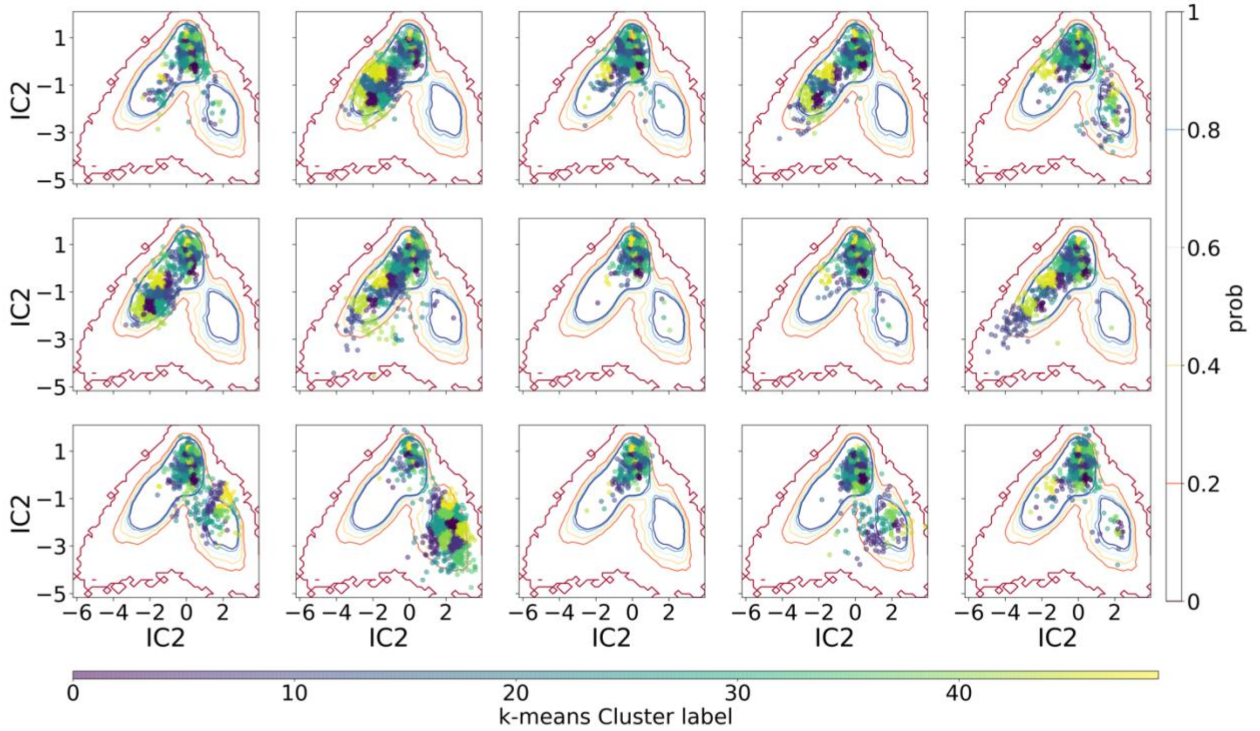


Figure I.3 - K-means clustering of the tICA results for a single toxin (PCS-HSA) system. The tICA subspace data is colored by the given K-means cluster (1-50 in the PCS system). The underlying contour plots represent the probability of the system being found in a given region of the tICA space.

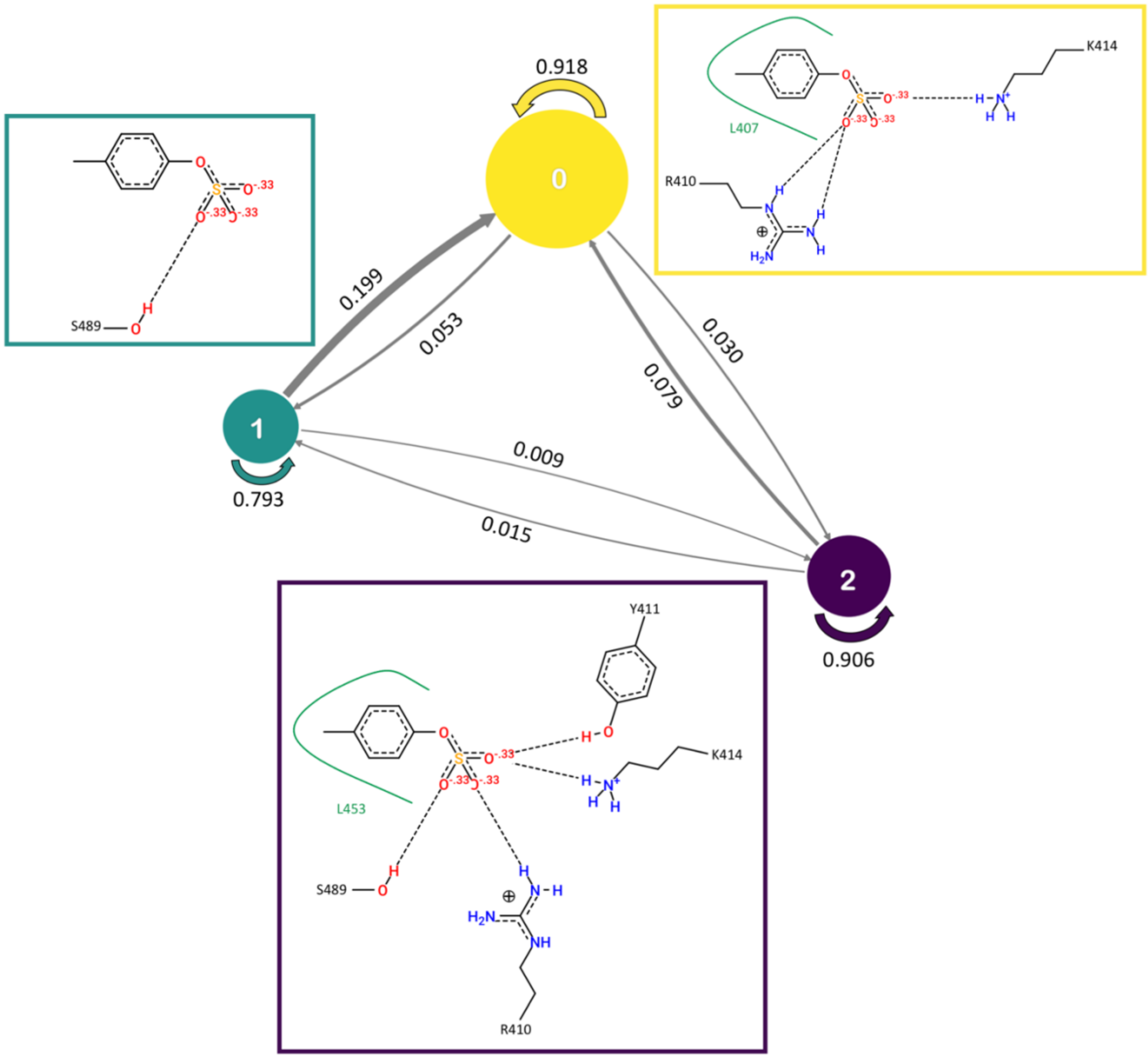


Figure I.4 - Pose View images of frame representative of each PCS-HSA binding mode accompanied by transition probabilities. Each Pose View image is boxed by the binding modes particular color. Hydrogen bonds are represented by dashed black lines, hydrophilic interactions are in green, and polar atoms are either red (oxygen) or blue (nitrogen). The transition probabilities are shown for transitions between modes with proportionally sized arrows. The size of each colored sphere is proportional to the binding modes total fraction of MD frames. It's worth reminding the reader that the images are not necessarily characteristic of the averaged structure, but rather a single snapshot ideally estimating certain interactions.

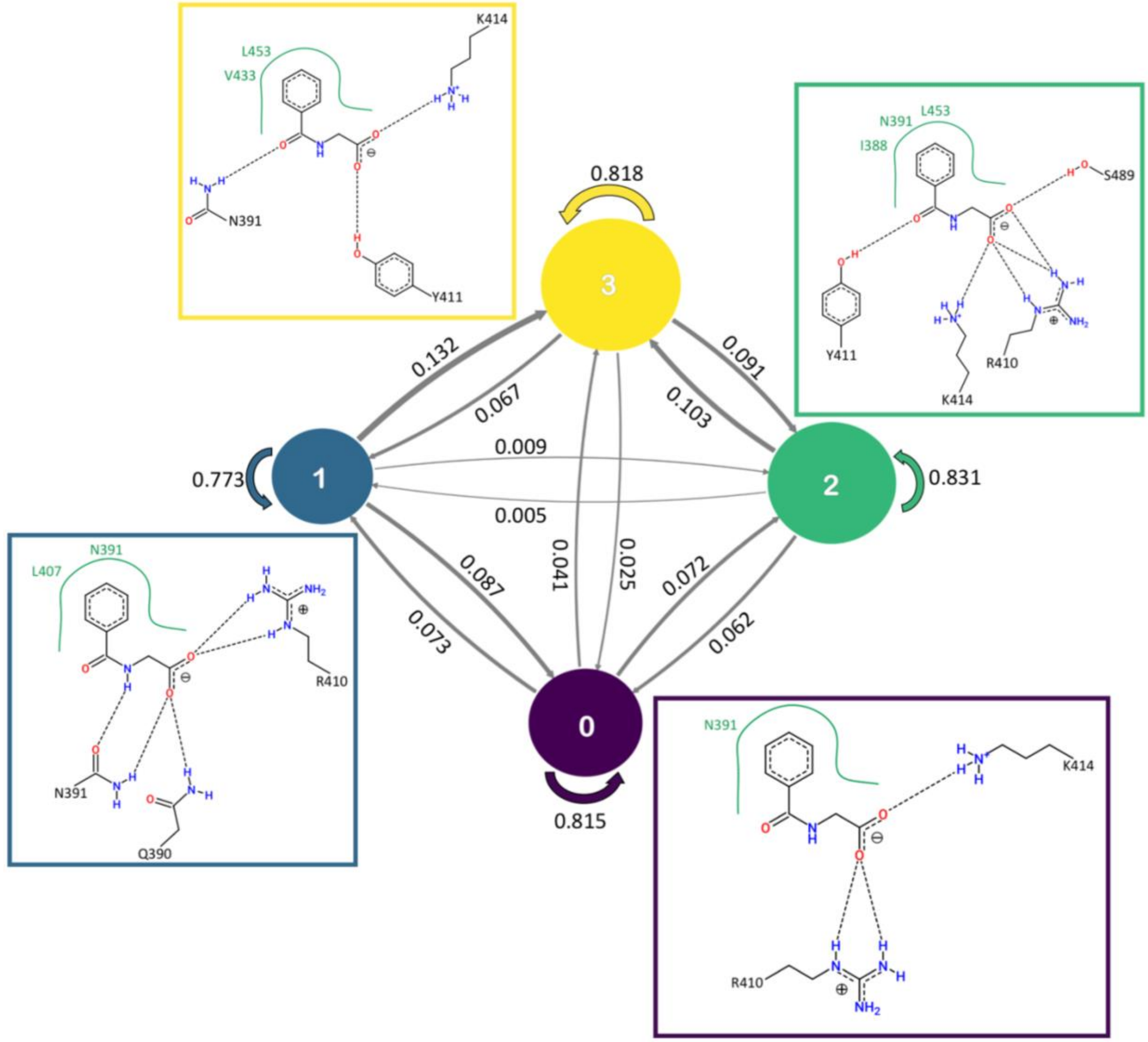


Figure I.5 - Pose View images of frame representative of each HIP-HSA binding mode accompanied by transition probabilities. Each Pose View image is boxed by the binding modes particular color. Hydrogen bonds are represented by dashed black lines, hydrophilic interactions are in green, and polar atoms are either red (oxygen) or blue (nitrogen). The transition probabilities are shown for transitions between modes with proportionally sized arrows. The size of each colored sphere is proportional to the binding modes total fraction of MD frames. It’s worth reminding the reader that the images are not necessarily characteristic of the averaged structure, but rather a single snapshot ideally estimating certain interactions.

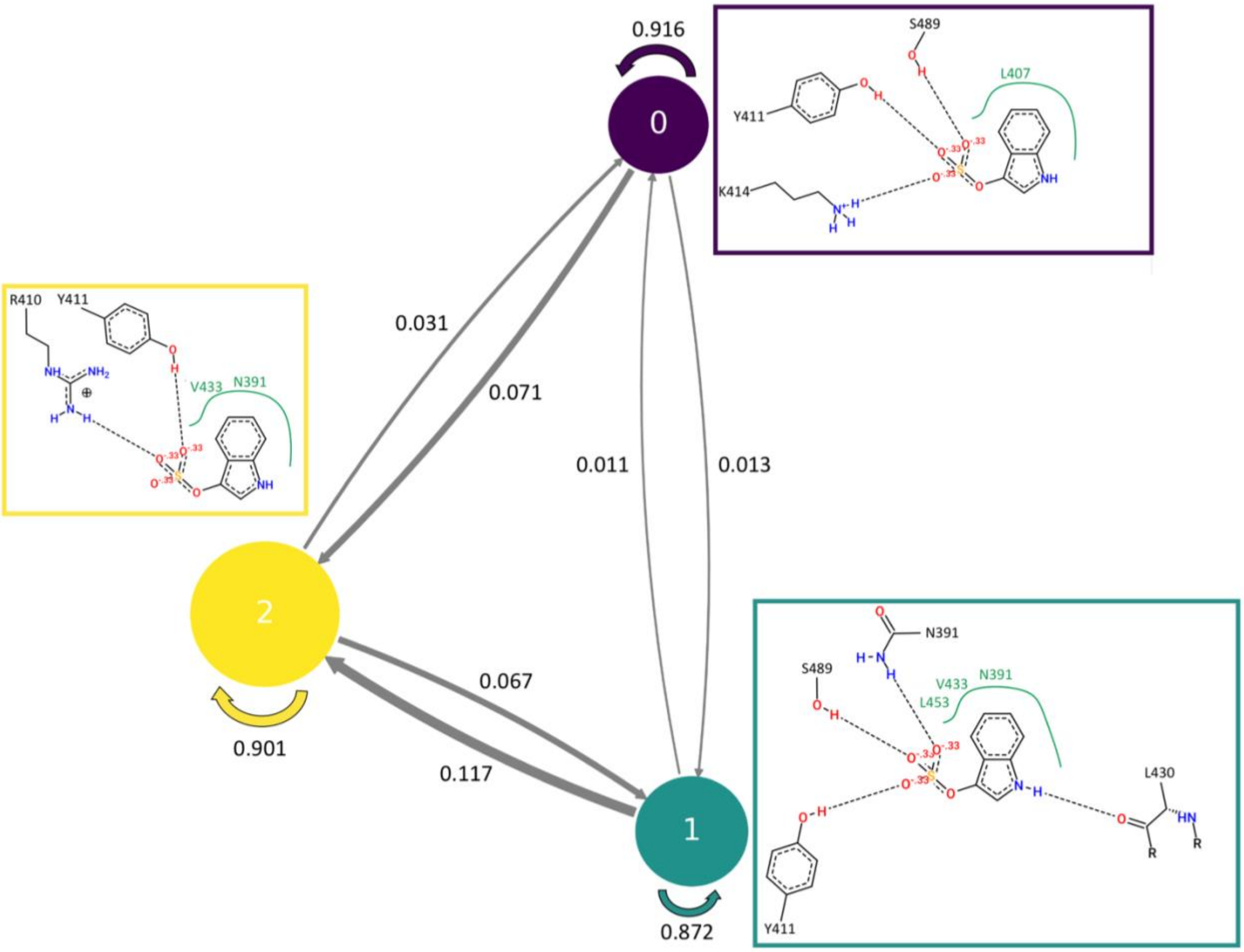


Figure I.6 - Pose View images of frame representative of each ISO-HSA binding mode accompanied by transition probabilities. Each Pose View image is boxed by the binding modes particular color. Hydrogen bonds are represented by dashed black lines, hydrophilic interactions are in green, and polar atoms are either red (oxygen) or blue (nitrogen). The transition probabilities are shown for transitions between modes with proportionally sized arrows. The size of each colored sphere is proportional to the binding modes total fraction of MD frames. It’s worth reminding the reader that the images are not necessarily characteristic of the averaged structure, but rather a single snapshot ideally estimating certain interactions.

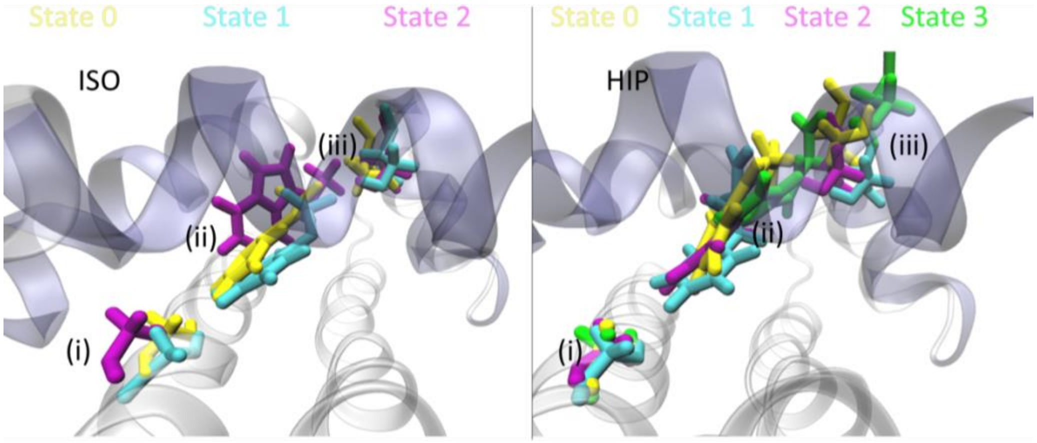


Figure I.7 - VMD images for frame representative of each binding mode. A set of three MD frames are visualized using VMD and color coded according to the particular binding mode for two toxin systems; ISO and HIP in the left and right panels respectively. Each frame contains a set of molecules to help orient the binding pocket: (i) Residue G434 deep in the binding pocket (ii) toxin position (iii) S489 at the mouth of the binding pocket. The three helices that make up the mouth of Sudlow site II are colored in light blue to further orient the pocket.

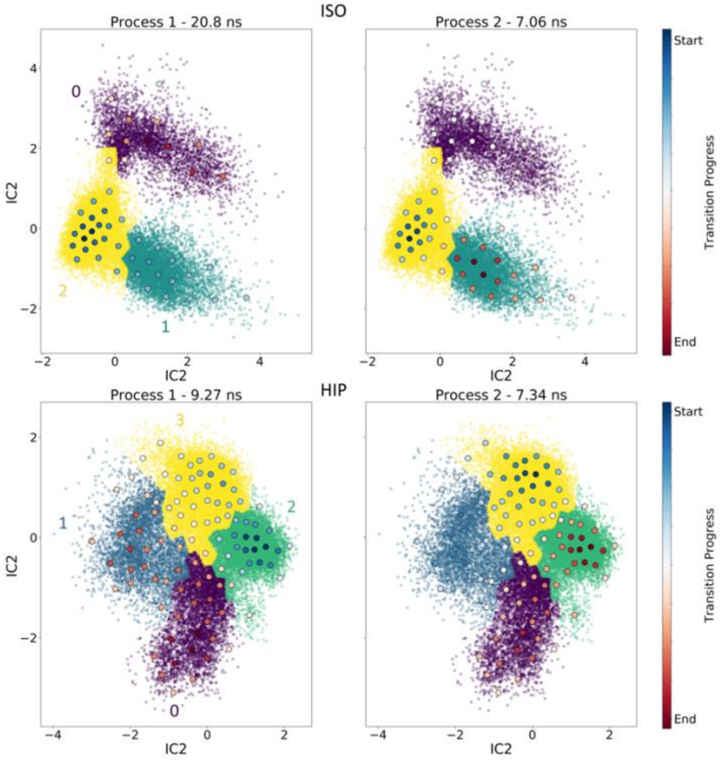


Figure I.8 - tIC plots of the coarse grained MSM overlaid by the projection of each eigenvector. Each of the 4 tIC plots details the results of the tICA of our 15 MD simulations after being first treated with k-means clustering and then coarse grained using PCCA+ clustering. Toxins are varied between ISO and HIP in upper and lower respectively. Each pair of plots for toxin differ in the MSM eigenvector projection overlaid atop which details the progression of the particular timescale discovered by our MSM.

Tables

Table I.1 - Coefficients of the linear transformation to the first 2 tICA components calculated for each toxin-HSA complex.

indoxyl sulfate			p-cresyl sulfate			indole-3-acetic acid			hippurate		
CVs	**tIC 1**	**tIC 2**	**CVs**	**tIC 1**	**tIC 2**	**CVs**	**tIC 1**	**tIC 2**	**CVs**	**tIC 1**	**tIC 2**
L387	0.05	0.11	**L387**	0.01	-0.06	**L387**	-0.03	0.04	**L387**	0.02	0.01
N391	0.14	0.18	**I388**	-0.04	-0.18	**I388**	-0.08	-0.06	**I388**	-0.27	-0.27
R410	0.98	-0.43	**N391**	-0.10	-0.06	**N391**	-0.17	0.03	**N391**	0.11	-0.18
Y411	0.01	0.23	**L407**	-0.05	0.15	**F403**	-0.14	0.06	**L407**	-0.07	0.03
K414	-0.27	-0.26	**R410**	-0.83	-0.50	**L407**	0.02	0.06	**R410**	-0.44	0.17
L430	0.04	0.16	**Y411**	0.38	-0.63	**R410**	-0.05	-1.14	**Y411**	-0.16	-0.45
V433	0.13	0.15	**K414**	-0.12	0.05	**Y411**	-0.14	0.15	**K414**	0.13	0.37
L453	0.1	0.10	**L430**	0.14	-0.11	**K414**	0.21	0.36	**L430**	0.18	-0.01
S489	-0.12	-0.12	**L453**	-0.11	0.01	**L430**	-0.07	-0.01	**V433**	-0.08	0.08
H1-H2	-0.02	-0.11	**R485**	0.06	-0.26	**V433**	0.01	0.01	**L453**	-0.17	-0.13
H1-H3	0.09	0.06	**F488**	-0.19	0.31	**G434**	-0.47	-0.07	**R485**	-0.26	-0.24
H2-H3	0.16	0.11	**S489**	-0.02	-0.05	**L453**	-0.06	-0.02	**F488**	-0.26	0.25
			H1-H2	-0.11	0.23	**R485**	-0.13	0.01	**S489**	0.01	0.08
			H1-H3	0.10	-0.18	**S489**	0.13	0.07	**H1-H2**	0.02	0.14
			H2-H3	0.04	-0.25	**H1-H2**	0.02	-0.10	**H1-H3**	-0.04	-0.10
						H1-H3	0.01	-0.11	**H2-H3**	-0.13	-0.11
						H2-H3	0.01	-0.14			

Table I.2 - Hyperparameters determined by GMRQ scoring each HSA-toxin system.

	indoxyl sulfate	**p-cresyl sulfate**	**indole-3-acetic acid**	**hippurate**
Number of tICA components	2	2	2	2
tICA lag time (ns)	0.005	0.02	0.002	0.01
Number of K-means clusters	50	50	100	100

Table I.3 - Top six correlations between order parameters and projection on the eigenvector modes of the ISO MSM.

1st Eigenvector		**2nd Eigenvector**		**3rd Eigenvector**	
Residue (ranked)	r	Residue (ranked)	r	Residue (ranked)	R
Eig3	0.56	R410	(-)0.84	Eig1	0.56
K414	0.53	Y411	(-)0.34	K414	0.41
V433	(-)0.52	N391	0.30	V433	(-)0.38
N391	(-)0.50	V433	0.29	L430	(-)0.36
L430	(-)0.46	Eig4	0.27	L453	(-)0.34
S489	0.43	L430	0.24	Y411	0.30

Table I.4 - Top six correlations between order parameters and projection on the eigenvector modes of the HIP MSM.

1st Eigenvector		**2nd Eigenvector**		**3rd Eigenvector**		**4th Eigenvector**	
Residue (ranked)	r	Residue (ranked)	r	Residue (ranked)	r	Residue (ranked)	r
R485	(-)0.59	F488	0.81	N391	0.26	Eig1	0.24
H2-H3	(-)0.49	N391	(-)0.81	I388	0.24	Y411	(-)0.21
L453	(-)0.46	K414	0.74	K414	(-)0.24	H2-H3	(-)0.18
H1-H3	(-)0.34	R410	0.72	F484	(-)0.22	L430	0.16
Y411	(-)0.33	I388	(-)0.64	Eig2	(-)0.22	L453	0.13
I388	(-)0.32	S489	0.63	R410	(-)0.19	V433	0.12

Appendix II – Supplemental materials for Chapter 4

Figures

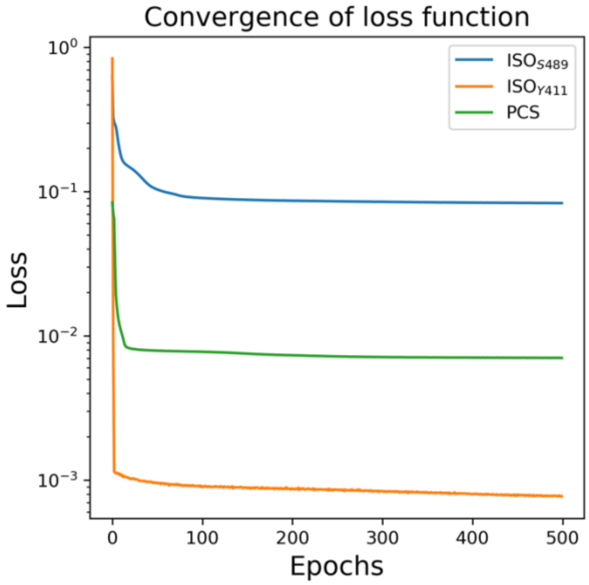


Figure II.1 – Convergence of loss function during neural network training. Each profile represents the loss function for the three systems studied, ISO_{S489}, ISO_{Y411}, and PCS.

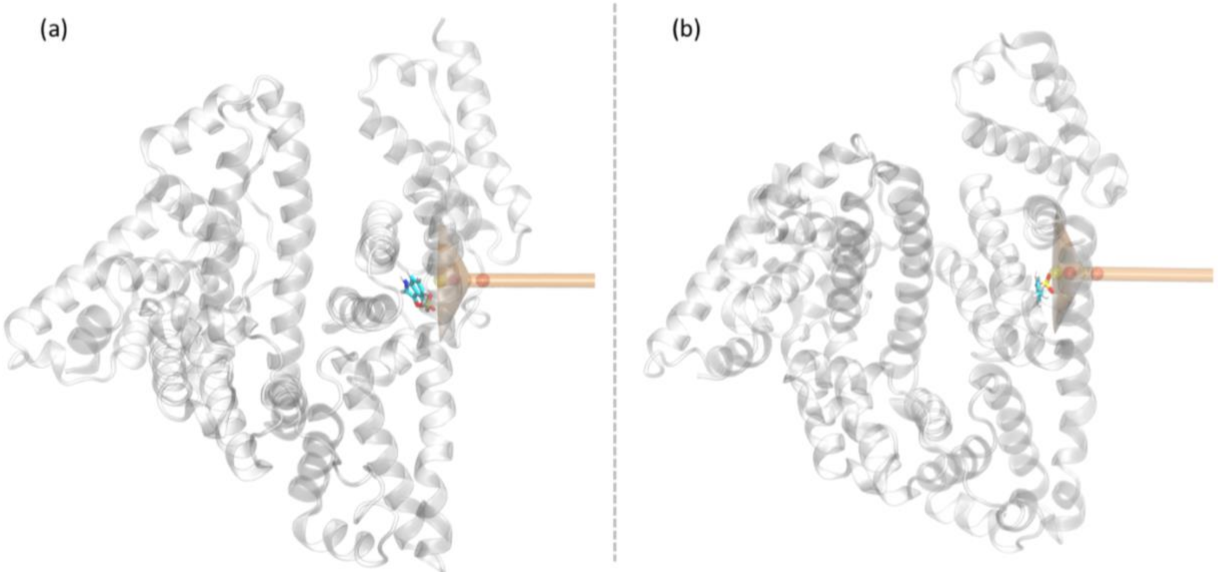


Figure II.2 – Initial MD frame and funnel position displayed in VMD. The ISO and PCS systems are shown in (a) and (b) respectively. Upper walls restraints are used to keep the toxin from moving too far into solvent during runs used in the calculation of unbinding free energies.

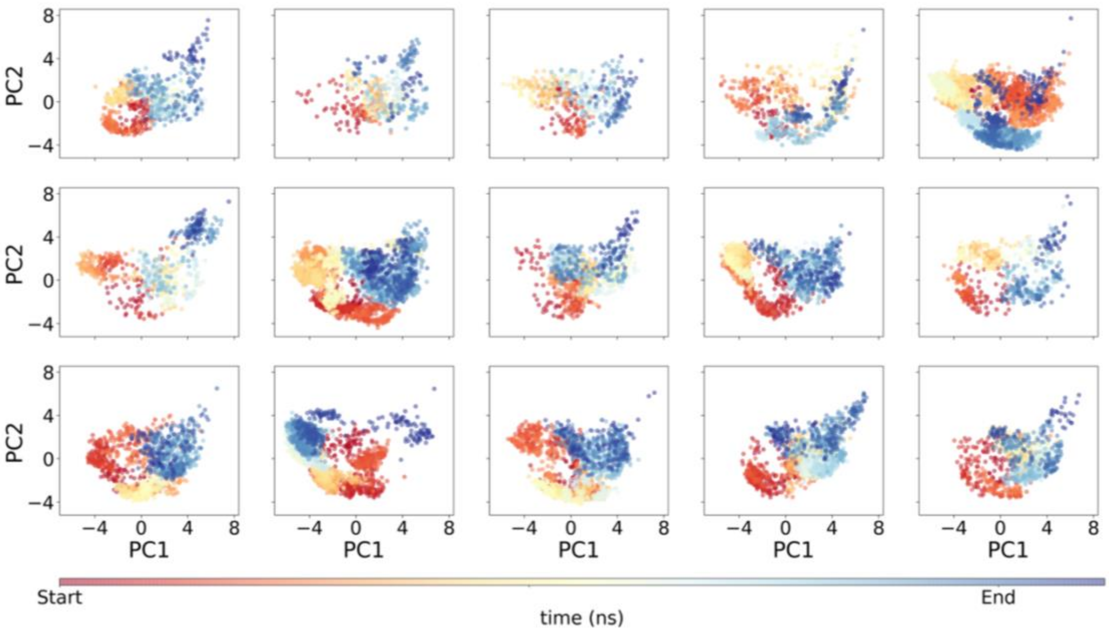


Figure II.3 - PCA data for 15 ISO$_{S489}$ trajectories colored by time. Note that coloring by time simply shows the movement through PCA space for each system not a comparison over time since each system has a unique total simulation time.

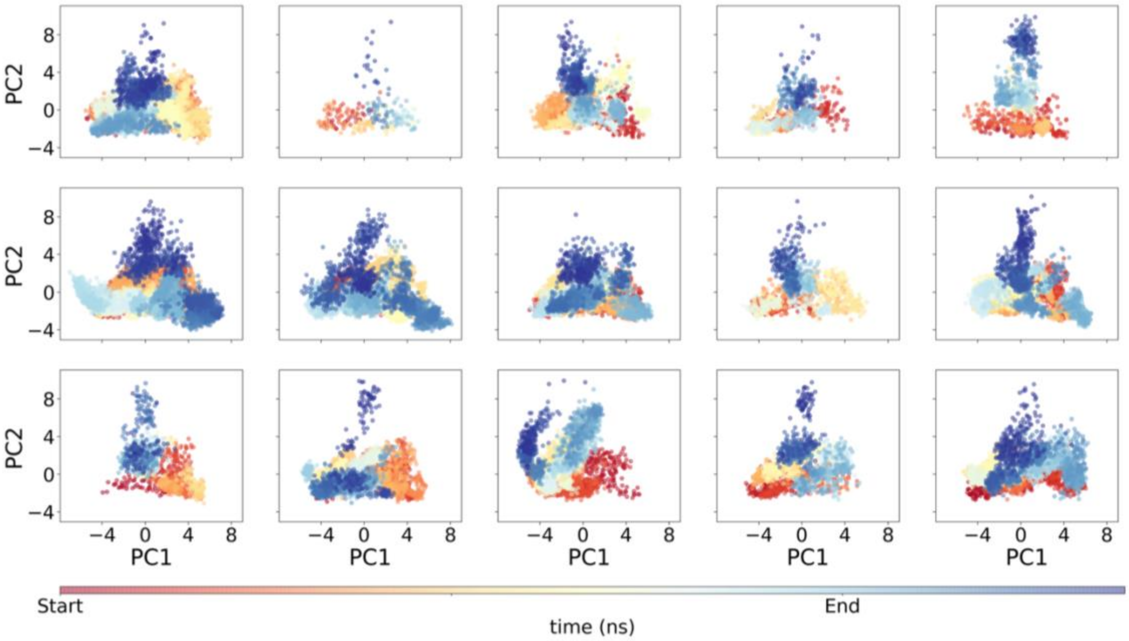


Figure II.4 - PCA data for 15 PCS trajectories colored by time. Note that coloring by time simply shows the movement through PCA space for each system not a comparison over time since each system has a unique total simulation time.

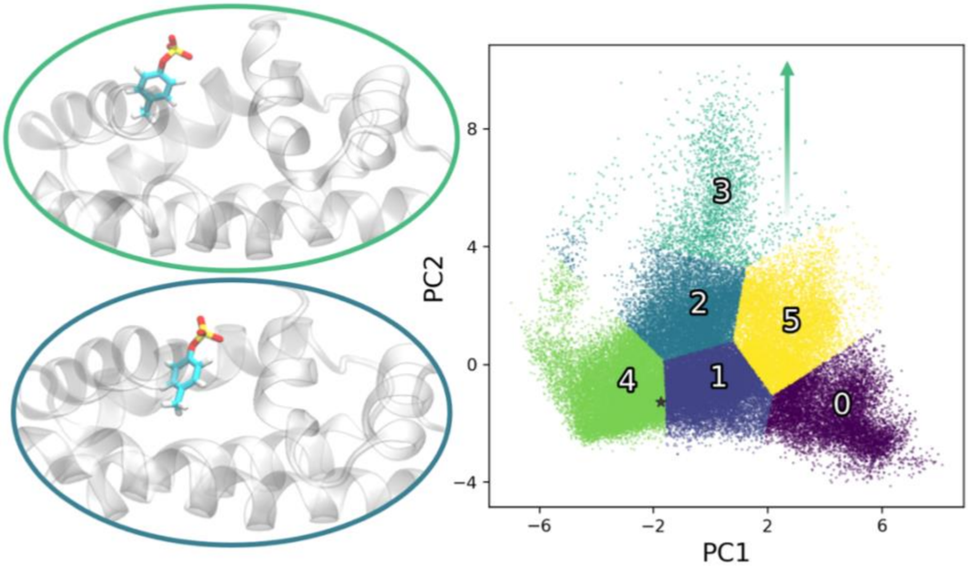


Figure II.5- PCA data for PCS system with increased K-means clusters. Example of K-means clustering when splitting the PCA space data into 6 clusters (rather than 4 displayed in section 4.3.4). The scatter plot shows the first 2 PC values for every 0.2 ns of simulation time colored and labeled by cluster with the label at the location of cluster center. A star marks the initial PCS pose and an arrow shows the unbinding direction in PC space. The accompanying VMD snapshots are representations of cluster 2 and 3 chosen as the frame closest to the cluster's center with outlines of the particular node's color.

Tables

Table II.1 – ANIMO selected order parameters for use in RAVE constructed CV.

$\boldsymbol{ISO^{a}_{S489}}$	$\boldsymbol{ISO^{a}_{Y411}}$	$\boldsymbol{PCS^{a}}$
L347-r	V343-s	E396-r
L347-s	L346-r	F403-s
E342-r	L347-r	F403-r
K402-r	K402-r	T412-r
F403-s	V426-r	E425-s
C437-r	R445-r	E425-r
V462-s	V462-r	S434-s
V462-r	V473-r	L453-s
C487-r	P384-r	V462-s
P384-s	F488-r	V473-r
P384-r		407Y-s
P416-s		K413-r
E442-s		P416-s
		P441-r

[a] Entries represent the distance between a pair of atoms, first either a C-α (blue) or sidechain atom of a residue (red) in Sudlow site II and followed by either a ring or the sulfate toxin atom, labeled r and s respectively.

Appendix III – Supplemental materials for Chapter 5

Figures

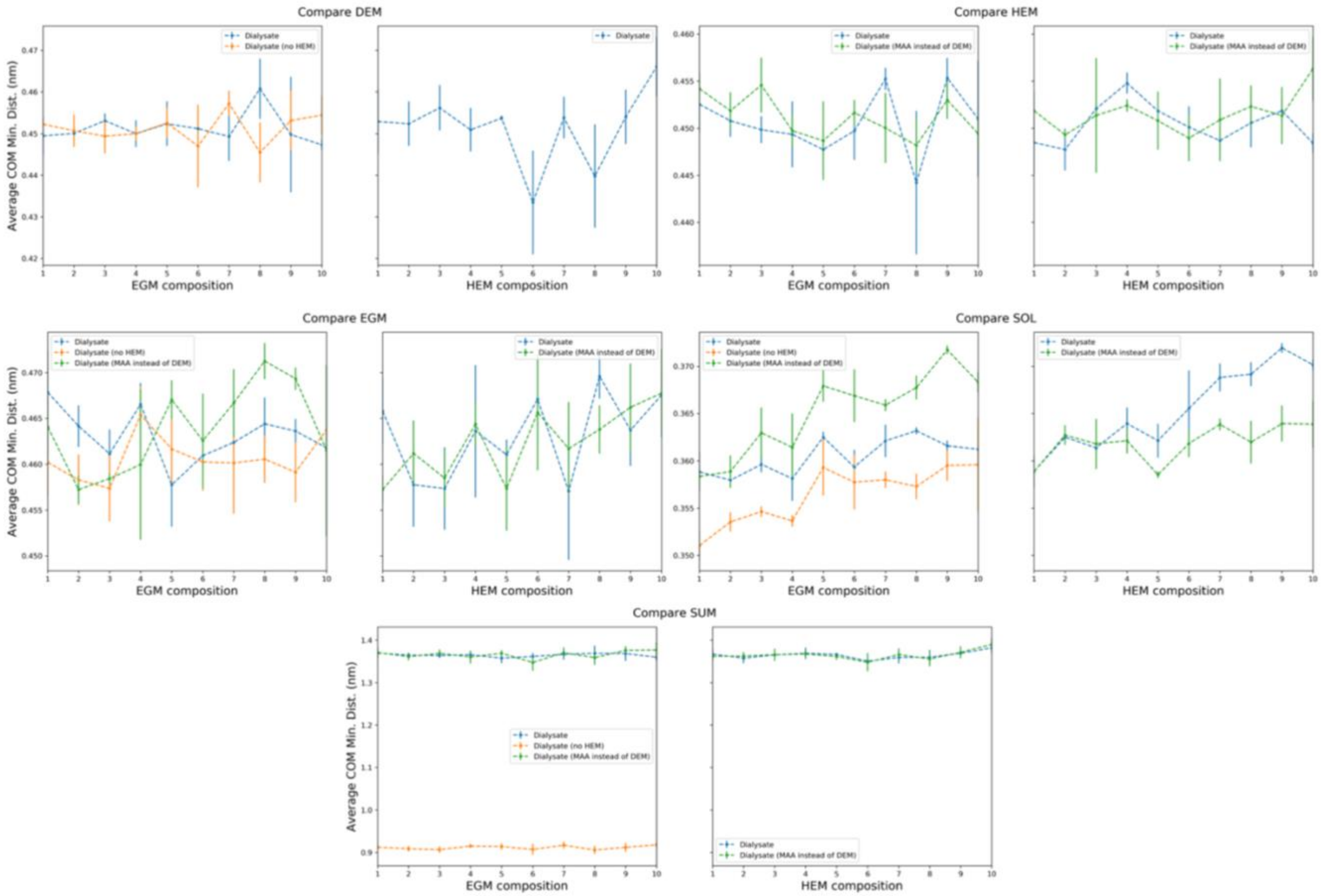


Figure III.1 - Average COM minimum distances between MIP component and ISO within a 0.5 nm cutoff. Each panel displays all possible results from our 3 cases: dialysate, dialysate without HEM, and dialysate with DEM replaced by MAA. The lack of any of the 3 systems is due to the particular components (signified by both the calculated COM distance and particular increasing component on the X-axis) not being present for particular systems.

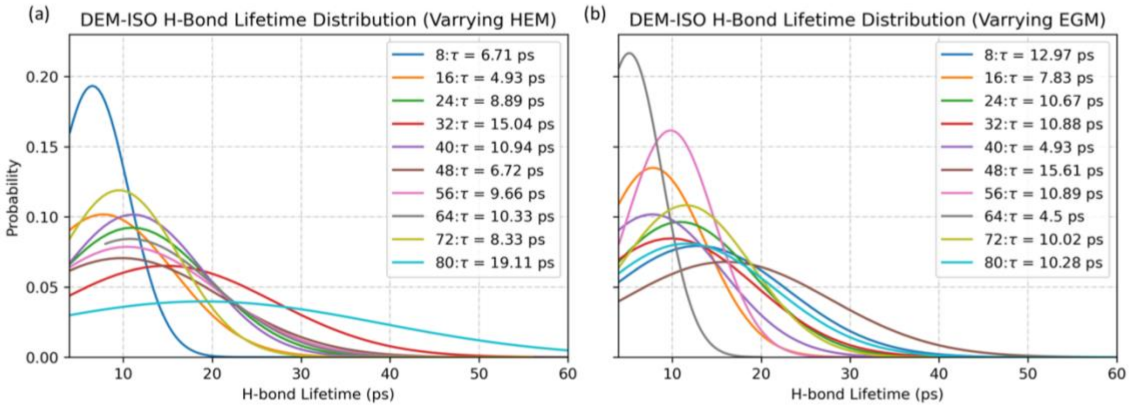


Figure III.2 – ISO-DEM HB lifetime distributions. The (a) and (b) panels show the DEM-ISO HB lifetime distributions while varying HEM and EGM respectively. Each lifetime reported in the legend is an average calculated across the full trajectory.

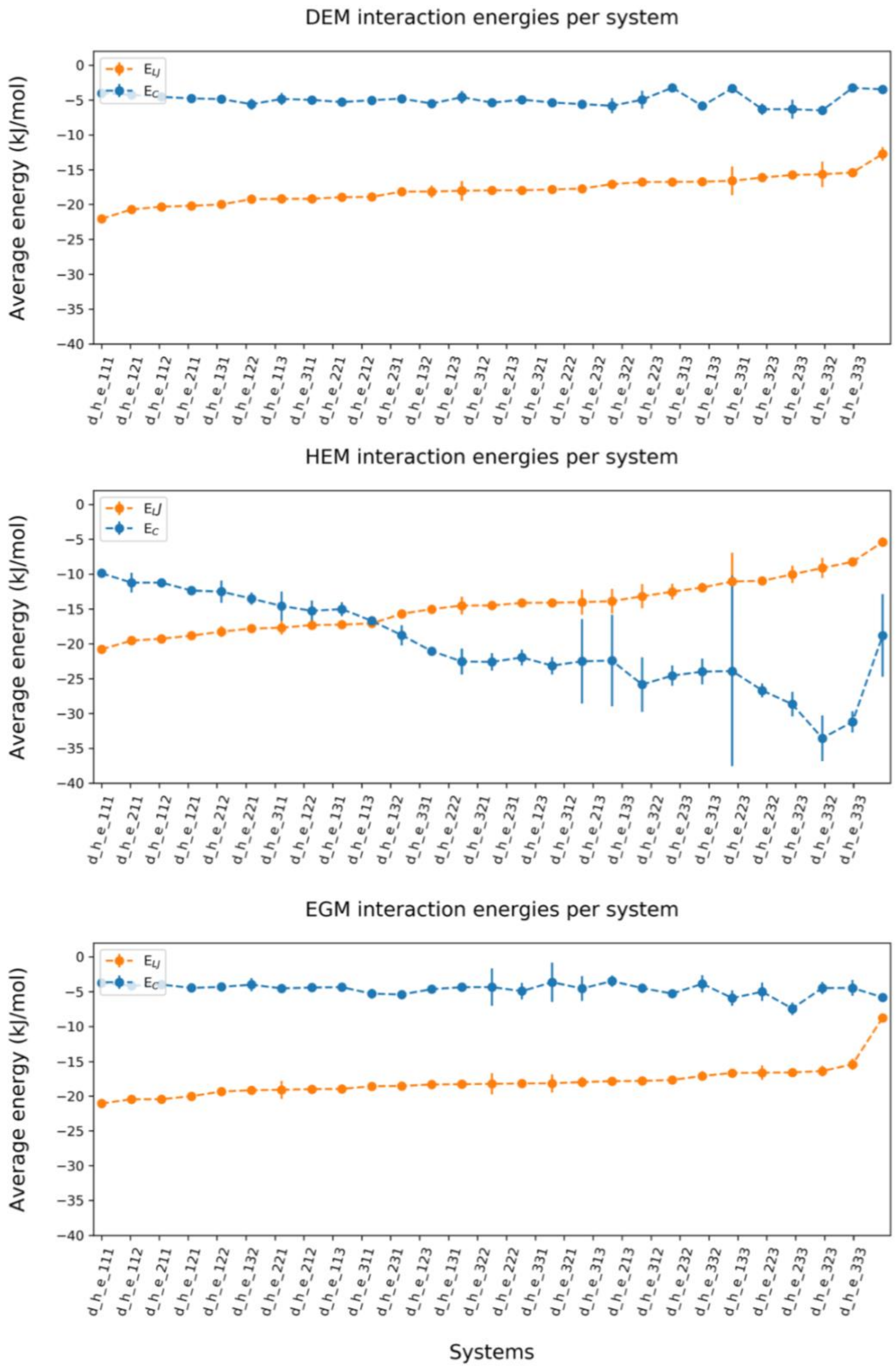


Figure III.3 - The interaction energies per MIP component in pseudo-binding pocket ranked by E_{LJ}. Average Coulombic (E_C) and Lennard-Jones (E_{LJ}) energies for each system are reported (sorted by E_C) per component. Error bars are calculated as standard deviations across 3 equal partitions of the trajectory.

Tables

Table III.1 - Component numbers for baseline MIP system.

Component	ISO	DEM	HEM	STY	EGM	MeOH	Water	K
Molecules	1	8	16	8	40	370	278	1