EXPERIMENTAL-COMPUTATIONAL INVESTIGATION OF CHIRAL SEPARATIONS WITH AMINO ACID-BASED SURFACTANTS

A Thesis

by

MAURO A. GARCIA, JR.

BS, Southwestern University, 2018

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

CHEMISTRY

Texas A&M University-Corpus Christi
Corpus Christi, Texas

May 2021

May 2021

EXPERIMENTAL-COMPUTATIONAL INVESTIGATION OF CHIRAL SEPARATIONS WITH AMINO ACID-BASED SURFACTANTS

A Thesis

by

MAURO A. GARCIA, JR.

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Eugene Billiot, PhD
Chair

Fereshteh Billiot, PhD
Co-Chair/Committee Member

Timothy Causgrove, PhD
Committee Member

May 2021

ABSTRACT

This thesis research investigates the use of amino acid-based micelles to separate enantiomeric compounds. These are investigated using both, experimental and computational methods.

The experimental portion of my thesis research investigates the effect of counterions on the chiral recognition of 1,1′-Binaphthyl-2,2′-diamine (BNA) and 1,1′-Binaphthyl-2,2′-diyl hydrogenphosphate (BNP) enantiomers when using an amino acid-based surfactant as the chiral pseudostationary phase in capillary electrophoresis. The effects of sodium counterions on the chiral recognition of binaphthyl derivatives were compared to that of using pH-dependent Lysine counterions at varying pH conditions. The enantiomeric separation of BNP and BNA enantiomers via capillary electrophoresis, using L-Undecyl-Leucine (und-Leu) as the chiral recognition medium, significantly improved the enantiomeric resolution in capillary electrophoresis at pH 7 when using Lysine counterions as compared to using sodium counterions. Therefore, this experimental project provides insight into the advantages of using cationic, pH-dependent counterions such as Lysine to significantly improve the chiral recognition of binaphthyl derivatives in capillary electrophoresis studies.

The computational portion of my thesis research focused on developing computational methodology to study the binding interactions between amino acid-based molecular micelles and chiral enantiomers. The computational method development plays a crucial role in helping the research advance towards the development of a predictive database. After developing a set of protocols for the computational research, the methods were then used to study the chiral separation mechanisms of Dansyl amino acids, including Dansyl-Leucine, Dansyl-Norleucine, Dansyl-

Tryptophan and Dansyl-Phenylalanine binding to poly-sodium *N*-undecanoyl-(L)-Leucylvalinate, poly(SULV). This study reveals that the computationally-calculated binding free energy values for Dansyl enantiomers binding to poly(SULV) are in agreement with the enantiomeric order produced in experimental MEKC studies. Furthermore, hydrogen bonding analyses was used to investigate and elucidate the molecular interactions that govern chiral recognition in these molecular systems.

DEDICATION

I would like to thank my family for their endless support on my academic journey. I am most thankful for every teacher and professor I have ever had and will have in the future. Finally, I would like to dedicate this thesis to the future scientists and leaders of the world, and especially to those who have been historically underrepresented in STEM- you belong, and you are capable.

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Eugene Billiot, Dr. Fereshteh Billiot and Dr. Timothy Causgrove for their guidance and support throughout the course of this research. Thank you for providing me with the laboratory resources to grow as a researcher, for pushing me to think critically, and allowing me the opportunity to continuously improve my writing and presentations skills. This research experience is one I will forever be thankful for.

I would like to thank my fellow graduate students and friends who made my time at Texas A&M University-Corpus Christi an enjoyable and unforgettable experience. I am most grateful for my fellow lab mates who have been encouraging throughout this entire process and for assisting me in collecting data for my thesis research.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES PAGE

LIST OF TABLES

TABLES PAGE

SUPPLEMENTAL INFORMATION

Supplemental Information for Appendix 1

Supplemental Information for Appendix 2

CHAPTER I: INTRODUCTION

1.1 Chirality

Chirality refers to the asymmetry of a molecule. One source of asymmetry arises in the molecule if it contains a tetrahedral carbon atom that is connected to four unique substituents. Thus, this carbon atom makes the molecule asymmetrical in nature, and is commonly referred as a chiral center. Additionally, a molecule can also be considered chiral if it has an asymmetrical plane. Figure 1.1 shows that enantiomers are chiral compounds that are nonsuperimposable, mirror images of each other, very similar to our left and right hands. Enantiomers are identical in atomic connectivity, but their spatial arrangement of the atoms differ.

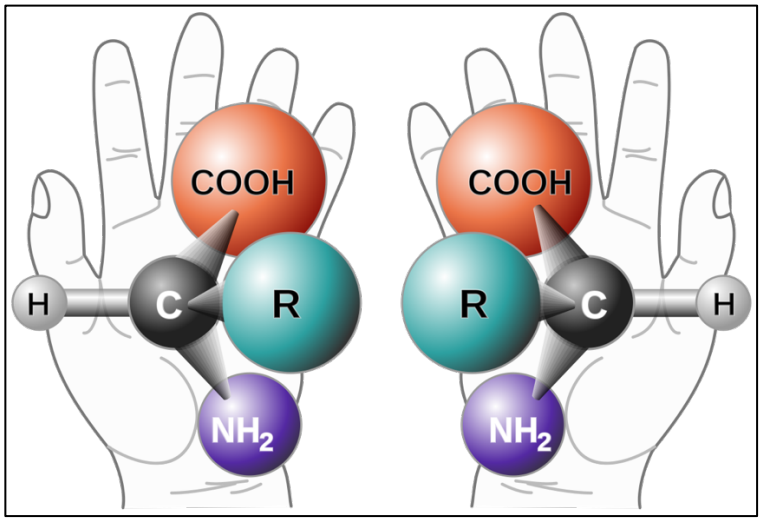


Figure 1.1. Enantiomers with similar connectivity, but different spatial orientations.

The enantiomers are nearly identical when comparing their physical properties. Some examples include melting and freezing points, density, mass, and solubility. However, it is still possible to distinguish and name the enantiomers. The Latin terms *rectus* (R) or *sinister* (S), meaning right-handed or left-handed, respectively, can be assigned to enantiomers based on their

spatial orientation. Alternatively, chiral compounds are also distinguishable due to their optical activity. Figure 1.2 shows that when an enantiomer is subjected to plane-polarized light, the asymmetry of the compound causes the plane-polarized light to be rotated either clockwise or counterclockwise, also known as dextrorotary [(D) or (+)] or levorotary [(L) or (-)], respectively. As such, one enantiomer of a chiral compound rotates the plane-polarized light clockwise, whereas the opposite enantiomer actually rotates the plane-polarized light counterclockwise.

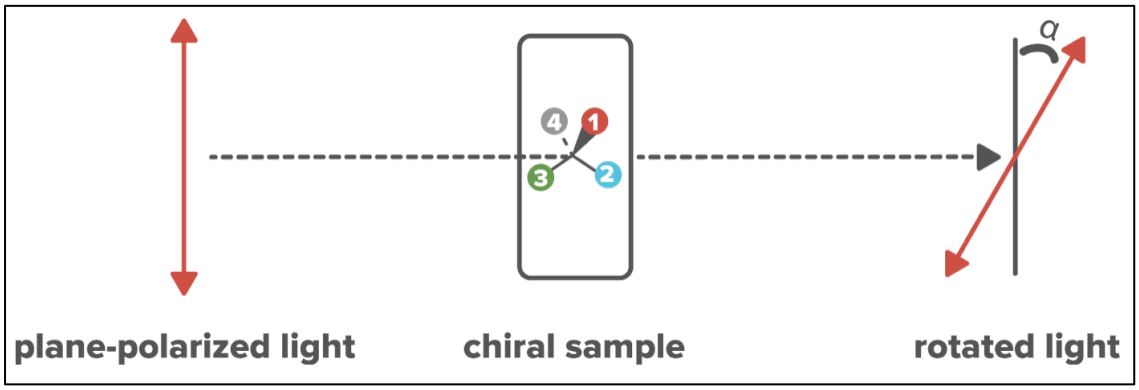


Figure 1.2. Rotation of plane-polarized light by an enantiomeric compound.

1.2 Effects of Chiral Drugs

In the late 1950s, a newly marketed drug, Thalidomide, was marketed in Europe and other parts of the world as a medication for treating morning sickness for pregnant women.[1] Soon after, mothers who took this drug gave birth to babies with congenital defects. Ultimately, Thalidomide, which was sold as a racemic mixture, caused the babies to be born with undeveloped and non-functional limbs. It was later determined that the (S) configuration of Thalidomide caused teratogenic effects, whereas the (R) configuration of Thalidomide had therapeutic effects for treating morning sickness.[1]

In response to this tragedy, the Kefauver-Harris Amendment was established in 1962, for the improvement of the Food and Drug Administration processes, which mandates that pharmaceutical companies must provide thorough testing and accurate information on chiral medications. This also enforced that the enantiomers of racemic drugs be separated into their enantiomerically purified forms, their respective R and S enantiomers. Each enantiomer of the drug must be separated and tested for its efficacy and safety. If both enantiomers are deemed effective and safe, then the pharmaceutical drug may be sold as a racemic mixture. However, if the enantiomers cause significantly different physiological effects, or if one is detrimental to the health of the patient, this requires that the drug be sold in its enantiomerically pure form. Due to this FDA mandate, a new era and scientific challenge was born: the mission to establish and effectively separate enantiomers. This calls for the establishment and optimization of enantiomeric separation processes.

1.3 Capillary Electrophoresis

Some commonly used chiral separation techniques are high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE), which utilize chiral stationary and pseudostationary phases, respectively.[2, 3] In using chiral chromatographic techniques, if the two enantiomers of a chiral compound, the R- and S-enantiomer, have different binding affinities to the chiral separation medium, then enantiomeric separations can occur. While both aforementioned techniques are able to effectively separate enantiomers on the basis of their binding affinity to the chiral separation medium, CE has several advantages over HPLC.[4] A much smaller sample size is required for CE as compared HPLC and often, CE yields a higher number of theoretical plates, ultimately producing better enantiomeric resolution in a shorter period of time.[3] In HPLC, the

chiral recognition medium is covalently linked to a type of solid support material such as silicon beads. Therefore, the user would need to purchase individual columns containing the desired stationary chiral recognition medium. Consequently, if the user wanted to switch out the chiral separation medium, it would entail switching out the column(s). However, in CE, the chiral selectors can be exchanged quite easily with another since it is part of the mobile phase, in which the chiral selector acts as a pseudostationary phase. Chiral selectors such as cyclodextrins, polysaccharides, crown ethers and chiral micelles have proven to be effective chiral selectors in CE.[5] The research described in this thesis proposal focuses on the latter, chiral micelles.

1.4 Amino Acid-Based Surfactants

Micelles are composed of aggregated surfactant units. Surfactants are composed of a hydrophobic tail and a hydrophilic head group, shown in yellow and dark blue, respectively, as shown in Figure 1.3.

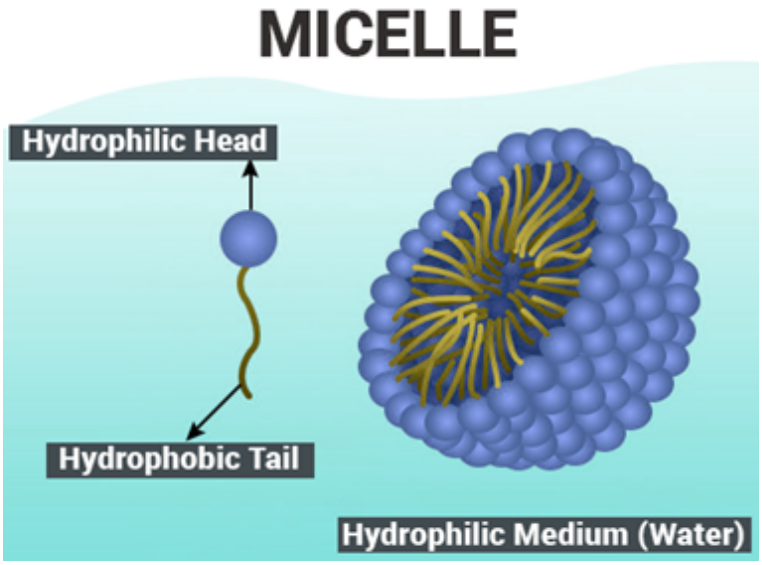
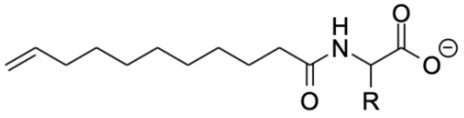


Figure 1.3. Structural feature of a micelle composed of aggregated surfactant units.

When a high enough concentration of surfactant is present in aqueous solution, this causes the surface tension to decrease, allowing for the hydrophobic regions of the individual surfactants to aggregate in solution, whereas the hydrophilic regions of the surfactants face the aqueous solution, as shown in Figure 1.3. The surfactant concentration at which micelles begins to form is commonly known as the critical micelle concentration (CMC).

In particular, the micelles studied in our research group are amino acid-based micelles. For these micelles, the surfactants are composed of a hydrocarbon chain and a hydrophilic, amino acid head group.[6] For this research, single and dipeptide surfactants were studied using experimental and computational methods, respectively.

Single amino acid-based surfactants studied in our research group are composed of an undecanoyl hydrocarbon chain, and the hydrophilic head group of the surfactant is composed of either the L-form of alanine, valine or leucine. A basic structure of a single amino acid-based surfactant used in our research group is shown in Figure 1.4.

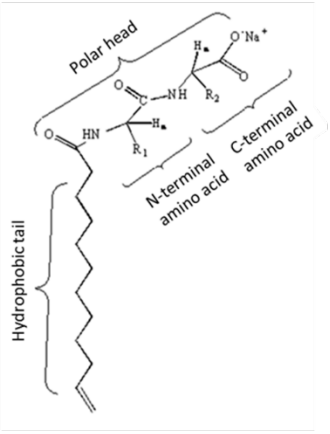


Surfactant	R- Group
SUA	CH_3
SUV	$CH_2(CH_3)_2$
SUL	$CH_2CH(CH_3)_2$

Abbreviations used above for single amino acid surfactants: S = sodium, U = undecanoyl, **A = alanine, V = valine, L = leucine**

Figure 1.4. Basic structure of a single amino acid-based surfactant.

The dipeptide surfactants that are researched in our group contain at least one chiral center using various dipeptide combinations of the L-form of alanine, valine and leucine, as well as the achiral amino acid glycine.[7] The basic structure(s) of the AABMs discussed in this thesis are shown in Figure 1.5.



Surfactant	R1	R2
SUGA	H	CH_3
SUGV	H	$CH_2(CH_3)_2$
SUGL	H	$CH_2CH(CH_3)_2$
SUAG	CH_3	H
SUAA	CH_3	CH_3
SUAV	CH_3	$CH_2(CH_3)_2$
SUAL	CH_3	$CH_2CH(CH_3)_2$
SUVG	$CH_2(CH_3)_2$	H
SUVA	$CH_2(CH_3)_2$	CH_3
SUVV	$CH_2(CH_3)_2$	$CH_2(CH_3)_2$
SUVL	$CH_2(CH_3)_2$	$CH_2CH(CH_3)_2$
SULG	$CH_2CH(CH_3)_2$	H
SULA	$CH_2CH(CH_3)_2$	CH_3
SULV	$CH_2CH(CH_3)_2$	$CH_2(CH_3)_2$
SULL	$CH_2CH(CH_3)_2$	$CH_2CH(CH_3)_2$

Abbreviations used in table above for dipeptide surfactants: S = sodium, U = undecanoyl, **G = glycine, A = alanine, V = valine, L = leucine**

Figure 1.5. Basic structure of dipeptide amino acid-based surfactants

1.5 Micellar Electrokinetic Chromatography

When chiral micelles are used as the pseudostationary phase in CE, this technique is known as micellar electrokinetic chromatography (MEKC). MEKC is a well-established technique for the enantiomeric separation of chiral compounds.[7-11] Figure 1.6 shows the main components in MEKC. The technique is described and summarized as follows. First, in general, the buffer solution is composed of a mixture of aqueous solvent, negatively charged amino acid-based micelles, and positively charged counterions. The column utilized in our research was an uncoated silica column. The silanol group on the inside surface of the capillary is negatively charged above pH 2. The buffer solution is introduced into the column, where the positively charged counterions will interact with the negatively charged silanol groups on the capillary wall. The positively charged counterions will also interact with the negatively charged micelles. A charge is applied from the inlet of the capillary to the outlet, causing the inlet to be positively charged and the outlet to be negatively charged. The counterions interact with the negatively charged column which cause a flow of the solution from the inlet to the outlet of the column; this is known as an electroosmotic

flow (EOF). While this process is occurring, all negatively charged species (the micelles) within the column will have a slight movement towards the inlet, whereas the positively charged species (the counterions, analyte, etc.) will move towards the outlet.

As mentioned previously, all micelles are moving against the EOF, when the racemic mixture is then introduced into the column, the different enantiomers are allowed to interact with the suspended chiral micelles. Eventually, everything will elute from the column due to the strong EOF. Due to the enantiomers having differences in their stereoconfiguration, they will innately have different interactions with the chiral micelles, thus allowing for a slight difference in their binding affinities. For example, one enantiomer may bind stronger to the chiral micelle than its opposite enantiomer, therefore we expect the enantiomer with the stronger binding affinity to the micelle to elute from the column last.

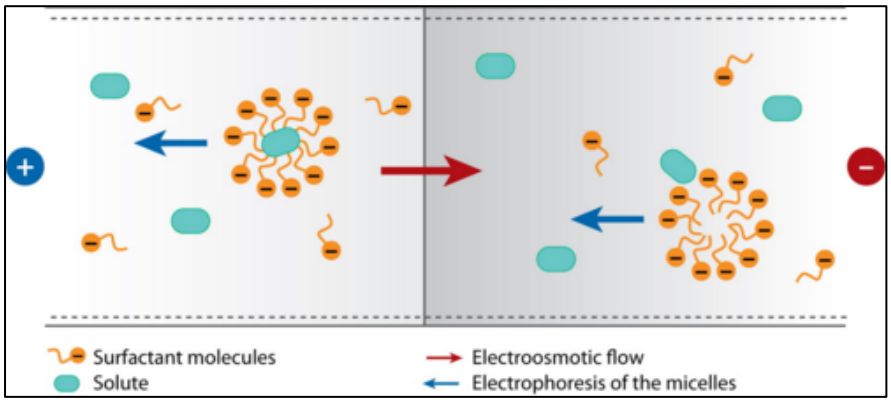


Figure 1.6. Visual representation of the electroosmotic flow and charge distribution in MEKC.

The interaction and binding of the enantiomers to the micelles directly affects their retention time and elution order. Although this is a chromatographic technique, it utilizes a UV-Vis detector to display the peaks of the enantiomers when they elute from the column as a function of time. For example, as a preliminary study to determine elution order, the R and S enantiomers of a compound are injected in a 1:2 ratio, providing enantiomeric excess of the S enantiomer. The

absorbance of the peak is directly proportional to the concentration of each enantiomer present. Therefore, if the taller peak elutes second (has a longer retention time), then we know that the S enantiomer binds more strongly to whichever micelle was being used in this particular study. When comparing the two peaks, the distance between the two peaks can be representative of how well a micelle is at separating the two enantiomers. Fortunately, when investigating different types of amino acid-based surfactants, there are certain ways that can be used to optimize enantiomeric separations, which is described in the following sections.

1.6 Single vs. Dipeptide Amino Acid-Based Surfactants

Dipeptide amino acid-based surfactants have shown significant advantages in achieving enhanced enantiomeric resolution in MEKC compared to single amino acid-based micellar systems. This is due, in part, to the increased number of chiral centers on the head group, thus allowing for more chiral interactions to occur, thus enhancing chiral selectivity. Additionally, when working with the single amino acids, as mentioned previously, our group has studied three: undecyl-Leucine, -Valine and -Alanine. When using dipeptide amino acid-based surfactants, it allows for more combinations of micelles to be investigated- a total of 15.

1.7 Monomer vs. Polymerized Amino Acid-Based Micelles

Previous research reveals that the dipeptide amino acid-based micelles act as better chiral selectors in their polymerized forms.[10] The polymerized surfactants are formed via free radical polymerization when the surfactant solution is subjected to gamma radiation at concentrations significantly above their critical micelle concentration (CMC), usually 5-10 times the CMC. As

shown in Figure 1.7, the gamma radiation causes the terminal alkenes to undergo free radical polymerization to form covalent bonds with neighboring surfactant units.

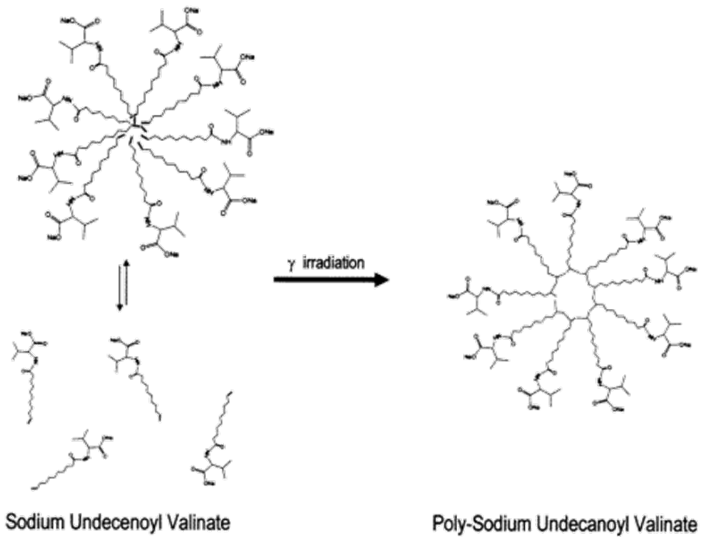


Figure 1.7. Polymerization process for amino acid-based surfactants.

These polymerized micelles are commonly referred to as amino acid-based molecular micelles (AABMMs). AABMMs have significant advantages over the traditional, non-polymerized micelles, including: AABMMs do not have a critical micelle concentration (CMC) and have reduced joule heating. The separation of enantiomers is often enhanced when using AABMMs as compared to the non-polymerized form. Figure 1.8 shows that the monomers of L-SULV and L-SUVL have lower resolution values when separating binaphthol (BOH) enantiomers when compared to their polymerized micellar form, denoted as poly(L-SULV) and poly(L-SUVL). Figure 1.8 also shows that the resolution for the separation of BOH is enhanced at lower concentrations of surfactants (w/v %) for the polymers, as compared to the monomers.

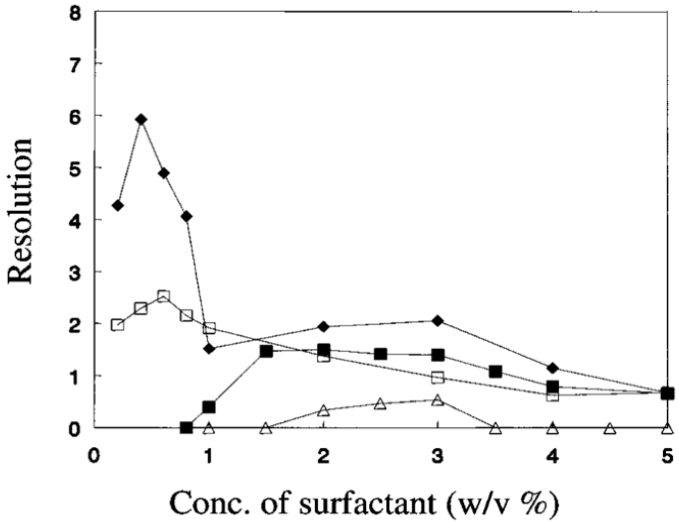


Figure 1.8. Comparison of the polymers of L-SULV and L-SUVL to the monomers for the separation of BOH. (a, ◆) Poly(L-SULV), (b, □) poly(L-SUVL), (c, ■) the monomer of L-SULV, and (d, △) the monomer of L-SUVL.

Therefore, it is worthwhile to further investigate AABMMs as a pseudostationary phase in MEKC due to their enhanced chiral selectivity at concentrations lower than the critical micelle concentration of the non-polymerized forms of the micelles, as previously mentioned and also displayed in Figure 1.8.

1.8 Effect of Dipeptide Order

Previous studies have shown that the alteration of the dipeptide polar head constituent of the AABMMs largely can have a huge effect on its formation and effectiveness as a chiral recognition medium.[7, 12] As shown in Figure 1.9, sodium N-undecyl (L,L) leucine-valine (SULV) acted as a better enantiomeric separation medium than sodium N-undecyl (L,L) valine-leucine (SUVL) when separating binaphthyl phosphate (BNP) enantiomers, as they had resolution values of approximately 7.5 and 0.5, respectively.[7]

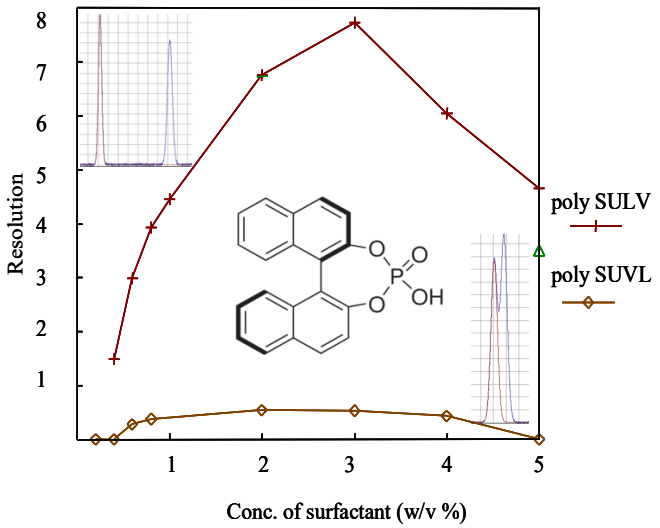


Figure 1.9. Effect of dipeptide order on the separation resolution of binaphthyl phosphate enantiomers.

This reasoning has led to further investigations regarding the properties that make chiral recognition and enantiomeric separation mediums optimal. However, very little is known about the exact structure of these micelles or their chiral recognition mechanisms. Further experimental studies of these AABMMs have been examined using steady-state fluorescence anisotropy, NMR, and MEKC.[9] These give insight to AABMM aggregation number and hydrogen bonding interactions between the micelles and the enantiomers, and their binding affinities, respectively.

1.9 Effect of Counterions in Chiral Separations

The charge of amino acids, part of the amino acid-based micelle composition, is greatly influenced by pH levels, which can be used as an advantage in separation chemistry because it allows for pH-dependent trends to be studied. Furthermore, the ability to impose different pH environments helps to study how the charge of the surfactant determines micelle formation and shape, overall net charge of the surfactant, how enantiomers interact with the surfactant, and the

effects it may have with various counterions. All of which greatly influence chiral recognition when separating enantiomers with amino acid-based micelles.

As mentioned previously the effects of amino acid order, steric hindrance, dihedral angles and the polymerization of micelles have been studied in terms of their role in chiral recognition. For many years, sodium has been extensively used as the counterion for the MEKC buffer solution. Most recently, the effects of pH-dependent counterions have been used to significantly improve the resolution of enantiomeric compounds in MEKC, specifically with arginine as the counterion.[13] The side chain of arginine contains amino groups which are greatly affected by pH. Recent studies have shown that when using pH-dependent counterions, the resolution can be significantly enhanced when separating binaphthyl derivatives.

1.10 Predictive Database

As mentioned previously, MEKC can be utilized to separate enantiomeric compounds when using amino acid-based micelles as the pseudostationary phase. Previous research reveals that the alteration of the experimental conditions such as dipeptide order, counterions, pH levels and/or temperature largely affect the formation and effectiveness of the AABMMs as a chiral recognition medium. Furthermore, the process of testing all wide-range possible testing conditions and combinations is intractable due to the cost, time and extensive experiments to perform. The need for understanding these systems is crucial for effectively separating chiral compounds and deeply understanding their enantioseparation mechanisms associated with them.

A cost-effective, time efficient and predictive model for accurately predicting optimal testing conditions for chiral separations in MEKC using amino acid-based surfactants would be ideal. Therefore, the goal of our research is to provide more data that will aid in the development

and use of a novel quantitative structure-enantioselective retention relationship (QSERR) database, created with the input of both experimental and computational data.[14] This will allow for the QSERR database to predict optimal experimental conditions for MEKC to optimally separate the enantiomers of interest. One of the most important steps in working towards the QSERR is to establish a reliable molecular modelling system to study the interactions between enantiomers and amino acid-based micelles.

1.11Molecular Modeling

Models are typically referred to as simplified systems. When applying this concept to chemical structures it is often referred to as molecular modeling, in which atoms are displayed as nodes or spheres. Bonds are often represented by sticks that connect the various atoms in the chemical model. Each atom can have different characteristics associated with it such as size, electronegativity and valence. Furthermore, the bonds in the system can have various levels of stiffness based on its hybridization. For example, a sp^3-hybridized carbon would have free rotation in its bond, compared to a sp-hybridized carbon, which is associated with a less flexible triple bond, in which a triple bond has less distance between the two atoms compared to two atoms sharing a single bond. Molecular modeling helps to (1) simplify structures and chemical properties, and (2) to help simplify computational calculations. One such way that molecular modeling can be simplified and studied is by examining the coulombic interaction potentials within the system.

1.12 Coulombic Interaction Potential

The Coulombic interactions can be quantified in terms of attraction and repulsion between the charged species within a molecular system. The attraction of the electrons in the valence shells

to the positively charged nuclei of another atom, as well as the repulsion between similarly charged ions can be further examined using Equation 1-1.

The Coulombic interaction potential (V_{ij}) between two particles can be calculated using:

Equation $1-1$

$$V_{ij} = V(r_{ij}) \frac{qi \cdot qj}{(r_{ij})}$$

The particle charges are identified as qi and qj, and the distance between the two particles is rij. Essentially, this can be used to describe the interaction potential intermolecularly and intramolecularly.

1.13Classical Mechanics vs. Quantum Mechanics

In order to process how a chemical system evolves as a function of time, classical mechanics can be applied to the molecular system using Newton's second law of motion: Force = mass * acceleration. Since molecular modeling is typically in a three-dimensional setting, position vectors (r), are taken into account to calculate the evolution of the chemical system using a derived version of Newton's second law of motion:

Equation $1-2$

$$-\frac{dV}{dr} = m\frac{d^2r}{dt^2}$$

Due to the nature of electrons containing both wave and particle characteristics, they cannot be studied using Newton's second law of motion since that only pertains to particles. In the case of quantum mechanics, the evolution of the system over time must be calculated as a wave function, which applies the time-dependent Schrodinger equation. However, the applications used for this particular research project focus on the use of classical mechanics, thus highlighting the diversity of computational equations and methods that can be used to study the evolution of a

chemical system over time, also known as a molecular dynamics (MD) simulation. Essentially, for classical mechanics calculations, Newton's second law is repeatedly calculated for each atom and its interactions with other atoms in the system.

1.14 Software for Molecular Dynamics Simulations

As mentioned previously, the evolution of a chemical system as a function of time, when taking into account the molecular forces and interactions occurring at an assigned pressure and temperature via *in silico* studies of the molecular system is known as a MD simulation. Various software exists to run classical mechanics-based MD simulations such as: GROningen MAchine for Chemical Simulation (GROMACS), Chemistry at Harvard Macromolecular Mechanics (CHARMM), and Assisted Model Building with Energy Refinement (AMBER).[15] The research described in this thesis focus on the use of the AMBER software to conduct molecular dynamics simulation studies of amino acid-based molecular micellar systems. Prior to actually running the MD simulations using the AMBER software, the molecular system must first be properly defined in terms of its charges, bond angles, etc.

1.15 Force Fields

An important step prior to running an MD simulation is to assign force fields to the molecular system of interest. Force fields are used to further simplify, categorize and define certain structural characteristics of a molecule to better evaluate the chemical system evolving over time. Similar functional groups and molecular trends allow for general trends to be made in regard to the chemical and physical properties. For example, all aromatic hydrogens have similar bond lengths, as well as other functional groups sharing common characteristics, such as: carbonyls,

amides, benzenes, and esters. These functional groups also display similar vibrational frequencies determined by IR Spectroscopy. Some molecules with repeating units, such as alkanes and alkenes display an increase in energy (kJ/mol) associated with the elongation of their molecular units. All of these characteristics can be simplified and assigned quantitative values that can be used to describe a chemical system, also known as molecular descriptors. For force fields, the parameters such as the bond length, bond angle, torsional strain and more are all assigned based on experimentally-determined trends in molecules.[16] The force fields that will be mentioned in the upcoming sections 1.15.1-1.15.2 are specific to the AMBER software.

1.15.1 Generalized AMBER Force Field

For this research, the General AMBER force field (GAFF) is applied to the micellar systems models. GAFF is a general force field, meaning a large array of organic atom types were studied experimentally, in which general trends were observed in their bond lengths, bond angles, charge distribution, and their hybridization.[16, 17] Table 1.1 displays the basic atom types in GAFF, and how the GAFF force field primarily serves to generalize the main organic elements, and halogens with their appropriate atom types, hybridizations, hypervalency and atom descriptions. GAFF is a specific force field that is generally used to define simple organic structures computationally. In particular, GAFF is used in our research to define the atoms in the hydrocarbon chain of our surfactants.

Table 1.1. Basic Atom Types in the GAFF.

Atom type	Description	Atom type	Description
c	sp2 C in C=O, C=S	o	sp2 O in C=O, COO-
c1	sp1 C	oh	sp3 O in hydroxyl group
c2	sp2 C, aliphatic	os	sp3 O in ether and ester
c3	sp3 C		
ca	sp2 C, aromatic		
n	sp2 N in amide	s2	sp2 S (p=S, C=S etc)
n1	sp1 N	sh	sp3 S in thiol group
n2	sp2 N with 2 subst.	ss	sp3 S in -SR and SS
	readl double bond	s4	hypervalent S, 3 subst.
n3	sp3 N with 3 subst.	s6	hypervalent S, 4 subst.
n4	sp3 N with 4 subst.		
na	sp2 N with 3 subst	hc	H on aliphatic C
nh	amine N connected	ha	H on aromatic C
	to the aromatic rings	hn	H on N
no	N in nitro group	ho	H on O
		hs	H on S
		hp	H on P
f	any F	p2	sp2 P (C=P etc)
cl	any Cl	p3	sp3 P, 3 subst.
br	any Br	p4	hypervalent P, 3 subst.
i	any I	p5	hypervalent P, 4 subst.

1.15.2 AMBER Protein ff19SB Force Field

The AMBER Protein ff19SB force field is used to identify and assign proper amino acid parameters. This specific force field was finetuned to accurately parameterize dihedral angles, and properly modify backbone and sidechains in peptide systems. The surfactants used in this particular study contain amino acid-based polar head constituents. Thus, the protein ff19SB force field was used to properly parameterize the dipeptide head group of the surfactants. GAFF was specifically made to be compatible and mixed with other force fields such as ff19SB.

1.16 Modelling the Aqueous Solvent

The transferable intermolecular potential with 3 points (TIP3P) was used to parameterize water as a 3-point system and is often used to solvate molecular systems computationally. Currently, many point models exist for water, each with their unique H-O-H parameters. Figure 1.10 shows different models of modeling a water molecule with either a 3-site, 4-site, 5-site and

6-site model. As such, either of these water models could be used as the computational solvent of the system.

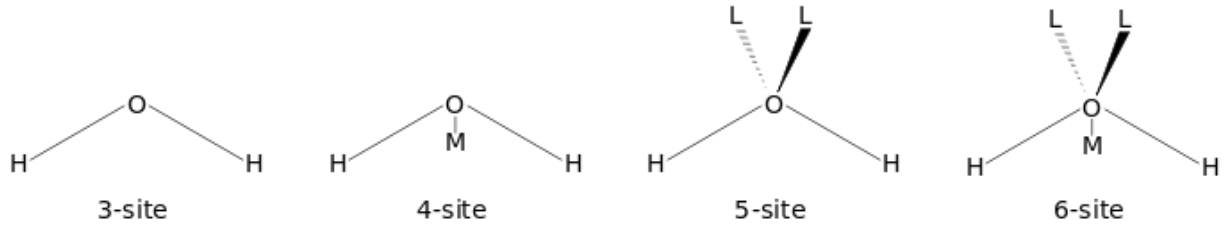


Figure 1.10. Graphical representation of the 3-point through 6-point sites for various solvation models. The M and L in the figure above represent negatively charged dummy atoms and lone pair electrons, respectively.

Due to the TIP3P water model only having 3 points and being the least flexible, it has the highest computational efficiency and allows for less calculations as compared to using the 6-point water system. For this reason, the TIP3P water model has been implemented into our molecular dynamics simulations. Additionally, monatomic counterions such as sodium, calcium and so forth can also be added to the solution to neutralize the system.

1.17 Files Used in AMBER

After generally discussing how force fields are assigned to molecules, it is important to describe the computation files and the terminology. The protein database (PDB) and mol2 files are input files for AMBER, which contain information about the atoms in the system and their coordinates. The force field files add more information to the chemical system, such as mass, bonding parameters and non-bonding parameters. Generally, the input files and force field files, in combination, should define the entire molecular system. However, in order to transfer this

cumulative information to be useful for initiating an MD simulation, these must be saved as a parameter topology file (prmtop) and input coordinate file (inpcrd). Again, for emphasis, the prmtop and inpcrd files contain all of the necessary descriptive information of the system to be able to run an MD simulation. Table 2 shows the terms that are associated for each of the various files, including PDB, mol2, lib/mol2, dat/frcmod, topology and coordinate files. Notice that Table 1.2 also shows that some files only contain one or few terms, therefore more than one file may be needed to define an entire system properly (complement checked boxes to define all terms). As such, the topology and coordinate files complement each other and help to define the entire system, therefore these files are necessary to initiate a molecular dynamics simulation.

Table 1.2. Files used in AMBER software, with their associated terms.

Terms	Input Files		Force Field Files		Output Files	
	PDB	mol2	lib/mol2	dat/frcmod	topology	coordinate
units	✔	✔	✔		✔	
names	✔	✔	✔		✔	
types		✔	✔	✔	✔	
chargs		✔	✔		✔	
connects	✔	✔	✔		✔	
coords	✔	✔	✔			✔
masses				✔	✔	
bond params				✔	✔	
nonb params				✔	✔	

1.18 Ligand Docking

Docking is a commonly used computational technique that best identifies the orientation of an analyte within a binding pocket of a larger molecule. The binding pocket is characterized by its hydrophilic and hydrophobic regions and is kept static during molecular docking. The analyte, on the other hand, is allowed to be flexible so that its optimal spatial orientation and charged interactions within the pocket can be determined and ranked in a variety of positions. Molecular docking has previously been used to study protein-ligand and analyte-micelle interactions. The Molecular Operating Environment (MOE) software package was used to identify the binding pockets on the micelles and dock ligand enantiomers into each of the identified pockets.[18] The MOE software contains a Site Finder module, which was used to identify the binding pockets of the micelles.[19, 20] Site Finder utilizes the alpha sphere method to identify molecular cavities and potential binding sites. Alpha spheres are placed within cavities where four receptor atoms sit at its spherical boundaries.[20] Non-polar regions with poor hydrogen bonding capabilities, as well as containing regions where hydrogen bonding interactions are likely to occur are represented by spheres. Therefore, the hydrophobicity and hydrophilicity of a binding pocket can be quantified.

MOE was then used to dock the enantiomers of various analytes into each of the binding pockets of the micelles. During the docking analysis process, the micelles are kept static, whereas the enantiomers of interest were dynamic in their movement and poses. The Triangle Matcher method within MOE was used to screen various poses of the enantiomer into the binding pocket.[18] The Triangle Matcher method screens ligand poses by aligning three atoms on the enantiomeric compound with three alpha spheres within the binding pocket.[21] For each pose that is screened, the London dG scoring function is used to calculate the binding free energy value. The poses are

then ranked based on their dG values, and the highest scoring enantiomer is used for studies in the MD simulation.[21]

1.19 Running a Molecular Dynamics Simulation

The pdb files of the micelles will be used in conjunction with the pdb files of the docked analytes to run molecular dynamics (MD) simulations of these solvated complexes. The analyte:micelle complex will be solvated with sodium counterions and TIP3P water molecules and Xleap will be used to save the parameter topology (prmtop) and input coordinate (inpcrd) files of these systems. As such, Xleap is a graphic interface that allows the user to create initial molecular modeling coordinates, and also has the capabilities to load various force fields and library files that can then be applied to the computational model, and Xleap can also be used to produce the parameter topology (prmtop) and input coordinate (inpcrd) files.

As mentioned in section 1.17, the prmtop and inpcrd files of the system are necessary to initiate an MD simulation. Now, we can discuss the general steps that are performed in an MD simulation. The steps are as follows: (1) Minimization, (2) Heating, (3) Equilibration, (4) Final production run.

In summary, the minimization step is used to adjust the structure to the applied force field, to relax possible steric clashing between molecules, as well as find an initial energy minimum between the solvent and the solute. After minimization, the system slowly heats to the temperature that was manually set by the user, typically 300K. After heating the system, the pressure of the system is equilibrated to the pressure that was also manually set by the user, typically 1 atm. After these steps, the production run begins and the molecules are then studied at a constant temperature and pressure, and this simulation occurs for as long as it was manually set for by the user.

Throughout the simulation, the coordinates of the molecules will change over time as different interactions are allowed to occur between attractive and repulsive charges in the system, causing changes in their coordinates. These coordinates are recorded to a trajectory file, which keeps track of the molecular motions of each atom in the system as its coordinates change. This trajectory file, therefore, contains information about how the molecular system evolves as a function of time. The trajectory file can then be inputted into various programs such as Visual Molecular Dynamics (VMD) to visualize how the molecular system evolves over time.

1.20 Post-Trajectory Analysis

Because the trajectory file contains the information about how the chemical system evolved over time, it allows for the trajectory file to be analyzed after the simulation is complete. You can think of the trajectory file as a movie, and the user can go back and review certain scenes of the movie. Additionally, certain calculations can be conducted to reveal primary interactions that occurred during the simulation.

1.20.1 Binding Free Energy Calculations

The trajectory file of the complex system can then be analyzed for its binding free energy. This can provide insight into the binding affinity of each enantiomer to the micelle and can then be compared to the experimental MEKC elution order. Please view section *4.3.3* for detailed information on the binding free energy calculations used in these studies.

1.20.2 Hydrogen Bond Analysis

The trajectory file of the complex system can then be analyzed for its hydrogen bonding interactions. This can provide insight into the exact interactions that govern chiral recognition in these molecular systems and can be compared to NMR studies for validation. Please view section *4.3.4* for detailed information on the binding free energy calculations used in these studies.

1.21 Experimental and Computational Experiments

As part of this thesis, the two main studies are developed into a computational and experimental project. Both of which inform and acquire more data that will eventually aid in helping the research advance towards the development of a QSERR database.[14]

1.21.1 Experimental Research

The purpose of this experimental research project was to further explore the effects of pH-dependent counterions on the chiral recognition abilities of the single amino acid-based surfactant, L-Undecyl Leucine. The effect of enantiomeric resolution was investigated via MEKC. This was investigated at varying pH levels and surfactant concentrations. Moreover, the effects of pH-dependent diamine counterions at surfactant to counterion ratios of 1:1 for the single amino acid surfactant were investigated to determine how chiral recognition is affected. The experiments were as follows: (1) synthesized the single amino acid-based surfactants, L-Undecyl Leucine. (2) prepared solutions containing the surfactant in the presence of either sodium or Lysine counterions at varying surfactant concentrations and pH levels, and (3) conducted MEKC experiments to separate binaphthyl derivatives using the created solutions in the running buffer, and (4)

determined and compared the effects of the sodium and Lysine counterions on the resolution (Rs) and retention factor (k') of the surfactants when separating binaphthyl derivatives via MEKC.

1.21.2 Computational Research

The purpose of this research was to develop the methodology and protocols for both creating and investigating amino acid-based molecular micelles (computationally), in which the micelles are able to collapse properly for all dipeptide surfactant combinations, and able to undergo MD simulations without crashing. In addition, it is essential to develop a computational system that accurately represents the molecular binding interactions that occur experimentally between the micelles and the enantiomers, in terms of the computational binding free energy values matching experimental MEKC elution order data. Furthermore, the development of computational methodology for molecular docking, developing protocols and automation scripts for MD simulations, and testing various binding free calculations and hydrogen bond analyses that are in agreement with experimental data were investigated, and developed in efforts to work towards the predictive QSERR model.

More specifically, the following was conducted for this research:

(1) Generated properly equilibrated polymerized amino acid-based molecular micelle structures, so that the hydrophilic head groups face the aqueous solution, and properly collapse so that the hydrophobic regions are buried within the micellar core,

(2) docked enantiomers from the structural library to each binding pocket of the micellar structures to obtain the initial coordinates of the best fit of the various enantiomers for MD simulations,

(3) developed an automation script to run molecular dynamics simulations for each micelle and analyte combination in order to speed up the data collection process, while also establishing

testing parameters in terms of how long each simulation should be conducted, the effects of temperature, running simulations with varying data collection intervals, etc.

(4) developed methodology for determining if the simulation ran properly via visualization software such as Visual Molecular Dynamics (VMD),

(5) performed and accessed the binding free energy calculations using Molecular Mechanics Generalized Born Surface Area (MMGBSA) and Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) methods, and

(6) determine if the binding free energy values match experimental MEKC elution data, as well as

(7) developed methodology for conducting hydrogen bond analysis using the AMBER software and included it into the automation script to analyze the molecular interactions that govern Enantioseparation in these micellar systems.

(8) determined if the hydrogen bond analyses match experimental NMR data.

1.22 References

1. Nguyen, L. A.; He, H.; Pham-Huy, C., Chiral drugs: an overview. *Int J Biomed Sci* **2006,** *2* (2), 85-100.
2. Malviya, R.; Bansal, V.; Pal, O.; Sharma, P., High performance liquid chromatography: A short review. *Journal of Global Pharma Technology* **2010,** *2*, 22-26.
3. Engelhardt, H.; Cuñat-Walter, M. A., Use of plate numbers achieved in capillary electrophoretic protein separations for characterization of capillary coatings. *Journal of Chromatography A* **1995,** *717* (1), 15-23.
4. Mantovani, V.; Galeotti, F.; Maccari, F.; Volpi, N., Recent advances in capillary electrophoresis separation of monosaccharides, oligosaccharides, and polysaccharides. *ELECTROPHORESIS* **2018,** *39* (1), 179-189.
5. Yu, R. B.; Quirino, J. P., Chiral Selectors in Capillary Electrophoresis: Trends During 2017–2018. *Molecules* **2019,** *24* (6), 1135.
6. Bordes, R.; Holmberg, K., Amino acid-based surfactants – do they deserve more attention? *Advances in Colloid and Interface Science* **2015,** *222*, 79-91.
7. Billiot, E.; Macossay, J.; Thibodeaux, S.; Shamsi, S. A.; Warner, I. M., Chiral Separations Using Dipeptide Polymerized Surfactants: Effect of Amino Acid Order. *Analytical Chemistry* **1998,** *70* (7), 1375-1381.

8. Shamsi, S. A.; Valle, B. C.; Billiot, F.; Warner, I. M., Polysodium N-Undecanoyl-l-leucylvalinate: A Versatile Chiral Selector for Micellar Electrokinetic Chromatography. *Analytical Chemistry* **2003,** *75* (3), 379-387.
9. Billiot, F. H.; McCarroll, M.; Billiot, E. J.; Rugutt, J. K.; Morris, K.; Warner, I. M., Comparison of the Aggregation Behavior of 15 Polymeric and Monomeric Dipeptide Surfactants in Aqueous Solution. *Langmuir* **2002,** *18* (8), 2993-2997.
10. Billiot, F. H.; Billiot, E. J.; Warner, I. M., Comparison of monomeric and polymeric amino acid based surfactants for chiral separations. *Journal of Chromatography A* **2001,** *922* (1), 329-338.
11. Billiot, A.; Fang, Y.; Morris, K., Characterization of Amino Acid Based Molecular Micelles with Molecular Modeling. *Open Journal of Physical Chemistry* **2019,** *09*, 221-240.
12. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., A Molecular Dynamics Simulation Study of Two Dipeptide Based Molecular Micelles: Effect of Amino Acid Order. *Open journal of physical chemistry* **2013,** *3* (1), 20-29.
13. Ramos, Z.; Rothbauer, G. A.; Turner, J.; Lewis, C.; Morris, K.; Billiot, E.; Billiot, F.; Fang, Y., Comparison of Chiral Recognition of Binaphthyl Derivatives with l-Undecyl-Leucine Surfactants in the Presence of Arginine and Sodium Counterions. *Journal of Chromatographic Science* **2019,** *57* (1), 54-62.
14. Montanari, C. A.; Cass, Q. B.; Tiritan, M. E.; Souza, A. L. S. d., A QSERR study on enantioselective separation of enantiomeric sulphoxides. *Analytica Chimica Acta* **2000,** *419* (1), 93-100.
15. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005,** *26* (16), 1668-1688.
16. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157-74.
17. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020,** *16* (1), 528-552.
18. Vilar, S.; Cozza, G.; Moro, S., Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem* **2008,** *8* (18), 1555-72.
19. Le Guilloux, V.; Schmidtke, P.; Tuffery, P., Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **2009,** *10*, 168-168.
20. Galli, C. L.; Sensi, C.; Fumagalli, A.; Parravicini, C.; Marinovich, M.; Eberini, I., A computational approach to evaluate the androgenic affinity of iprodione, procymidone, vinclozolin and their metabolites. *PloS one* **2014,** *9* (8), e104822-e104822.
21. Cuzzolin, A.; Sturlese, M.; Malvacio, I.; Ciancetta, A.; Moro, S., DockBench: An Integrated Informatic Platform Bridging the Gap between the Robust Validation of Docking Protocols and Virtual Screening Simulations. *Molecules* **2015,** *20* (6).

CHAPTER II: CHIRAL RECOGNITION OF BINAPHTHYL DERIVATIVES WITH L-UNDECYL LEUCINE SURFACTANTS IN THE PRESENCE OF SODIUM AND LYSINE COUNTERIONS

2.1 Abstract

This study investigates the effect of counterions on the chiral recognition of 1,1′-Binaphthyl-2,2′-diamine (BNA) and 1,1′-Binaphthyl-2,2′-diyl hydrogenphosphate (BNP) enantiomers when using an amino acid-based surfactant as the chiral pseudostationary phase in capillary electrophoresis. The effects of sodium counterions on the chiral recognition of binaphthyl derivatives were compared to that of using pH-dependent Lysine counterions at varying pH conditions.

The enantiomeric separation of BNP and BNA enantiomers via capillary electrophoresis, using L-Undecyl-Leucine (und-Leu) as the chiral recognition medium, significantly improved the enantiomeric resolution in capillary electrophoresis at pH 7 when using Lysine counterions as compared to using sodium counterions. More specifically, at a surfactant concentration of 45 mM, Lysine counterions at pH 7 significantly increased the enantiomeric resolution of BNA and BNP by 6-fold and 1.1-fold, respectively, in capillary electrophoresis experiments compared to using sodium as the counterion. Furthermore, the retention factor of BNA and BNP enantiomers also increased approximately 3.5-fold and 4-fold, respectively, in the presence of Lysine counterions as compared to using sodium counterions. The running buffer in capillary electrophoresis was increased to pH 11, and capillary electrophoresis experiments determined that when separating BNA and BNP enantiomers, the resolution and retention factors were nearly identical when comparing the effects of the sodium and Lysine counterions. This signifies the important role of Lysine's positive net charge on chiral recognition.

This study provides insight into the advantages of using cationic, pH-dependent counterions such as Lysine to significantly improve the chiral recognition of binaphthyl derivatives when using anionic surfactants as the pseudostationary phase in capillary electrophoresis.

2.2 Introduction

Chirality is ubiquitous in nature. Two of the most important simple class of chiral compounds include amino acids, and sugars. Sugars and amino acids serve as the building blocks for a large percentage of biological compounds.[1-6] This suggests the profound influence of chirality on fundamental physiological and biological processes. Furthermore, stereospecific reactions play an essential role in drug metabolism, cell membrane stability and gene expression.[7-12]

Due to the aforementioned stereospecific reactions that occur within the body, it is not surprising that more than 50% of marketed drugs are chiral.[13-16] These synthesized pharmaceutical drugs, however, frequently yield racemic mixtures, meaning it contains both enantiomers. This is of major concern because each enantiomer of a drug may exhibit different pharmacological effects. For example, Thalidomide was originally marketed as a racemic mixture. However, this drug caused a global tragedy in the late 1950s, contributing to an alarming rise in teratogenic effects in newborns due to one enantiomer being harmful.[17, 18] In response to this tragedy, the Food and Drug Administration has mandated that each enantiomer of a chiral drug be evaluated for its respective physiological effects prior to being marketed.[13, 16, 19-23] This gave rise to a new scientific challenge: to establish and optimize enantiomeric separation processes and techniques.

Since then, many techniques have been established to separate enantiomeric compounds. Two of the more common techniques are high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE).[19, 21, 24-27] CE typically yields a high number of theoretical plates compared to HPLC and thus CE often yields better enantiomeric separations in a shorter period of time.[28, 29] In addition, CE allows for a quick and easy way to change the chiral recognition medium since the chiral recognition medium is part of the mobile phase acting as a pseudo-stationary phase. This is in contrast to HLPC which requires the purchase and installation of different analytical columns if one wishes to change chiral recognition medium. As with HPLC, a wide variety of chiral recognition media exist including but not limited to cyclodextrins, crown ethers and chiral micelles.[21, 24-27] This research focuses on the latter, chiral micelles.

In particular, the chiral micelles described in this thesis are amino acid-based micelles (AABMs). It is worth noting that when micelles are used as the pseudostationary phase in CE, this technique is known as micellar electrokinetic chromatography (MEKC). MEKC is a well-established technique for the enantiomeric separation of chiral compounds.[21, 24-26, 30-34] Amino acid-based surfactants are composed of a non-polar hydrocarbon chain and an amino acid head group.[35-40] The charge on the amino acid head groups can be greatly influenced by pH. The ability to impose different pH environments aids in studying how the charge of the surfactant, analytes, and counterions may affect the physiochemical properties of the micelles, as well as its ability to act as an effective chiral separation medium.[30, 34] All of which greatly influence the chiral recognition ability of the AABMs.

Previously, the effects of amino acid order, steric hindrance, dihedral angles and the polymerization of AABMs have been studied experimentally and computationally to investigate

their contributions to chiral recognition.[21, 30-33] For many years, sodium has been extensively used as the counterion for the MEKC buffer solution.[21, 22, 30-34] Most recently, the effects of pH-dependent counterions have been used to significantly improve the resolution of enantiomeric compounds in MEKC.[30, 34] As previously reported from our research group, pH-dependent counterions, such as Arginine have been demonstrated to significantly improve the enantiomeric resolution of various binaphthyl derivatives, when compared to using monatomic sodium as the counterion.[34]

To further investigate the effects of pH-dependent counterions on chiral selectivity, this study investigates the role of Lysine as the counterion on the chiral separations of BNP and BNA enantiomers at pH 7 and 11 at varying concentrations of und-Leu surfactant. The structures of these surfactants, analytes, and Lysine are provided in Figure 2.1 below.

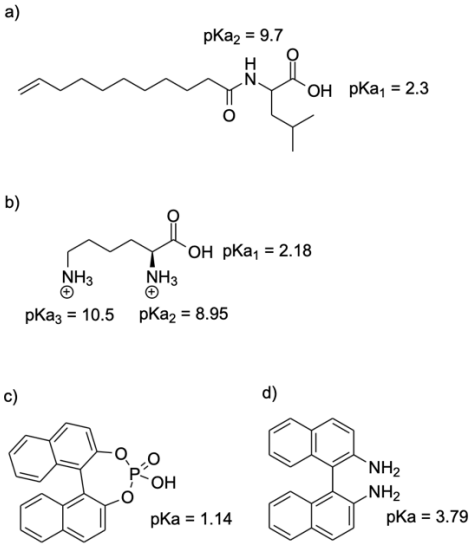


Figure 2.1. Structure of (a) L-Undecyl-Leucine surfactant (b) Lysine (c) BNP (d) BNA.

As shown in that figure, Lysine contains two amine group; in addition to the one the chiral hydrogen, lysine has another amine group on the side chain. The pKa values are approximately 9 on the chiral hydrogen and ~ 10.5 on the side chain amine group. As shown in Figure 2.2, Lysine has various charge states at different pH levels.

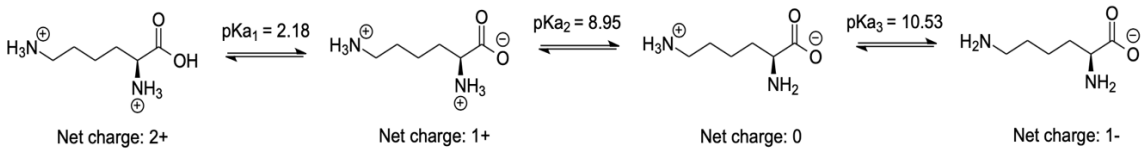


Figure 2.2. Representation of the charged states and isoelectric point of Lysine.

The various charge states shown in Figure 2.2 can have a significant effect on the physicochemical properties and chiral recognition ability of the micelles formed. However, due to the solubility limitations of the micelles, we cannot work below pH 7, nor above pH 11 due to the limitations of capillary electrophoresis. Therefore, we limited our experiments to the useful pH extremes of our system, pH 7 and 11.

2.3 Methods

2.3.1 Chemicals

Leucine, Lysine and racemic mixtures of binaphthyl derivatives [1,1′-Binaphthyl-2,2′-diyl hydrogenphosphate (BNP) and 1,1′-Binaphthyl-2,2′-diamine (BNA)] were purchased from Sigma-Aldrich (St. Louis, MO). The undecyl L-Leucine surfactant was synthesized from the N-hydroxysuccinimide ester of undecylenic acid according to a previously reported procedure.[41, 42]

2.3.2 Capillary Electrophoresis

Chiral separations were performed using a Hewlett-Packard (HP) 3D CE model #G7100A. The fused silica capillary [effective length of 45 cm (to detection window), 50-µm i.d., with a total length of 56 cm] was purchased from Agilent Technologies (Lake Jackson, TX) and mounted in an HP capillary cartridge. The temperature of the cartridge was maintained at 25°C throughout these experiments. Solutions of 45 mM undecyl L-leucine with Lysine and sodium were prepared in a 5 mM sodium borate buffer and pH was adjusted to values of 7 and 11 with the use of NaOH and HCl. These solutions were diluted to concentrations ranging from 15 to 45 mM and filtered through a 0.45-µm syringe filter before use. Analyte standards were prepared in 1:1 methanol–water at 0.1 mg/mL. Samples were injected for 5 s at 10 mbar pressure. Separations were performed at +30 kV, with UV detection at 230 nm.

2.4 Results

In this study, the chiral recognition of und-Leu at varying concentrations in the presence of sodium and L-Lysine counterions was investigated at pH 7 and 11. We separated enantiomers of BNP and BNA with varying concentrations of surfactant ranging from 15 to 45 mM, with 5 mM concentration intervals. As shown in Figure 2.3a, pH 7 provided a better resolution for separating BNP enantiomers in the presence of Lysine as compared to sodium. The retention factor (k') values shown in Figure 2.3b indicate that at pH 7, enantiomers of BNP interacted stronger when Lysine was used as the counterion compared to sodium.

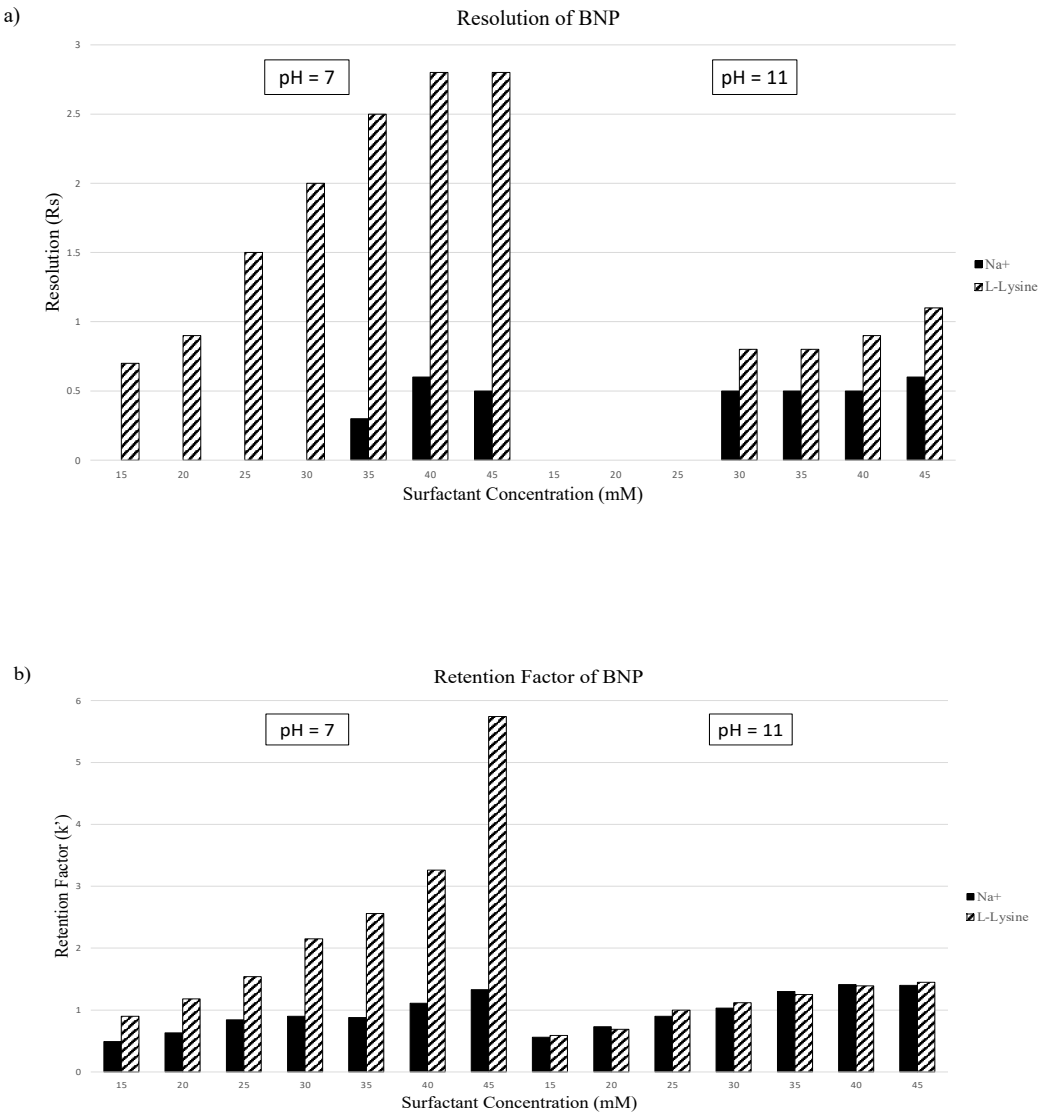


Figure 2.3. Comparison of the (a) resolution of BNP enantiomers and (b) k' of BNP enantiomers in the presence of Na+ and L-Lysine counterions, at concentrations of und-Leu ranging from 15 mM to 45 mM (in 5 mM increments), at pH 7 and 11.

At pH 7 and a surfactant concentration of 45 mM, the k' value for BNP was ~5.8, whereas this value was ~1.2 in the presence of sodium. As pH increased to 11, the retention factor for the enantiomers of BNP were similar in the presence of both counterions, Lysine and sodium.

The resolution and retention factors for enantiomers of BNA in the presence of Lysine and sodium counterions are shown in Figure 2.4. At pH 7, with und-Leu concentration of 15 mM, in the presence of lysine, the enantiomers of BNA were separated with a resolution of ~1.8, whereas at the same concentration and pH, a resolution of ~0.4 was observed in the presence of sodium.

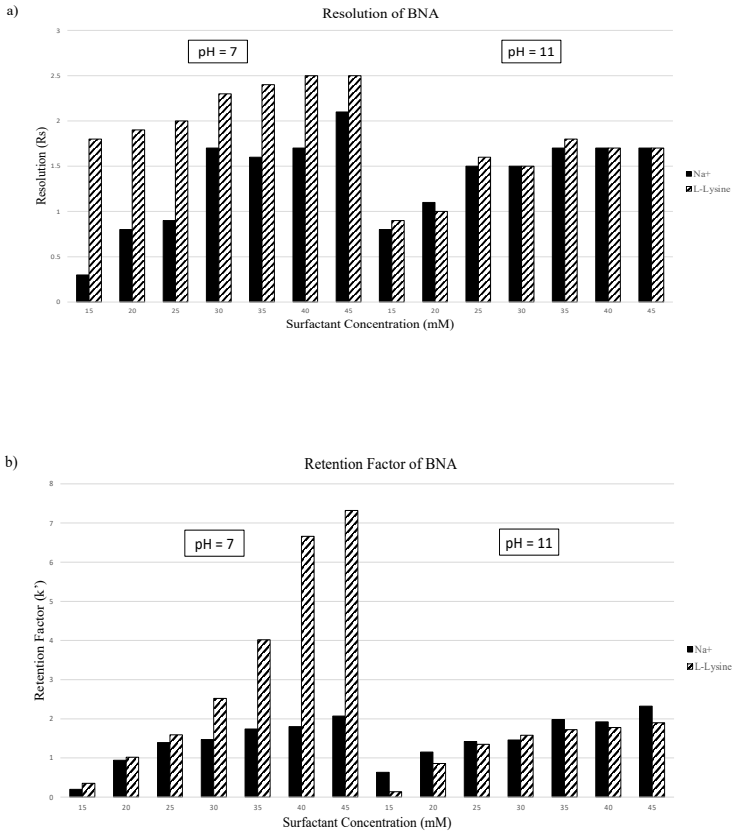


Figure 2.4. Comparison of the (a) resolution of BNA enantiomers and (b) k' of BNA enantiomers in the presence of Na+ and L-Lysine counterions, at concentrations of und-Leucine ranging from 15 mM to 45 mM (in 5 mM increments), at pH 7 and 11.

Similar to BNP, at higher pH levels, the resolution and retention factors for BNA were approximately the same for both, sodium and Lysine counterions. Shown in Figure 2.5a-b below is an electropherogram comparing the separation of BNP enantiomers at pH 7 and 25 mM in the presence of sodium and Lysine. As can be seen in Figure 2.5a, no separation was achieved with sodium as the counterion but as shown in Figure 2.5b, a resolution of ~ 1.5 was achieved when Lysine was used as the counterion.

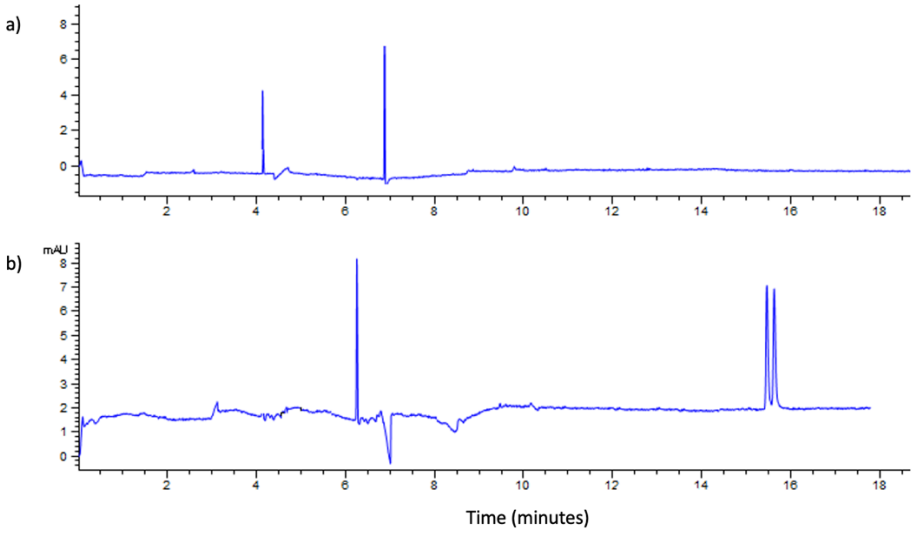


Figure 2.5 Electropherogram of BNP at pH 7 and surfactant concentration of 25 mM comparing separations in the presence of (a) sodium counterions and (b) Lysine counterions.

2.5 Discussion

Chiral recognition of amino acid-based surfactants is strongly dependent on molecular interactions such as: hydrogen bonding, electrostatic attraction, steric hindrance and the hydrophobic effect. Surfactants with amino acid head groups contain amide and carboxylic acid

moieties, in which the pH is expected to affect the electrostatic and hydrogen bonding capabilities of the polar head constituents of the surfactants. In a previous study, our research group investigated the effects of chiral recognition of und-Leu surfactant in the presence of Arginine and sodium counterions.[34] In the aforementioned study, the BNP enantiomers were separated via MEKC at pH 7 in the presence of Arginine and sodium, with resolutions of ~4.1 and 0.6, respectively. When sodium was used as the counterion, baseline resolution of BNP was not observed. In contrast, the BOH enantiomers were separated slightly better in the presence of sodium. Results of that study suggests that the counterion plays a significant role in chiral recognition with amino acid-based surfactants. This motivates the current investigation of comparing the effect of pH on chiral recognition of BNP and BNA enantiomers with und-Leu surfactants in presence of Lysine and sodium counterions.

As previously discussed, Lysine contains two amine groups; one connected to the chiral carbon and another on the side chain. As also previously mentioned, at lower pH levels both of the amine moieties are protonated, providing Lysine with a net positive charge, as shown in Figure 2.2. This net positive charge on Lysine allows it to act as a counterion for the negatively charged amino acid-based surfactants. The presence of amine groups on Lysine allows for hydrogen bonds and electrostatic attraction to occur with the BNP and BNA enantiomers, as well as with the surfactant polar head group. These intermolecular interactions cannot occur with sodium counterions as it is monatomic and does not contain hydrogen bonding moieties. As previously reported, the amount of Lysine molecules bound to the surfactant (fraction bound, f_b) changes significantly as a function of pH.[30, 34] At pH 7, ~ 37% of the Lysine molecules are bound to the surfactant. This number drops to ~ 3% at pH 11.[30] As seen in Figure 2.2, Lysine has a net positive charge at pH 7, allowing for stronger electrostatic interactions between Lysine and the negatively

charged surfactants. However, this attraction is significantly reduced as it is subjected to higher pH levels.

2.5.1 Separation of BNP Enantiomers

The separation of the BNP enantiomers via MEKC with und-Leu surfactants in the presence of Lysine and sodium counterions were compared at pH 7 and 11. The resolution and k' values of the BNP enantiomers are shown in Figures 2.3a-b.

Enantiomeric separation improves as the surfactant concentration increases from 15 to 45 mM in the presence of Lysine counterions at pH 7. At the same pH, the best resolution observed with sodium counterions was ~0.6 at 40 mM. This resolution was better in the presence of Lysine, which was ~1.1 at 45 mM. The separation of BNP enantiomers significantly decreased at pH 11 in the presence of both, sodium and Lysine counterions. As previously mentioned, a possible contributing factor is that at lower pH levels, Lysine counterions have a net positive charge, allowing for stronger electrostatic attraction to occur intermolecularly with the negatively charged surfactant head groups, compared higher pH levels.

The enantiomers of BNP were separated with a resolution of ~ 2.0 at a surfactant concentration of 30 mM at pH 7 in the presence of Lysine, whereas no baseline separation was observed at the same pH and concentrations in the presence of sodium. Comparing retention factors in Figure 2.3b shows that the enantiomers of BNP interact stronger with the und-Leu surfactants in the presence of Lysine counterions than that of sodium counterions. At pH 7 and 45 mM surfactant concentration of und-Leu, a k' value of 5.8 was observed with Lysine present, which is significantly higher than the k' value of 1.1 when sodium was utilized as the counterion.

Baseline resolution of BNP enantiomers was achieved at pH 7 with a surfactant concentration of 25 mM in the presence of Lysine, producing a k' value of ~1.54. In contrast, baseline separation was not observed in the presence of sodium counterions. As previously reported, the physical properties of und-Leu in the presence of Lysine and sodium are quite different. The critical micelle concentration (CMC) was determined to be ~17-18 mM in the presence of either counterion. However, at pH 7, the hydrodynamic radius (R_h) of und-Leu is larger in the presence of Lysine, than that of sodium. The R_h values were reported to be ~12.8 Å and ~10.9 Å for Lysine and sodium counterions, respectively. The difference in chiral recognition in the presence of Lysine may be due to the interactions of this counterion with the charged head groups of the surfactants. Lysine counterions are positively charged at pH 7, which may participate in electrostatic attractions and hydrogen bonding interactions with the micelle at this pH, attributing to the improvement in chiral selectivity of the BNP enantiomers, as shown in Figure 2.6.

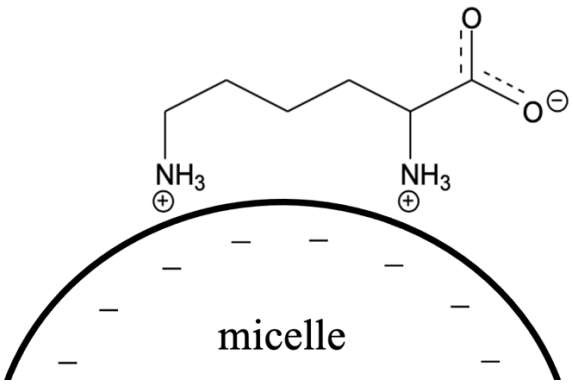


Figure 2.6. Positively charged Lysine interacting with anionic micelle. At pH 7, Lysine has a net charge of +2, therefore it is able to act as a counterion, bridging anionic surfactant head groups. Intermolecular bonding is not evident with sodium as it is monatomic.

Contrarily, at pH 11, Lysine has an overall net charge of zero thus losing its charged properties to act as an effective counterion. Due to the loss of positive charges, we therefore expect it to have less electrostatic and hydrogen bonding interactions with the negatively-charged surfactants. This is due to Lysine losing its effectiveness as a counterion at higher pH levels, which causes sodium ions to then act as the predominant counterion. Moreover, the interaction of BNP enantiomers with und-Leu are similar in the presence of sodium or neutralized Lysine. Further evidence of similar interactions is shown in Figure 2.3b, where similar retention factors for BNP were observed when comparing the presence of sodium or Lysine counterions.

2.5.2 Separation of BNA Enantiomers

Similar to BNP, the BNA enantiomers exhibited better separation in the presence of Lysine at pH 7. At pH 7, resolutions of 1.8, 2.0, 2.4 and 2.5 were observed at 15, 25, 35 and 45 mM, respectively, in the presence of Lysine counterions, as presented in Figure 2.4.

These resolution values far exceed those observed in the presence of sodium at the same pH level. As seen in Figure 2.4a, the BNA enantiomers were separated with resolutions of 0.4, 0.9, 1.6 and 2.1 at concentrations of 15, 25, 35 and 45 mM, respectively in the presence of sodium counterions. As previously mentioned, the CMC of und-Leu is approximately the same (~17-18 mM) in the presence of sodium and Lysine. However, at pH 7, baseline separation of BNA enantiomers was observed at 15 mM. Furthermore, the k' values indicate that BNA enantiomers do not bind strongly to und-Leu at this surfactant concentration. Overall, the interaction of BNA enantiomers is very effective at low pH levels and low surfactant concentrations. Below the CMC, Lysine counterions still interact with und-Leu surfactants. At pH 7, Lysine molecules are positively

charged, therefore, electrostatic attraction between the positively charged amine moieties of Lysine and the negatively charged und-Leu draw these two molecules closer together. This allows Lysine to form hydrogen bonds with the surfactant polar head groups thus providing a chiral cavity that can improve its chiral selectivity of BNA enantiomers.

2.5.3 Effect of Counterion Amino Acid Chirality on Chiral Recognition

To determine if the chirality of Lysine played a role in chiral recognition, two experiments were performed. In one experiment, 50 mM L-Lysine was used without und-Leu, and no chiral recognition was observed. In another experiment, 20 mM D-lysine and 20 mM L-und-leu were used as the running buffer in MEKC. No difference in chiral recognition of BNA enantiomers was observed when either D- or L-Lysine was utilized as the counterion. We hypothesize that Lysine counterions may form small aggregates that provide an environment for selective chiral recognition of BNA enantiomers. If this hypothesis holds true, then using D-Lysine as the counterion should make a difference in the resolution of BNA enantiomers. Therefore, aggregates formed by D-Lysine should act in an opposite manner of L-Lysine, resulting in either reversal of enantiomeric order or reduction of resolution value. This was not observed when D-Lysine was used as the counterion.

2.6 Conclusion

The enantiomeric resolution of BNP significantly improved in the presence of Lysine counterions as compared to sodium counterions at pH 7 when using und-Leu surfactants as the

chiral recognition medium. Most notably, when using a surfactant concentration of 45 mM in the presence of Lysine counterions at pH 7, the enantiomeric resolution increased approximately 6-fold compared to that of when using sodium counterions. Furthermore, at the same conditions, the retention factor increased approximately 4-fold when using Lysine counterions, as compared to using sodium. However, with experimental conditions at pH 11, the enantiomeric resolution and retention factor are nearly identical when separating BNP enantiomers with either Lysine or sodium counterions.

At pH 7, the enantiomeric resolution of BNA enantiomers was achieved using und-Leu surfactants in the presence of Lysine counterions. This improved the enantiomeric resolution values as compared to when using sodium counterions. For example, at a surfactant concentration of 45 mM in the presence of Lysine counterions at pH 7, the enantiomeric resolution increased approximately 1.1-fold compared to that of when using sodium counterions. Furthermore, the retention factor at those same aforementioned conditions increased approximately 3.5-fold when using Lysine counterions, as compared to using sodium. When separating BNA enantiomers using 45 mM concentration of und-Leu surfactant at pH 11, the enantiomeric resolution and retention factors are nearly identical to that of when using sodium counterions.

In conclusion, when separating chiral compounds such as BNA and BNP with und-Leu surfactants in the presence of pH-dependent counterions such as Lysine via MEKC, the enantiomeric resolution and retention factors significantly improved when compared to using sodium counterions in these studies. Therefore, this study provides insight to further optimize chiral separation conditions using pH-dependent counterions as opposed to monatomic counterions such as sodium.

2.7 References

1. Gou, K.; Wang, Y.; Xie, L.; Guo, X.; Guo, Y.; Ke, J.; Wu, L.; Li, S.; Li, H., Synthesis, structural properties, biosafety and applications of chiral mesoporous silica nanostructures. *Chemical Engineering Journal* **2020**, 127862.
2. Ma, Y.; Shi, L.; Yue, H.; Gao, X., Recognition at chiral interfaces: From molecules to cells. *Colloids and Surfaces B: Biointerfaces* **2020,** *195*, 111268.
3. Wu, Q.; Lv, H.; Zhao, L., Applications of carbon nanomaterials in chiral separation. *TrAC Trends in Analytical Chemistry* **2020,** *129*, 115941.
4. Tohala, L.; Oukacine, F.; Ravelet, C.; Peyrin, E., Chiral Resolution Capabilities of DNA Oligonucleotides. *Analytical Chemistry* **2015,** *87* (11), 5491-5495.
5. Liang, R.-P.; Wang, X.-N.; Liu, C.-M.; Meng, X.-Y.; Qiu, J.-D., Facile preparation of protein stationary phase based on polydopamine/graphene oxide platform for chip-based open tubular capillary electrochromatography enantioseparation. *Journal of Chromatography A* **2014,** *1323*, 135-142.
6. Kondepudi, D., Chapter 1 - Chiral Asymmetry in Nature. In *Chiral Analysis (Second Edition)*, Polavarapu, P. L., Ed. Elsevier: 2018; pp 3-28.
7. Battisti, U. M.; Citti, C.; Larini, M.; Ciccarella, G.; Stasiak, N.; Troisi, L.; Braghiroli, D.; Parenti, C.; Zoli, M.; Cannazza, G., "Heart-cut" bidimensional achiral-chiral liquid chromatography applied to the evaluation of stereoselective metabolism, in vivo biological activity and brain response to chiral drug candidates targeting the central nervous system. *Journal of Chromatography A* **2016,** *1443*, 152-161.
8. Du, Z.; Liu, C.; Song, H.; Scott, P.; Liu, Z.; Ren, J.; Qu, X., Neutrophil-Membrane-Directed Bioorthogonal Synthesis of Inflammation-Targeting Chiral Drugs. *Chem* **2020,** *6* (8), 2060-2072.
9. Ding, F.; Peng, W.; Peng, Y.-K.; Liu, B.-Q., Estimating the potential toxicity of chiral diclofop-methyl: Mechanistic insight into the enantioselective behavior. *Toxicology* **2020,** *438*, 152446.
10. Lalitha, S.; Sampath Kumar, A.; Stine, K. J.; Covey, D. F., Chirality in Membranes: First Evidence that Enantioselective Interactions Between Cholesterol and Cell Membrane Lipids Can Be a Determinant of Membrane Physical Properties. *Journal of Supramolecular Chemistry* **2001,** *1* (2), 53-61.

11. Dong, L.; Gong, J.; Wang, Y.; He, J.; You, D.; Zhou, Y.; Li, Q.; Liu, Y.; Cheng, K.; Qian, J.; Weng, W.; Wang, H.; Yu, M., Chiral geometry regulates stem cell fate and activity. *Biomaterials* **2019,** *222*, 119456.

12. Wang, M.; Jiang, W.; Liu, X.; Wang, J.; Zhang, B.; Fan, C.; Liu, L.; Pena-Alcantara, G.; Ling, J.-J.; Chen, J.; Zhu, T. F., Mirror-Image Gene Transcription and Reverse Transcription. *Chem* **2019,** *5* (4), 848-857.

13. Nguyen, L. A.; He, H.; Pham-Huy, C., Chiral drugs: an overview. *Int J Biomed Sci* **2006,** *2* (2), 85-100.

14. Rentsch, K. M., The importance of stereoselective determination of drugs in the clinical laboratory. *J Biochem Biophys Methods* **2002,** *54* (1-3), 1-9.

15. McConathy, J.; Owens, M. J., Stereochemistry in Drug Action. *Prim Care Companion J Clin Psychiatry* **2003,** *5* (2), 70-73.

16. Brooks, W. H.; Guida, W. C.; Daniel, K. G., The significance of chirality in drug design and development. *Curr Top Med Chem* **2011,** *11* (7), 760-770.

17. Burley, D. M.; Lenz, W., THALIDOMIDE AND CONGENITAL ABNORMALITIES. *The Lancet* **1962,** *279* (7223), 271-272.

18. Melchert, M.; List, A., The thalidomide saga. *The International Journal of Biochemistry & Cell Biology* **2007,** *39* (7), 1489-1499.

19. Liu, J.-T.; Liu, R. H., Enantiomeric composition of abused amine drugs: chromatographic methods of analysis and data interpretation. *Journal of Biochemical and Biophysical Methods* **2002,** *54* (1), 115-146.

20. Wright, A. K.; Batsomboon, P.; Dai, J.; Hung, I.; Zhou, H.-X.; Dudley, G. B.; Cross, T. A., Differential Binding of Rimantadine Enantiomers to Influenza A M2 Proton Channel. *Journal of the American Chemical Society* **2016,** *138* (5), 1506-1509.

21. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., A Molecular Dynamics Simulation Study of Two Dipeptide Based Molecular Micelles: Effect of Amino Acid Order. *Open journal of physical chemistry* **2013,** *3* (1), 20-29.

22. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Ingle, J. A.; Krause, K. B.; Lewis, C. R.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., Using molecular dynamics simulations to identify the key factors responsible for chiral recognition by an amino acid-based molecular micelle. *Journal of Dispersion Science and Technology* **2019,** *40* (5), 716-727.

23. FDA's policy statement for the development of new stereoisomeric drugs. *Chirality* **1992,** *4* (5), 338-40.

24. Konjaria, M.-L.; Scriba, G. K. E., Enantioseparation of alanyl-phenylalanine analogs by capillary electrophoresis using negatively charged cyclodextrins as chiral selectors. *Journal of Chromatography A* **2020,** *1632*, 461585.

25. Lee, S.; Kim, S.-J.; Bang, E.; Na, Y.-C., Chiral separation of intact amino acids by capillary electrophoresis-mass spectrometry employing a partial filling technique with a crown ether carboxylic acid. *Journal of Chromatography A* **2019,** *1586*, 128-138.

26. Sánchez-Hernández, L.; Castro-Puyana, M.; Marina, M. L.; Crego, A. L., Recent approaches in sensitive enantioseparations by CE. *Electrophoresis* **2012,** *33* (1), 228-42.

27. Zhu, X.; Chen, C.; Chen, J.; Xu, G.; Du, Y.; Ma, X.; Sun, X.; Feng, Z.; Huang, Z., Synthesis and application of tetramethylammonium-carboxymethylated-β-cyclodextrin: A novel ionic liquid in capillary electrophoresis enantioseparation. *Journal of Pharmaceutical and Biomedical Analysis* **2020,** *180*, 113030.

28. Datta, R., Theoretical evaluation of capillary electrophoresis performance. *Biotechnology Progress* **1990,** *6* (6), 485-493.

29. Engelhardt, H.; Cuñat-Walter, M. A., Use of plate numbers achieved in capillary electrophoretic protein separations for characterization of capillary coatings. *Journal of Chromatography A* **1995,** *717* (1), 15-23.

30. Lewis, C.; Hughes, B. H.; Vasquez, M.; Wall, A. M.; Northrup, V. L.; Witzleb, T. J.; Billiot, E. J.; Fang, Y.; Billiot, F. H.; Morris, K. F., Effect of pH on the Binding of Sodium, Lysine, and Arginine Counterions to l-Undecyl Leucinate Micelles. *Journal of Surfactants and Detergents* **2016,** *19* (6), 1175-1188.

31. Billiot, A.; Fang, Y.; Morris, K., Characterization of Amino Acid Based Molecular Micelles with Molecular Modeling. *Open Journal of Physical Chemistry* **2019,** *09*, 221-240.

32. Billiot, E.; Macossay, J.; Thibodeaux, S.; Shamsi, S. A.; Warner, I. M., Chiral Separations Using Dipeptide Polymerized Surfactants: Effect of Amino Acid Order. *Analytical Chemistry* **1998,** *70* (7), 1375-1381.

33. Haddadian, F.; Billiot, E. J.; Shamsi, S. A.; Warner, I. M., Chiral separations using polymeric dipeptide surfactants: effect of number of chiral centers and steric factors. *Journal of Chromatography A* **1999,** *858* (2), 219-227.

34. Ramos, Z.; Rothbauer, G. A.; Turner, J.; Lewis, C.; Morris, K.; Billiot, E.; Billiot, F.; Fang, Y., Comparison of Chiral Recognition of Binaphthyl Derivatives with l-Undecyl-Leucine Surfactants in the Presence of Arginine and Sodium Counterions. *Journal of Chromatographic Science* **2019,** *57* (1), 54-62.

35. Tackie-Otoo, B. N.; Ayoub Mohammed, M. A., Experimental investigation of the behaviour of a novel amino acid-based surfactant relevant to EOR application. *Journal of Molecular Liquids* **2020,** *316*, 113848.

36. Fawzy, A.; Abdallah, M.; Zaafarany, I. A.; Ahmed, S. A.; Althagafi, I. I., Thermodynamic, kinetic and mechanistic approach to the corrosion inhibition of carbon steel by new synthesized amino acids-based surfactants as green inhibitors in neutral and alkaline aqueous media. *Journal of Molecular Liquids* **2018,** *265*, 276-291.

37. Pinazo, A.; Manresa, M. A.; Marques, A. M.; Bustelo, M.; Espuny, M. J.; Pérez, L., Amino acid–based surfactants: New antimicrobial agents. *Advances in Colloid and Interface Science* **2016,** *228*, 17-39.

38. Perinelli, D. R.; Casettari, L.; Cespi, M.; Fini, F.; Man, D. K. W.; Giorgioni, G.; Canala, S.; Lam, J. K. W.; Bonacucina, G.; Palmieri, G. F., Chemical–physical properties and cytotoxicity of N-decanoyl amino acid-based surfactants: Effect of polar heads. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2016,** *492*, 38-46.

39. Pinazo, A.; Pons, R.; Pérez, L.; Infante, M. R., Amino Acids as Raw Material for Biocompatible Surfactants. *Industrial & Engineering Chemistry Research* **2011,** *50* (9), 4805-4817.

40. Clapés, P.; Rosa Infante, M., Amino Acid-based Surfactants: Enzymatic Synthesis, Properties and Potential Applications. *Biocatalysis and Biotransformation* **2002,** *20* (4), 215-233.

41. Ramos, Z.; Rothbauer, G. A.; Turner, J.; Lewis, C.; Morris, K.; Billiot, E.; Billiot, F.; Fang, Y., Comparison of Chiral Recognition of Binaphthyl Derivatives with l-Undecyl-Leucine Surfactants in the Presence of Arginine and Sodium Counterions. *J Chromatogr Sci* **2019,** *57* (1), 54-62.

42. Billiot, F. H.; Billiot, E. J.; Warner, I. M., Depth of penetration of binaphthyl derivatives into the micellar core of sodium undecenoyl leucyl-leucinate surfactants. *J Chromatogr A* **2002,** *950* (1-2), 233-9.

CHAPTER III: COMPUTATIONAL METHOD DEVELOPMENT

3.1 Purpose of Computational Studies

It is crucial to develop a QSERR model that will efficiently predict optimal chiral separation conditions as it will be time and cost-efficient when trying to separate chiral compounds in the future. Due to the aforementioned advantages of polymerized micelles over traditional micelles, as well as the advantages of dipeptide surfactants over single amino acid surfactants, the computational studies described in this thesis focus on studying the amino acid-based molecular micelles, which are polymerized dipeptide surfactants. A significant issue that needs to be addressed in order to successfully develop a predictive and valid QSERR model is to first ensure that the computational studies accurately reflect the experimental conditions of these systems. More specifically, the computational amino acid-based molecular micelle structures must closely reflect that of the actual structures of these micelles. Although the exact structures of these AABMMs are not well-known, we know that the hydrophobic regions of the surfactants should face the micellar core, and the hydrophilic amino acid head group should be facing the aqueous solution. In addition, the AABMMs should generally form a somewhat spherical structure.

Furthermore, when investigating these AABMMs in conjunction with various enantiomers, it is important that these interactions accurately reflect observations from experimental studies. The experimental MEKC elution order data provides insight into relative binding free energy values of the enantiomers to the AABMMs. The experimental NMR studies can provide insight into the primary molecular interactions that are occurring between the AABMMs and the enantiomers. However, even if the AABMM structures are computationally built to be representative of the actual AABMM structures, there is still the task of properly defining the AABMMs in terms of applying and mixing various force fields so as to fine-tune the computational

systems to accurately predict experimental observations. Furthermore, there are many software packages that may be used to conduct MD simulations, and there are also multiple ways of computationally calculating the binding free energy values. As such, there are a variety of ways to better finetune the computational system so that it may accurately reflect experimental observations.

3.2 Automation of Molecular Dynamics Simulation: Overview

After the AABMMs have been created with a method that produces representative structures of the micelles, we ran molecular dynamics simulations to determine if the computational binding free energy values produced within this system match the experimental MEKC elution order data. However, in order to analyze all combinations of AABMMs with enantiomers, it is important to automate this process to (1) reduce human error, but most importantly (2) to be more efficient with the time, as there are many AABMM:enantiomers combinations to be evaluated via MD simulations. To automate the system, we must know the general overview of what the process looks like. As such, the general automation overview is shown in Figure 3.1, and summarized as follows: (1) design the AABMMMs that properly represents the micelle and is able to undergo simulations without crashing; (2) dock the AABMMs with a set of enantiomeric compounds; (3) run MD simulations of the enantiomers in conjunction with the AABMM that it was docked to; (4) conduct post-trajectory analysis to obtain binding free energy values, using both, the MMGBSA and MMPSA calculations.

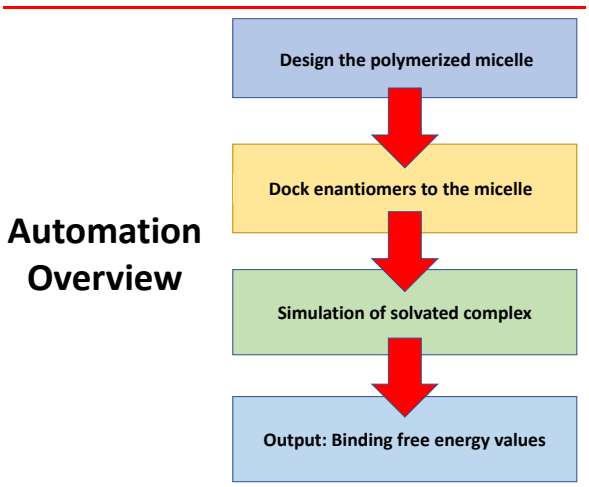


Figure 3.1. Automation overview for running simulations and evaluating accuracy.

This thesis aims to provide a general framework for the method development of building and optimizing the molecular modeling aspects of these AABMM systems, which is crucial in working towards the development of the QSERR model. As mentioned previously there are many combinations of conditions to test, therefore, it is essential to automate the computational studies to allow it to be as time-effective as possible. Figure 3.1 shows a big picture of the steps that need to be developed. Figure 3.1 shows that one of the first major goals is to computationally design the polymerized micelles, to then dock various enantiomers to these micelles, and then run simulations of the micelle:enantiomer solvated complex, and then run binding free energy calculations on these systems being studied.

As such, the first major step is to computationally develop the AABMMs to be accurate representations of how they actually exist experimentally. The method development and optimization processes will be described in detail in the following sections, beginning with the first step mentioned in Figure 3.1 above, to properly design the polymerized micelles to be as close to reality as possible.

3.3 Molecular Models of the AABMMs

In working towards the development of the QSERR model, we must first develop the accurate representations of the AABMMs. As such the polymerized, dipeptide combination of the L- form of Alanine, Leucine, Valine and achiral Glycine yield a total of 15 AABMM structures, must be properly made. The previous attempts for designing the computational AABMM structures are presented in the following sections.

The general *in silico* development of the AABMMs and the exact process used to study the enantioseparation mechanisms will be described and summarized in the following sections. The detailed computational files and protocols will be referenced throughout the following sections, per their respective appendixes.

3.3.1 Single Chain Method

Much of the molecular dynamics software that exists is predominantly used to study proteins, which are bioactive chains of amino acid residues that are covalently linked via peptide bonds. As such, when the proteins are simulated in a solvated environment, they tend to fold from a linear chain structure into its globular shape. The hydrophobic regions are buried and hidden from the aqueous solvent whereas the hydrophilic, charged regions face the aqueous solvent. Applying this same thought process, our research group previously modeled the AABMMs as a single linear chain of amino acid-based surfactants that were covalently linked at their terminal carbons on the surfactant tail, as shown in Figure 3.2.

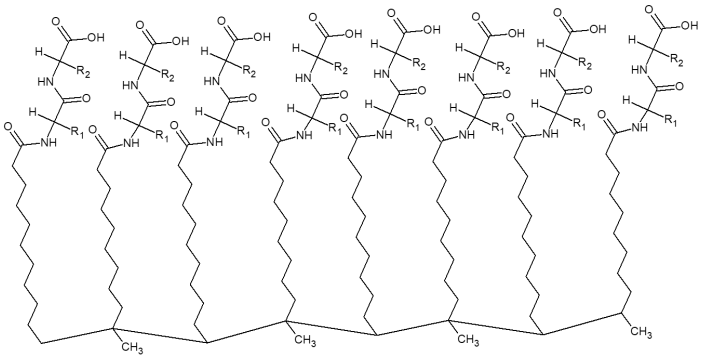


Figure 3.2. Representation of an AABMM using a single linear chain of surfactants.

The general steps for creating the single, polymerized chain are summarized as follows. The initial monomer surfactant unit was manually created in Xleap, where the antechamber program was used to assign the GAFF force field to the surfactant unit and also assigns the surfactant with a net charge of -1. This molecule was then brought into Xleap, where the terminal carbon at the end of the tail had a hydrogen removed. The adjacent carbon (2^{nd} to last carbon) then had a hydrogen removed, which was trans to the previously removed hydrogen atom. This is necessary to form covalent bonds between those adjusted carbon atoms with other surfactant units. The number of surfactant units used to build the linear chain AABBMs was based on aggregation numbers that were experimentally observed and reported in a previous publication.[1] The number is dependent on the surfactant being used and should be adjusted for each micelle that is being created *in silico*. The polymerized linear chains were made for each AABMM, then each were solvated with a TIP3P water solvent and sodium counterions. The prmtop and inpcrd files of the solvated micelle structures were then saved via Xleap. The next step would be to run a molecular dynamics simulation on these AABMMs for a prolonged simulation time to allow for the micelle structure to settle into a thermodynamically favorable structure when in an aqueous environment; this is commonly known as an equilibrated structure. The equilibrated structure can be determined

by running a root-mean-square deviation (RMSD) on the trajectory file that was produced from the simulation. This allows the user to see how the atom coordinates change over time when given a reference structure. The exact details of the RMSD protocol and details will be discussed in further detail in section 3.10.

One of the issues that was observed when analyzing the single chain AABMMs is that some of the equilibrated structures were not representative of what a AABMM would actually look like, as some had hydrophilic regions buried within the hydrophobic core of the AABMM, and also had hydrophobic regions that faced the aqueous solvent. When observing the Leucine-based surfactants: undecyl-(L,L)-leucine-glycine (ULG), undecyl-(L,L)-leucine-alanine (ULA), undecyl-(L,L)-leucine-valine (ULV), and undecyl-(L,L)-leucine-leucine (ULL), as shown in Figure 3.3, the equilibrated structures of these AABMMs seem to be closely representative of the micelle structure that is desired.

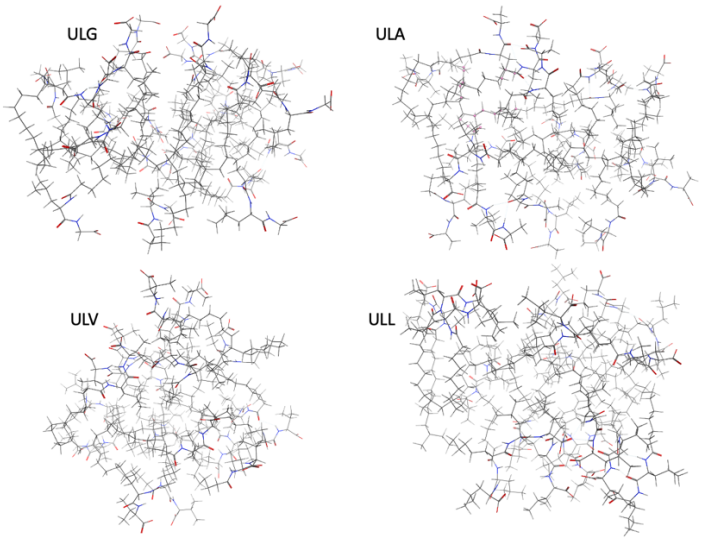


Figure 3.3. Leucine-based AABMMs built with the single linear chain method.

When observing the Alanine-based surfactants: undecyl-(L,L)-alanine-glycine (UAG), undecyl-(L,L)-alanine-alanine (UAA), undecyl-(L,L)-alanine-valine (UAV), and undecyl-(L,L)-alanine-leucine (UAL), as shown in Figure 3.4, there seems to be proper representation of the AABMMs, with minor discrepancies seen in ULA and UAV.

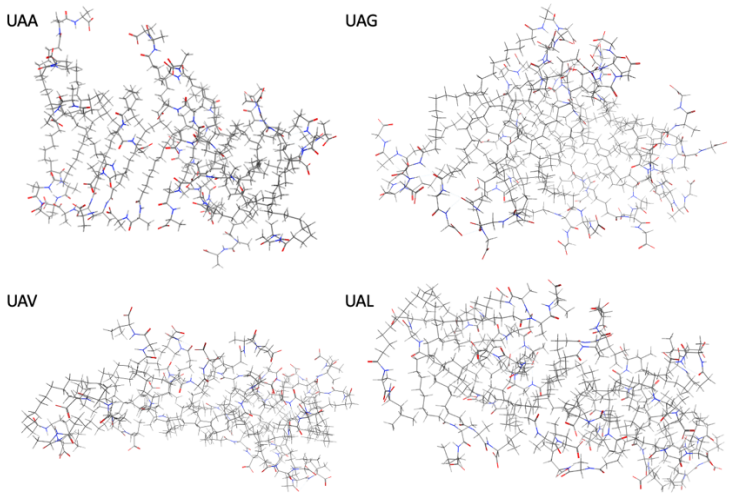


Figure 3.4. Alanine-based AABMMs built with the single linear chain method.

When observing the Valine-based surfactants: undecyl-(L,L)-valine-glycine (UVG), undecyl-(L,L)-valine-alanine (UVA), undecyl-(L,L)-valine-valine (UVV), and undecyl-(L,L)-valine-leucine (UVL), as shown in Figure 3.5, there seems to be proper representation of the AABMMs.

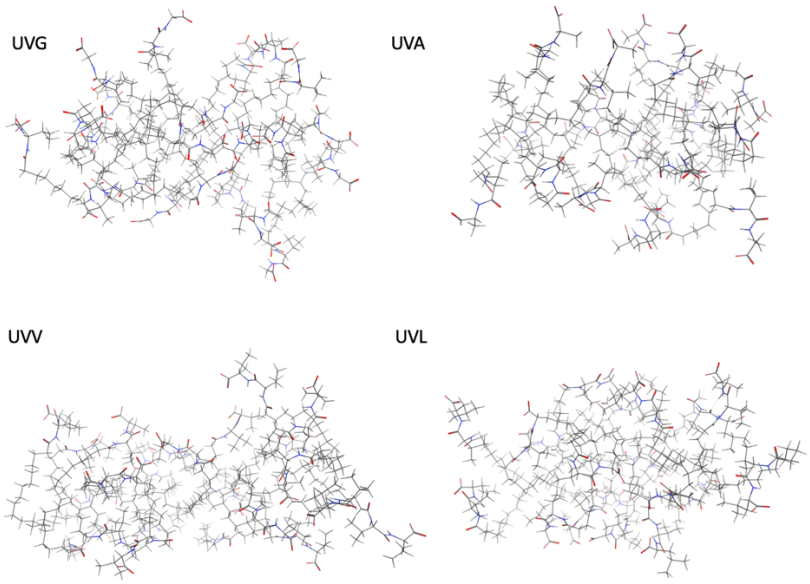


Figure 3.5. Valine-based AABMMs built with the single linear chain method.

However, when observing the Glycine-based surfactants: undecyl-(L,L)-glycine-alanine (UGA), undecyl-(L,L)-glycine-valine (UGV), and undecyl-(L,L)-glycine-leucine (UGL), as shown in Figure 3.6 there seems to be a major issue with the UGA micelle.

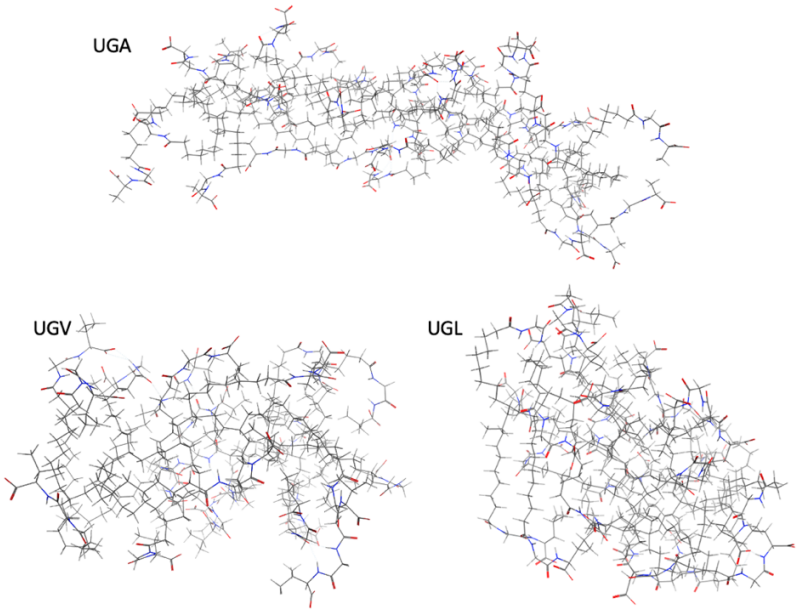


Figure 3.6. Glycine-based AABMMs built with the single linear chain method.

As mentioned previously, it is crucial to develop a method for that appropriately builds all 15 AABMM structures properly. Therefore, a new method for building the AABMM structures computationally was deemed necessary. Our group has tried a few different approaches, which are discussed in detail in the following sections.

3.3.2 Interlocked Rings Method

To address this issue, our research group attempted to create the AABMMs using an interlocked ring approach. This interlocked ring method consists of creating two polymerized rings and interlocking the two, in hopes that this would help to form spherical AABMMs which had hydrophobic regions on the interior of the micellar core, and the hydrophilic regions facing the aqueous solution. The method for building the AABMM interlocked rings is briefly described as follows. For the interlocked rings, we have used that same aggregation number that was used for the single chain but have split the monomer units equally between two separate linear chains. Each of the two linear chains were connected from one carbon atom at one end of the carbon backbone to the other carbon atom at the other end of the linear chain, thus creating a polymer ring, as shown in Figure 3.7 below.

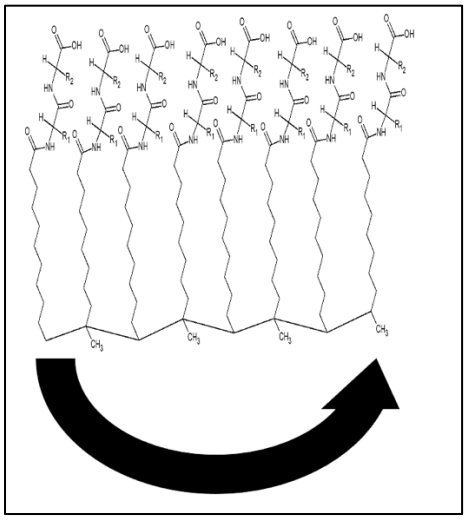


Figure 3.7. Formation of a single ring polymer by connecting the outermost carbon atom on the carbon backbone to the opposite carbon atom on the other end of the carbon backbone.

When the two separate linear chains are formed into rings in the same fashion, they can be formed in an interlocked fashion in hopes that upon equilibration the molecular micelle will then collapse properly with the hydrophilic residues facing the aqueous solution and the hydrophobic core facing away from the aqueous solution, as shown in Figure 3.8. The structures for the AABMMs seemed to have been resolved in terms of addressing the issues with the hydrophobic and hydrophilic regions of the micelles being improperly represented. Because the UGA micelle that was built with the linear chain had significant discrepancies when reviewing the UGA micelle, this AABMM was used to compare the different methods for building the micelles.

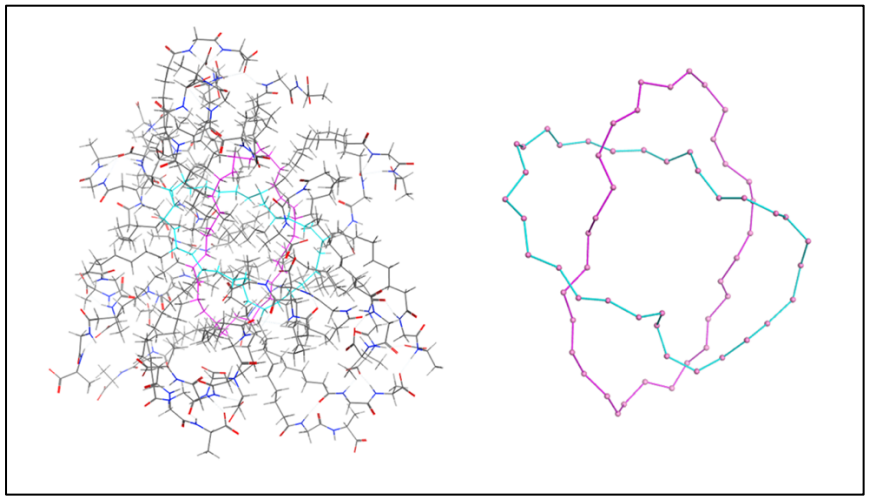


Figure 3.8. Representation of UGA constructed with interlocked rings; the figure on the right displays the carbon backbone of the interlocked rings for UGA.

Although the structures collapsed properly with the interlocked ring systems, it had issues and failed multiple times when running simulations for some of the AABMM systems. Therefore, this micelle structure was not able to be used, as we hypothesized that steric hindrance and the energetics of the systems contributed to the simulations crashing. One of the issues to be investigated in my research efforts was to develop a computational method for building the polymerized micelle structures that were able to collapse properly for all dipeptide surfactant combinations and able to undergo MD simulations without crashing. Therefore, another method for developing the AABMMs computationally was investigated as part of my thesis research which was to use the Single Ring Method.

3.3.3 Single Ring Method

Due to the aforementioned problems of steric hindrance with the interlocked rings, a single polymer ring consisting of the aggregation number that was experimentally collected for each

respective molecular micelle was created, using the GAFF force fields to define the entire AABMM. The same process that was used to create the single linear chain was essentially applied, except that the two opposing ends of the linear chain are now connected to form a single ring. After obtaining the equilibrated AABMM structures, it was determined that they yielded favorable, equilibrated micelle structures that represent micelles. Because UGA had difficulties collapsing properly with the linear chain method, it was important to compare this micelle, in particular, across all methods used to build the micelles. For comparative purposes, the equilibrated structures of UGA that were built with three different methods: linear chain, interlocked rings, and single ring can be seen in Figure 3.9.

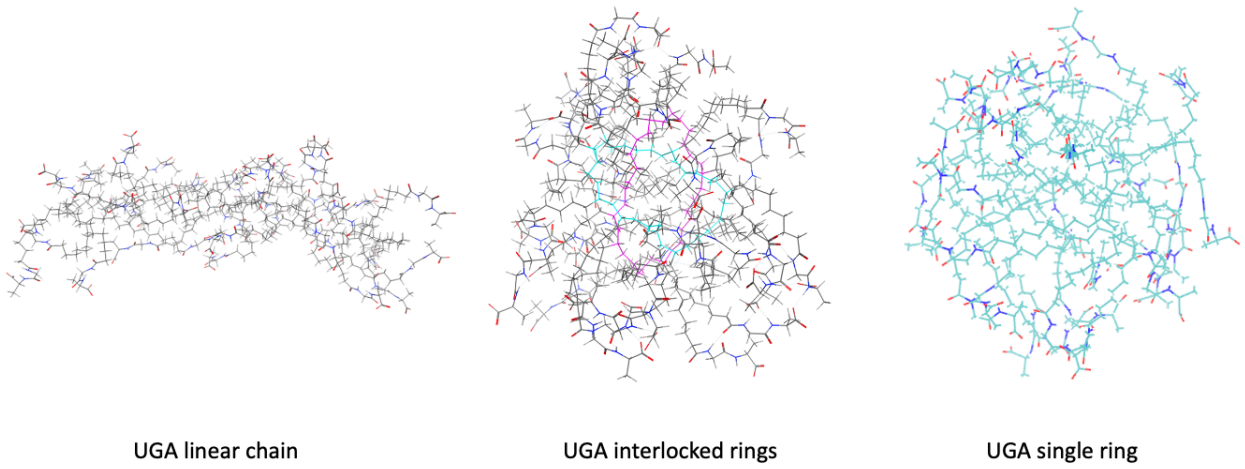


Figure 3.9. Comparison of the UGA micelle structures built with varying methods.

Using a single ring polymer ameliorates issues associated with steric hindrance because the single ring would allow for more rotational freedom as opposed to two, small, interlocked rings. After building the AABMMs using the single ring method, I then determined that the micelles were able to undergo simulations for at least 480 ns. This allows us to proceed with the

next step of the process, which is to dock the micelles with various enantiomers. My research focused on enantiomers binding to the polySULV micelle.

3.4 Ligand Docking

The next step after designing the AABMMs was to then dock the polySULV micelle with a library of chiral compounds using the Molecular Operating Environment (MOE) software. The protocol for docking the AABMMs with enantiomers is described in detail in section 3.10. Various enantiomers were docked to the pockets which were identified by the Site Finder module in MOE. Figure 3.10 shows the set of chiral compounds studied in this research: beta blockers such as Alprenolol, Propranolol and Atenolol along with Binaphthyl derivatives such as Binaphthyl amine (BNA), Binaphthyl Phosphate (BNP) and Binaphthol (BOH) and also Dansyl derivatives such as Dansyl-Leucine, Dansyl-Norleucine, Dansyl-Phenylalanine and Dansyl-Tryptophan.

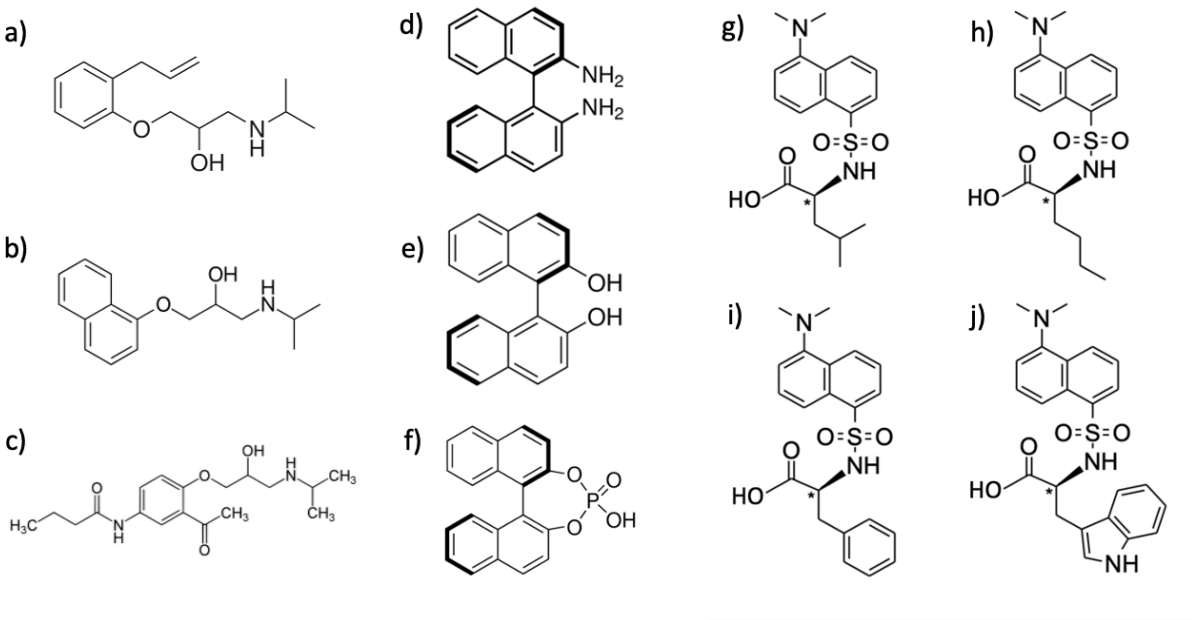


Figure 3.10. Chemical structures of Alprenolol, Propranolol, Atenolol, BNA, BOH, BNP, Dansyl Leucine, Dansyl Norleucine, Dansyl Phenylalanine and Dansyl Tryptophan (a-j, respectively).

3.5 Determining Run Conditions

3.5.1 ΔG vs. simulation time studies

After docking the analytes to the polySULV micelle, the PDB files of the docked analyte and the PDB files of the AABMM were loaded into Xleap and then solvated with TIP3P water molecules and sodium ions. This complex can then be saved as a prmtop and inpcrd file, which are the files necessary to initiate the simulation of the AABMM:enantiomer complex. The next step is to determine the conditions by which the simulations will be ran at, and to then calculate the binding free energy values of these systems using both, the MMGBA and MMPSA values, and seeing how they compare. The first few questions that needed to be addressed were: (1) How long should we run the simulations for? (2) Do the MMGBSA and MMPBSA calculations change significantly with the time of simulation, and how do these calculations compare?

To determine how long we should run the simulations of the complexed systems for, I first conducted a 240 ns simulation of Dansyl (L) Phenylalanine binding to the single ring polyULV in an aqueous environment. Table 3.1, shown below, provide information about the MMGBSA (GB) and MMPBSA (PB) calculations at various frames throughout the 240 ns simulation.

Table 3.1. The average GB and PB values between certain frames of the simulation of Dansyl-L-Phenylalanine in conjunction with the single ring SULV micelle.

Frames	**Time (ns)**	**GB** (kJ·mol^{-1})	**PB** (kJ·mol^{-1})
1-10,000	10	-14.6579	2.3551
10-20,000	20	-13.9402	-0.4243
20-40,000	40	-12.1333	-3.5779
40-60,000	60	-12.2104	-4.1140
60-80,000	80	-11.3724	-3.2704
80-100,000	100	-11.7772	-3.3198
100-120,000	120	-13.8895	-4.6923
120-140,000	140	-15.2325	-8.0817
140-160,000	160	-14.1389	-4.6913
160-180,000	180	-15.7546	-6.9681
180-200,000	200	-16.3712	-6.5123
220-240,000	240	-15.1663	-3.3624

As such, the information from Table 3.1 has been plotted for visual purposes and is shown in Figure 3.11 below. Figure 3.11 shows how the GB and PB values fluctuate over time at certain segments of the simulation. In general, for this particular study between polySULV and Dansyl-L-Phenylalanine, we see that GB calculations of the binding free energy appear to be more negative (strong binding) than the values presented by the PB calculations. This was a preliminary investigation to compare the two calculations.

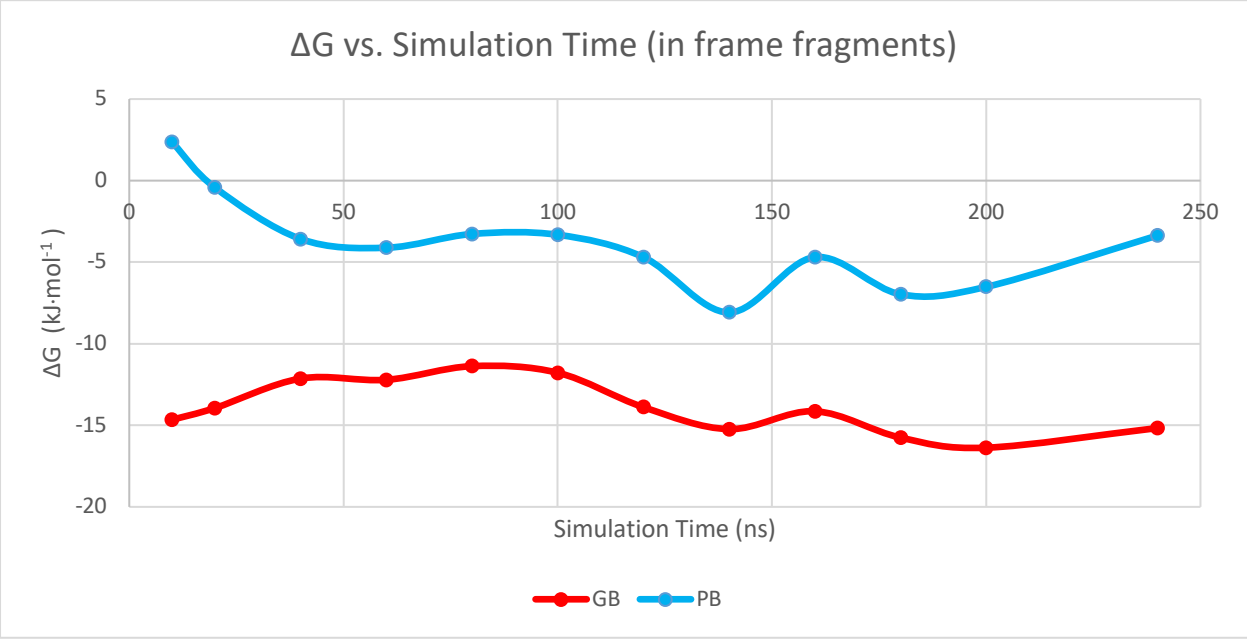


Figure 3.11. The GB and PB values calculated for the single ring polySULV to Dansyl-L-Phenylalanine complex at various frame segments.

This study helped to show how the binding free energy values compared for this specific study. However, to investigate how long the simulations should be ran for the solvated complex of the micelle to the enantiomer, we investigated how the binding free energy values changed over time.

The solvated complex of single ring polySULV and R-BNP was conducted for an 80 ns simulation. The binding free energy values were studies from frame 0 to frame 500 (0.1 ns simulation), from frame 0 to frame 3000 (0.6 ns), and so forth for various simulation times to identify optimal simulation time for studying these complexes. Table 3.2. shows the various time lengths for simulations that were investigated along with the GB and PB energy calculations at those particular run times.

Table 3.2. Single ring polySULV and R-BNP binding free energy values calculate via GB and PB were investigated as a function of simulation time.

Time (ns)	frames	nstlim	GB (kJ·mol^{-1})	PB (kJ·mol^{-1})
0.1	500	50,000	-1.0594	5.0277
0.6	3,000	300,000	0.0802	4.3767
2.0	10,000	1,000,000	0.1014	4.1370
6.0	30,000	3,000,000	0.6349	3.6143
9.0	45,000	4,500,000	0.7437	5.1801
12.0	60,000	6,000,000	-0.3465	5.6182
16.0	80,000	8,000,000	-1.7512	4.7945
20.0	100,000	10,000,000	-2.9959	3.8774
40.0	200,000	20,000,000	-7.5226	0.8665
60.0	300,000	30,000,000	-8.5821	0.2489
80.0	400,000	40,000,000	-8.1859	1.1394

As such, the information from Table 3.2 has been plotted for visual purposes and is shown in Figure 3.12 below. Figure 3.12 shows the effects of simulation time on the binding free energy values of R-BNP binding to poly(SULV).

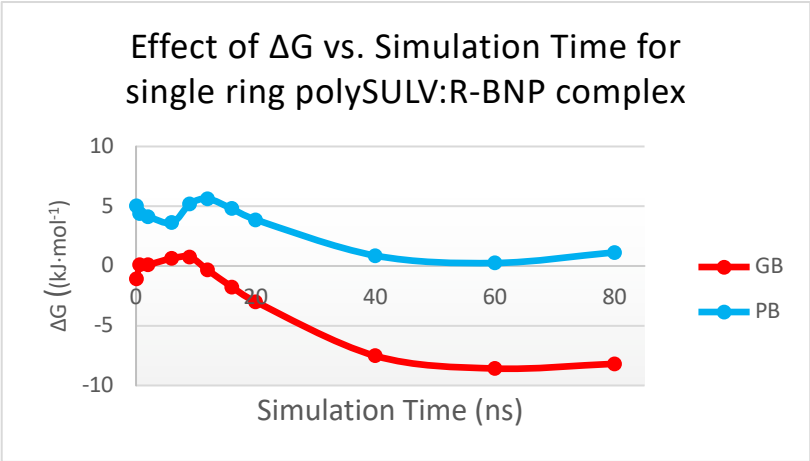


Figure 3.12. The effects of simulation time on the binding free energy values of R-BNP binding to poly(SULV).

As such, Figure 3.12 reveals that simulation time significantly affects the binding free energy calculations at the beginning of the simulation since the complexed system has not reach equilibrium. We see that the GB and PB values become steady after 40 ns and stays somewhat linear throughout 80 ns. Since we have many simulations to run, we decided to use 60 ns as the set simulation time to be conducted for all complexed systems.

3.5.2 Interval Studies

After determining that all enantiomer to micelle complexes would be simulated for 60 ns, the next step was to determine how to optimize the time of data collection for the various GB and PB calculations of the system. The trajectory file therefore contains all frames that were collected from the 60 ns simulation. When running the GB and PB calculations on the trajectory file, the user is able to change the intervals of the frames being analyzed. For example, if I ran a simulation for 1000 frames, and the interval is set to 1, then it would analyze every frame of that simulation. However, if the interval to 100, it would then analyze every 100th frame thus making our binding

free energy calculations much faster. Therefore, it would be best to optimize the intervals for the GB and PB calculations so that they can be calculated faster without sacrificing accuracy. Therefore, I asked (1) How does changing the interval value affect the GB and PB values? (2) Which interval would be most time-efficient to use for the simulations?

To investigate this, I used the 60 ns trajectory file for S-Propranolol binding to polySULV and calculated the GB and PB values. Table 3.3 shows how the GB and PB values change when analyzing the trajectory file using various intervals of 1, 2, 3, 4, 5 and 7.

Table 3.3. S-Propranolol binding to polySULV using various intervals for the GB and PB calculations.

Intervals for Binding Free Energy Calculations	GB values ($kJ \cdot mol^{-1}$)	PB values ($kJ \cdot mol^{-1}$)
int=1	-11.6458	-2.5493
int=2	-11.6128	-2.514
int=3	-11.6343	-2.5563
int=4	-11.6494	-2.5929
int=5	-11.6776	-2.6054
int=7	-11.6875	-2.6163
Average	-11.651233	-2.5723667
Std. Dev.	0.02759425	0.03568341

Table 3.3 shows that the intervals do not impact the GB and PB values significantly, as indicated by the standard deviations (green) and averages (yellow). As such, we have used an interval of 10 for the subsequent studies.

For all subsequent studies, the AABMM:enantiomer complexes underwent 60 ns simulations with the energy calculations being conducted at an interval of 10. The exact details about the computational conditions are further explained in Chapter 5.

3.6 Visual Molecular Dynamics

After the MD simulations were conducted for the polySULV in conjunction with the docked analytes (both R and S enantiomer) mentioned in *section 3.4*, the trajectory files were then used to visualize the simulations using the Visual Molecular Dynamics (VMD) software. The protocol for analyzing the trajectory file via VMD was created and can be found in section 3.10. This helps to visually identify any anomalies that may have occurred throughout the simulations and may provide explanation for any unusual or unexpected binding free energy values.

For example, Figure 3.13 shows the anomaly with the cyclic portion of an analyte being penetrated by part of the micelle, causing the binding free energy values to be skewed and inaccurate. Therefore, in this case, the next step would be to attempt the simulations with a different docked position.

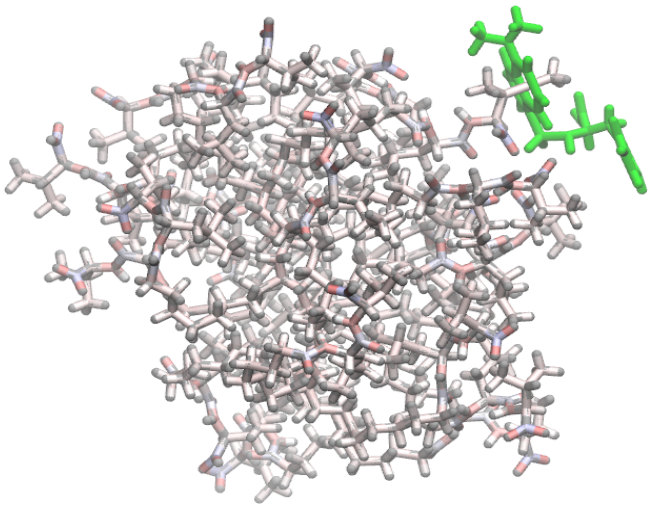


Figure 3.13. The conjugated pi system of the analyte is penetrated by the ULV group – this gave a substantially greater binding energy value than the other pockets investigated and thus had the greatest fraction population value.

In addition to the situation mentioned previously with the unusual, docked configuration, Figure 3.14 below shows an instance where an analyte is docked into a pocket of the micelle at the beginning of the micelle, then flies away from the pocket at frame 2529 deeming the calculated binding free energy values after that frame to be inaccurate and not representative of the binding free energy value that is output by our computational system.

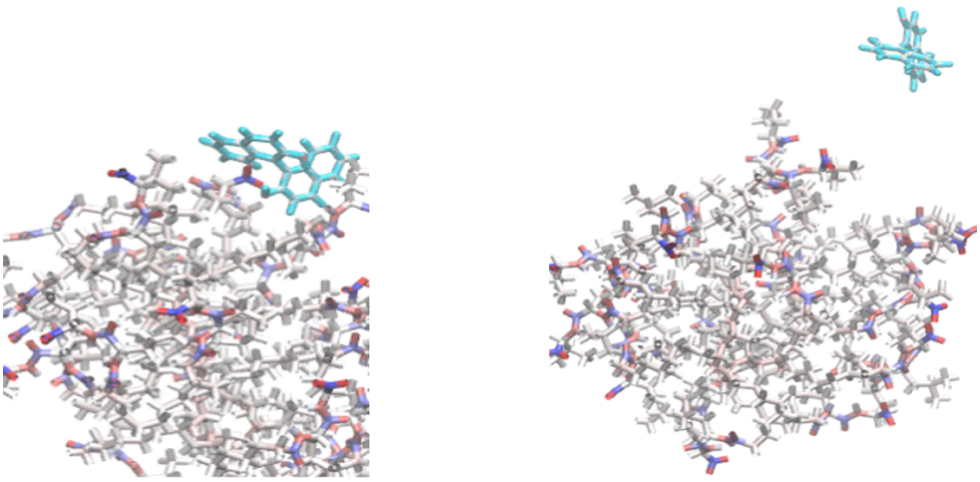


Figure 3.14. Analyte begins in its initial docked position at the starting point (frame 0) of the simulation, then leaves the binding pocket at frame 2529.

Therefore, the VMD software allows us to analyze each trajectory file and ensure no anomalies occurred, prior to confirming if the simulation was acceptable according to these standards.

3.7 Addressing Preliminary Issues

To address the issue of analytes leaving the binding pocket, tests with varying restraint weights were investigated to try to prevent the analyte from doing so. The restraint weights were meant to restrain the analyte within the binding pocket. Table 3.4 shows that although increasing the restrain weight of the analyte, R-BOH, kept the analyte in the binding pocket of polySULV, the analyte, however, became very rigid and lost all flexibility. Therefore, we did not continue with restraints on our system for subsequent studies. The studies proceeded forward and if an

analyte flew out of the pocket, we determined that it would not want to bind to that binding pocket anyway.

Table 3.4. Energy calculation values for R-BOH binding to a pocket of polySULV with varying restraint weights.

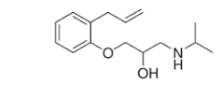
Restraint Weight	GB ($kJ \cdot mol^{-1}$)	PB ($kJ \cdot mol^{-1}$)	VMD Check
10	-7.6566	-2.3074	Remains in close proximity but still leaves binding pocket.
50	-6.1288	-0.3135	Leaves the binding pocket and moves
100	-12.5837	-2.2943	Does not leave the binding pocket; analyte very rigid and no flexibility.
250	-9.6406	-0.9086	Does not leave the binding pocket; analyte extremely rigid and no flexibility.

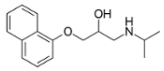
3.8 Binding Free Energy Results

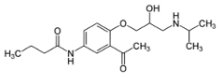
As previously mentioned in *section 3.4*, the analytes being studied in conjunction with the polySULV micelle were docked, and then investigated for their binding free energy values. The first class of compounds were binaphthyl derivatives binding to the three binding pockets of polySULV. After the binding free energy values were calculated for each enantiomer in each binding pocket of polySULV, the fractional populations were determined, which informs the research about where the analyte is more likely to bind. These results are then compared to the experimental MEKC data to determine if they agree with computational binding free energy values.

3.8.1 Neutral Beta Blockers

The first class of analytes studied were beta blockers, which are Alprenolol, Propranolol and Atenolol. Figure 3.15 shows the binding free energy values calculated with MMGBSA for each enantiomer of the beta blockers when bound to each of the three binding pockets of polySULV. Figure 3.15 is shown below, and the data will be described in detail in the subsequent sections.







Compound Name	(kJ·mol^{-1}) Binding Free Energy Values			(kJ·mol^{-1}) Percentage Population			Match?
	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	
*R-Alprenolol	-11.6309	-15.4364	-11.544	15.23%	70.07%	14.71%	
S-Alprenolol	-12.2967	-13.8947	-11.9491	26.54%	50.38%	23.08%	
R-Propranolol	-2.775	-9.2156	-7.0626	5.04%	66.79%	28.16%	
*S-Propranolol	-12.9284	-13.0125	-11.5081	38.46%	39.78%	21.76%	
R-Atenolol	-11.2559	-9.356	-13.3400	26.50%	12.37%	61.13%	
*S-Atenolol	-9.775	-13.5368	-12.8107	11.23%	50.80%	37.96%	

* indicates the second eluting enantiomer

Figure 3.15. Binding free energy values and percentage populations for each enantiomer of the beta blockers when bound to the three binding pockets of polySULV. The asterisk indicates which enantiomer elutes from the MEKC column second in experimental studies. The green squares indicate if the computational binding free energy values and their respective percentage population match the elution order seen in MEKC studies.

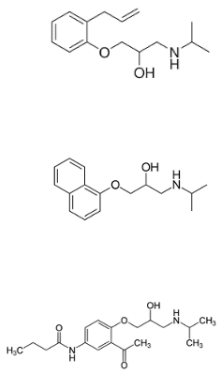
These beta blockers had a net charge of zero, so they were studied in their neutral forms. The first beta blocker being examined is Alprenolol, in which experimental MEKC elution order determines that the R enantiomer binds stronger to polySULV than its S enantiomer. The computational binding free energy values were analyzed further to see if they matched the experimental MEKC elution order. When analyzing R-Alprenolol bound to binding pockets 1-3 of polySULV, the binding free energy values were -11.6309, -15.4463 and -11.5440 $kJ \cdot mol^{-1}$, respectively. As such, the percentage population for pockets 1-3, were 15.23%, 70.07% and 14.71%, respectively. Therefore R-Alprenolol has favorable interactions with pocket 2 of polySULV with a binding free energy value of -15.4463 $kJ \cdot mol^{-1}$ and a percent population of 70.07%.

When analyzing S-Alprenolol bound to binding pockets 1-3 of polySULV, the binding free energy values were -12.2967, -13.8937 and -11.9491 $kJ \cdot mol^{-1}$, respectively. As such, the percentage population for pockets 1-3, were 26.54%, 50.38% and 23.08%, respectively. Therefore S-Alprenolol has favorable interactions with pocket 2 of polySULV with a binding free energy value of -13.8937 $kJ \cdot mol^{-1}$ and a percent population of 50.38%. It is evident that both, R- and S-Alprenolol favor pocket 2 of polySULV, therefore when comparing the two, it is evident that R-Alprenolol binds stronger to polySULV and thus matches experimental MEKC elution order.

The same process was applied when studying the other beta blockers, such as Propranolol and Atenolol as seen in Figure 3.15. The data indicates that the computational binding free energy values matched that of experimental MEKC elution order data. As such, the computational studies matched MEKC elution order data for 100% of the neutral beta blockers being investigated.

3.8.2 Positively Charged Beta Blockers

It is necessary to include both, neutral and charged analytes, in order studies to assess a wider range of compounds to develop a broader class of analytes to be studied with the QSERR database in the future. Therefore, the beta blockers: Alprenolol, Propranolol and Atenolol, were studied in their positively charged states. Figure 3.16 shows the binding free energy values calculated with MMGBSA for each enantiomer of the beta blockers when bound to each of the three binding pockets of polySULV. Figure 3.16 is shown below, and the data will be described in detail in the subsequent sections.



	(kJ·mol^{-1}) Binding Free Energy Values			(kJ·mol^{-1}) Percentage Population			
Compound Name	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Match?
*R-Alprenolol	-16.7991	-18.9067	-17.3878	21.76%	50.68%	27.56%	
S-Alprenolol	-13.9338	-21.053	-17.059	4.57%	79.43%	16.00%	
R-Propranolol	-17.8349	-17.5609	510.3287	52.74%	47.26%	0.00%	
*S-Propranolol	-14.9985	660.0806	-14.1359	58.56%	0.00%	41.44%	
R-Atenolol	-25.7980	-23.6331	-14.7024	69.87%	29.32%	0.82%	
*S-Atenolol	-20.6035	-18.5365	-12.1265	68.04%	29.69%	2.27%	

* indicates the second eluting enantiomer

Figure 3.16. Binding free energy values and percentage populations for each enantiomer of the positively charged beta blockers when bound to the three binding pockets of polySULV. The asterisk indicates which enantiomer elutes from the MEKC column second in experimental studies. The red squares indicate that the computational binding free energy values and their respective percentage population did not match the elution order seen in MEKC studies.

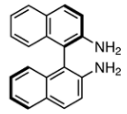
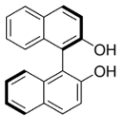
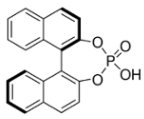
These beta blockers had a net charge of +1. The first beta blocker being examined is Alprenolol, in which experimental MEKC elution order determines that the R enantiomer binds stronger to polySULV than its S enantiomer. The computational binding free energy values were analyzed further to see if they matched the experimental MEKC elution order. When analyzing R-Alprenolol bound to binding pockets 1-3 of polySULV, the binding free energy values were -16.7991, -18.9067, and -17.3878 $kJ \cdot mol^{-1}$, respectively. As such, the percentage population for pockets 1-3, were 21.76%, 50.68%, and 27.56%, respectively. Therefore R-Alprenolol has favorable interactions with pocket 2 of polySULV with a binding free energy value of -17.9067 $kJ \cdot mol^{-1}$ and a percent population of 50.68%.

When analyzing S-Alprenolol bound to binding pockets 1-3 of polySULV, the binding free energy values were -13.9338, -21.0530, and -17.0590 $kJ \cdot mol^{-1}$, respectively. As such, the percentage population for pockets 1-3, were 4.57%, 79.43% and 16.00%, respectively. Therefore S-Alprenolol has favorable interactions with pocket 2 of polySULV with a binding free energy value of -21.0530 $kJ \cdot mol^{-1}$ and a percent population of 79.43%. It is evident that both, R- and S-Alprenolol favor pocket 2 of polySULV, therefore when comparing the two, it is evident that S-Alprenolol binds stronger to polySULV and therefore does NOT match experimental MEKC elution order.

The same process was applied when studying the other positively charged beta blockers, such as Propranolol and Atenolol as seen in Figure 3.16. The data indicates that the computational binding free energy values did NOT match experimental MEKC elution order data. As such, the computational studies matched MEKC elution order data for 0% of the positively beta blockers being investigated. This indicates that the computational system may be better suited for neutral analytes, however, further studies are necessary prior to making this determination.

3.8.3 Binaphthyl Derivatives

It is necessary to include both, analytes with chiral centers and those with chiral planes, in order studies to assess a wider range of compounds to develop a broader class of analytes to be studied with the QSERR database in the future. Therefore, the binaphthyl derivatives: BNA, BOH and BNP, which were studied with their net charge of 0, 0, and -1, respectively. Figure 3.16 shows the binding free energy values calculated with MMGBSA for each enantiomer of the binaphthyl derivatives when bound to each of the three binding pockets of polySULV. Figure 3.16 is shown below, and the data will be described in detail in the subsequent sections.

		(kJ·mol^{-1}) Binding Free Energy Values			(kJ·mol^{-1}) Percentage Population			
	Compound Name	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Match?
	*R-BNA	-9.2569	-8.3429	-8.8533	39.31%	27.25%	33.44%	
	S-BNA	-10.4596	-8.0866	-7.2795	60.05%	23.18%	16.77%	
	S-BOH	-12.457	-9.9645	-10.57	54.44%	20.03%	25.53%	
	*R-BOH	-15.337	-7.9261	-14.182	59.52%	3.05%	37.44%	
	*S-BNP	-9.4947	-9.8987	-8.5983	34.80%	40.92%	24.29%	
	R-BNP	-8.7583	-8.9873	-10.259	25.50%	27.95%	46.55%	

* indicates the second eluting enantiomer

Figure 3.17. Binding free energy values and percentage populations for each enantiomer of the binaphthyl derivatives when bound to the three binding pockets of polySULV. The asterisk indicates which enantiomer elutes from the MEKC column second in experimental studies. The green and red squares indicate that the computational binding free energy values and their respective percentage population either matched or did not match, respectively, with the elution order seen in MEKC studies.

One binaphthyl derivate being examined is BNA, in which experimental MEKC elution order determines that the R enantiomer binds stronger to polySULV than its S enantiomer, thus causing the R enantiomer to elute from the column second. The computational binding free energy values were analyzed further to see if they matched the experimental MEKC elution order. When analyzing R-BNA bound to binding pockets 1-3 of polySULV, the binding free energy values were -9.2569, -8.3429, and -8.8533 $kJ \cdot mol^{-1}$, respectively. As such, the percentage population for pockets 1-3, were 39.31%, 27.25%, and 33.44%, respectively. Therefore R-BNA has favorable interactions with pocket 1 of polySULV with a binding free energy value of -9.2569 $kJ \cdot mol^{-1}$ and a percent population of 39.31%.

When analyzing S-BNA bound to binding pockets 1-3 of polySULV, the binding free energy values were -10.4596, -8.0866, and –7.2795 $kJ \cdot mol^{-1}$, respectively. As such, the percentage population for pockets 1-3, were 60.05%, 23.18%, and 16.77%, respectively. Therefore S-BNA has favorable interactions with pocket 1 of polySULV with a binding free energy value of -10.4596 $kJ \cdot mol^{-1}$ and a percent population of 60.05%. It is evident that both, R- and S- Alprenolol favor pocket 1 of polySULV, therefore when comparing the two, it is evident that S-BNA binds stronger to polySULV and therefore does NOT match experimental MEKC elution order.

The same process was applied when studying the other binaphthyl derivatives, shown in Figure 3.16. The data indicates that the computational binding free energy values did NOT match experimental MEKC elution order data for the BOH, but not BNP. As such, the computational studies matched MEKC elution order data for 50% of the neutral binaphthyl derivatives and 0% of the charged derivatives being investigated.

3.8.4 Dansyl Derivatives

In addition to the beta blockers and binaphthyl derivatives being studied, Dansyl derivatives were also investigated and showed that 100% of the neutral Dansyl derivatives matched experimental MEKC elution order data. Since Dansyl derivatives have not been deeply investigated in our group, *Chapter 4* of this thesis has been dedicated to studying the binding free energy values, and investigating which interactions govern chiral separations using hydrogen bond analyses. Please see *Chapter 4* for the full, detailed investigation.

3.8.5 Summarized Results

It is important to investigate any trends that may be evident when using GAFF force fields to define our entire AABMM system. Figure 3.18 summarized the neutral compounds that were investigated and recorded as green if the computational binding free energy data agreed with experimental MEKC elution order data, and red if it did not. As seen in Figure 3.18, 100% of neutral Dansyl derivatives, 100% of neutral beta blockers and 50% of neutral binaphthyl derivatives had computational binding free energy values that matched experimental MEKC elution order data. This evidence may show promise with using a GAFF-defined computational system for accurately investigating neutral compounds.

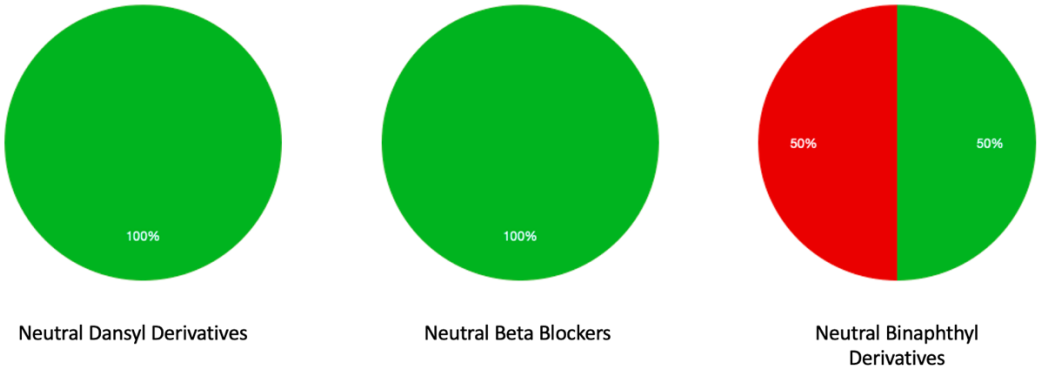


Figure 3.18. Neutral compounds where the computational binding free energy data, either did or did not match experimental MEKC elution order. The green portions represent compounds that matched experimental MEKC elution order data, and red portions represent those that did not.

Figure 3.19 shown below, summarizes the charged compounds that were investigated and recorded as green if the computational binding free energy data agreed with experimental MEKC elution order data, and red if it did not. As seen in Figure 3.19, 100% of negative aliphatic Dansyl derivatives matched. However, 50% of negatively charged aliphatic Dansyl derivatives, 100% positively charged beta blockers and 100% of negatively charged binaphthyl derivatives did NOT have computational binding free energy values that matched experimental MEKC elution order data. This evidence may show promise with using a GAFF-defined computational system does not work well for charged enantiomeric compounds.

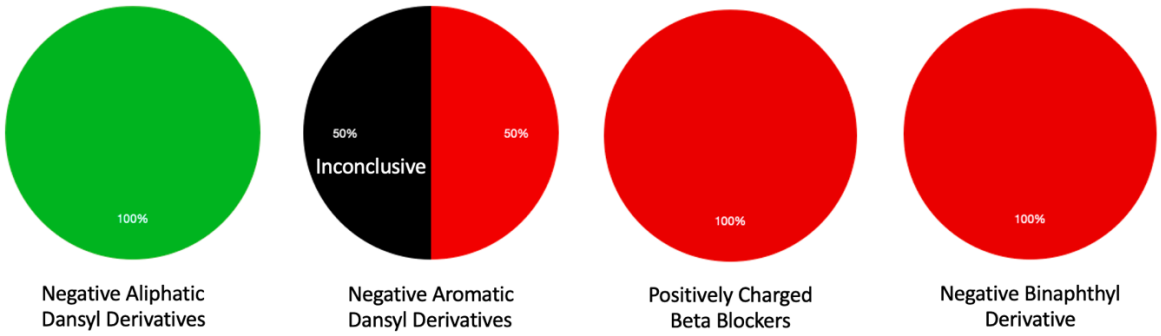


Figure 3.19. Charged compounds where the computational binding free energy data, either did or did not match experimental MEKC elution order. The green portions represent compounds that matched experimental MEKC elution order data, and red portions represent those that did not. Black portions indicate inconclusive data.

3.9 Refining the System

Prior to moving forward with the research, the results mentioned above needed to be addressed. The first issue to address is why most of the binding free energy values for the charged analytes binding to polySULV did not match experimental MEKC order. It has been shown that the enantiomeric order affects the elution order, therefore we conducted temperature studies to see if this had any effect on producing desired results. The second issue was to investigate using different force fields to better define our surfactants.

3.9.1 Temperature Studies

A small experiment was conducted to see if running simulations at colder temperatures would help to improve the matching of the computational values with experimental MEKC elution order data. The BOH enantiomer:polySULV complex underwent 60 ns simulations to determine if temperature played a significant role in affecting the GB values. Figure 3.20 shows GB values

being calculated for R-BOH (orange) and S-BOH (blue) when binding to polySULV. We can see that the enantiomeric order is greatly affected at many of the temperatures. For example, in Figure 3.20, we see that if we compared the enantiomers at temperatures of 270, 285 and 290 K, then the data would show that S-BOH binds stronger to polySULV, however at temperatures of 275, 280, 295 and 300, the enantiomeric order would be reversed. This did not provide conclusive data as to which temperatures to be using so we continue to use 300K for all subsequent studies to keep it consistent with experimental conditions.

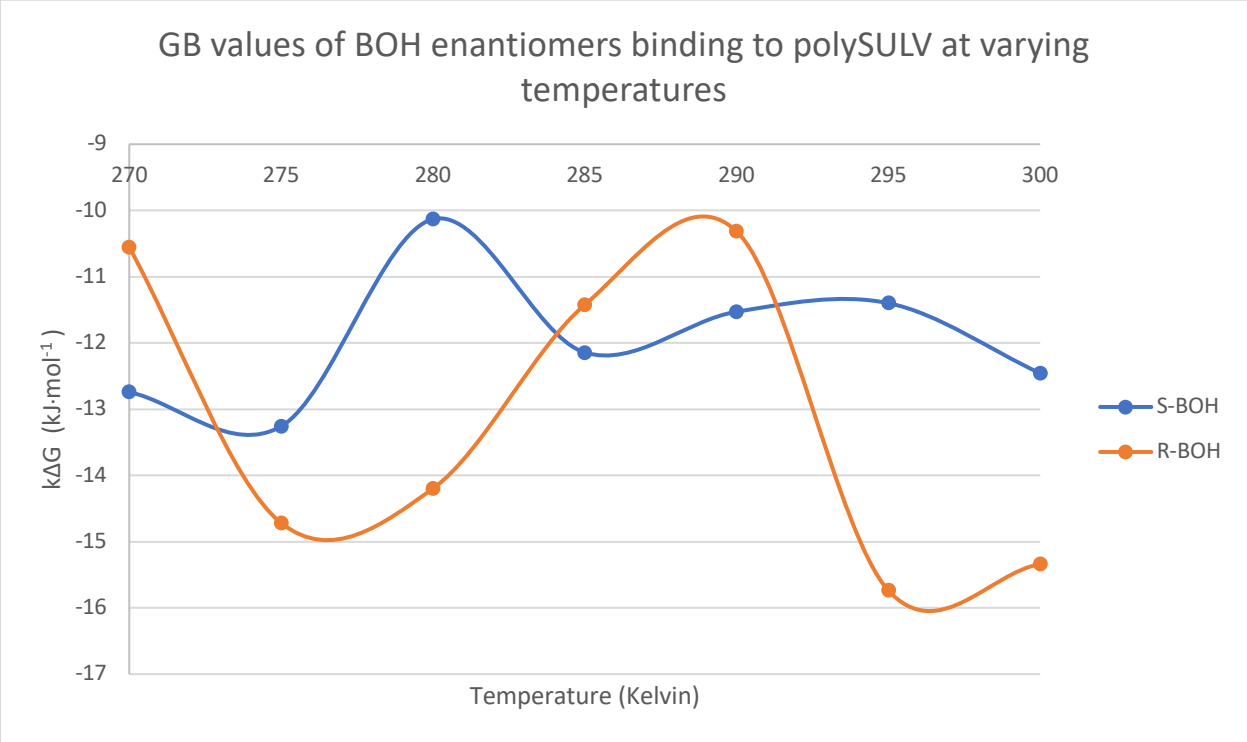


Figure 3.20. Temperature effects on the binding free energy values of BOH enantiomers binding to polySULV.

3.9.2 Mixing Force Fields

Another issue that may affect the binding free energy values for the charged species is to use different force fields to better define the amino acid head groups. The AMBER force field called ff19SB is often used to define amino acid residues in proteins, therefore, this research focuses on developing methods for defining the amino acid head groups with the ff19SB force field, and to define the surfactant tail constituents with the GAFF force fields. However, mixing force fields is quite challenging as the force fields have a difficult time communicating with each other when simultaneously used to defining a single macromolecule, such as the micelle. The subsequent section focuses on the use of developing a method for properly creating AABMMs with mixed force fields, along with very detailed protocols about the steps needed to be performed computationally to build and analyze these computational studies.

3.10 Protocols for Developing the Mixed Force Field Micelles for Future Studies

The general outline for developing the AABMMs with GAFF force fields defining the tail of the surfactant and the amino acid head group being defined by the protein ff19SB force fields are shown by Figure 3.21 below. Each step has an Appendix associated with it, which contains the detailed information necessary for a user to follow in order to conduct this research. This serves as a basic guideline for incoming research students to follow.

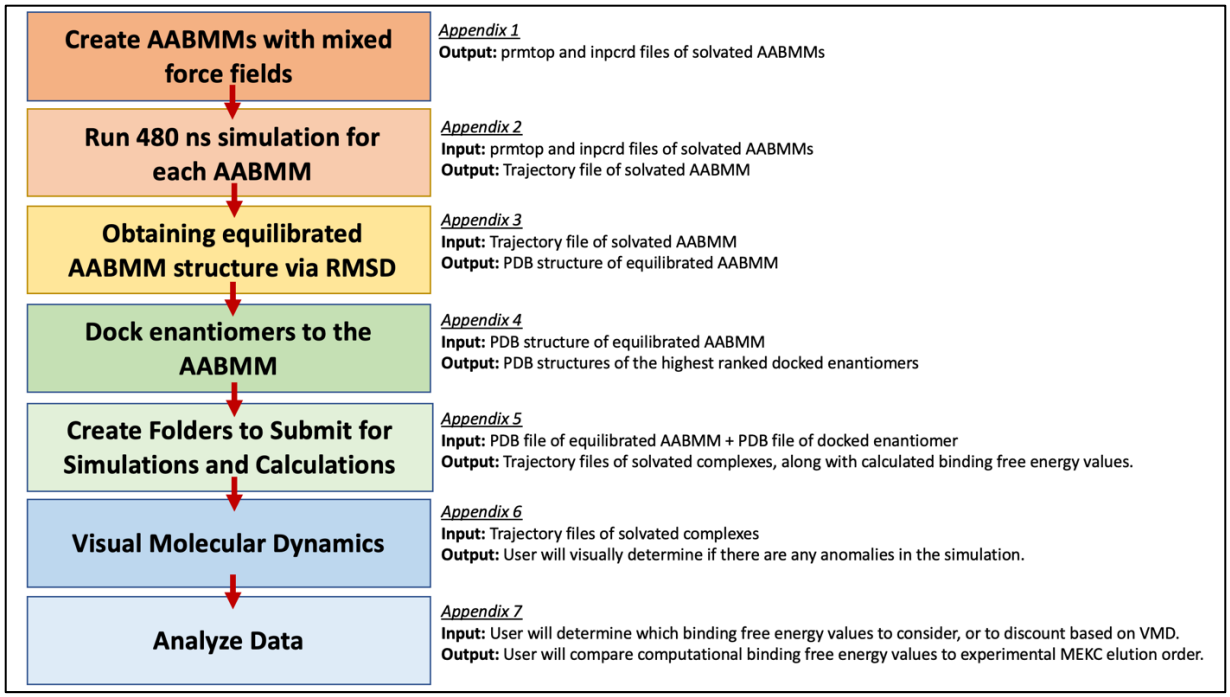


Figure 3.21. Overview of the steps for developing the AABMMs with mixed force fields, and for studying the binding interactions between the micelle and various enantiomers is displayed. The appendix for each step is denoted.

3.11 References

1. Billiot, F. H.; McCarroll, M.; Billiot, E. J.; Rugutt, J. K.; Morris, K.; Warner, I. M., Comparison of the Aggregation Behavior of 15 Polymeric and Monomeric Dipeptide Surfactants in Aqueous Solution. *Langmuir* **2002,** *18* (8), 2993-2997.

CHAPTER IV: CHIRAL RECOGNITION OF DANSYL DERIVATIVES WITH AN AMINO ACID-BASED MOLECULAR MICELLE: A MOLECULAR DYNAMICS INVESTIGATION

4.1 Abstract

In this study, the chiral separation mechanisms of Dansyl amino acids, including Dansyl-Leucine (Dans-Leu), Dansyl-Norleucine (Dans-Nor), Dansyl-Tryptophan (Dans-Trp) and Dansyl-Phenylalanine (Dans-Phe) binding to poly-sodium *N*-undecanoyl-(L)-Leucylvalinate, poly(SULV), were investigated using molecular dynamics simulations.

Micellar electrokinetic chromatography (MEKC) has previously shown that when separating the enantiomers of these aforementioned Dansyl amino acids, the L- enantiomers binds stronger to poly(SULV) than the D- enantiomers. This study aims to investigate the molecular interactions that govern chiral recognition in these systems using computational methods.

This study reveals that the computationally-calculated binding free energy values for Dansyl enantiomers binding to poly(SULV) are in agreement with the enantiomeric order produced in experimental MEKC studies. The L- enantiomers of Dans-Leu, Dans-Nor, Dans-Trp, and Dans-Phe binding to their preferred binding pockets in poly(SULV) yielded binding free energy values of -21.8938, -22.1763, -21.3329 and -13.3349 $kJ \cdot mol^{-1}$, respectively. The D-enantiomers of Dans-Leu, Dans-Nor, Dans-Trp, and Dans-Phe binding to their preferred binding pockets in poly(SULV) yielded binding free energy values of -14.5811, -15.9457, -13.6408, and -12.0959 $kJ \cdot mol^{-1}$, respectively. Furthermore, hydrogen bonding analyses was used to investigate and elucidate the molecular interactions that govern chiral recognition in these molecular systems.

4.2 Introduction

More than half of all pharmaceutical drugs currently in use are chiral, and the synthesis of these drugs often yield a racemic mixture, containing both enantiomers of the compound.[1] While the physical and chemical properties of the enantiomers are similar, they often produce very different pharmacological effects.[1-3] In many cases, one enantiomer may produce the desired medicinal effect, whereas the other may cause adverse health effects. For example, (+)-Ethambutol is primarily used to treat Tuberculosis, whereas its enantiomer, (-)-Ethambutol, causes blindness.[4] Due to the enantiomers of chiral drugs often producing vastly different effects, the United States Food and Drug Administration has mandated that each enantiomer of a chiral drug be tested for enantiomeric purity prior to being marketed.[5] Due to the enantiomers sharing nearly identical physical properties such as boiling point, density, mass and solubility, this often makes it difficult to separate enantiomers based on those aforementioned properties. [6] However, the enantiomers differ in stereoconfiguration, which allows them to interact differently with chiral separation mediums, thus allowing for enantioseparation.[5]

Some commonly used chiral separation techniques are high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE), which utilize chiral stationary and pseudostationary phases, respectively.[7, 8] In using chiral chromatographic techniques, if the two enantiomers of a chiral compound, the R- and S-enantiomer, have different binding affinities to the chiral separation medium, then enantiomeric separations can occur. While both aforementioned techniques are able to effectively separate enantiomers on the basis of their binding affinity to the chiral separation medium, CE has several advantages over HPLC. A much smaller sample size is required for CE as compared HPLC and often, CE yields a higher number of theoretical plates, ultimately producing better enantiomeric resolution in a shorter period of time.[9, 10] In HPLC, the

chiral recognition medium is covalently linked to a type of solid support material such as silicon beads.[11] Therefore, the user would need to purchase individual columns containing the desired stationary chiral recognition medium. Consequently, if the user wanted to switch out the chiral separation medium, it would entail switching out the column(s). However, in CE, the chiral recognition medium can be exchanged quite easily with another chiral medium since it is part of the mobile phase, in which the chiral medium acts as the pseudostationary phase.

Chiral selectors such as cyclodextrins, polysaccharides, crown ethers and chiral micelles have proven to be effective chiral selectors in CE.[7, 12, 13] This research will focus on the latter, chiral micelles- more specifically, a class of chiral micelles known as amino acid-based micelles. When micelles are used as the pseudostationary phase in CE, the technique is commonly known as Micellar Electrokinetic Chromatography (MEKC).[8, 10, 14, 15]

Amino acid-based micelles are composed of surfactant units, each containing a hydrocarbon chain; one end of the chain contains a terminal alkane, and the opposite side of the chain is connected to an amino acid head group. Amino acid-based surfactants have advantages over other chiral selectors in that amino acids are ubiquitous therefore deeming it cheaper than other options; these surfactants are also biodegradable and environmentally-friendly.[16] These surfactants have been extensively studied using single amino acid and dipeptide head groups, in which the dipeptide micellar systems have shown significant advantages in achieving enhanced enantiomeric resolution in MEKC compared to single amino acid based micellar systems.[10] This is due in part to the increased number of chiral centers on the head group, thus allowing for enhanced chiral selectivity. In addition, previous research reveals that the dipeptide amino acid-based micelles act as better chiral selectors in their polymerized forms.[15, 17] The polymerized surfactants are formed via free radical polymerization when the surfactant solution is subjected to

gamma radiation at concentrations significantly above their critical micelle concentration (CMC), usually 5-10 times the CMC. This gamma radiation causes the terminal alkenes to undergo free radical polymerization to form covalent bonds with neighboring surfactant units, as shown in Figure 4.1a. below.

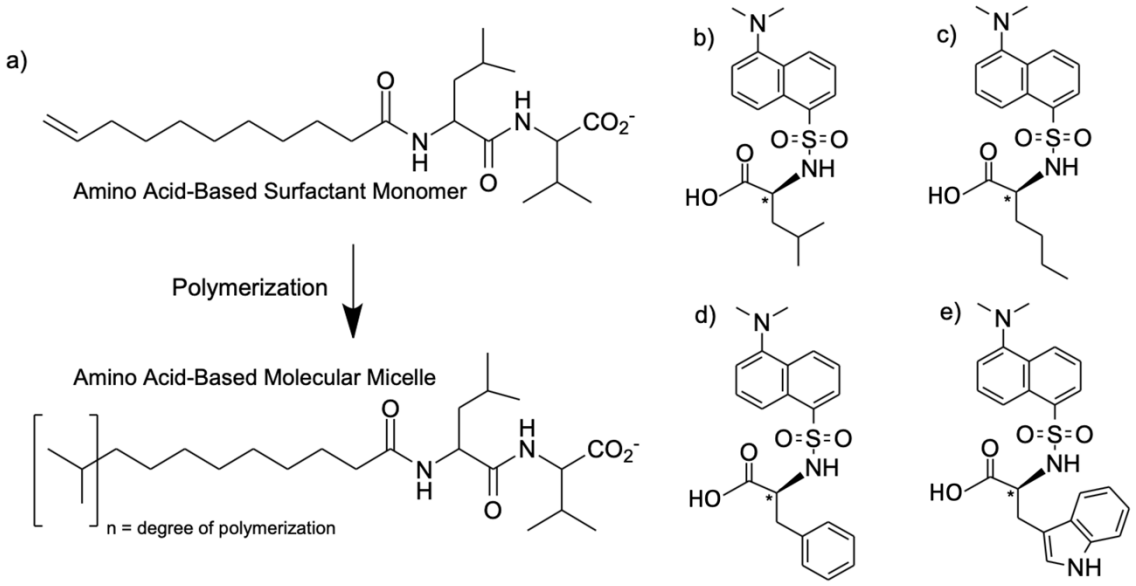


Figure 4.1. Chemical structures of (a) Poly-(sodium Undecyl-(L)-Leucine-Valine), (b) Dansyl Leucine, (c) Dansyl Norleucine, (d) Dansyl Phenylalanine, (e) Dansyl Tryptophan. The four Dansyl amino acids (b-e) contain an asymmetric center denoted by an asterisk.

These polymerized micelles are commonly referred to as amino acid-based molecular micelles (AABMMs).[17] AABMMs have significant advantages over the traditional, non-polymerized micelle, including: AABMMs do not have a critical micelle concentration (CMC), reduced joule heating and most notably enhancing the enantiomeric resolution.[15]

This manuscript focuses on one particular AABMM, poly-sodium *N*-undecanoyl-(L)-Leucylvalinate, or poly(SULV). Previous literature reports that when studying a wide-range of chiral compounds, enantiomeric separation was achieved for more than 75% of the compounds when using poly(SULV) as the pseudostationary phase in MEKC.[14] Chiral recognition with

poly(SULV) has been examined using a variety of techniques. One of those techniques is Nuclear Overhauser Effect Spectroscopy Nuclear Magnetic Resonance (NOESY NMR), which has been used to study the primary site of interaction(s) of chiral analytes to molecular micelles.[18] Molecular dynamics (MD) simulations have also previously been used to investigate AABMMs, such as poly(SULV) to further investigate the factors that contribute to chiral recognition.[5, 17] Some of the chiral analytes previously studied with poly (SULV) include, beta blockers and binaphthyl compounds.[5, 19] This manuscript will focus on examining the binding of various Dansyl amino acids to the AABMM poly SULV.

Due to the intractable number of testing conditions and combinations that exist between: AABMMs, chiral analytes, counterions, surfactant concentration, and pH levels, it is important to develop a predictive model that is ultimately more cost-effective, efficient and requires less time than experimental methods. This investigation is part of a long-term project aimed at developing a Quantitative Structure-Enantioselective Retentions Relationship (QSERR) model.[20, 21] More specifically, the QSERR model will be developed to study and predict the best MEKC conditions for chiral selectivity using various AABMMs that contain at least one chiral center using various dipeptide combinations of the L-form of alanine, valine and leucine as well as the achiral amino acid glycine. To build this predictive model, experimental data from Nuclear Magnetic Resonance (NMR) studies and MEKC will be combined with the insights gained from the MD simulation studies.

Here we report an MD simulation study of Dansyl amino acids, including Dansyl-Leucine (Dans-Leu), Dansyl-Norleucine (Dans-Nor), Dansyl-Tryptophan (Dans-Trp) and Dansyl-Phenylalanine (Dans-Phe) binding to poly(SULV) to further the knowledge of a previous study that examined the same aforementioned systems with experimental MEKC and NOESY NMR.[22]

The chemical structures of these compounds are shown in Figure 4.1b-e. The Dansyl amino acid analytes examined in this study can be divided into two main groups: aliphatic (Dans-Leu and Dans-Nor), and aromatic (Dans-Phe and Dans-Trp). This data is important to incorporate into the QSERR model in the future because it expands the depth and diversity of the analyte training set by introducing chiral analytes with varying characteristics; this will ultimately improve the predictions of the QSERR model. Combining experimental MEKC and NOESY NMR data with computational MD simulation studies will allow for the elucidation of these chiral separation mechanisms and helps to further the knowledge base that is necessary to develop a predictive QSERR database.

4.3 Methods

4.3.1 Building the poly(SULV) micelle computationally

Molecular modeling and MD simulation methods employed in this research project have been reported in detail in previous work from our group.[5, 17-19, 22] The methods for this research project are summarized as follows. First, a monomer unit of *N*-undecanoyl-(L)-Leucylvalinate (ULV) molecule was built with an overall net charge of -1 within the graphical interface, Xleap, within the AMBER18 software package (see Figure 4.1a).[23] The *N*-undecanoyl and (L)-Leucylvalinate, represent the surfactant's hydrocarbon tail and dipeptide head group, respectively. The ULV monomer unit then serves as the template surfactant for building the polymerized micelle. Previous fluorescence quenching experiments have shown that poly(SULV) contains on average approximately 20 surfactant monomer units.[24] For this reason, molecular modelling of poly(SULV) was made by covalently linking 20 ULV surfactants at their terminal carbon atom on *N*-undecanoyl in a linear fashion. The linear chain of 20 ULV surfactants units thus yields an

overall net charge of -20. The two terminal surfactant units on either ends of the linear chain, denoted as ULV1 and ULV20, were then covalently linked at their terminal carbon atom on *N*-undecanoyl to ultimately form a polymer ring, which is representative of the poly(SULV) molecular micelle. Experiments in our group have determined that the polymer ring allows for the proper equilibration of all amino acid-based molecular micelles with the dipeptide combinations of the L-form of alanine, valine and leucine, and achiral glycine, which is necessary to proceed with the long-term study and development of the QSERR model.

After the poly(SULV) molecular micelle was created, AMBER18 was then used to run a 480.0 ns simulation on a system containing the poly(SULV) molecular micelle and 20 sodium counterions, all of which were solvated with TIP3P water molecules within a volume of a 10 Angstrom truncated octahedron. The average structure of poly(SULV) was then calculated using the following procedure.

Step 1) The root mean squared deviation (RMSD) of the 480 ns MD simulation was determined. The RMSD is used to monitor the coordinate changes of a system over time in comparison to a reference set of coordinates. In this case, the reference structure was the initial frame. The RMSD values were calculated with Equation 4.1.

Equation 4.1:

$$RMSD = \sqrt{\frac{\sum_{i=0}^{N}[m_i * (X_i - Y_i)^2]}{M}}$$

The number of atoms in the system, the mass of each atom, the coordinate vector for the target atom, the coordinate vector for the reference atom, and the total mass of the system are denoted as N, m_i, X_i, Y_i and M, respectively.[25, 26] The RMSD protocol has previously been used in other studies from our group and is summarized as follows.[5, 17-19, 27, 28] Throughout the 480.0 ns

simulation, the poly(SULV) micelle structure will begin to change over time as it tries to settle into its energetically-favorable structure. The RMSD plot is composed of a Y-axis representing the RMSD value, and an X-axis representing the simulation time. The RMSD value will initially increase at a rapid rate at the beginning of the simulation then proceed to equilibrate and level off. The leveled-off region in the RMSD plot denotes which frames in the trajectory file contain representative structures of the equilibrated micelle.

Step 2) After determining which frames were associated with the equilibrated region, the average structure was determined from those frames, ultimately yielding an average, equilibrated structure of poly(SULV). This average structure was then used as the reference structure for the subsequent RMSD analysis on the 480 ns trajectory file.

Step 3) To obtain the final equilibrated poly(SULV) molecular micelle structure that was used for this study, the RMSD was calculated on all frames of the 480 ns trajectory file with the reference frame being the average theoretical structure obtained from Step 2. From here, the frame containing the lowest RMSD value is the representative poly(SULV) molecular micelle structure. The poly(SULV) molecular micelle structure was then stripped of its aqueous solvent consisting of water and sodium ions. The non-solvated, equilibrated structure of poly(SULV) was then used for the ligand docking analyses.

4.3.2 Ligand Docking

The Molecular Operating Environment (MOE) software package was used to identify the binding pockets on the representative poly(SULV) and dock ligand enantiomers into each of the identified pockets.[29] The MOE software contains a Site Finder module, which was used to identify the binding pockets of poly(SULV); Site Finder utilizes the alpha sphere method to identify

molecular cavities and potential binding sites.[19, 30] Alpha spheres are placed within cavities where four receptor atoms sit at its spherical boundaries.[31] Non-polar regions with poor hydrogen bonding capabilities are represented by white spheres. Regions where hydrogen bonding interactions are likely to occur are represented by red spheres.[19] Therefore, the hydrophobicity and hydrophilicity of a binding pocket can be quantified. Site Finder identified three binding pockets within poly(SULV).

MOE was then used to dock the L- and D- enantiomers of Dans-Leu, Dans-Nor, Dans-Phe and Dans-Trp into each of the three binding pockets of poly(SULV). During the docking analysis process, poly(SULV) was kept static, whereas the Dansyl enantiomers being investigated were dynamic in their movement and poses. The Triangle Matcher method within MOE was used to screen various poses of the enantiomer into the binding pocket. The Triangle Matcher method screens ligand poses by aligning three atoms on the enantiomeric compound with three alpha spheres within the binding pocket.[32] For each pose that is screened, the London dG scoring function is used to calculate the binding free energy value.[30] The poses are then ranked based on their dG values, and the highest scoring enantiomer is used for studies in the MD simulation. The L- and D- enantiomers of each of the four Dansyl amino acids were docked and scored into the three binding pockets of poly(SULV) resulting in a total of 24 MD simulations to be conducted and analyzed.

MD simulations were conducted on each enantiomer in all three binding pockets of poly(SULV) using AMBER18. The AMBER GAFF force field was employed in these simulations. The GAFF force field was used to describe the chemical parameters of the surfactant using the antechamber package within AMBER18.[33] Each MD simulation of the enantiomer to poly(SULV) complex was solvated with 20 sodium ions and solvated with TIP3P water molecules

within a volume of a 10 Angstrom truncated octahedron. The MD simulation is summarized as follows. The first step in the MD simulation is initiated with a 50 ps minimization step. This was followed by a 20 ps warm-up step, which allowed the system time to increase to 300 K. The system was then allowed to equilibrate to a pressure of 1 atm with a time of 20 ps. The final production run for each MD simulation was conducted for 60 ns. Each simulation was monitored for anomalies with the Visual Molecular Dynamics (VMD) software to ensure that the simulation ran properly. This validation method was used to ensure that the enantiomer being investigated did not leave the pocket during the simulation or do anything else unexpected.

4.3.3 Binding Free Energy Analyses

The binding free energy values ($kJ \cdot mol^{-1}$) for each enantiomer: poly(SULV) molecular complex were calculated using post-trajectory analysis. The trajectory files of each MD simulation for each molecular complex were used to calculate the $\Delta G_{binding}$ values using the Molecular Mechanics–Generalized Born Solvent Accessibility (MM-GBSA) method.[35, 36] Equation 4.2 shows that the $\Delta G_{binding}$ values are calculated for each of the 24 molecular systems by taking the difference between the sum of the separate free energy values of poly(SULV) and the respective enantiomers, with its difference with the poly(SULV):enantiomer complex.[5, 18, 19, 28]

Equation 4.2:

$$\Delta G_{binding} = \Delta G_{complex} - (\Delta G_{poly(SULV)} + \Delta G_{enantiomer})$$

The binding free energy value of each enantiomer in the three pockets of poly(SULV) were used to calculate the percent occupancy, P_i, which calculates where a given enantiomer is most likely to occupy based on its $\Delta G_{binding}$ values in each of the binding pockets. The percent population was calculated with Equation 4.3.[5]

Equation 4.3:

$$P_i = \frac{e^{-\frac{G_i}{K_B * T}}}{\sum_{i=1}^{N} e^{-\frac{G_i}{K_B * T}}} * 100$$

The binding free energy of the enantiomer in the ith pocket of poly(SULV), Boltzmann's constant, and Kelvin temperature are represented by G_i, K_B, T, respectively.

4.3.4 Hydrogen Bond Analyses

The CPPTRAJ utility in AMBER18 was used to perform hydrogen bond analyses on each of the 24 poly(SULV):enantiomer complexes.[25] Hydrogen bonding interactions occur when a hydrogen atom is covalently linked to heavy atoms, such as fluorine, oxygen or nitrogen, and that same hydrogen also shares intermolecular interactions with other nearby heavy atoms (usually fluorine, oxygen or nitrogen). The heavy atom that the hydrogen atom is covalently linked to is known as the donor atom, whereas the acceptor atom is the nearby atom that the hydrogen atom is having intermolecular interactions with. The CPPTRAJ utility is able to analyze each trajectory file and track the hydrogen bonds that are being broken and formed throughout the 60 ns simulation. It does so by following two main geometric criteria that are characteristic for hydrogen bonds: (1) the distance between the donor-to-acceptor heavy atoms must be within 3 Å of each other, and (2) the donor-hydrogen-acceptor angle cutoff must be ± 30°.[5]

4.4 Results and Discussion

Previously reported data shows that when using poly(SULV) as the pseudostationary phase in MEKC experiments in conjunction with the L- and D- enantiomers of Dans-Leu, Dans-Nor, Dans-Trp and Dans-Phe, the L- enantiomer of each of these compounds eluted second.[22] This indicates that the L- enantiomers of each of the Dansyl amino acids studied have stronger binding interactions with poly(SULV) than the D- enantiomers. The previous study also investigated the interactions between the Dansyl amino acids and poly(SULV) using NOESY NMR. However, due to spectral overlap in the corresponding NOESY NMR spectra between the Dansyl amino acids and poly(SULV), the NOE interactions were not fully determined.[22] By investigating the interactions between the Dansyl amino acids and poly(SULV) via MD simulation studies, molecular-level insight will be able to provide information about the molecular interactions that govern chiral recognition in these molecular systems.

As mentioned previously, the MOE software package identified poly(SULV) as having three binding pockets, which can be found in Figure 4.2.

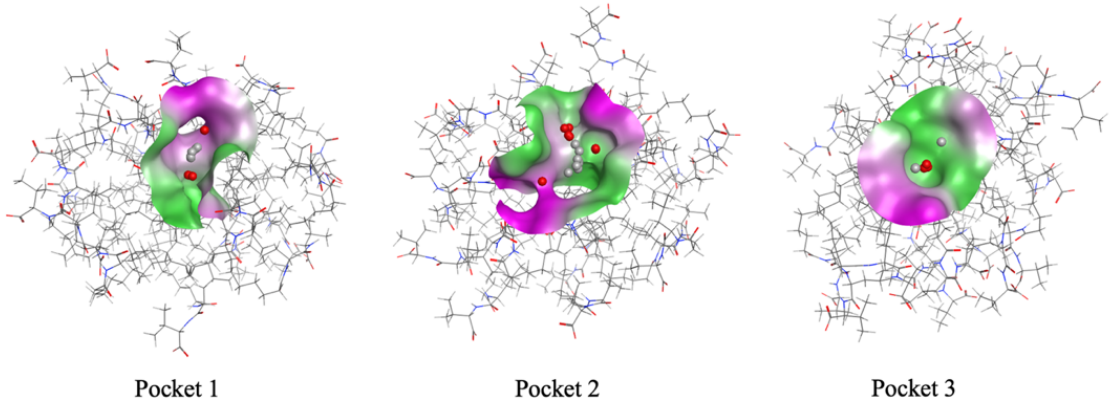


Figure 4.2. Poly(SULV) binding pockets identified via MOE analysis. Three binding pockets were identified and populated with alpha centers. The lipophilic and hydrophilic regions of the pocket are denoted by green and purple, respectively.

Additionally, each pocket in relation to each other in poly(SULV) is shown in Figure 4.3.

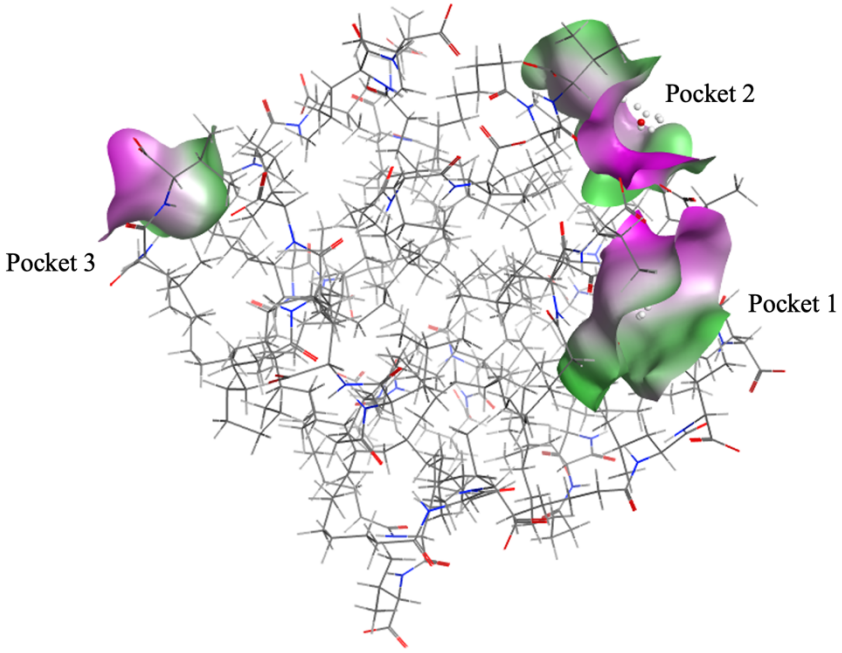


Figure 4.3. Spatial orientation of the three Poly(SULV) binding pockets in respect to each other.

The binding free energy values and the occupancy percentages for each pocket with each of the Dansyl amino acid enantiomers are summarized in Table 4.1 shown below. This will be discussed in the subsequent sections.

Table 4.1. Binding free energy values and pocket percent occupied for each of the Dansyl amino acid enantiomers binding to poly(SULV).

	Pocket 1	Pocket 2	Pocket 3
Dansyl (L) Leucine			
$\Delta G_{binding}$ (kJ·mol^{-1})	-21.8938	-8.4391	-9.4166
Percent occupied	98.84%	0.50%	0.66%
Dansyl (D) Leucine			
$\Delta G_{binding}$ (kJ·mol^{-1})	-14.5811	0.1686	-9.3757
Percent occupied	87.84%	1.28%	10.89%
Dansyl (L) Norleucine			
$\Delta G_{binding}$ (kJ·mol^{-1})	-22.1763	-13.6191	-12.2257
Percent occupied	95.17%	3.07%	1.76%
Dansyl (D) Norleucine			
$\Delta G_{binding}$ (kJ·mol^{-1})	-15.9457	-7.8426	-14.2665
Percent occupied	64.57%	2.50%	32.93%
Dansyl (L) Tryptophan			
$\Delta G_{binding}$ (kJ·mol^{-1})	-18.8091	-11.5680	-21.3329
Percent occupied	26.27%	1.44%	72.29%
Dansyl (D) Tryptophan			
$\Delta G_{binding}$ (kJ·mol^{-1})	-8.6631	-13.6408	-11.7956
Percent occupied	8.42%	62.00%	29.58%
Dansyl (L) Phenylalanine			
$\Delta G_{binding}$ (kJ·mol^{-1})	-12.4774	-10.2752	-13.3349
Percent occupied	35.41%	14.64%	49.95%
Dansyl (D) Phenylalanine			
$\Delta G_{binding}$ (kJ·mol^{-1})	-11.5032	-	-12.0959
Percent occupied	44.08%	0.00%	55.92%

The binding free energy data for each Dansyl amino acid will be summarized in the following sections.

Furthermore, previous studies also showed that experimental NOESY NMR data was inconclusive in determining which interactions governed chiral selectivity due to spectral overlap between the NOE intensities of the Dansyl amino acids and poly(SULV).[22] Therefore, hydrogen bond analyses will be used to rationalize why the L- enantiomers of the Dansyl amino acids bind stronger to poly(SULV) than the D- enantiomers, while providing molecular-level insight into the

specific bonding interactions that govern chiral selectivity in these systems. The trajectory files for each molecular system will be analyzed for its intermolecular hydrogen bonds that are formed throughout the simulation. The hydrogen bonding analyses will provide detailed information about the acceptor atoms and donor atoms that are participating in hydrogen bonding.[5, 18, 19] Additionally, the highest number of consecutive frames that a hydrogen bond is maintained will be denoted as the 'max lifetime' for that specific hydrogen bond; this provides insight into the frequency of the hydrogen bond being formed.[37] The hydrogen bond occupancy represents the percentage of frames that a hydrogen bond is maintained throughout an entire simulation. Table 4.2 below displays bonds that had greater than a 10% hydrogen bond occupancy throughout the simulations.

Table 4.2. Hydrogen bonds formed between Poly(SULV) and aliphatic Dansyl amino acids, consisting of Dans-Leu and Dans-Nor.

Acceptor atom	Donor atom	Max Lifetime	Hydrogen Bond Occupancy
Dansyl-(L)-Leucine pocket one			
Dans-Leu C=O	MM chain 16: Leu NH	4895	46.36%
MM chain 18: Val C=O	Dans-Leu NH	1320	17.83%
MM chain 18: Leu C=O	Dans-Leu NH	345	10.41%
Dansyl-(D)-Leucine pocket one			
MM chain 17: Leu C=O	Dans-Leu NH	1131	33.68%
Dans-Leu C=O	MM chain 20: Leu NH	1737	14.22%
Dansyl-(L)-Norleucine pocket one			
MM chain 16: Leu C=O	Dans-Nor NH	2838	61.41%
MM chain 17: Val CO_2^-	Dans-Nor OH	132	45.66%
Dans-Nor C=O	MM chain 17: Leu NH	4358	39.47%
MM chain 17: Val CO_2^-	Dans-Nor OH	99	25.71%
Dans-Nor C=O	MM chain 17: Val NH	4026	21.26%
Dansyl-(D)-Norleucine pocket one			
MM chain 17: Val CO_2^-	Dans-Nor OH	215	33.28%

All hydrogen bond occupancies less than 10% were considered negligible and therefore did not contribute greatly to chiral selectivity. In the following sections, the binding free energy

values for each Dansyl amino acid will be rationalized using the data collected from hydrogen bonding analyses.

4.4.1 Dans-Leu binding to poly(SULV)

Dansyl-(L)-Leucine and Dansyl-(D)-Leucine were docked into each of the three binding pockets identified in poly(SULV) as shown in Figure 4.4a-f.

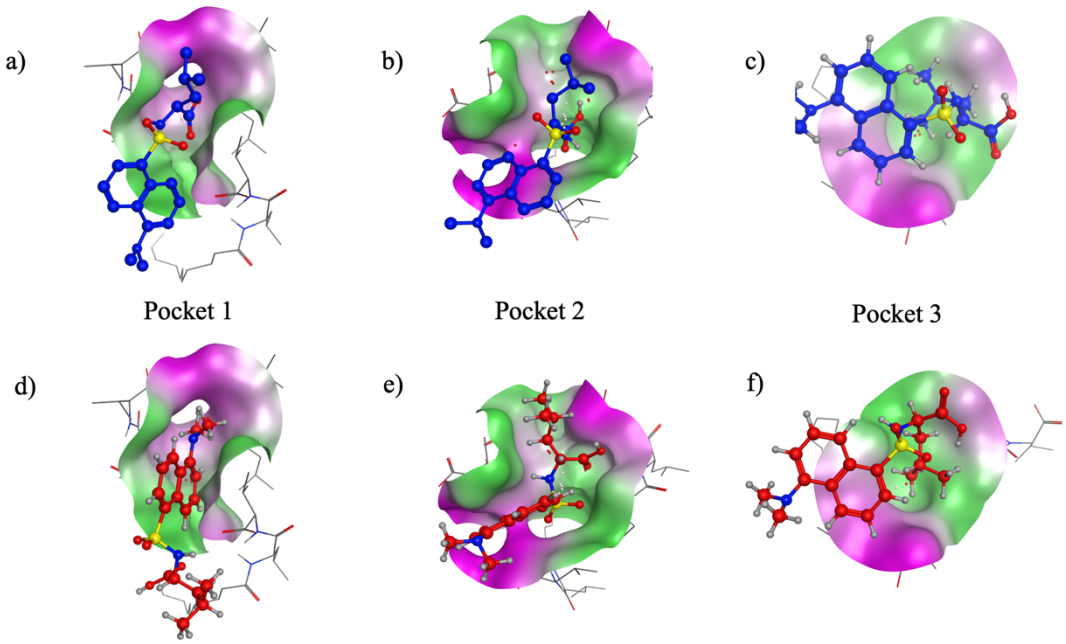


Figure 4.4. The highest scoring docked position of Dansyl-(L)-Leucine (blue) in each binding pocket of Poly(SULV) is displayed in (a) Pocket 1 (b) Pocket 2 and (c) Pocket 3. The highest scoring docked position of Dansyl-(D)-Leucine (red) in each binding pocket of Poly(SULV) is displayed in (d) Pocket 1 (e) Pocket 2 and (f) Pocket 3.

Table 4.1 shows that for Dansyl-(L)-Leucine, the binding free energy values for poly(SULV) pockets one, two and three are: -21.8938, -8.4391 and -9.4166 kJ·mol^{-1}, respectively. The percent occupancies can be calculated by employing Equation 4.3 to this data set. The percent occupancy for Dansyl-(L)-Leucine in pockets one, two and three of poly(SULV) are: 98.84%, 0.50% and 0.66%, respectively. Based on this data, it is evident that Dansyl-(L)-Leucine clearly

favors pocket one as shown by its binding free energy value and percent occupancy compared to the other binding pockets. The data indicates that Dansyl-(L)-Leucine does not favor pockets two and three, as this enantiomer has less than 1% occupancy in those locations based on its unfavorable binding free energy values, when compared to pocket one. Therefore, pockets two and three can be considered so low they are nearly negligible.

Table 4.1 indicates that the binding free energy values of Dansyl-(D)-Leucine to pockets one, two and three of poly(SULV) are -14.5811, 0.1686 and -9.3757 $kJ \cdot mol^{-1}$, respectively. The percent occupancy for Dansyl-(D)-Leucine in pockets one, two and three of poly(SULV) are 87.84%, 1.28% and 10.89%, respectively. It is evident that of the three pockets available, Dansyl-(D)-Leucine significantly favors pocket one, whereas this enantiomer does not favor pockets two or three. Since Dansyl-(L)-Leucine had the highest percent occupancy in pocket one with a binding free energy value of -21.8938 $kJ \cdot mol^{-1}$, this will be compared to the pocket and binding free energy values that Dansyl-(D)-Leucine favored most. In this case, Dansyl-(D)-Leucine also had the highest percent occupancy in pocket one with a binding free energy value of -14.5811 $kJ \cdot mol^{-1}$. When comparing the two binding free energy values, it is evident that Dansyl-(L)-Leucine has a more negative or stronger binding interaction to poly(SULV) when compared to Dansyl-(D)-Leucine. Therefore, analysis of the MD simulations show that Dansyl-(L)-Leucine would elute second, thus matching experimental MEKC data showing that the L- enantiomer of Dans-Leu elutes second when using poly(SULV) as the pseudostationary phase.

As confirmed by the binding free energy values in Table 4.1, Dansyl-(L)-Leucine binds stronger to poly(SULV) than Dansyl-(D)-Leucine. It is evident that both of the Dans-Leu enantiomers prefer pocket one of poly(SULV) when compared to the remaining two pockets. It is, therefore, important to evaluate and compare the hydrogen bonds that are formed between both

enantiomers and poly(SULV) to fully understand why Dansyl-(L)-Leucine binds stronger to poly(SULV) than Dansyl-(D)-Leucine. Table 4.2 outlines the hydrogen bonding interactions formed between poly(SULV) and both enantiomers of Dans-Leu.

For Dansyl-(L)-Leucine, three major hydrogen bonds were identified. The carbonyl on Dans-Leu acted as the acceptor atom, which interacted with the Leucine N-H of poly(SULV). This bond had the longest max lifetime of 4895 frames, as well as the highest hydrogen bond occupancy of 46.36%. The second most prominent hydrogen bond formed between the Valine carbonyl on poly(SULV) which acted as the acceptor atom for Dans-Leu N-H; this had a max lifetime of 1320 frames, and a hydrogen bond occupancy of 17.83%. The third most prominent hydrogen bond formed between the Leucine carbonyl on poly(SULV) and the Dans-Leu N-H with a max lifetime of 345 frames and a hydrogen bond occupancy of 10.41%. Based on this data, it is evident that the key interaction is the bond formed between the carbonyl on Dans-Leu and the poly(SULV) Leucine N-H. This suggests that Dansyl-(L)-Leucine preferentially binds to the inner amino acid of poly(SULV), thus interacting more within the core of the micelle rather than its exterior amino acid, Valine.

In Dansyl-(D)-Leucine, two major hydrogen bonds were identified throughout the simulation. The most prominent hydrogen bond formed between the Leucine carbonyl on poly(SULV) and the Dans-Leu N-H with a max lifetime of 1131 frames and a hydrogen bond occupancy of 33.68%. The second most prominent hydrogen bond for Dansyl-(D)-Leucine was formed between the carbonyl on Dans-Leu and the poly(SULV) Leu N-H bond with a max lifetime of 1737 frames and a hydrogen bond occupancy of 14.22%. While this bond had a longer max lifetime it was not sustained quite as long throughout the simulation as the other aforementioned hydrogen bond interaction. It is evident that the hydrogen bond formed between the Leucine

carbonyl on poly(SULV) and the Dans-Leu N-H contributed the most to the binding free energy value, and similar to Dansyl-(L)-Leucine, it also preferentially binds to the inner amino acid head group of poly(SULV).

When comparing the hydrogen bonding analysis data for Dansyl-(L)-Leucine and Dansyl-(D)-Leucine, it is evident that more hydrogen bonds are able to be formed with the L- enantiomer than the D- enantiomer, and for a longer period of time. More specifically, Dansyl-(L)-Leucine was able to form three strong intermolecular hydrogen bonds with poly(SULV) compared to the two intermolecular hydrogen bonds formed with Dansyl-(D)-Leucine in pocket one of poly(SULV). Therefore, the hydrogen bonding interactions justify the stronger binding free energy values of Dansyl-(L)-Leucine to poly(SULV) than that of Dansyl-(D)-Leucine. For both enantiomers, the preferred binding interactions are near the inner amino acid head group of poly(SULV), suggesting its significance in separating the Dans-Leu enantiomers.

4.4.2 Dans-Nor binding to poly(SULV)

Dansyl-(L)-Norleucine and Dansyl-(D)-Norleucine were docked into each of the three binding pockets identified in poly(SULV) as shown in Figure 4.5a-f.

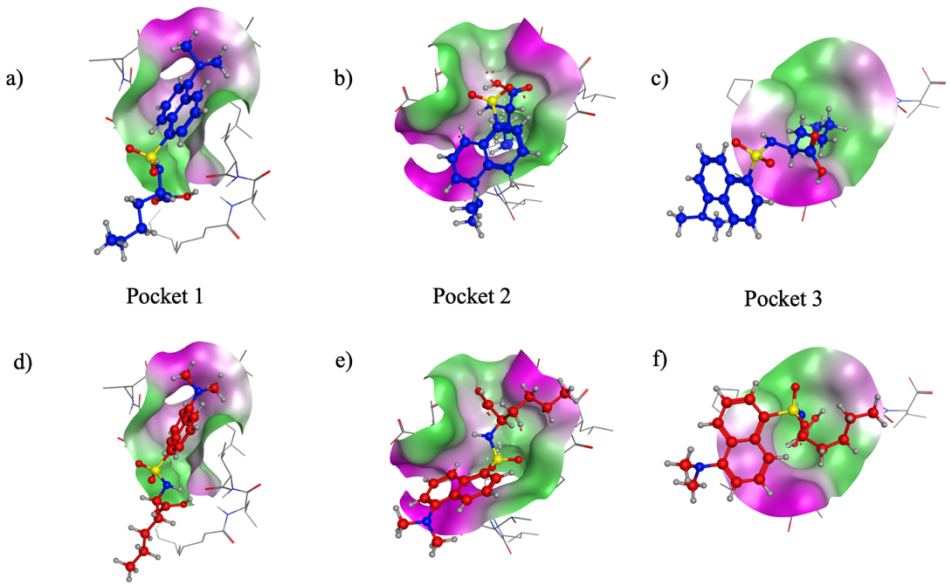


Figure 4.5. The highest scoring docked position of Dansyl-(L)-Norleucine (blue) in each binding pocket of Poly(SULV) is displayed in (a) Pocket 1 (b) Pocket 2 and (c) Pocket 3. The highest scoring docked position of Dansyl-(D)-Norleucine (red) in each binding pocket of Poly(SULV) is displayed in (d) Pocket 1 (e) Pocket 2 and (f) Pocket 3.

Table 4.1 shows that for Dansyl-(L)-Norleucine, the binding free energy values for poly(SULV) pockets one, two and three are: -22.1763, -13.6191 and -12.2257 $kJ \cdot mol^{-1}$, respectively. The percent occupancy for Dansyl-(L)-Norleucine in pockets one, two and three of poly(SULV) are: 95.17%, 3.07% and 1.76%, respectively. Based on this data, it is evident that Dansyl-(L)-Norleucine favors pocket 1 as shown by its binding free energy value and percent occupancy compared to the other binding pockets. The data indicates that Dansyl-(L)-Norleucine does not favor binding pockets two and three, as this enantiomer has a range of approximately 1-3% for occupying those binding pockets based on its unfavorable binding free energy values.

Table 4.1 also indicates that the binding free energy values of Dansyl-(D)-Norleucine to binding pockets one, two and three of poly(SULV) are -15.9457, -7.8426 and -14.2665 $kJ \cdot mol^{-1}$, respectively. The percent occupancy for Dansyl-(D)-Norleucine in pockets one, two and three of poly(SULV) are 64.57%, 2.50% and 32.93%, respectively. It is evident that of the three pockets

available, Dansyl-(D)-Norleucine favors pocket one, whereas this enantiomer does not favor pocket three. Approximately one-third of the simulation time, Dansyl-(D)-Norleucine might spend time in pocket two, however it significantly favors pocket one.

Since Dansyl-(L)-Norleucine had the highest percent occupancy in pocket one with a binding free energy value of -22.1763 $kJ \cdot mol^{-1}$, this will be compared to the pocket and binding free energy values that Dansyl-(D)-Norleucine favored most. In this case, Dansyl-(D)-Norleucine also had the highest percent occupancy in pocket one with a binding free energy value of -15.9457 $kJ \cdot mol^{-1}$. When comparing the two binding free energy values of both enantiomers in their most favored pockets, it is evident that Dansyl-(L)-Norleucine has a more negative or stronger binding interaction to poly(SULV) as compared to its enantiomer, Dansyl-(D)-Norleucine. Therefore, analysis of the MD simulations show that Dansyl-(L)-Norleucine would elute second, which is consistent with experimental MEKC data when using poly(SULV) as the pseudostationary phase to separate the enantiomers of Dans-Nor. As confirmed by the binding free energy values in Table 4.1, Dansyl-(L)-Norleucine binds stronger to poly(SULV) than Dansyl-(D)-Norleucine. It is also evident that both of the Dans-Nor enantiomers prefer pocket one of poly(SULV). It is, therefore, important to evaluate and compare the hydrogen bonds that are formed between both enantiomers and poly(SULV) to fully understand why Dansyl-(L)-Norleucine binds stronger to poly(SULV) than Dansyl-(D)-Norleucine. Table 4.2 outlines the hydrogen bonding interactions formed between poly(SULV) and both enantiomers of Dans-Nor. The data is summarized as follows.

The most prominent hydrogen bond in Dansyl-(L)-Norleucine with poly(SULV) pocket one was formed between the carbonyl of Leucine in poly(SULV) with the N-H bond on Dans-Nor. This bond had a max lifetime of 2838 frames and a hydrogen bond occupancy of 61.41%. Dansyl-(L)-Norleucine also had an additional four other hydrogen bonds formed, in which the details on

their interactions are outline in Table 4.2. Thus, Dansyl-(L)-Norleucine had a total of five intermolecular hydrogen bonds within pocket one of poly(SULV), which had preferential binding to the inner amino acid head group of poly(SULV), which is Leucine.

Dansyl-(D)-Norleucine only had one prominent hydrogen bond, which was between the terminal carboxylate group on the outer amino acid, Valine, in poly(SULV) and the O-H group on Dans-Nor. This bond had a max lifetime of 215 with a hydrogen bond occupancy of 33.28%. Since this had a lower binding energy value, it is possible that its conformation did not allow it to penetrate as deeply into the micellar core to interact with the inner amino acid head group.

In conclusion, Dansyl-(L)-Norleucine was able to form significantly more hydrogen bonds and for longer periods of time than that of Dansyl-(D)-Norleucine in pocket one of poly(SULV). More specifically, Dansyl-(L)-Norleucine was able to form a total of five intermolecular hydrogen bonds with poly(SULV), whereas Dansyl-(L)-Norleucine only formed one. This ultimately contributes to the stronger binding free energy value seen in Dansyl-(L)-Norleucine when compared to the D- enantiomer binding to poly(SULV).

4.4.3 Dans-Trp binding to poly(SULV)

Dansyl-(L)-Tryptophan and Dansyl-(D)-Tryptophan were docked into each of the three binding pockets identified in poly(SULV) as shown in Figure 4.6a-f.

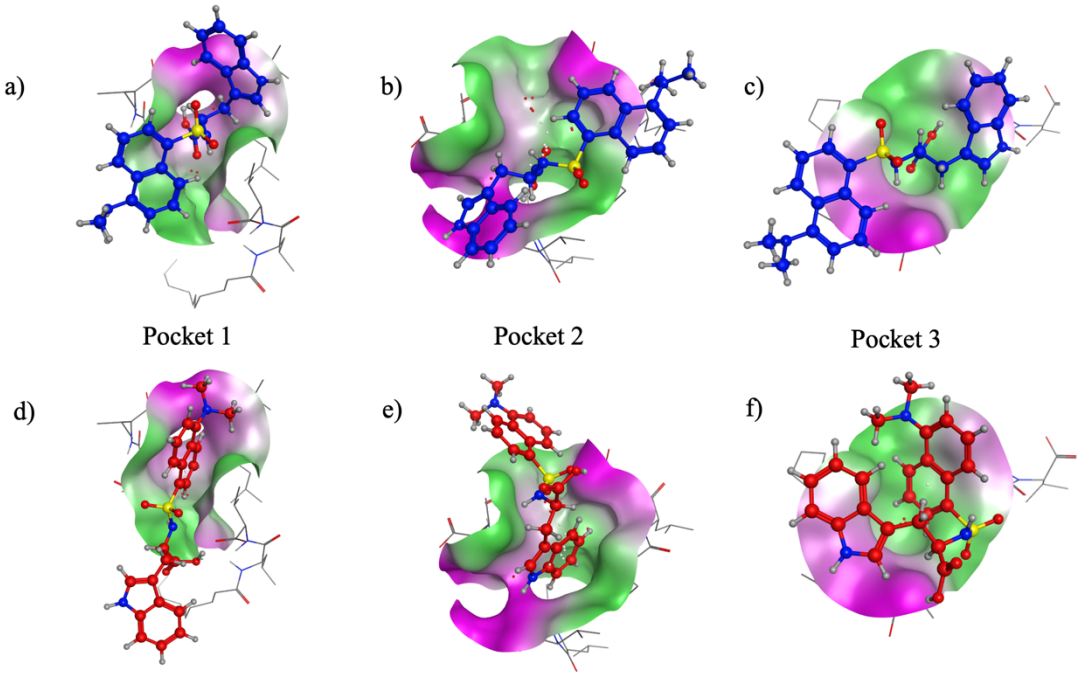


Figure 4.6. The highest scoring docked position of Dansyl-(L)-Tryptophan (blue) in each binding pocket of Poly(SULV) is displayed in (a) Pocket 1 (b) Pocket 2 and (c) Pocket 3. The highest scoring docked position of Dansyl-(D)-Tryptophan (red) in each binding pocket of Poly(SULV) is displayed in (d) Pocket 1 (e) Pocket 2 and (f) Pocket 3.

Table 4.1 shows that for Dansyl-(L)-Tryptophan, the binding free energy values for poly(SULV) pockets one, two and three are: -18.8091, -11.5680 and -21.3329 $kJ \cdot mol^{-1}$, respectively. The percent occupancy for Dansyl-(L)-Tryptophan in pockets one, two and three of poly(SULV) are: 26.27%, 1.44% and 72.29%, respectively. Based on this data, it is evident that Dansyl-(L)-Tryptophan favors pocket three as shown by its binding free energy value and percent occupancy compared to the other binding pockets. The data indicates that Dansyl-(L)-Tryptophan does not favor binding pocket two as this enantiomer has a less than 2% occupying for this site. Dansyl-(L)-Tryptophan would spend approximately a quarter of its simulation time in pocket one. On the other hand, most of the time this enantiomer favors pocket three based on its binding percent occupancy.

Table 4.1 indicates that the binding free energy values of Dansyl-(D)-Tryptophan to binding pockets one, two and three of poly(SULV) are -8.6631, -13.6408 and -11.7956 $kJ \cdot mol^{-1}$, respectively. The percent occupancy for Dansyl-(D)-Tryptophan in pockets one, two and three of poly(SULV) are 8.42%, 62.00% and 29.58%, respectively. It is evident that of the three total pockets available, Dansyl-(D)-Tryptophan favors pocket two, whereas this enantiomer does not favor pocket one and shows some favorability for pocket three.

Dansyl-(L)-Tryptophan had the highest percent occupancy in pocket three with a binding free energy value of -21.3329 $kJ \cdot mol^{-1}$, therefore, this will be compared to the pocket and binding free energy values that Dansyl-(D)-Tryptophan favor most. In this case, Dansyl-(D)-Tryptophan had the highest percent occupancy (62%) in pocket two with a binding free energy value of -13.6408 $kJ \cdot mol^{-1}$. When comparing the two binding free energy values of both enantiomers in their favored poly(SULV) pockets, it is evident that Dansyl-(L)-Tryptophan has a more negative or stronger binding interaction to poly(SULV) than that of the Dansyl-(D)-Tryptophan enantiomer. This data is consistent with experimental MEKC elution data because when separating Dans-Trp enantiomers using poly(SULV) as the pseudostationary phase, the L- enantiomer elutes second.

As confirmed by the binding free energy values in Table 4.1, Dansyl-(L)-Tryptophan binds stronger to poly(SULV) than Dansyl-(D)-Tryptophan. It is also evident that Dansyl-(L)-Tryptophan prefers pocket three of poly(SULV), whereas Dansyl-(D)-Tryptophan prefers pocket two. It is, therefore, important to evaluate and compare the hydrogen bonds that are formed between both enantiomers in those respective pockets of poly(SULV) to fully understand why Dansyl-(L)-Tryptophan binds stronger to poly(SULV) than Dansyl-(D)-Tryptophan. Table 4.3 outlines the hydrogen bonding interactions formed between poly(SULV) and both enantiomers of Dans-Trp. The data is summarized as follows.

Table 4.3. Hydrogen bonds formed between Poly(SULV) and aromatic Dansyl amino acids, consisting of Dans-Trp and Dans-Phe.

Acceptor atom	Donor atom	Max Lifetime	Hydrogen Bond Occupancy
Dansyl-(L)-Tryptophan pocket three			
MM chain 13: Val CO_2^-	Dans-Trp NH	600	61.66%
MM chain 13: Val CO_2^-	Dans-Trp OH	386	61.14%
MM chain 13: Val CO_2^-	Dans-Trp NH	656	35.82%
MM chain 13: Val CO_2^-	Dans-Trp OH	333	35.11%
Dans-Trp OH	MM chain 13: Leu NH	1955	18.76%
Dansyl-(D)-Tryptophan pocket two			
Dans-Trp C=O	MM chain 6: Leu NH	4704	58.90%
MM chain 4: Val CO_2^-	Dans-Trp OH	179	15.45%
Dans-Trp C=O	MM chain 6: Leu NH	971	10.99%
Dansyl-(L)-Phenylalanine pocket three			
MM chain 13: Leu C=O	Dans-Phe OH	9903	33.01%
Dansyl-(D)-Phenylalanine pocket three			
MM chain 12: Val CO_2^-	Dans-Phe NH	1148	~3.83%
MM chain 14: Val CO_2^-	Dans-Phe OH	594	~1.98%
MM chain 14: Val CO_2^-	Dans-Phe NH	583	~1.94%

For Dansyl-(L)-Tryptophan, a total of five intermolecular hydrogen bonds were formed with pocket three of poly(SULV). The most prominent hydrogen bonding interactions occurred between the terminal carboxylate group on the outer amino acid, Valine, in poly(SULV) and the N-H bond on Dans-Trp. This hydrogen bond had a max lifetime of 600 frames and a hydrogen bond occupancy of 61.66%. Another strong hydrogen bond worth mentioning is the one formed between the terminal carboxylate group on the outer amino acid, Valine, in poly(SULV) and the O-H moiety on Dans-Trp. This bond had a max lifetime of 386 frames, with a hydrogen bond occupancy of 61.14%. In contrast, Dansyl-(D)-Tryptophan formed 3 intermolecular hydrogens. The most prominent was the bond formed between the carbonyl on Dans-Trp and the Leu N-H

group on poly(SULV). This bond had a max lifetime of 4704 frames and a hydrogen bond occupancy of 58.90%.

Therefore, Dansyl-(L)-Tryptophan was able to form significantly more hydrogen bonds and for longer periods of time in pocket three than that of Dansyl-(D)-Tryptophan in pocket two of poly(SULV). Dansyl-(L)-Tryptophan was able to form a total of five intermolecular hydrogen bonds (four of which had bond occupancies of greater then 30%) with poly(SULV), whereas Dansyl-(D)-Tryptophan only formed three with only one of those higher a higher occupancy than 30%. This ultimately contributes to the stronger binding free energy value seen in Dansyl-(L)-Tryptophan when compared its D- enantiomer binding to poly(SULV).

4.4.4 Dans-Phe binding to poly(SULV)

Dansyl-(L)-Phenylalanine and Dansyl-(D)-Phenylalanine were docked into each of the three binding pockets identified in poly(SULV) as shown in Figure 4.7a-f.

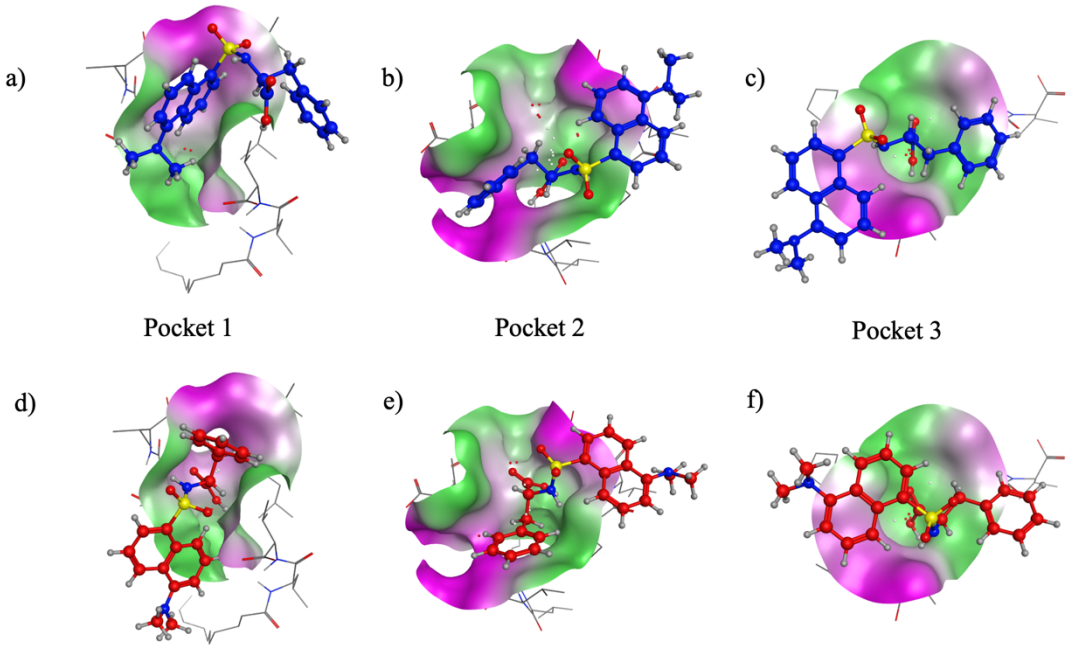


Figure 4.7. The highest scoring docked position of Dansyl-(L)-Phenylalanine (blue) in each binding pocket of Poly(SULV) is displayed in (a) Pocket 1 (b) Pocket 2 and (c) Pocket 3. The highest scoring docked position of Dansyl-(D)- Phenylalanine (red) in each binding pocket of Poly(SULV) is displayed in (d) Pocket 1 (e) Pocket 2 and (f) Pocket 3.

Table 4.1 shows that for Dansyl-(L)-Phenylalanine, the binding free energy values for poly(SULV) pockets one, two and three are: -12.4774, -10.2752 and -13.3349 kJ·mol^{-1}, respectively. The percent occupancy for Dansyl-(L)-Phenylalanine in pockets one, two and three of poly(SULV) are: 35.41%, 14.64% and 49.95%, respectively. Based on this data, it is evident that Dansyl-(L)-Phenylalanine favors pocket three, followed closely by pocket one as shown by those binding free energy values and percent occupancies as compared to pocket two. Dansyl-(L)-Phenylalanine ultimately favors pocket three.

Table 4.1 indicates that the binding free energy values of Dansyl-(D)-Phenylalanine to binding pockets one and three of poly(SULV) are -11.5032 and -12.0959 kJ·mol^{-1}, respectively. Pocket two was considered unfavorable and thus negligible as the top scoring docked analyte continuously left the pocket when visualizing the trajectory files for those MD simulations. In that case, the second best scored analyte was then tested and also continuously left the binding pocket during simulations. The simulations were viewed using the VMD software. The percent occupancy for Dansyl-(D)-Phenylalanine in pockets one, two and three of poly(SULV) are 44.08%, 0.00% and 55.92%, respectively. It is evident that of the three total pockets available, Dansyl-(D)-Phenylalanine favors pocket three based on its % occupancy and binding free energy value.

Dansyl-(L)-Phenylalanine had the highest percent occupancy in pocket three with a binding free energy value of -13.3349 kJ·mol^{-1}, therefore, this will be compared to the pocket and binding free energy values that Dansyl-(D)-Phenylalanine favor most. In this case, Dansyl-(D)-Phenylalanine also had its highest percent occupancy in pocket three with a binding free energy value of -12.0959 kJ·mol^{-1}. When comparing the two binding free energy values of both enantiomers in their favored poly(SULV) pockets, it is evident that Dansyl-(L)-Phenylalanine has a more negative or stronger binding interaction to poly(SULV) than that of the Dansyl-(D)-

Phenylalanine enantiomer. Therefore, the L- enantiomer of Dans-Phe would elute second in MEKC studies based on the computational binding free energy values. This data is also consistent with experimental MEKC data because when separating Dans-Phe enantiomers using poly(SULV) as the pseudostationary phase, the L- enantiomer elutes second.

As confirmed by the binding free energy values in Table 4.1, Dansyl-(L)-Phenylalanine binds stronger to poly(SULV) than Dansyl-(D)-Phenylalanine. It is also evident that both enantiomers of Dansyl-Phe prefer pocket three of poly(SULV). It is, therefore, important to evaluate and compare the hydrogen bonds that are formed between both enantiomers in pocket three of poly(SULV) to fully understand why Dansyl-(L)-Phenylalanine binds stronger to poly(SULV) than Dansyl-(D)-Phenylalanine. Table 4.3 outlines the hydrogen bonding interactions formed between poly(SULV) and both enantiomers of Dans-Phe. The data is summarized as follows.

One major hydrogen bond is notable between Dansyl-(L)-Phenylalanine and poly(SULV), which is formed between the Leu carbonyl on poly(SULV) and the O-H moiety on Dans-Phe. This bond has a max lifetime of 9903 frames and a hydrogen bond occupancy of 33.01%. In contrast, Dansyl-(D)-Phenylalanine did not have any significant hydrogen bonds as all of its hydrogen bond occupancies were below 4.00%.

Therefore, Dansyl-(L)-Phenylalanine was able to form significantly more hydrogen bonds and for longer periods of time in pocket three of poly(SULV) than that of Dansyl-(D)-Phenylalanine. Dansyl-(L)-Tryptophan was able to form one intermolecular hydrogen bond with a relatively long max lifetime, whereas the D- enantiomer did not form any significant hydrogen bonding interactions with pocket three of poly(SULV). This ultimately contributes to the stronger

binding free energy value observed in Dansyl-(L)-Phenylalanine when compared its D- enantiomer binding to poly(SULV).

4.5 Conclusion

The binding free energy values for the Dansyl amino acids examined in this study to poly(SULV) were all in agreement with the enantiomeric order determined by MEKC, in which the L- enantiomer of each Dansyl amino acid interacted stronger with poly(SULV) than the D-enantiomer. Spectral overlap in NOESY NMR spectra ultimately led to inconclusive results on the exact interactions that occurred between the Dansyl amino acids binding to poly(SULV). Therefore, MD simulations were able to provide molecular-level insight into the hydrogen bonding interactions that govern the chiral separation in these systems. In conclusion, the binding free energy and hydrogen bond analyses studies on poly(SULV) in conjunction with the enantiomers of Dans-Leu, Dans-Nor, Dans-Trp and Dans-Phe were able to provide insight into the molecular interactions that govern chiral selectivity. Furthermore, the computational hydrogen bond analyses supported the computational binding free energy values, which helped to rationalize the reasoning for why the L- enantiomers of each Dansyl amino acid had stronger binding interactions with the amino acid-based molecular micelle, poly(SULV), as opposed to its D- enantiomer. In conclusion, this computation investigation was in agreement with experimental MEKC enantiomer order data, and able to provide insight into the chiral separation mechanisms of these systems. This study also shows that when experimental NMR studies provide inconclusive results of binding interactions due to spectral overlap, MD simulations may be able to supplement information about the underlying chemical interactions.

4.6 References

1. Nguyen, L. A.; He, H.; Pham-Huy, C., Chiral drugs: an overview. *Int J Biomed Sci* **2006,** *2* (2), 85-100.

2. Ananthi, N., Role of Chirality in Drugs. *Organic & Medicinal Chemistry International Journal* **2018,** *5* (3).

3. Stolarska, M.; Bocian, W.; Sitkowski, J.; Naumczuk, B.; Bednarek, E.; Popławska, M.; Błażewicz, A.; Kozerski, L., Cathinones - Routine NMR methodology for enantiomer discrimination and their absolute stereochemistry assignment, using R-BINOL. *Journal of Molecular Structure* **2020,** *1219*, 128575.

4. Chhabra, N.; Aseri, M.; Padmanabhan, D., A review of drug isomerism and its significance. *International Journal of Applied and Basic Medical Research* **2013,** *3* (1), 16-18.

5. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Ingle, J. A.; Krause, K. B.; Lewis, C. R.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., Using molecular dynamics simulations to identify the key factors responsible for chiral recognition by an amino acid-based molecular micelle. *Journal of Dispersion Science and Technology* **2019,** *40* (5), 716-727.

6. Can Başer, K. H.; Özek, T., Chapter 22 - Analysis of Essential Oils and Fragrances by Gas Chromatography. In *Gas Chromatography*, Poole, C. F., Ed. Elsevier: Amsterdam, 2012; pp 519-527.

7. Yu, R. B.; Quirino, J. P., Chiral Selectors in Capillary Electrophoresis: Trends During 2017–2018. *Molecules* **2019,** *24* (6), 1135.

8. Ramos, Z.; Rothbauer, G. A.; Turner, J.; Lewis, C.; Morris, K.; Billiot, E.; Billiot, F.; Fang, Y., Comparison of Chiral Recognition of Binaphthyl Derivatives with l-Undecyl-Leucine Surfactants in the Presence of Arginine and Sodium Counterions. *Journal of Chromatographic Science* **2019,** *57* (1), 54-62.

9. Engelhardt, H.; Cuñat-Walter, M. A., Use of plate numbers achieved in capillary electrophoretic protein separations for characterization of capillary coatings. *Journal of Chromatography A* **1995,** *717* (1), 15-23.

10. Billiot, E.; Macossay, J.; Thibodeaux, S.; Shamsi, S. A.; Warner, I. M., Chiral Separations Using Dipeptide Polymerized Surfactants: Effect of Amino Acid Order. *Analytical Chemistry* **1998,** *70* (7), 1375-1381.

11. Malviya, R.; Bansal, V.; Pal, O.; Sharma, P., High performance liquid chromatography: A short review. *Journal of Global Pharma Technology* **2010,** *2*, 22-26.
12. Escuder-Gilabert, L.; Martín-Biosca, Y.; Medina-Hernández, M. J.; Sagrado, S., Cyclodextrins in capillary electrophoresis: Recent developments and new trends. *Journal of Chromatography A* **2014,** *1357*, 2-23.
13. Mantovani, V.; Galeotti, F.; Maccari, F.; Volpi, N., Recent advances in capillary electrophoresis separation of monosaccharides, oligosaccharides, and polysaccharides. *ELECTROPHORESIS* **2018,** *39* (1), 179-189.
14. Shamsi, S. A.; Valle, B. C.; Billiot, F.; Warner, I. M., Polysodium N-Undecanoyl-l-leucylvalinate: A Versatile Chiral Selector for Micellar Electrokinetic Chromatography. *Analytical Chemistry* **2003,** *75* (3), 379-387.
15. Billiot, F. H.; Billiot, E. J.; Warner, I. M., Comparison of monomeric and polymeric amino acid based surfactants for chiral separations. *Journal of Chromatography A* **2001,** *922* (1), 329-338.
16. Bordes, R.; Holmberg, K., Amino acid-based surfactants – do they deserve more attention? *Advances in Colloid and Interface Science* **2015,** *222*, 79-91.
17. Billiot, A.; Fang, Y.; Morris, K., Characterization of Amino Acid Based Molecular Micelles with Molecular Modeling. *Open Journal of Physical Chemistry* **2019,** *09*, 221-240.
18. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Gladis, A. A.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., A Molecular Dynamics Simulation Study of the Association of 1,1'-Binaphthyl-2,2'-diyl hydrogenphosphate Enantiomers with a Chiral Molecular Micelle. *Chem Phys* **2014,** *439*, 36-43.
19. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Hoffman, C. B.; Gladis, A. A.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., Molecular dynamics simulation and NMR investigation of the association of the β-blockers atenolol and propranolol with a chiral molecular micelle. *Chemical Physics* **2015,** *457*, 133-146.
20. Lipkowitz, K. B., CHAPTER 5 - Analysis of chiral chromatographic separations by molecular modeling. In *Chiral Analysis*, Busch, K. W.; Busch, M. A., Eds. Elsevier: Amsterdam, 2006; pp 97-129.

21. Montanari, C. A.; Cass, Q. B.; Tiritan, M. E.; Souza, A. L. S. d., A QSERR study on enantioselective separation of enantiomeric sulphoxides. *Analytica Chimica Acta* **2000,** *419* (1), 93-100.
22. Valle, B. C.; Morris, K. F.; Fletcher, K. A.; Fernand, V.; Sword, D. M.; Eldridge, S.; Larive, C. K.; Warner, I. M., Understanding Chiral Molecular Micellar Separations Using Steady-State Fluorescence Anisotropy, Capillary Electrophoresis, and NMR. *Langmuir* **2007,** *23* (2), 425-435.
23. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005,** *26* (16), 1668-1688.
24. Billiot, F. H.; McCarroll, M.; Billiot, E. J.; Rugutt, J. K.; Morris, K.; Warner, I. M., Comparison of the Aggregation Behavior of 15 Polymeric and Monomeric Dipeptide Surfactants in Aqueous Solution. *Langmuir* **2002,** *18* (8), 2993-2997.
25. Roe, D. R.; Cheatham, T. E., 3rd, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013,** *9* (7), 3084-95.
26. Maiorov, V. N.; Crippen, G. M., Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol* **1994,** *235* (2), 625-34.
27. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., A Molecular Dynamics Simulation Study of Two Dipeptide Based Molecular Micelles: Effect of Amino Acid Order. *Open journal of physical chemistry* **2013,** *3* (1), 20-29.
28. Morris, K. F.; Billiot, E. J.; Billiot, F. H.; Gladis, A. A.; Lipkowitz, K. B.; Southerland, W. M.; Fang, Y., A molecular dynamics simulation study of the association of 1,1′-binaphthyl-2,2′-diyl hydrogenphosphate enantiomers with a chiral molecular micelle. *Chemical Physics* **2014,** *439*, 36-43.
29. Vilar, S.; Cozza, G.; Moro, S., Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem* **2008,** *8* (18), 1555-72.
30. Galli, C. L.; Sensi, C.; Fumagalli, A.; Parravicini, C.; Marinovich, M.; Eberini, I., A computational approach to evaluate the androgenic affinity of iprodione, procymidone, vinclozolin and their metabolites. *PLoS One* **2014,** *9* (8), e104822.

31. Le Guilloux, V.; Schmidtke, P.; Tuffery, P., Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **2009,** *10*, 168-168.

32. Cuzzolin, A.; Sturlese, M.; Malvacio, I.; Ciancetta, A.; Moro, S., DockBench: An Integrated Informatic Platform Bridging the Gap between the Robust Validation of Docking Protocols and Virtual Screening Simulations. *Molecules* **2015,** *20* (6).

33. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157-74.

34. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020,** *16* (1), 528-552.

35. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T., End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical Reviews* **2019,** *119* (16), 9478-9508.

36. Ylilauri, M.; Pentikäinen, O. T., MMGBSA As a Tool To Understand the Binding Affinities of Filamin–Peptide Interactions. *Journal of Chemical Information and Modeling* **2013,** *53* (10), 2626-2633.

37. Chanda, J.; Bandyopadhyay, S., Hydrogen bond lifetime dynamics at the interface of a surfactant monolayer. *J Phys Chem B* **2006,** *110* (46), 23443-9.

CHAPTER V: FUTURE WORK

In conclusion, working towards the development of the QSERR database, it is essential to investigate chiral separations further and to collect more data from both the computational and experimental studies.

Experimentally, since we have observed significant advantages in using diamine counterions in terms of improving enantiomeric resolution in MEKC, it is worth investigating further other pH dependent counterions and collecting more experimental data.

Computationally, there are many options to explore in terms of further optimizing and improving the computational efforts of this research. Further optimization of the computational micellar systems can be investigated in some of the following ways: (1) investigate different force fields that are applied to the system to see if this improves the molecular interactions that may match experimental data; (2) investigate different methods for calculating the binding free energy values between the complex (micelle:enantiomer); (3) try utilizing software that conducts quantum mechanics-based simulations versus the classical mechanics-based simulations that are currently used in our group. Although it is computationally more expensive/taxing, it might provide a more fine-tuned analysis of the computational systems being investigated.

LIST OF APPENDICES

APPENDIX PAGE

APPENDIX 1

Protocol for creating the amino acid-based molecular micelles with mixed force fields:

Step 1. Creating the surfactant tail with the appropriate charges and atoms.

Download the undecanal (UNA) crystallographic information file (.cif) from the internet using this link: https://www.iucr.org/cgi-bin/newiucrsearch?query=undecanal&submit=Search

Find and click on the **UNA.cif** file. You will now save this file to a new directory, in which you will now be working in for the rest of this module.

You are going to run the following antechamber command, which will process the source code in our **UNA.cif** file and convert it to an automatic configuration (.ac) file. In our case, the **UNA.ac** file. The ac file contains information about the appearance of its 3-D structure.

Within the directory you are working in, you will type the following in the command line:

- gedit antechamber_cif_ac

A blank file template should have opened. You will type the following into the screen:

antechamber -fi ccif -i **UNA.cif** -bk UNA -fo ac -o **UNA.ac** -c bcc -at amber

Save and close the file.

Next, you will create a prepgen backbone mainchain (.mc) file that identifies the atoms which need to be omitted and sets the net charge of the molecule.

Within the directory you are working in, you will type the following in the command line:

- gedit prepgen.mc

A blank file template should have opened. You will type the following into the screen:

TAIL_NAME C1 OMIT_NAME H1 OMIT_NAME H111 OMIT_NAME H101 POST_TAIL_TYPE C CHARGE 0.0

Save and close the file.

Next, you will now create a prepgen command to call in the prepgen.mc file, which will produce a .prepin file of UNA. Within the directory you are working in, you will type the following in the command line:

- gedit prepgen

A blank file template should have opened. You will type the following into the screen:

```
prepgen -i UNA.ac -o UNA.prepin -m prepgen.mc -rn UNA
```

Save and close the file.

You will now make the prepgen file executable by typing the following into the command prompt:

- chmod +x prepgen

You will now execute the file by typing the following into the command prompt:

- ./prepgen

You have now successfully created the UNA.prepin file that is needed to run the *parmchk2* command.

Side note: The surfactant tail will be made with the gaff2 force filed, whereas the surfactants head group that is composed of amino acids will be made with the protein.ff19SB force field. The two different force fields do not communicate well with each other within a single molecule. Previously, we have realized that when loading a molecule with two different force fields, the bond(s) that connect the two different force fields are often missing. Therefore, it is necessary to create a force field modification (.frcmod) file that will help the two force fields to better communicate where bonds must be made. The *parmchk2* program figures out what parameters will be needed to cross communicate bonds, bond angles, dihedral angles and other various bonding parameters.

Next, you will create a file called parmchk2 by typing the following into the command line:

- gedit parmchk2

A blank file template should have opened. You will type the following into the screen:

```
parmchk2 -i UNA.prepin -f prepi -o frcmod.UNA -a Y \
-p $AMBERHOME/dat/leap/parm/parm19.dat
```

Step 2. Generating the individual surfactant monomer units using an automation script.

Next, you will create a file that will be called in to run an automation script. Begin by typing the following into the command prompt:

- gedit generate_monomer_units.in

A blank file template should have opened.

For this step, type the text from *Supplemental Information*, **Section A.1-A.2** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Now that you have created the generate_monomer_units.in file, you will now create an automation script that will be used to run it.

Begin by typing the following into the command prompt of the directory you are working in:

- gedit RUN_generate_monomer_units.in

A blank file template should have opened. You will type the following into the screen:

xleap -f generate_monomer_units.in > generate_monomer_units.out

Save and close.

You will now make the RUN_generate_monomer_units.in an executable command by typing the following into the command prompt of the directory you are working in:

- chmod +x RUN_generate_monomer_units.in

You will now run execute the command by typing the following into the command prompt of the directory you are working in:

- ./RUN_generate_monomer_units.in

After running this automation script, you now have the library and mol2 files for each individual surfactant monomer unit.

You must now use the .mol2 files created above to obtain the .frcmod files for the individual surfactant units. This can be accomplished using the *parmchk2* program. To do so, type the following into the command prompt of the directory you are working in:

- gedit parmchk2_create_monomer_frcmod

A blank file template should have opened. Type the following into the screen:

For this step, type the text from *Supplemental Information*, **Section B** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

To make this file executable, type the following into the command prompt of the directory you are working in:

- chmod +x parmchk2_create_monomer_frcmod

Now, execute this command by typing the following into the command prompt of the directory you are working in:

- ./parmchk2_create_monomer_frcmod

You have now created the .frcmod files for all surfactant units.

Now that you created the .lib and .frcmod files for the individual monomer units, we must now use these files to create the polymerized amino acid-based molecular micelles.

Begin by typing the following into the command prompt of the directory you are working in:

- gedit load_monomer_files.in

A blank file template should have opened. Type the following into the screen:

For this step, you will type the text from *Supplemental Information*, **Section C.1-C.2** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Now that you have created the load_monomer_files.in file, you will now create an automation script that will be used to run it.

Begin by typing the following into the command prompt of the directory you are working in:

- gedit RUN_ load_monomer_files.in

A blank file template should have opened. Type the following into the screen:

```
xleap -f load_monomer_files.in > load_monomer_files.out
```

Save and close.

To make this file executable, type the following into the command prompt of the directory you are working in:

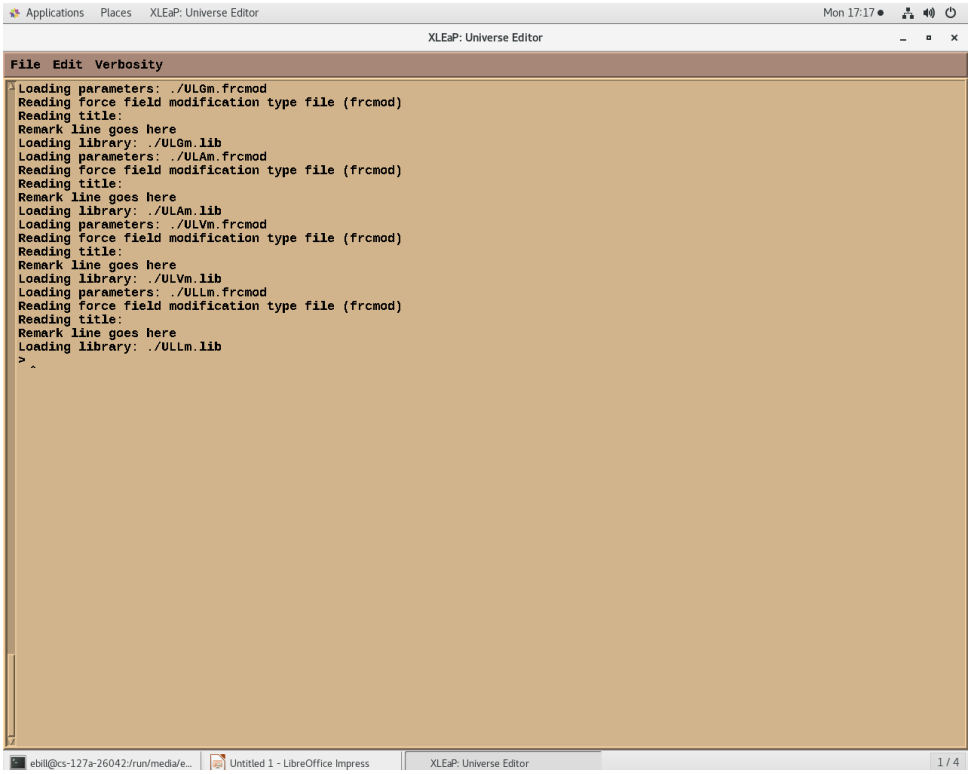
- chmod +x RUN_ load_monomer_files.in

Now, execute this command by typing the following into the command prompt of the directory you are working in:

- ./RUN_ load_monomer_files.in

You have now loaded all files necessary to build the polymerized amino acid-based molecular micelles.

Your screen should now be showing the xleap terminal, as shown below:



To avoid repetition in steps, this protocol will create the polymerized amino acid-based molecular micelle, ULVp. Essentially, the remaining micelles can be produced following the same steps.

First, we will look at previously reported data to determine how many surfactant units we will use to build each amino acid-based polymerized micelle. The results from the studies are summarized below:

Table 1. Approximate number of repeat units for each amino acid-based molecular micelle.

Surfactant	# repeat units
UGA	32
UGV	24
UGL	24
UAG	30
UAA	26
UAV	24
UAL	24
UVG	22
UVA	20
UVV	22
UVL	20
ULG	20
ULA	20
ULV	20
ULL	20

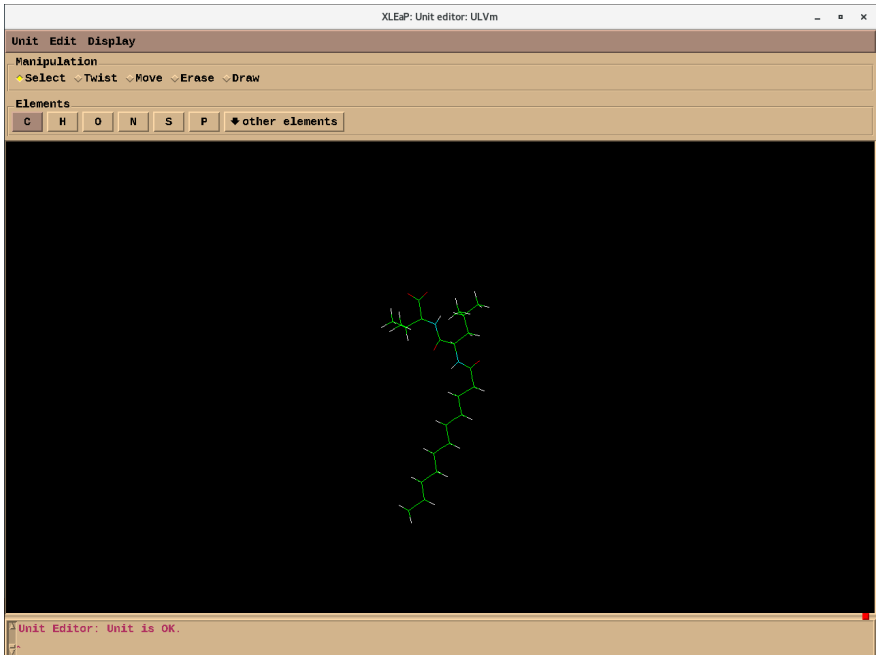
Key: Undecyl = U, Glycine = G, Alanine = A, Leucine = L, Valine = V.

Since we are building the polymerized micelle of ULV, we will be sure to use 20 repeat units, as suggested in the data presented above.

We will now continue working through the xleap terminal. To properly follow this protocol, anything after ">"will need to be typed into the xleap terminal!

>edit ULVm

The graphical interface will pop up, as shown below:



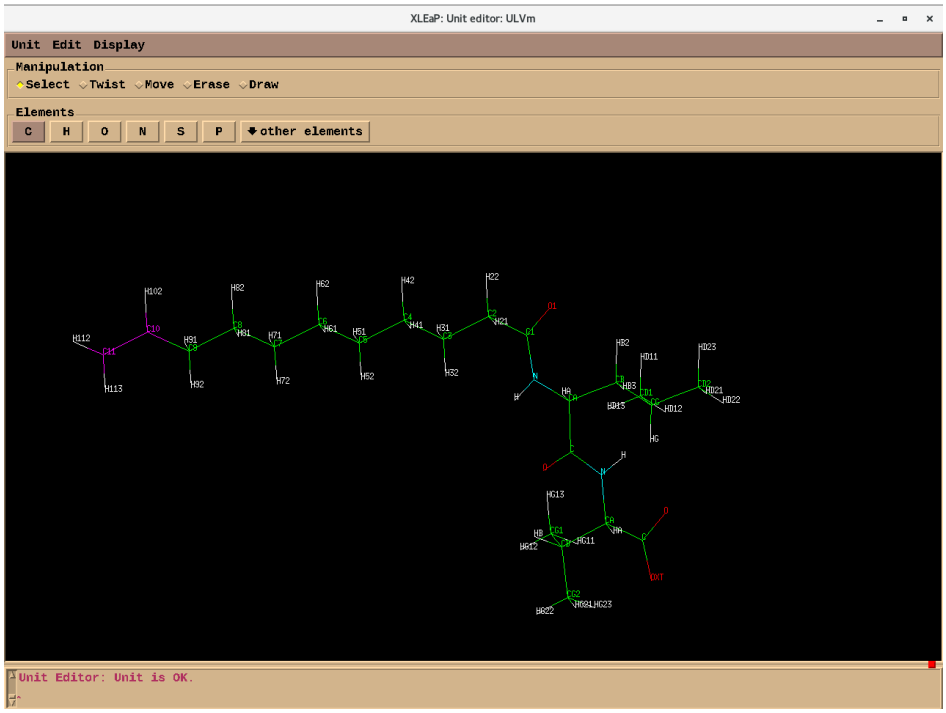
At the top left, left-click and hold “Unit” then slide the mouse down and select “Check unit.” The bottom of the screen should indicate “Unit Editor: Unit is OK.” This gives us confirmation to move forward in the process.

Again, at the top left, you will left-click and hold “Unit” then slide the mouse down and select “Calculate net charge.” In this case, it should be approximately -1, and this also gives us confirmation to move forward in the process.

Next, at the top left, left-click “Display” then slide the mouse down and select “Names.”

At the top left, left-click the “Select” option.

Locate atoms C11 and C10 at the terminal end of the surfactant’s tail. You may maneuver around the surfactant and change its viewing angles using the left-click button on the mouse to rotate the molecule. Once you have located atoms C11 and C10, left-click on the atoms. These two carbon atoms should be highlighted in purple, as shown below:



As we are going to make the polymer, it makes it easier to see the parent chain of the polymer when it is selected/highlighted as purple.

At the top left, left-click and hold “Unit” then slide the mouse down and select “Close.”

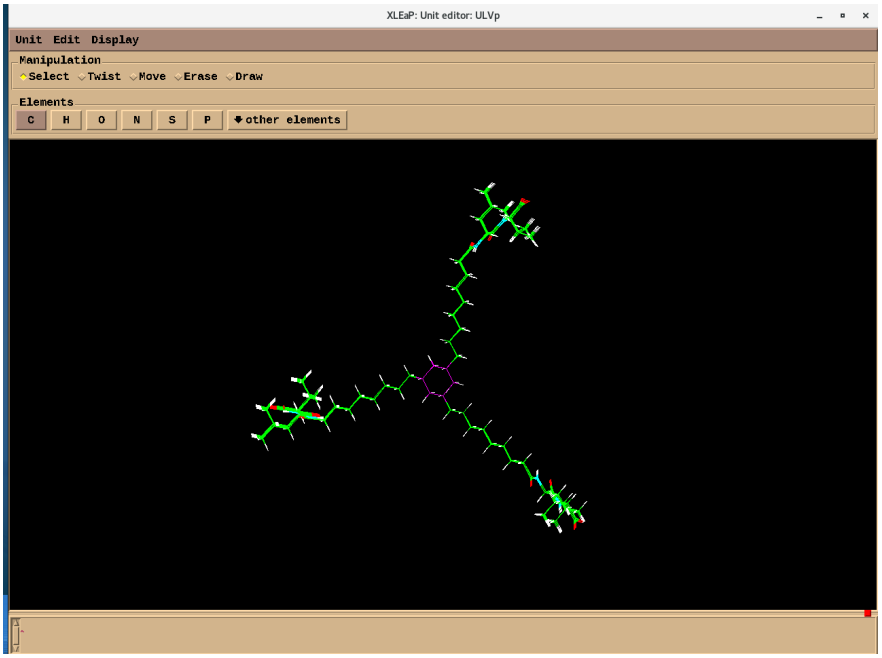
Pro-tip: Always avoid closing the xleap terminal with the top right X, as this will completely shut down the module and all work will be lost.

>ULVp = sequence {ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm ULVm}

**Remember, we are using twenty ULV units as suggested from the summarized data in Table 1 above.

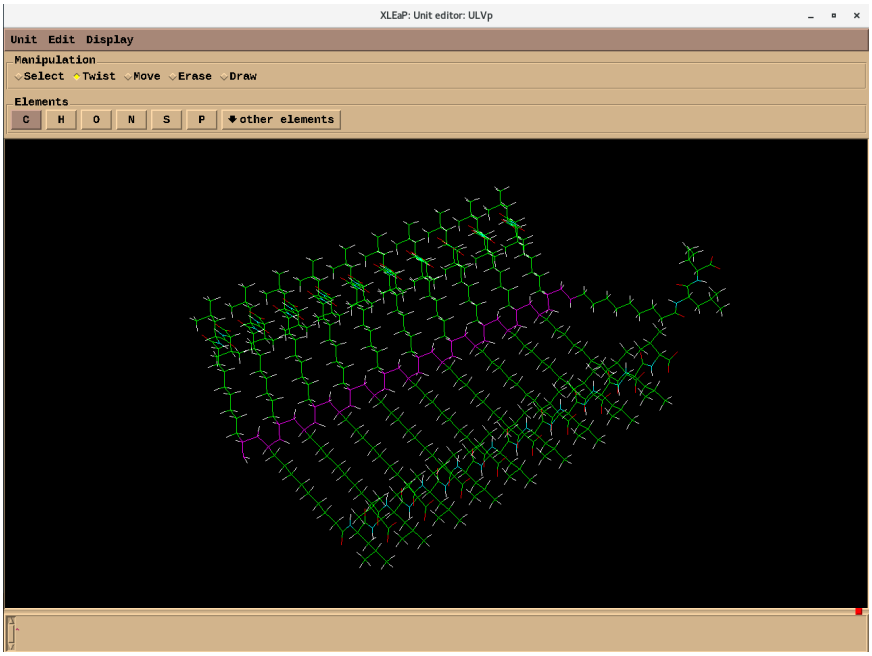
>edit ULVp

You will see a stacked version of the polymer as shown below:



Since the parent chain is already selected and highlighted, we can now left-click “Twist” at the top left (located directly to the right of “Select.”)

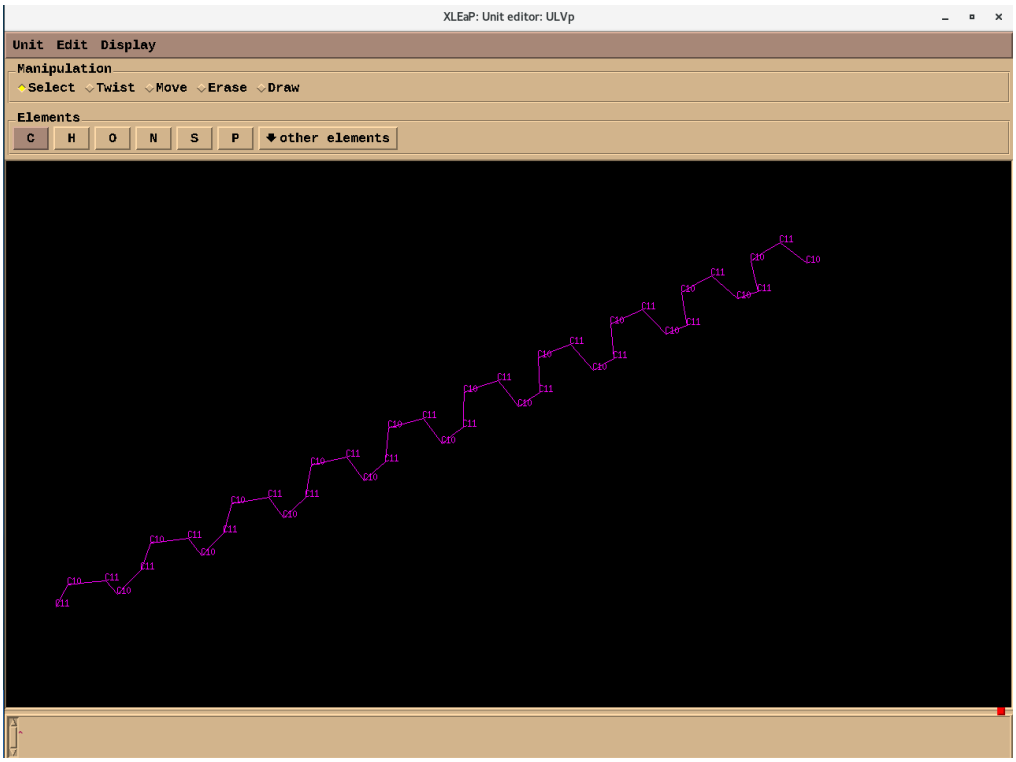
You will now left-click in the black area surrounding the micelle and slowly slide the mouse to the left or right until you have the best linear image of the micelle that you can see, as shown below:



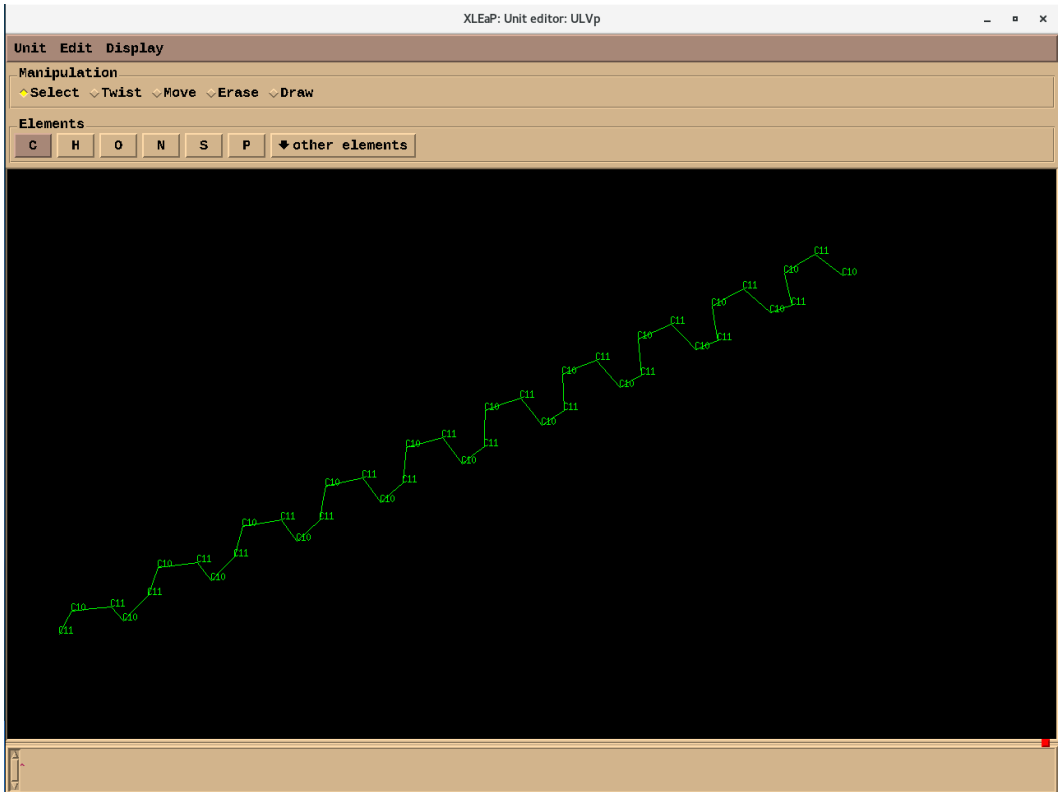
Next, at the top left, left-click “Edit” then slide the mouse down and select “Show selection only.” You will now see the parent chain of the polymer highlighted in purple.

At the top left, left-click the “Select” option.

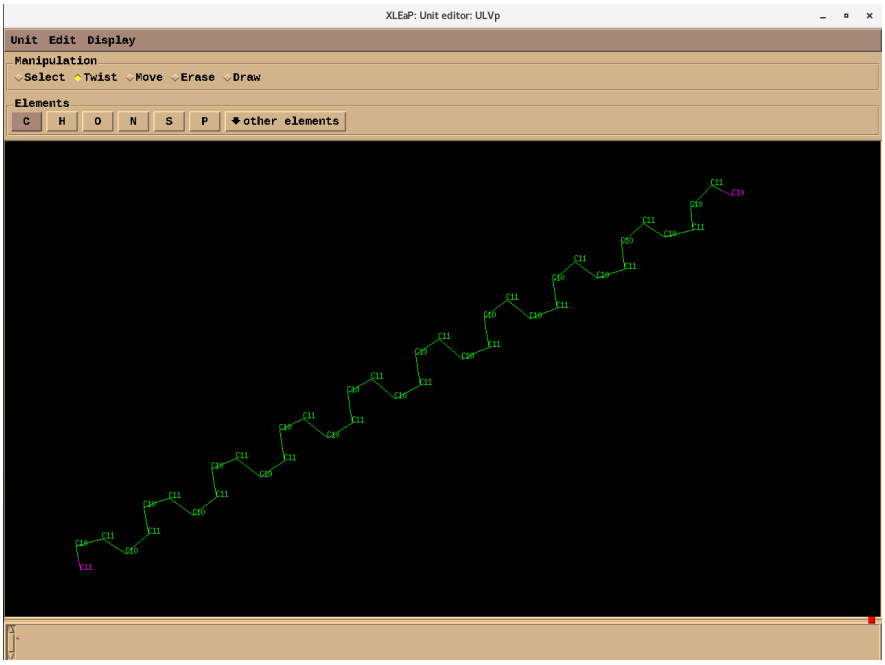
Next, at the top left, left-click “Display” then slide the mouse down and select “Names.” You should now see the following on your screen:



Press and hold the “Shift” key on the keyboard while double left-clicking in the black area of the xleap screen. The polymer chain should now be green, as shown below:



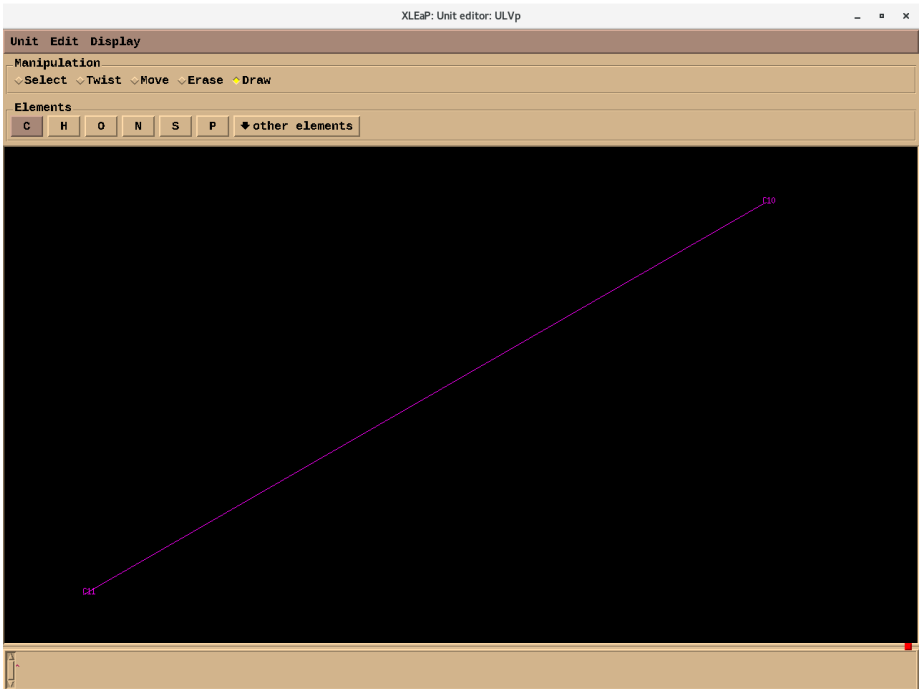
Select the two opposite ends of the parent chain in the polymer, which should be C11 on one end, and C10 on the other. You should see the following:



At the top left, left-click “Edit” then slide the mouse down and select “Show selection only.” You should now only see two C10 and C11 atoms at the two ends of the polymer linear chain.

Left-click the “Draw” option at the top left of the xleap screen.

Now, left-click on the C11 atom and drag it to C10 atom to create a bond between them. You should see the following:

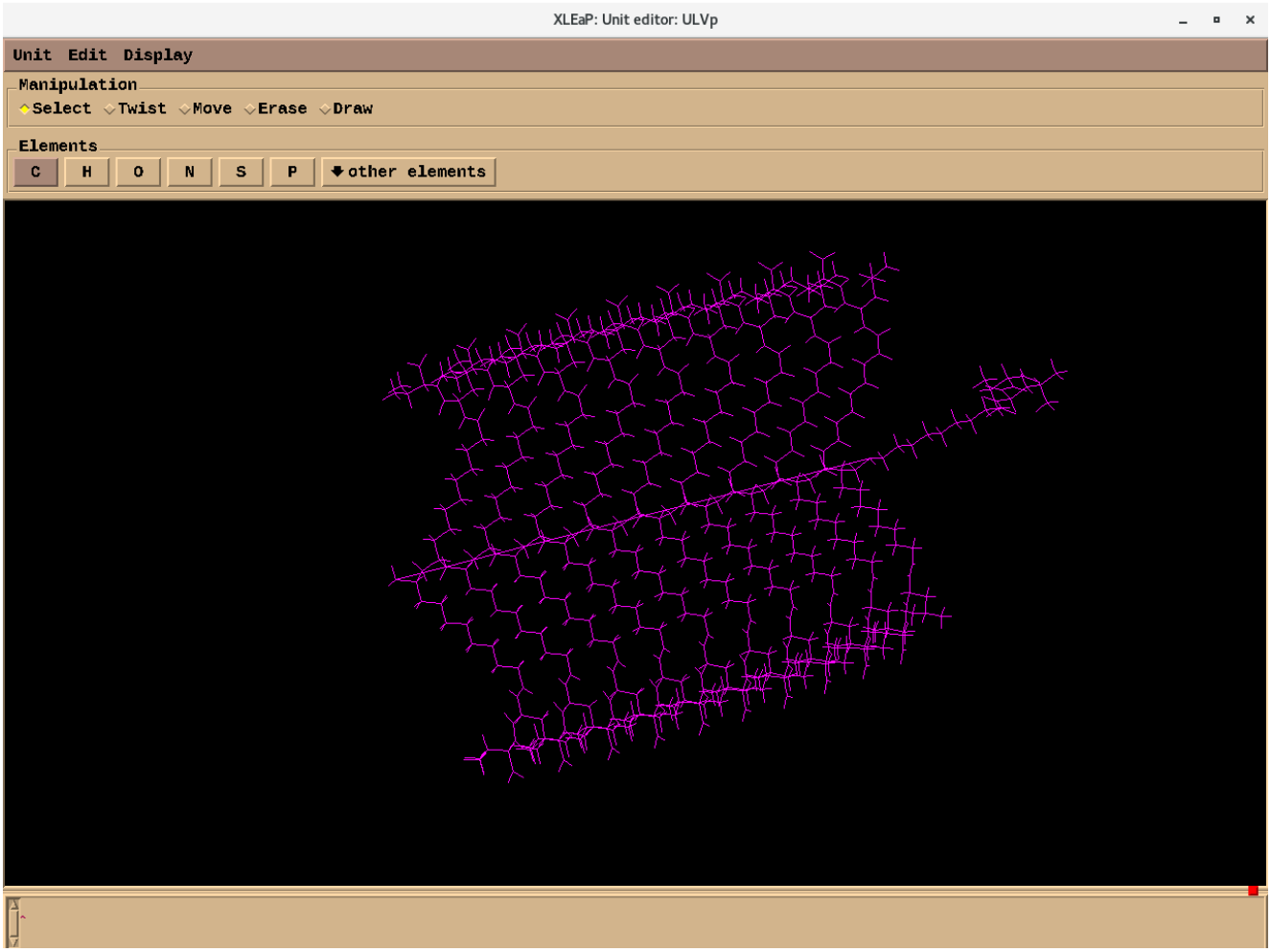


A linear ring polymer has now been made into a “ring” polymer.

At the top left, left-click “Edit” then slide the mouse down and select “Show everything.”

Again, at the top left, left-click the “Select” option. To remove the atom names from the screen, left-click “Display” then slide the mouse down and select “Names.”

By clicking and holding, both the middle button and right-click button of the mouse, then sliding up or down to zoom in and/or out. Zoom out enough to see the entire micelle, then select the entire micelle. You should see the following:



At the top left, left-click “Edit” then slide the mouse down and select “Relax selection.” Do this multiple times until it can no longer be relaxed.

Note: Sometimes the micelle will disappear, and the only way to fix it is to just create the polymer again. :]

Once the micelle cannot be relaxed any further, click on “Unit” at the top left, then slide down and select “Close.”

We will now save the files for this polymer in both, its unsolvated and non-ionized form, as well as its solvated and ionized form.

We need the solvated and ionized parameter topology and input coordinate files of the micelle(s) to run the molecular dynamics simulations.

In the xleap terminal, we will now save files for the non-ionized and unsolvated micelle.

>saveamberparm ULVp ULVp.prmtop ULVp.inpcrd
>saveoff ULVp ULVp.lib
>savepdb ULVp ULVp.pdb

>savemol2 ULVp ULVp.mol2 0

Now, we will ionize and solvate micelle:

>SolvateOct ULVp TIP3PBOX 10.0
>addIons ULVp Na+ 0

Now, we will save files for the ionized and solvated micelle:

>saveamberparm ULVp ULVp_water.prmtop ULVp_water.inpcrd
>saveoff ULVp ULVp_water.lib
>savepdb ULVp ULVp_water.pdb
>savemol2 ULVp ULVp_water.mol2 0

Now, we just need to repeat this process for each polymer. :]

In the case for the ULVp micelle, both the ULVp_water.prmtop and ULVp_water.inpcrd files will be used to initiate the molecular dynamics simulations.

Please refer to the “Conducting A Molecular Dynamics Simulation” Protocol for information on how to start the MD simulation.

APPENDIX 2
Running the 480 ns simulation on each amino acid-based molecular micelle:

You should have completed the previous protocol titled, "Protocol for creating the amino acid-based molecular micelles."

In the previous protocol, you should have created the: ***p_water.prmtop and ***p_water.inpcrd files for each surfactant.

The three asterisks above are representative of the various three-letter surfactants (e.g. ULG, ULA, ULL, ULV, etc.)

These two files for each surfactant are necessary to initiate molecular dynamics simulations. For example, ULVp_water.prmtop and ULVp_water.inpcrd are necessary to initiate the molecular dynamics simulations on the ULV amino acid-based molecular micelle.

Note: Molecular dynamics (MD) simulations submitted to the HPC are not able to exceed 24 hours, in which the 480 ns simulations inevitable will. Therefore, we perform the 480 ns simulations on the AMBER18 and AMBER20 computers since there are no time limits associated with them.

Therefore, this protocol will examine how to submit the 480 ns simulations of each amino acid-based molecule micelle to either the AMBER18 or AMBER20 computers.

The AMBER20 computer has two GPUs. Therefore, we are able to run two simulations in parallel. Therefore, for AMBER20, we can separate the MD simulations needed to be submitted into two queues.

First, open a terminal.

Within the terminal, type the following into the command prompt:

- gedit Minimize.in

Within this file, you will type in the following:

For this step, type the text from *Supplemental Information*, **Section A** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Within the terminal, type the following into the command prompt:

- gedit MD-warmup.in

Within this file, you will type in the following:

For this step, type the text from *Supplemental Information*, **Section B** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Within the terminal, type the following into the command prompt:

- gedit MD-equilibrate.in

Within this file that just opened, you will type in the following:

For this step, type the text from *Supplemental Information*, **Section C** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Within the terminal, type the following into the command prompt:

- gedit MD-final.in

Within this file that just opened, you will type in the following:

For this step, type the text from *Supplemental Information*, **Section D** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Now that you have created the Minimize.in, MD-warmup.in, MD-equilibrate.in and MD-final.in files, you will now create a script that will be used to call in these files in order to run the MD simulations.

You will now create three separate folders, each titled:

- AMBER18
- AMBER20_BATCH1
- AMBER20_BATCH2

Copy and paste the Minimize.in, MD-warmup.in, MD-equilibrate.in and MD-final.in files into each of these folders.

You will open the AMBER20_BATCH1 folder in a terminal and type the following:

- gedit command_min_warm_equi_final_BATCH1.in

For this step, type the text from *Supplemental Information*, **Section E.1-E.4** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Now, make the command_min_warm_equi_final_BATCH1.in files executable using the following command prompt:

- chmod +x command_min_warm_equi_final_BATCH1.in

Now, all of the ***p_water.prmtop and ***p_water.inpcrd files that are in the command_min_warm_equi_final_BATCH1.in will need to be transferred to the same AMBER20_BATCH1 folder.

After all of the files are there, you will now execute the command using the following command prompt:

- ./command_min_warm_equi_final_BATCH1.in

_____

Again, you can add any files you want to run to this automation script as long as its respective ***p_water.prmtop and ***p_water.inpcrd files are in the folder AND has been placed in the command_min_warm_equi_final_BATCH1.in file.

Now that you have simulations running on one of the GPUs, you can submit to another GPU within the AMBER20 computer. You will open the AMBER20_BATCH2 folder in a terminal and type the following:

- gedit command_min_warm_equi_final_BATCH2.in

For this step, type the text from *Supplemental Information*, **Section F.1-F.3** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Now, make the command_min_warm_equi_final_BATCH2.in files executable using the following command prompt:

- chmod +x command_min_warm_equi_final_BATCH2.in

Now, all of the ***p_water.prmtop and ***p_water.inpcrd files that are in the command_min_warm_equi_final_BATCH2.in will need to be transferred to the same AMBER20_BATCH2 folder.

After all of the files are there, you will now execute the command using the following command prompt:

- ./command_min_warm_equi_final_BATCH2.in

Again, you can add any files you want to run to this automation script as long as its respective ***p_water.prmtop and ***p_water.inpcrd files are in the folder AND has been placed in the command_min_warm_equi_final_BATCH2.in file.

Now that you have simulations running on both GPUs on the AMBER20 computer, you can also submit to the single GPU on the AMBER18.

You will open the AMBER18 folder in a terminal and type the following:

- gedit command_min_warm_equi_final_AMBER18.in

For this step, type the text from *Supplemental Information*, **Section G.1-G.2** into this screen that just opened. The *Supplemental Information* section can be found at the end of this protocol.

Save and close.

Now, make the command_min_warm_equi_final_AMBER18.in files executable using the following command prompt:

- chmod +x command_min_warm_equi_final_AMBER18.in

Now, all of the ***p_water.prmtop and ***p_water.inpcrd files that are in the command_min_warm_equi_final_AMBER18.in will need to be transferred to the same AMBER18 folder.

After all of the files are there, you will now execute the command using the following command prompt:

- ./command_min_warm_equi_final_AMBER18.in

APPENDIX 3

RMSD Protocol to obtain equilibrated structures of AABMM

*Re-image the file to make the calculations faster!! (cut down by 10 fold)

Step 1: First RMSD Run

gedit rmsd_calculation.in

trajin mic_water-new-final_480ns_04_REIMAGED.traj

rmsd :1-180 out ULVmicelle_water_rmsd.dat

Save and close

cpptraj mic_water.prmtop rmsd_calculation.in

Open output file (.dat) in Excel Spreadsheet and plot in on a scatter plot to determine most equilibrated range (e.g. 800,000 to 120,000,000)

Step2: Average → Reference Structure

gedit rmsd_calculation_average.in

trajin mic_water-new-final_480ns_04_REIMAGED.traj 20000 30000 1

average ULVmicelle_avg_20000_30000.traj

Save and close

cpptraj mic_water.prmtop rmsd_calculation_average.in

Step 3: Second RMSD Run Using Step 2's Output as Reference Structure

gedit rmsd_calculation_average-last_frames.in

trajin mic_water-new-final_480ns_04_REIMAGED.traj 20000 30000 1

reference ULVmicelle_avg_20000_30000.traj [ref1]

rmsd :1-180 ref [ref1] out ULVmicelle_20000_30000_rmsd_average_final.dat

Save and close

cpptraj mic_water.prmtop rmsd_calculation_avg_last_frames.in

Open the output .dat file in Excel and sort by the RMSD column from smallest → largest and use the lowest number (e.g. 4509) as the frame for step 4. (20,000 + 4,509 = 24,509)

Step 4: Final Structure

gedit pdb_REP1_last_step.in

trajin mic_water-new-final_480ns_04_REIMAGED.traj 24509 24509 1

trajout ULVmicelle_frame_24509.pdb 24509 24509 1 pdb

#go
cpptraj mic_water.prmtop pdb_REP1_last_step.in

Step 5: Fixing the pdb file

Gedit ULVmicelle_frame_24509.pdb

Remove the water and counterions from the pdb file

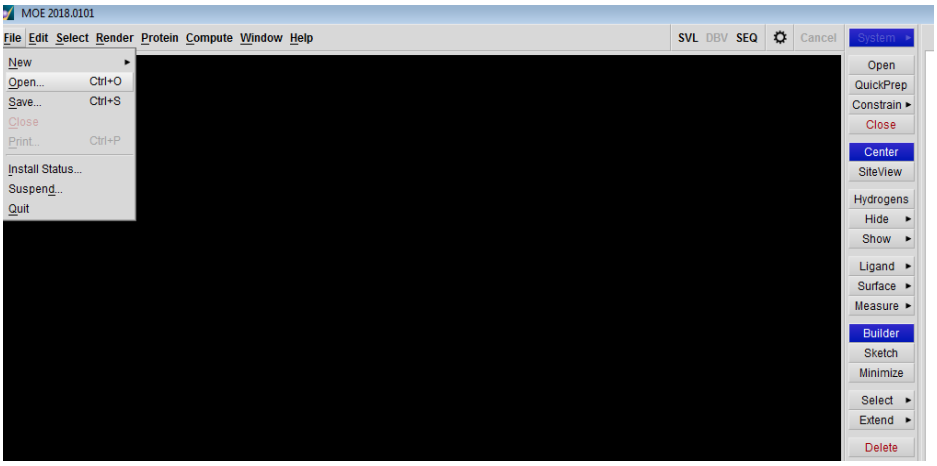
Erase the "END" and leave the pdb file with "TER" at the end and rename it for organization purposes to: **ULVmicelle_frame_204509_TER.pdb**

This PDB is now ready for docking in the Molecular Operating Environment (MOE) software! :)

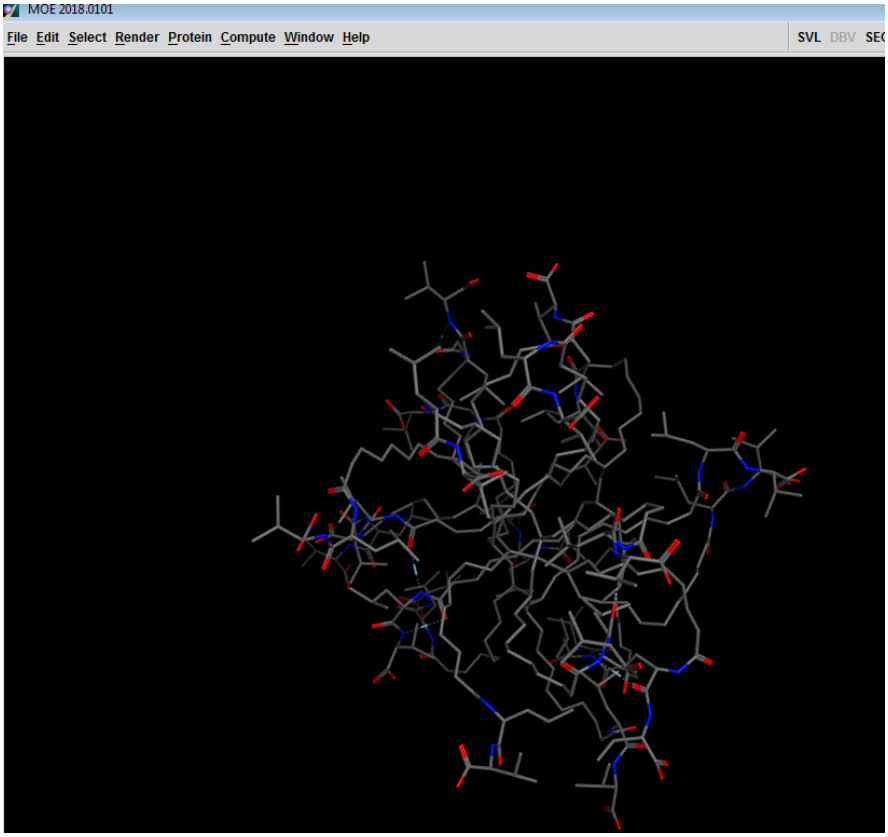
APPENDIX 4

Ligand Docking Protocol:

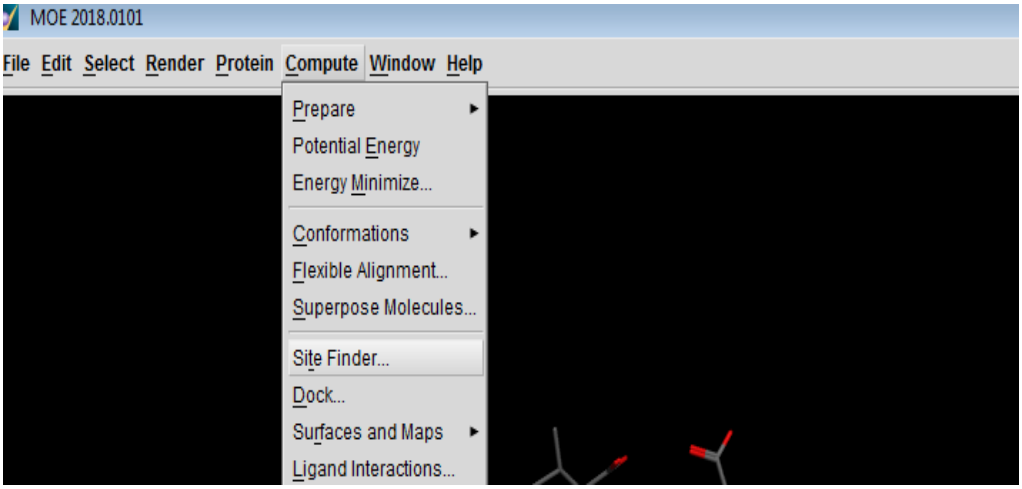
1. Load your equilibrated micelle file by clicking on File then Open and selecting the .pdb file of the structure that was pulled from the AMBER18 computer.



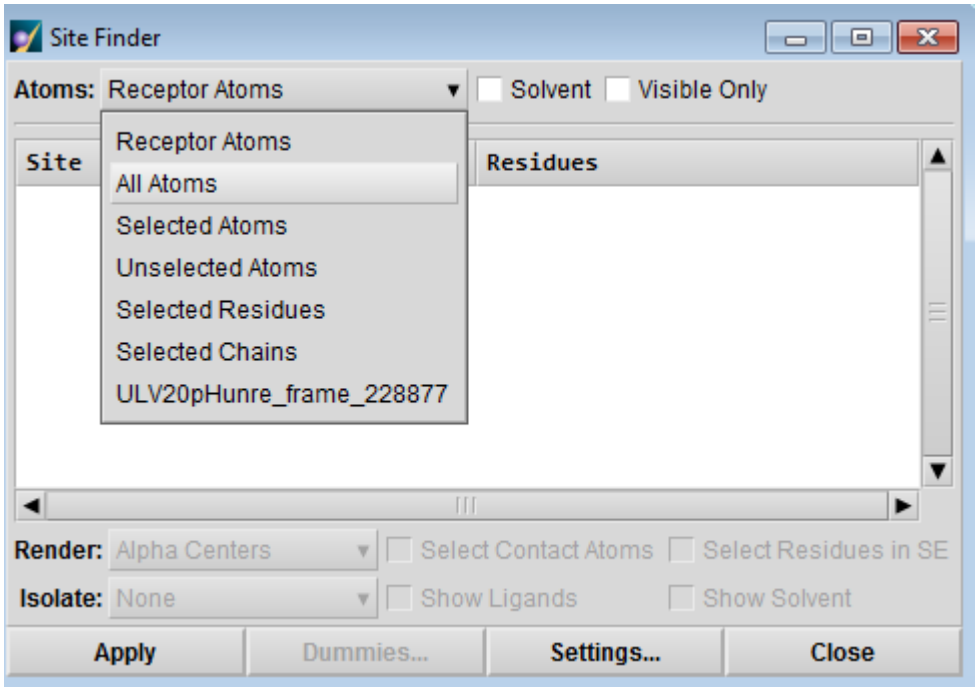
2. Your micelle file is then loaded within MOE.



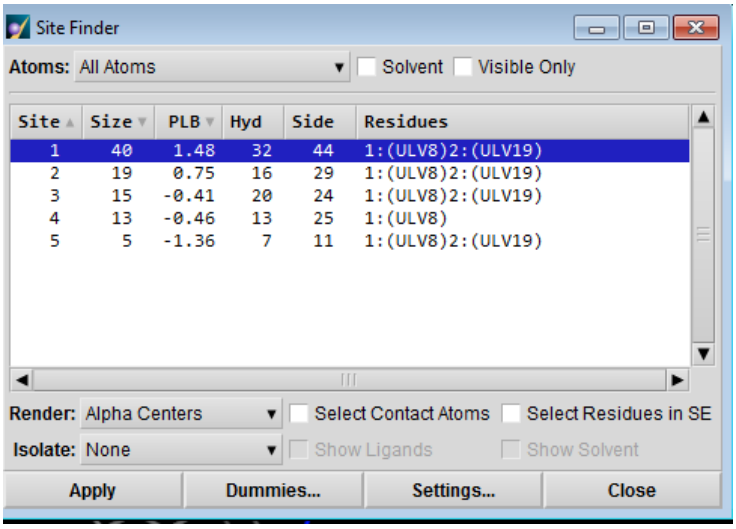
3. Now click on the “Compute” button on the top of the screen and select “Site Finder”



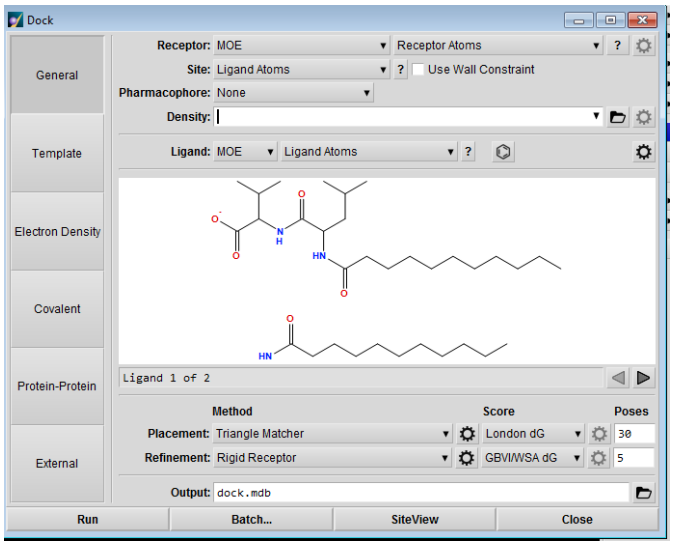
4. After clicking on “Site Finder” the following screen should appear. Select “All Atoms” under “Receptor Atoms” the click *Apply*.



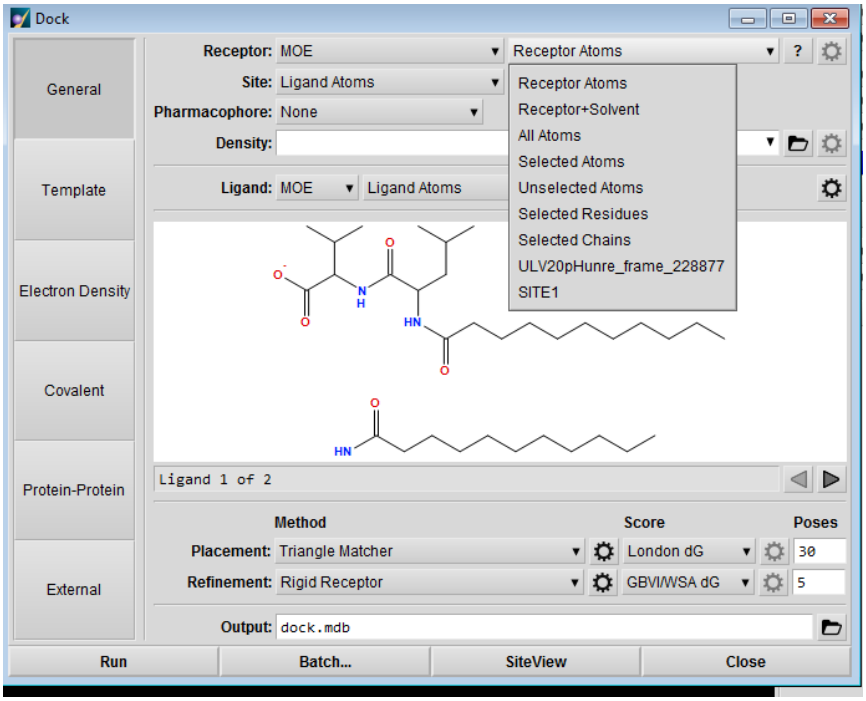
5. Now another screen should appear such as the one shown below. Now select the site "pocket" and click on the "Dummies" button. A pop-up screen will appear, click "Create dummy atoms at alpha sphere centers" and click yes.



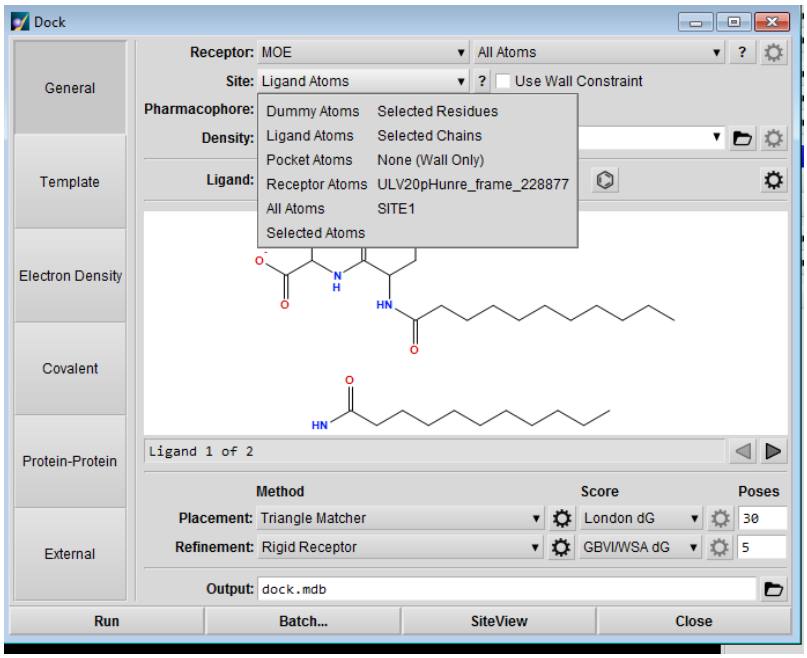
6. Now click on the Compute button on the top of the screen and click on the "Dock" button and a screen similar to the one below will appear.



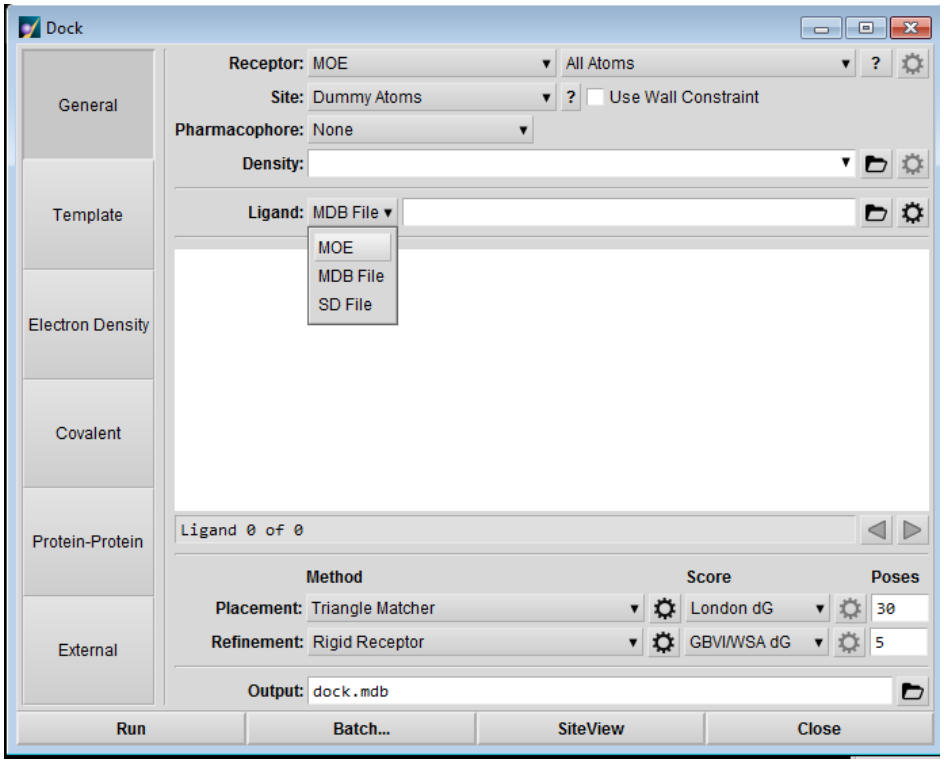
7. Under the "Receptor Atoms" button click on All Atoms.



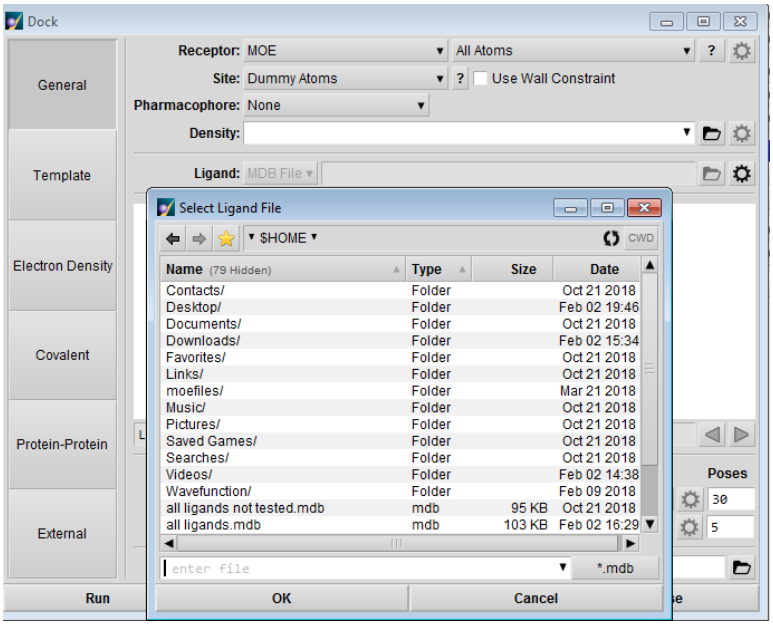
8. Now click on the “Site” button and select “Dummy Atoms.”



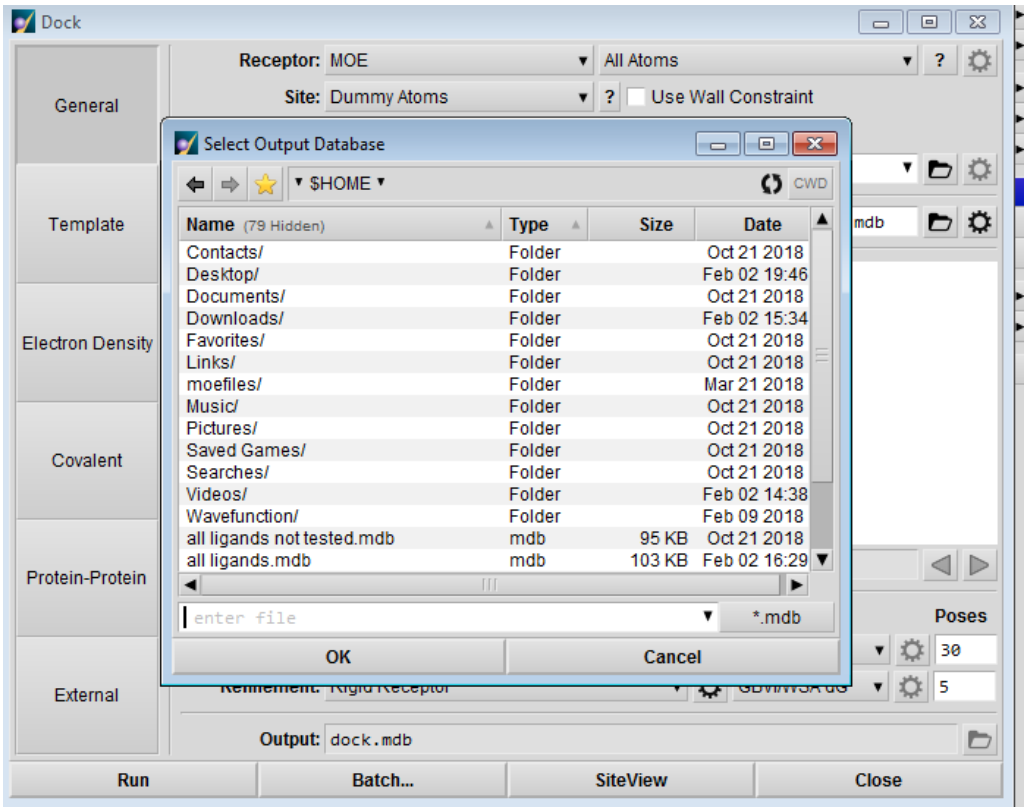
9. Next, click on the “Ligand” button and select “MDB file”.



10. Now click on the *icon that looks like a folder* to select your MDB database file.



11. Now choose the folder in which you wish to save your output file and name that output file to something that makes sense to you.



12. Now you are ready to click “Run.”

APPENDIX 5

Creating Folders to Submit to HPC for Simulations/Energy Calculations

Step 1. To start, you will need to create a directory called "Startup_Files"

Step 2. Make sure to put the following files within the "Startup_Files" folder you just made:

- energycalc0.in
- energycalc0.sbatch
- MD-equilibrate.in
- MD-final.in
- MD-warmup.in
- Minimize.in
- mmpbsa.in
- mmgbsa.in
- reimage_NEW0.in
- step1thru8_with_reimage0.in
- step1thru8_with_reimage0.sbatch
- tleap0.in
- UNA.prepin
- UNA.frcmod

Step 3. If you are running simulations on all 15 dipeptide combinations, you will then create the following folders:

- ULG_Startup_Files
- ULA_Startup_Files
- ULL_Startup_Files
- ULV_Startup_Files
- … and so forth for all 15 dipeptide combinations.

Step 4. Proceed by copying and pasting all of the bulleted files from Step 2 into each of the folders you just made in Step 3.

Step 5. For each surfactant folder you will additionally add: the equilibrated polymerized micelle .pdb structure, the surfactant's monomer .lib file and the surfactant's monomer .frcmod files.

For example:
If I am performing step 5 for ULV_Startup_Files, I will locate and add the following files in it:

- ULVp_frame_6919_AMBER_TER.pdb
- ULVm.lib
- ULVm.frcmod

You will then perform the same for the rest of the folders; be sure to add the correct files.

Step 6. Now, will create the individual analyte folders for each of the micelle folders. First, we will begin by right-clicking on the folder called **ULV_Startup_Files**, and then open it in the terminal. Type the following into the command line:

- Pwd

**This will identify and provide you with your directory pathway. In this example, mine is: /data/WinterBreak_2020/ULV/ULV_StartupFiles/ Therefore, whenever you see me using this pathway, you will, instead, use yours.

Step 7. Now, type the following into the command line:

- gedit make_analyte_folders.in

Step 8. Now, we will create a sequence that will create each analyte of our choosing for the ULV folder specifically. You will do this by typing the following into the file you just gedited:

```
mkdir OOA

cp energycalc0.sbatch energycalc0.in MD-equilibrate.in MD-final.in MD-warmup.in Minimize.in
mmgbsa.in mmpbsa.in reimage_NEW0.in step1thru8_with_reiamge0.sbatch
step1thru8_with_reimage0.in tleap0.in ULVm.lib ULVm.frcmod
ULVp_frame_6919_AMBER_TER.pdb /data/WinterBreak_2020/ULV/ULV_StartupFiles/OOA

cd OOA
sed 's/xxx/OOA/g' tleap0.in > tleap.in
sed 's/xxx/OOA/g' reimage_NEW0.in > reimage_NEW.in
sed 's/xxx/OOA/g' step1thru8_with_reimage0.in > step1thru8_with_reimage.in
sed 's/xxx/OOA/g' energycalc0.in > energycalc.in
sed 's/xxx/OOA/g' energycalc0.sbatch > energycalc.sbatch
cd ..
```

**side note: the yellow highlighted portions above will obviously be changed to match whichever folder you are working in; since I am making ULV folders, I will use ULV micelle files.

**You will essentially just find and replace the three letter code (eg. OOA to OOB) and then copy and paste it. You will repeat this process for all analytes you want within the folder so the sequence will become long-ish but is more efficient in the long-term scope of things.

Save and exit.

Step 9. Executing the script you just made. First we must make it executable so go to the command line of the terminal and type the following:

- chmod +x make_analyte_folders.in
- ls
- ./ make_analyte_folders.in

Step 10. We must now make all files within the folder compatible with the files we have placed in them so that they run correctly. We must look at the aggregation number used to build the micelles. For example, ULV was created using 20 monomer units. The 20 units are each composed of 3 subunits (undecyl, Leucine and Valine), equaling 60 units. Once we add the analyte, we are adding one additional unit, so we must denote in the files that we want to run and analyze all 61 units. Therefore, it is necessary to go into the following files and set the number of units to 61:

- reimage_NEW0.in
- energycalc0.in
- energycalc0.sbatch
- step1thru8_with_reimage0.in

Step 11. We must also make the file names compatible for each micelle, therefore the "receptor" is the micelle. In the case for ULV, we must edit the following files:

- tleap0.in

Within this file, there will be a section denoted with "#Creating receptor files"

We must change "loadoff ULLm.lib" and "loadamberparams ULLm.frcmod" to "loadoff ULVm.lib" and "loadamberparams ULVm.frcmod"

In regards to the receptor, be sure to change the micelle pdb file accordingly. For example, change "receptor = loadpdb ULLmicelle_frame_TER.pdb" to the ULV micelle, in this case: "receptor = loadpdb ULVmicelle_frame_TER.pdb"

Step 12. Now that all of these folders are made, you can copy all of those folders into a directory called "ULV_Pocket1", do the same thing again and create a "ULV_Pocket2" folder. Do this consecutively for the amount of pockets each micelle has based on SiteFinder at MOE (this would have been done already).

Step 13. Now, you will dock pocket 1 of ULV in MOE, obtain its docked structure (eg. OOA.pdb) file add it to that folder. Do this for each docked analyte and move it into its respective folder(s).

Step 14. You are now ready to move these folders into the HPC via Filezilla and submit for simulations and energy calculations.

APPENDIX 6

Visual Molecular Dynamics (VMD) Protocol:

Step 1: Once a run has completed, you will transfer the following files from the HPCM system to the the desktop (example):

Step 2: Open the terminal and type in:

cd $AMBERHOME

Step 3: You will then type:

./vmd

Step 4: Four window tabs should open. Click on the window titled “Molecule File Browser.”

Click the “Browse...” button and navigate to the “OOA_complex_water.prmtop” file

Make sure the “Determine file type” is set to AMBER7 Parm, then click “Load.”

Step 5: Click on the “OOA_complex_water.prmtop” line that pops up in the “VMD Main” window tab, then click back to the “Molecule File Browser” window tab.

Step 6: Click “Browse...” and navigate to the “OOA_complex_water-new-final_60ns_10.traj” file.

Make sure the “Determine file type” is set to “AMBER Coordinates with Periodic Box”

Set the frame range to the range you would like to view, usually the entire trajectory for this case:

First: 1
Last: 3000
Stride: 1

Click “Load”

Step 7: In the “Graphics Representations” tab click within the “Selected Atoms” line and erase the word “all.”

Step 8: Click “not”

APPENDIX 7

Protocol for Analyzing the Data:

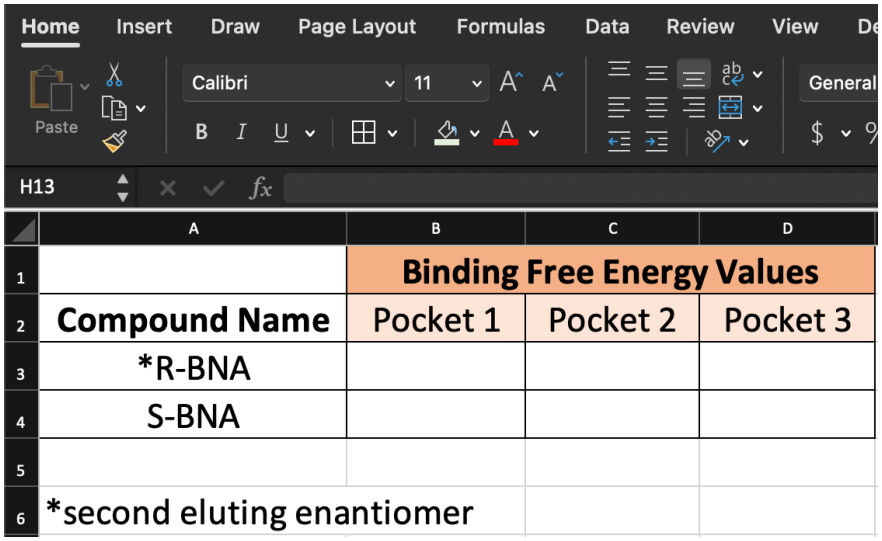
After obtaining the binding free energy values from the previous protocol, you will open a Microsoft Excel Spreadsheet file.

This protocol aims to provide the user with a guide for analyzing the binding free energy values and comparing them to experimental MEKC elution order data for the respective AABMM being used.

Step 1: For each set of enantiomers, determine which experimentally elutes from the MEKC column using the specific micelle you are using.

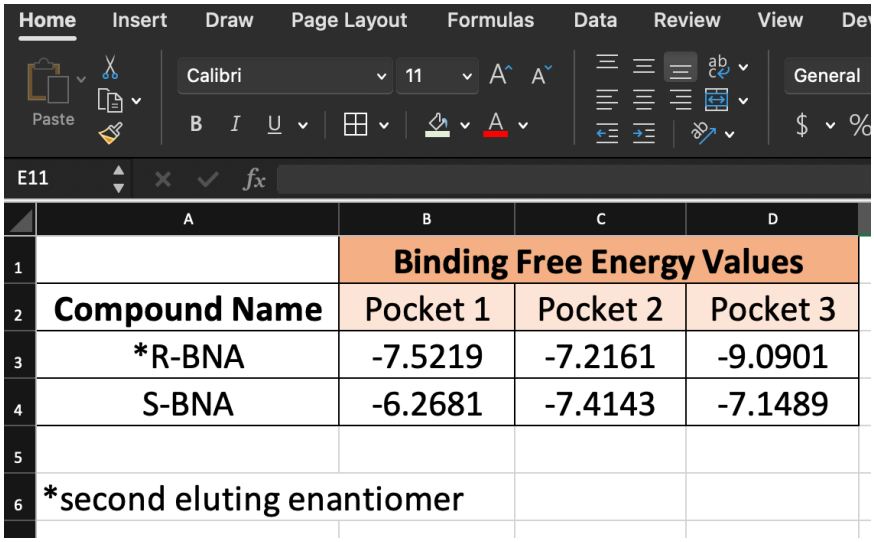
For example, when using polySULV as the pseudostationary phase in MEKC, and when separating BNA enantiomers, R-BNA elutes from the column second, whereas S-BNA elutes first. This tells us that experimentally R-BNA has a stronger binding affinity to the polySULV therefore we want to determine if the computational binding free energy values match with this assessment.

Step 2: For whichever micelle you are investigating, determine the number of binding pockets it has. This was determined in Appendix 4. For example, if you load the polySULV micelle into MOE and Site Finder identifies 3 binding pockets, you will be sure to include three pockets in your Excel Spreadsheet as follows.



	A	B	C	D
1		Binding Free Energy Values		
2	Compound Name	Pocket 1	Pocket 2	Pocket 3
3	*R-BNA			
4	S-BNA			
5				
6	*second eluting enantiomer			

Step 3: Now, from Appendix 5, you should have all binding free energy values for each enantiomer binding to each pocket of the micelle. Enter those values into their respective boxes in the Excel Spreadsheet as follows.



	A	B	C	D
1		Binding Free Energy Values		
2	Compound Name	Pocket 1	Pocket 2	Pocket 3
3	*R-BNA	-7.5219	-7.2161	-9.0901
4	S-BNA	-6.2681	-7.4143	-7.1489
5				
6	*second eluting enantiomer			

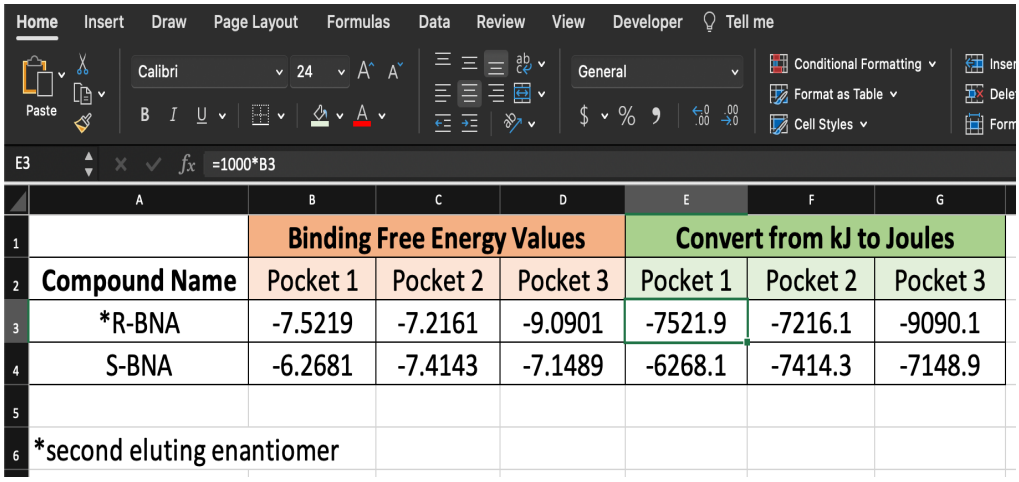
**General Note: The binding free energy value of each enantiomer in the three pockets of poly(SULV) were used to calculate the percent occupancy, P_i, which calculates where a given enantiomer is most likely to occupy based on its $\Delta G_{binding}$ values in each of the binding pockets.

The percent population was calculated with the following equation:

$$P_i = \frac{e^{-\frac{G_i}{K_B}*T}}{\sum_{i=1}^{N} e^{-\frac{G_i}{K_B}*T}} * 100$$

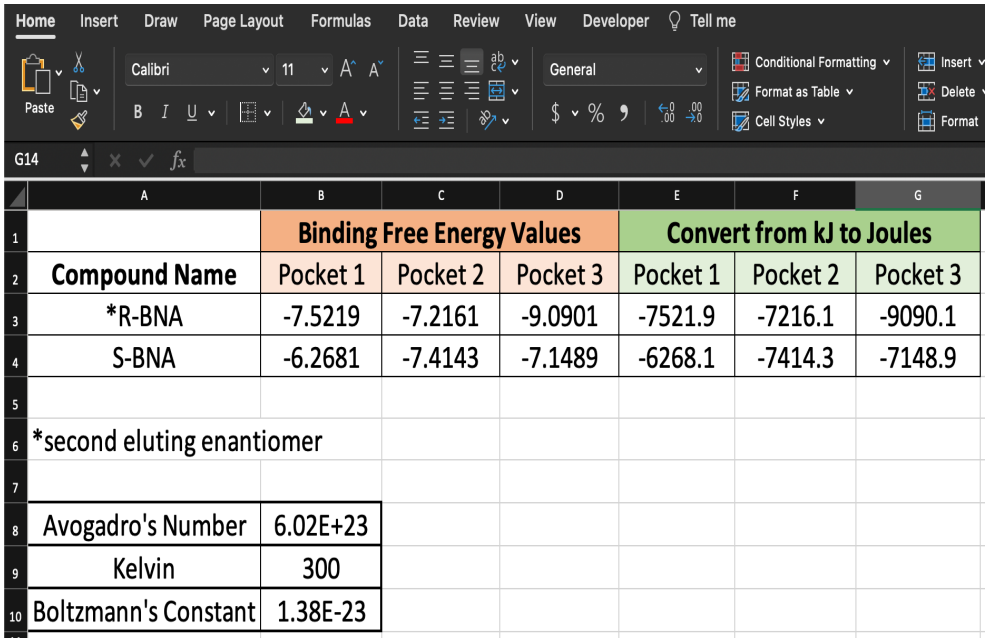
The binding free energy of the enantiomer in the i^{th} pocket of poly(SULV), Boltzmann's constant, and Kelvin temperature are represented by G_i, K_B, T, respectively. This protocol will provide a step-by-step procedure for performing this calculation properly and to also properly analyze your data.

Step 3: Convert from kilojoule to Joules. In the Excel sheet, you will see that I clicked in cell E1, and typed "=1000*B3" into the E1 cell, then clicked enter. I then dragged and copied so that cells E3, F3, G3, E4, F4 and G4 have the same equation.



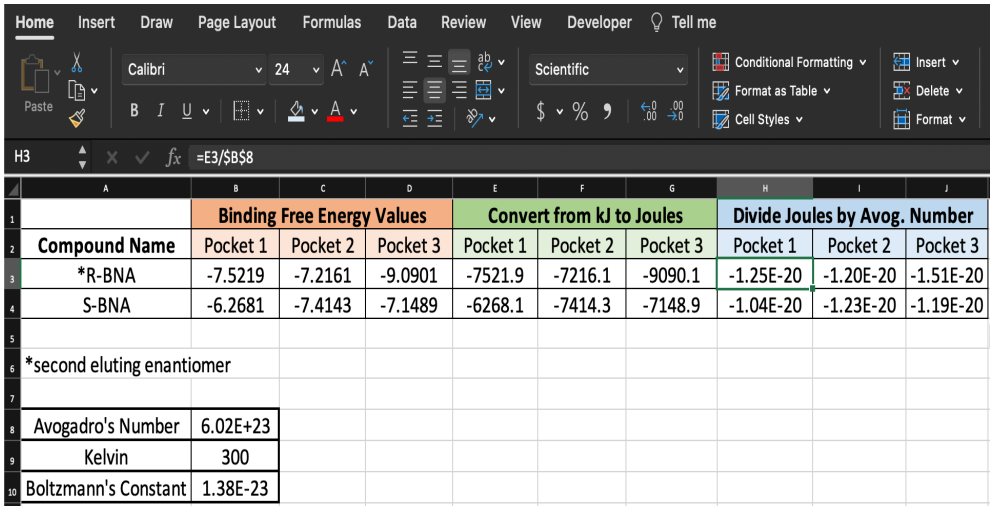
	A	B	C	D	E	F	G
1		Binding Free Energy Values			Convert from kJ to Joules		
2	Compound Name	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3
3	*R-BNA	-7.5219	-7.2161	-9.0901	-7521.9	-7216.1	-9090.1
4	S-BNA	-6.2681	-7.4143	-7.1489	-6268.1	-7414.3	-7148.9
5							
6	*second eluting enantiomer						

Step 4: Add fixed constants into the Excel Sheet, as shown below:



	A	B	C	D	E	F	G
1		Binding Free Energy Values			Convert from kJ to Joules		
2	Compound Name	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3
3	*R-BNA	-7.5219	-7.2161	-9.0901	-7521.9	-7216.1	-9090.1
4	S-BNA	-6.2681	-7.4143	-7.1489	-6268.1	-7414.3	-7148.9
5							
6	*second eluting enantiomer						
7							
8	Avogadro's Number	6.02E+23					
9	Kelvin	300					
10	Boltzmann's Constant	1.38E-23					

Step 5: Divide Joules value by Avogadro's number (6.02E+23), as shown below. This was performed by clicking into H3, then typing "=E3/B8" then dragging and copying the equation to the H3, I3, J3, H4, I4 and J4 cells.

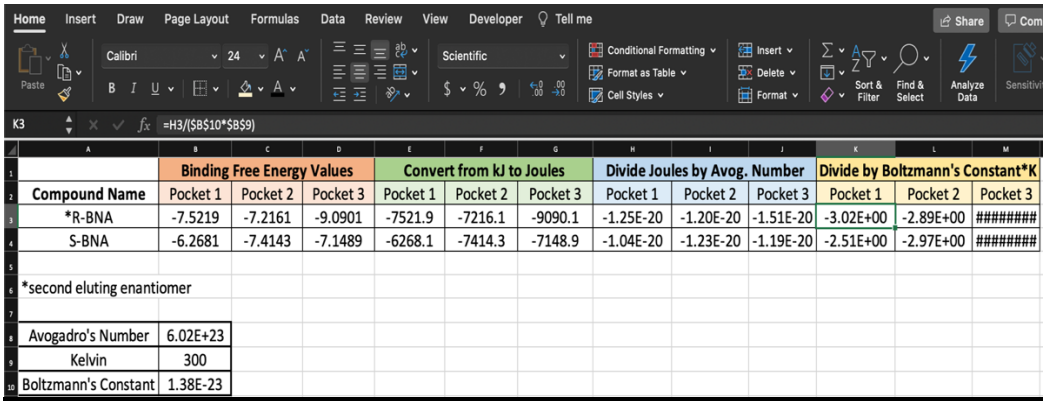


	Binding Free Energy Values			Convert from kJ to Joules			Divide Joules by Avog. Number		
Compound Name	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3
*R-BNA	-7.5219	-7.2161	-9.0901	-7521.9	-7216.1	-9090.1	-1.25E-20	-1.20E-20	-1.51E-20
S-BNA	-6.2681	-7.4143	-7.1489	-6268.1	-7414.3	-7148.9	-1.04E-20	-1.23E-20	-1.19E-20

*second eluting enantiomer

Avogadro's Number	6.02E+23
Kelvin	300
Boltzmann's Constant	1.38E-23

Step 6: Divide the by Boltzmann's Constant multiplied by the temperature in Kelvin. This was performed by clicking into cell K3, then typing "=H3/(B!*B9)" then dragged and copied into cells L3, M3, K4, L4, and M4. ***Notice there are #### because I made up values, for this example, but when you actually do this, your numbers will be fine.*

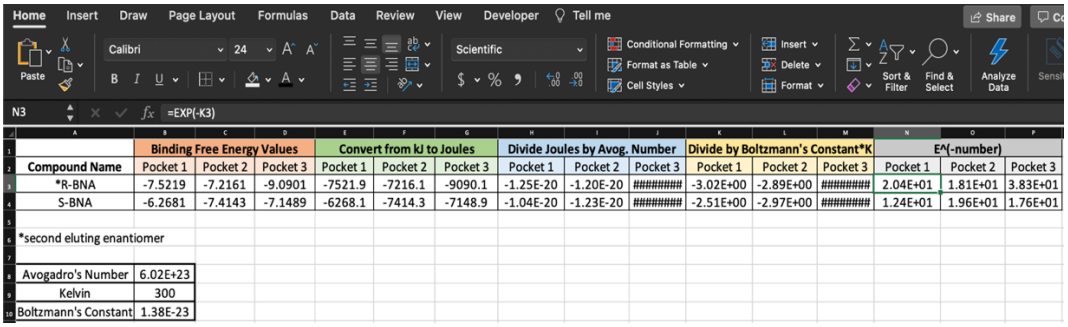


	Binding Free Energy Values			Convert from kJ to Joules			Divide Joules by Avog. Number			Divide by Boltzmann's Constant*K		
Compound Name	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3
*R-BNA	-7.5219	-7.2161	-9.0901	-7521.9	-7216.1	-9090.1	-1.25E-20	-1.20E-20	-1.51E-20	-3.02E+00	-2.89E+00	########
S-BNA	-6.2681	-7.4143	-7.1489	-6268.1	-7414.3	-7148.9	-1.04E-20	-1.23E-20	-1.19E-20	-2.51E+00	-2.97E+00	########

*second eluting enantiomer

Avogadro's Number	6.02E+23
Kelvin	300
Boltzmann's Constant	1.38E-23

Step 7: Take the e^ of the number calculated in Step 6 as shown below. This is done by clicking into cell N3, then typing "=EXP(-K3)" then click enter. Copy and drag into cells O3, P3, N4, O4 and P4.

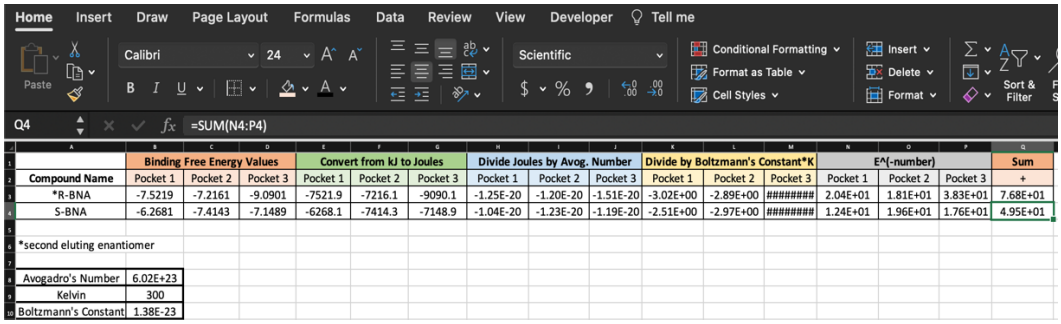


Compound Name	Binding Free Energy Values			Convert from kJ to Joules			Divide Joules by Avog. Number			Divide by Boltzmann's Constant*K			E^(-number)		
	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3
*R-BNA	-7.5219	-7.2161	-9.0901	-7521.9	-7216.1	-9090.1	-1.25E-20	-1.20E-20	#########	-3.02E+00	-2.89E+00	#########	2.04E+01	1.81E+01	3.83E+01
S-BNA	-6.2681	-7.4143	-7.1489	-6268.1	-7414.3	-7148.9	-1.04E-20	-1.23E-20	#########	-2.51E+00	-2.97E+00	#########	1.24E+01	1.96E+01	1.76E+01

*second eluting enantiomer

Avogadro's Number	6.02E+23
Kelvin	300
Boltzmann's Constant	1.38E-23

Step 8: Take the sum of Pocket 1-3 of the values from the gray-colored area, as shown below. This was done by clicking into cell Q3 and typing "=SUM(N4:P4)" then clicking enter. Cell Q3 was clicked and dragged to paste the same equation into Q4.

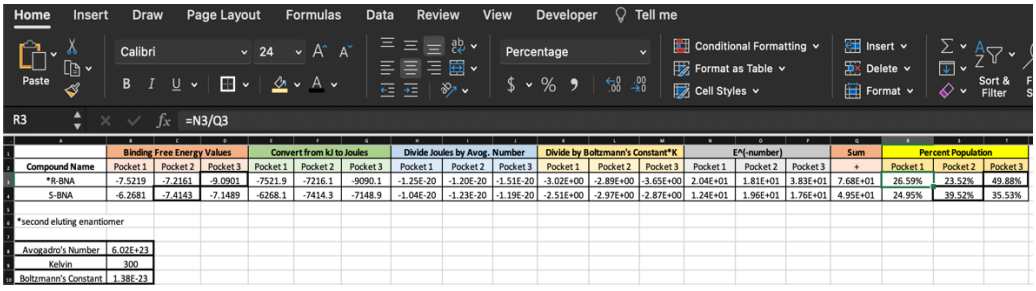


Compound Name	Binding Free Energy Values			Convert from kJ to Joules			Divide Joules by Avog. Number			Divide by Boltzmann's Constant*K			E^(-number)			Sum
	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	+
*R-BNA	-7.5219	-7.2161	-9.0901	-7521.9	-7216.1	-9090.1	-1.25E-20	-1.20E-20	-1.51E-20	-3.02E+00	-2.89E+00	#########	2.04E+01	1.81E+01	3.83E+01	7.68E+01
S-BNA	-6.2681	-7.4143	-7.1489	-6268.1	-7414.3	-7148.9	-1.04E-20	-1.23E-20	-1.19E-20	-2.51E+00	-2.97E+00	#########	1.24E+01	1.96E+01	1.76E+01	4.95E+01

*second eluting enantiomer

Avogadro's Number	6.02E+23
Kelvin	300
Boltzmann's Constant	1.38E-23

Step 9: Determine the percent population. This is performed by clicking into cell R3 then typing "=N3/Q3" then clicking enter then dragging it down into cell R4. Next, click into cell S3 then typing "=O3/Q3" then clicking enter then dragging it down into cell S4. Finally, click into cell T3 then typing "=P3/Q3" then clicking enter then dragging it down into cell T4.



Compound Name	Binding Free Energy Values			Convert from kJ to Joules			Divide Joules by Avog. Number			Divide by Boltzmann's Constant*K			E^(-number)			Sum	Percent Population		
	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	Pocket 1	Pocket 2	Pocket 3	+	Pocket 1	Pocket 2	Pocket 3
*R-BNA	-7.5219	-7.2161	-9.0901	-7521.9	-7216.1	-9090.1	-1.25E-20	-1.20E-20	-1.51E-20	-3.02E+00	-2.89E+00	-3.65E+00	2.04E+01	1.81E+01	3.83E+01	7.68E+01	26.59%	23.52%	49.88%
S-BNA	-6.2681	-7.4143	-7.1489	-6268.1	-7414.3	-7148.9	-1.04E-20	-1.23E-20	-1.19E-20	-2.51E+00	-2.97E+00	-2.87E+00	1.24E+01	1.96E+01	1.76E+01	4.95E+01	24.95%	39.52%	35.53%

*second eluting enantiomer

Avogadro's Number	6.02E+23
Kelvin	300
Boltzmann's Constant	1.38E-23

Now, look at which pocket each enantiomer prefers by looking at the percent population. These are blocked with black outline in the percent population section, so we compare their highest percent occupation against each other's binding free energy values. For example, now we can compare the binding free energy value of R-BNA in Pocket 3 to the binding free energy value of S-BNA pocket 2. We look at R-BNA's free energy value in pocket 3, which is -9.001 kJ/mol, and compare that to S-BNA's free energy value in pocket 2 which is -7.4143 kJ/mol. Since R-BNA elutes second according to experimental MEKC elution order data, we can confirm that the computational binding free energy calculations agree with experimental data.