Improved force fields for the simulation of biological systems

by

Kyung-Shin Suh

B.S., Ohio State University, 2003
M.S., Northern Illinois University, 2008

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

Computer simulation is a common tool used to investigate the behavior of a wide variety of systems. The accuracy, and therefore the usefulness, of a simulation is largely dependent on the quality of the underlying force field (FF) used to model the interactions between atoms and molecules. Therefore, it is essential to validate the capacity of any FF to accurately depict the properties of the system under study. Various studies have identified problems with excessive solute aggregation in solutions, especially for intrinsically disordered proteins/peptides (IDP), when using traditional FFs. Here, we use an established approach for developing new improved FFs for the simulation of solutions of biological interest. The Kirkwood-Buff (KB) Force Field approach was developed to compare experimentally derived KB integrals with those provided by molecular dynamic simulations. Here we examine a variety of intermolecular interactions in aqueous solutions, together with the behavior of peptides at various interfaces, using KB theory and molecular simulations. First, we describe force fields for the simulation of charged and neutral zwitterionic amino acids, sulfates, alkyl sulfates, and sodium dodecyl sulfate (SDS) in aqueous solution, that are specifically developed to reproduce the KB Integrals derived from experimental data. Additional studies of selected small peptides both in the presence and absence of SDS micelles using these new FFs provided conformational changes dependent on the presence of SDS micelles and in agreement with experiment. Further studies of Amylin, an IDP, at various interfaces such as air/water, water/micelle, and lipid/water, also suggested significant conformational changes. Again, the application of these new FFs for studying interface behavior yielded results in good agreement with experimental data.

Improved force fields for the simulation of biological systems

by

Kyung-Shin Suh

B.S., Ohio State University, 2003
M.S., Northern Illinois University, 2008

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Paul E. Smith

Copyright

Abstract

Computer simulation is a common tool used to investigate the behavior of a wide variety of systems. The accuracy, and therefore the usefulness, of a simulation is largely dependent on the quality of the underlying force field (FF) used to model the interactions between atoms and molecules. Therefore, it is essential to validate the capacity of any FF to accurately depict the properties of the system under study. Various studies have identified problems with excessive solute aggregation in solutions, especially for intrinsically disordered proteins/peptides (IDP), when using traditional FFs. Here, we use an established approach for developing new improved FFs for the simulation of solutions of biological interest. The Kirkwood-Buff (KB) Force Field approach was developed to compare experimentally derived KB integrals with those provided by molecular dynamic simulations. Here we examine a variety of intermolecular interactions in aqueous solutions, together with the behavior of peptides at various interfaces, using KB theory and molecular simulations. First, we describe force fields for the simulation of charged and neutral zwitterionic amino acids, sulfates, alkyl sulfates, and sodium dodecyl sulfate (SDS) in aqueous solution, that are specifically developed to reproduce the KB Integrals derived from experimental data. Additional studies of selected small peptides both in the presence and absence of SDS micelles using these new FFs provided conformational changes dependent on the presence of SDS micelles and in agreement with experiment. Further studies of Amylin, an IDP, at various interfaces such as air/water, water/micelle, and lipid/water, also suggested significant conformational changes. Again, the application of these new FFs for studying interface behavior yielded results in good agreement with experimental data.

Table of Contents

List of Figures

List of Supplementary Figure

List of Tables

Acknowledgements

I would like to express my profound gratitude to my principal advisor, Prof. Paul E. Smith, for his unwavering support, insightful guidance, patience, and encouragement throughout my doctoral journey. I extend my heartfelt thanks to the members of my doctoral committee, Prof. Christine M. Aikens, Prof. Daniel Higgins, and Prof. Jeffrey R. Comer, for their invaluable contributions and constructive feedback. I would like to thank Dr. Elizabeth Ploetz for providing the amino acid code for the KB analysis, obtaining dihedral alkyl sulfate torsional parameters, and the idea of studying Amylin at different interfaces.

Dedication

To my family

Chapter 1 – Introduction

1.1 A brief history of computational chemistry

Theoretical and Computational Chemistry concentrates on the development and application of theoretical approaches in chemistry and related disciplines such as biology/biochemistry, physics, etc. This field of study has grown rapidly with the development of improved computer technologies and the development of new mathematical algorithms and software. The origin of computational chemistry begins from attempting to solve the time dependent Schrödinger equation by hand-cranked calculators[1]

$$\hat{H}(r,t)\psi(r,t) = i\hbar\frac{\partial}{\partial t}\psi(r,t) \qquad 1.1$$

where $\psi(r,t)$ is the wave function of the system, which provides the probabilities for the outcome of any measurement of the system's properties, r denotes the spatial coordinates (x, y, z), t is time, $\hbar$ is the reduced Planck constant given by $\hbar = h / 2\pi$. The Hamiltonian operator $(\hat{H})$ depends upon time as does the wavefunction.

Experimental results for the energy levels and atomic spectra of helium and hydrogen were amongst the first to be accurately predicted by solutions to the Schrödinger equation. During the mid-1950s, after the invention of the computer during World War II, chemists examined molecular behavior by numerical approximation of the Schrödinger equation employing computational methods.[2] The subsequent development of accurate basis sets, good estimations of electron correlation, and methods to derive energy with respect to nuclear position in 1960s, all led to a speed up in the development of computational chemistry.[3]

While the many-body electronic structure and potential energy surfaces associated with nuclear motion are the focus of quantum chemistry (QM), the molecular mechanics (MM) approach has been developed to provide more efficient estimates of the total energy for larger

systems of atoms. Here, the energy of a chemical system is evaluated by the application of additive terms involving distances between atoms, bond angles, and dihedral angles to calculate the energy of systems such as liquids, solid materials, and biological macromolecules.[4] These energy calculations are often combined with energy minima searching approaches in MM methods to determine relative energies of different configurations of molecules, although this particular research did not focus extensively on minima searching. Instead, the work primarily involved using the gradient of the energy function, known as force fields, to provide atomic forces for calculating the dynamic path of an ensemble of molecules in the MD approach.

MD simulation began in the late 1950s with simple systems of gases.[5] Bovine pancreatic trypsin inhibitor, a small globular protein, was selected to perform the first biomolecular MD simulation in the late 1970s.[6] In the intervening years, there were numerous significant advancements in the field. Key developments included improvements in simulation algorithms, the introduction of new force fields, and the application of MD simulations to increasingly complex biological systems. These innovations laid the groundwork for a deeper understanding of molecular behavior and paved the way for more accurate and comprehensive simulations. By 2013, this culminated in Martin Karplus, Michael Levitt, and Arieh Warshel being awarded the Nobel Prize for the development of multiscale models for complex chemical systems. A brief history of the development of computational chemistry, including other substantial advances, is shown in Figure 1.1, and further details are discussed in section 1.3.[7]

Computational chemistry has now become an important part of research in chemistry used to help understand the structures and properties of molecules, and the mechanisms and selectivity of reactions. Computer simulations play a major role in connecting our microscopic view of many systems with the macroscopic properties they display in the laboratory.

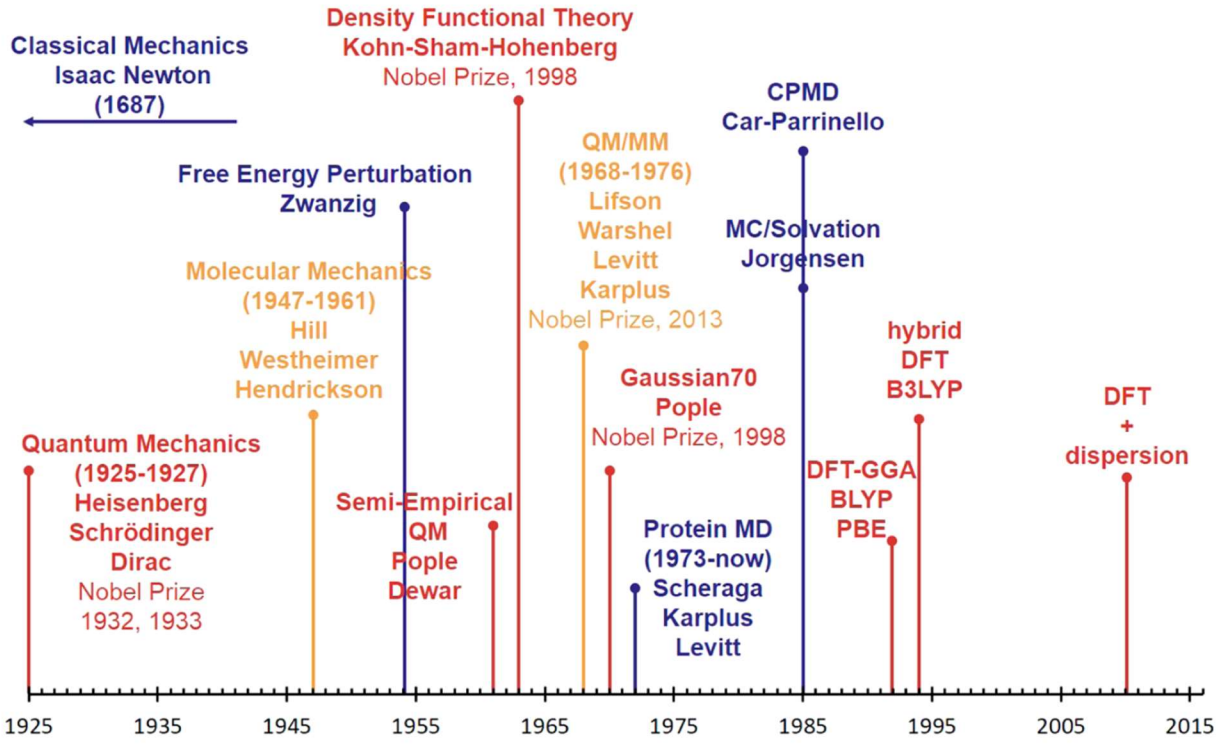


Figure 1.1 Milestones in computational organic and biological chemistry. This figure is reprinted with permission from Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. Accounts of Chemical Research 2017, 50 (3), 539-543. DOI: 10.1021/acs.accounts.6b00532. Copyright American Chemical Society (2017) [7]

Currently, the field of computational chemistry incorporates a variety of modelling techniques covering multiple time and size scales. Several representative modelling methods are listed in Figure 1.2,[8] together with their suitable ranges of application. As Figure 1.2[8] demonstrates, there are currently representative modelling methods that can be employed for computational investigations of physicochemical properties of biomolecules. The time scale and size of a system are the most important factors for the choice of a particular computational method.

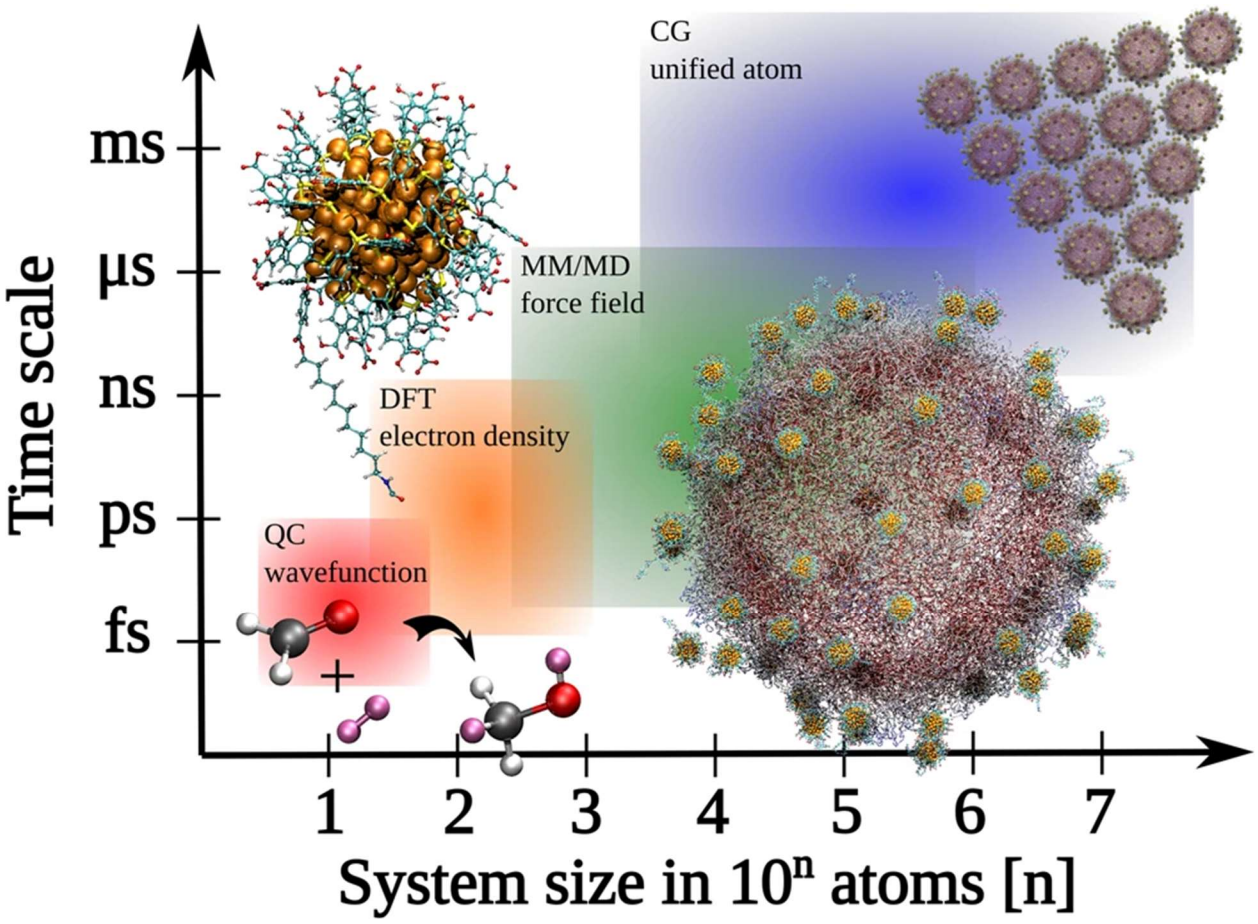


Figure 1.2 System size and time scale ranges for various computational methods: Quantum Calculations (QC), Density Functional Theory (DFT), Molecular Mechanics/Molecular Dynamics (MM/MD), and Coarse-Grained (CG) models. This figure is This figure is reprinted with permission from Malola, S.; Hakkinen, H. Prospects and challenges for computer simulations of monolayer-protected metal clusters. Nature Communications 2021, 12 (1), Editorial Material. DOI: 10.1038/s41467-021-22545-x. This is an open access article distributed under the terms of the Creative Commons CC BY license. (2021)[8]

Electronic structure methods provide accurate information about molecules, yet full QM calculations for multi-electron systems are extremely computationally expense, and therefore approximations are required for many calculations. Consequently, QM calculations are generally used for systems containing less than hundreds of atoms, and are restricted to calculate the ground-state structure and properties of the systems.[9] Calculations based on Density functional theory (DFT)[10] can reduce the complexity of the ground-state systems by using the total electron density of the system as a basic variable. Using effective DFT codes, with immensely parallelized

supercomputers, DFT methods can handle calculations involving thousands of atoms. However, the calculation of time dependent properties are generally limited to short time scales to the order of 100 ps.[11] The breaking or forming of chemical bond can be observed on this time scale, as well as configurational changes of small molecules can be studied in detail. However, more complex systems such as macromolecules, condensed phase systems, multi-phase interfaces, and their dynamic behaviors are difficult to study with QM or DFT methods. To calculate beyond the limits of the DFT system size and time scale requires the pre-parametrized atomistic interaction potentials provided by a classical force field. The optimization of structures with MM and the related MD simulation of systems with atomistic detail can be performed with up to millions of atoms in the system, and for up to microsecond time scales. Unfortunately, with MM and MD simulation methods the breaking and forming bond cannot be studied. However, these approaches are very useful for the study of non-bonded interactions in condensed phases, such as protein folding and other conformational changes in aqueous solution. Even larger system sizes and time scales can be studied with coarse-grained (CG) methods,[12] which simplify the system further by treating groups of atoms as single particles. This allows for the simulation of mesoscale systems on the millisecond time scale.[13]

1.2 Molecular dynamics simulation

Molecular dynamics simulation is a computational technique used to studying the motion of atoms and molecules which are allowed to interact for a fixed period of time, thereby providing a picture of the dynamic evolution of the system. The interaction potential is the main quantity of interest for MD simulations as this provides the forces that determine the atomic motions. Since most MD simulations are focused on the motion of atoms, and not electrons, the nuclear motion

of all the component atoms (and molecules) are assumed to follow the laws of classical mechanics.[14] The general process of performing a MD simulation is illustrated in Figure 1.3.

The motion for a system of N interacting atoms is governed by classical mechanics equations using the atomic forces. The force on any atom is obtained from evaluation of the force between non-bonded (intermolecular) atom pairs, together with forces from bonded (intramolecular) interactions and any restraining/external forces that may be required for the calculation.[15] The particle motions can be simulated by mathematically solving Newton's equations of motion as shown in Equation 1.2.,

$$F_i = -\frac{\partial V}{\partial r_i} = \sum_j F_{ij} = m_i \frac{d\vec{v}_i}{dt} = m_i \frac{d^2\vec{r}_i}{dt^2} \qquad 1.2$$

where the force F_{ij} arises from all other j particles acting on particle i, m_i is mass of the i th particle, V is potential energy and r_i is the vector describing the particle position. $\vec{v}_i$ is velocity of ith particle.

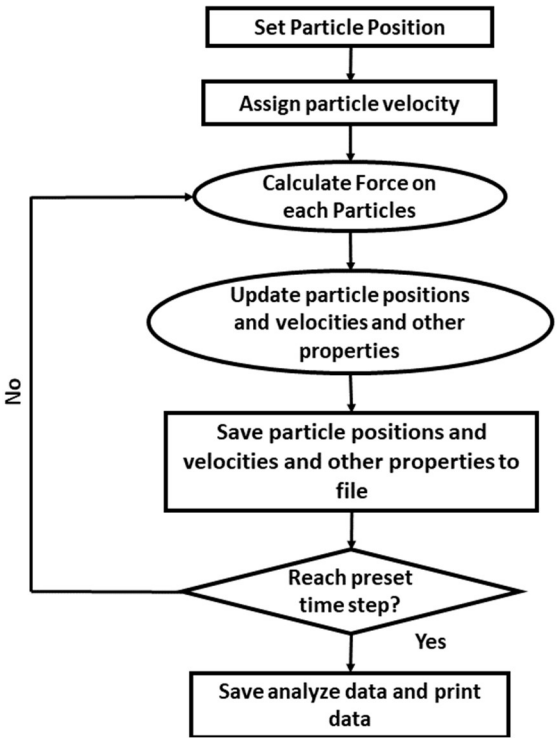


Figure 1.3 Flow chart for a general MD simulation

The simulation proceeds in small steps (Δt) during which all the atoms move, and new forces are determined for this new set of coordinates or configuration of the system. It is important to choose a proper time step because only a limited part of phase space will be monitored if the time step is too small, whereas long time steps would miss physiochemically important observations and ultimately cause numerical instability. Generally, a time step at least one order of magnitude smaller than the fastest motion present in the system is required. The average vibrational frequencies of atoms range from less than 10^{13} Hz to roughly 10^{14} Hz, therefore timesteps of 1~2 fs (femtosecond) are applied for normal classical MD simulations.

1.2.1 Equations of Motion

Kinetic and thermodynamic properties can be obtained from numerical solutions of Newton's equation of motion to determine the time evolution of the system. Newton's equation of motion has to be numerically integrated to determine the position $r_i\ (t + \Delta t)$ of a given object at any time $(t + \Delta t)$ when the atom's position and velocity are known at time t. The Leapfrog integrator is one of the most popular integrators and is broadly used in MD simulation. This integrator is generally known as the Verlet Ordinary Differential Equation (ODE) integration algorithm which is essentially a force integrator.[16] The forces come from the potential function. Acceleration, velocity and position can then be determined using this algorithm as described in equation 1.3 and 1.4. This integration is mathematically not so difficult, yet it can be computationally intensive because this needs to be done for many finite time steps. The leapfrog integrator can be described by the following equations,

$$\mathrm{v}\left(t + \frac{1}{2}\Delta t\right) = \mathrm{v}\left(t - \frac{1}{2}\Delta t\right) + \mathrm{a}(t)\Delta t \qquad 1.3$$

$$\mathrm{r}(t + \Delta t) = \mathrm{r}(t) + \mathrm{v}\left(t + \frac{1}{2}\Delta t\right)\Delta t \qquad 1.4$$

where r(t) is current position and a(t) is the current acceleration, which are the accumulated quantities, together with intermediate velocity $v(t - 1/2\,\Delta t)$. Equation 1.3 is first applied, and then the intermediate velocities $v(t + 1/2\,\Delta t)$ are obtained from the velocity "leaping over" the coordinates. The current velocities can then be calculated from,

$$\mathrm{v}(t) = \frac{1}{2}\left(\mathrm{v}\left(t + \frac{1}{2}\Delta t\right) + \mathrm{v}\left(t - \frac{1}{2}\Delta t\right)\right) \qquad 1.5$$

These steps are then repeated to provide the total simulation time.

1.2.2 Periodic boundary condition and minimum image

The representation of macroscopic (infinite) systems in computer simulations requires some care. In particular, finite size systems can suffer from the presence of surface effects. Finite system and surface effects can be significantly reduced by the use of periodic boundary conditions (PBCs). PBC essentially use replicas of a central system or cell to provide an effective macroscopic system in much the same way as a unit cell provides a picture of a macroscopic crystal. The use of PBCs diminishes problems with interatomic potential changes at the cell boundaries and more realistic (bulk system) simulations are generated by including interactions with the neighboring cell across the cell boundaries.[15] To employ PBCs the 3-dimensional surface of a central simulation box is specified and, as the simulation proceeds, if a molecule exists the central box on one side, it is translated by one unit cell such that it is replaced on the other side of the box, with the same velocity. This process is described in Figure 1.4.

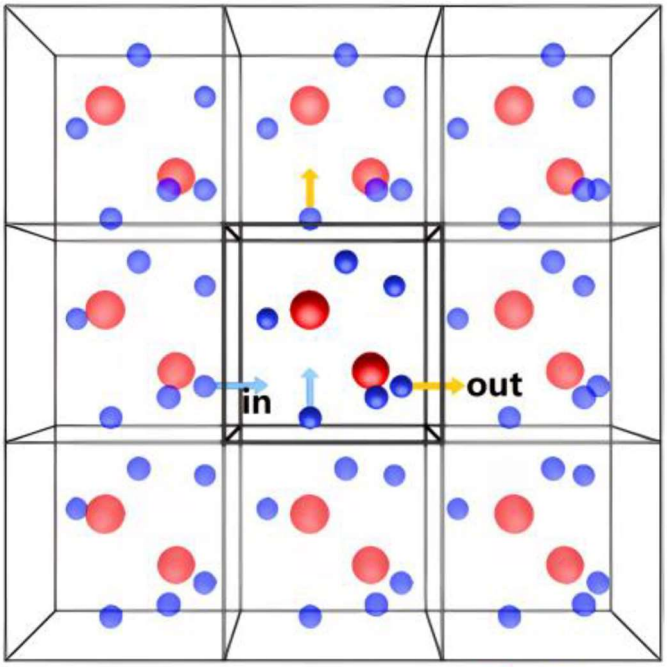


Figure 1.4 Periodic boundary conditions This figure is reprinted from Chen, Y. L.; Liu, H. X.; Gao, T. Z.; Wei, H. Simulation of the Irradiation Cascade Effect of 6H-SiC Based on Molecular Dynamics Principles. Micromachines 2023, 14 (2). DOI: 10.3390/mi14020455. This is an open access article distributed under the terms of the Creative Commons CC BY license (2023).

The infinite copies of the central simulation box are called images. The simulation of molecules using a central box under PBCs are performed using the minimum image convention to ensure that each molecule interacts with nearest image of any of the other molecules in the system. Using this approach, a small finite periodic system can be used to represent the infinite macroscopic system of interest. Generally, PBCs do not have a significant influence on the equilibrium thermodynamic properties and solution structure far from phase transition and when using short range potentials.

Nonbonded pairwise interactions between pairs of particles which are separated by large distances are usually truncated. The interaction between pairs of particles with large distance separations can be ignored if the nonbonded pairwise potentials decay quickly with separation distance. In this case a simple cutoff scheme is typically implemented in an attempt to save computational effort. Pair interactions that do not decay fast enough with distance can be

problematic for simulations. The Coulomb interaction, which applies between atoms with partial atomic charges that mimic the electron distribution in molecules, is the most common long-ranged interaction that appears in computer simulations. Coulombic interactions are generally treated with the Ewald method, which computes the electrostatic potential energy between all charged atoms in the central box, together with the interaction with those charged atoms and all periodic images of them.

Ewald methods are frequently utilized to handle long-range electrostatic interactions, particularly in systems with periodic boundary conditions.[17] These methods split the interaction into two contributions: a short-range component, calculated in real space, and a long-range component, calculated in reciprocal or Fourier space. While this approach determines the electrostatic energy of the entire (replicated) system, computations in Fourier space can be time-consuming, especially for large systems.[18] To mitigate this issue and improve computational efficiency, the particle-mesh Ewald (PME) method has been developed.[19] The PME method is a variant of the Ewald technique that incorporates a Fast Fourier Transform (FFT) algorithm for the reciprocal space computations, significantly accelerating the process.[20] The PME method assigns charges to a grid by interpolation, resulting in a more manageable number of grid points for the Fourier calculations.[21] The primary advantage of the PME method is that it essentially eliminates the need for a cutoff for the electrostatic interactions, thereby eliminating potential cutoff artifacts, and enabling a more accurate representation of these interactions throughout the system.[22] This makes it especially beneficial for systems with periodic boundary conditions where long-range interactions play a significant role.[23] As such, the PME method has become the method of choice for most molecular dynamics simulations involving long-range electrostatic forces.[24]

1.2.3 Temperature/pressure coupling

Molecular dynamics (MD) simulations are generally conducted within the constraints of a specific thermodynamic ensemble, each representing a unique set of maintained conditions.[25] The choice of ensemble largely depends on the nature of the system under study and the experimental conditions that the simulation seeks to emulate. One such ensemble is the microcanonical, or NVE, ensemble (where N represents the number of particles, V stands for volume, and E denotes the total energy), and for which N, V, and E take fixed values. The simplicity of the NVE ensemble, and its strict adherence to Newton's laws of motion, makes it the most straightforward ensemble to implement in simulations. However, it is often the case that experimental conditions maintain temperature or pressure constant, rather than energy. As such, the canonical (NVT) and isothermal-isobaric (NPT) ensembles—where T signifies temperature and P signifies pressure—are frequently more appropriate for comparison with experimental systems. In these ensembles, temperature and/or pressure are kept constant via thermostatic and barostatic controls, respectively. In more complex systems where particle number fluctuations need to be considered, the grand canonical ensemble (μVT) might be employed. This ensemble, characterized by fixed chemical potential (μ), volume, and temperature, permits exchanges of particles with a reservoir, offering an alternative view of the system's potential states.

A thermostat permits energy to enter and exit through the simulated system to maintain a constant temperature. Most thermostats achieve this by altering the velocities of subgroups of particles. Temperature control can be achieved by various approaches. Velocity rescaling is the simplest method that rescales the velocities at each step. The Berendsen thermostat uses a weak coupling approach.[26] This thermostat is similar to coupling the simulated system to a fictitious heat bath kept at an fixed temperature. In Stochastic methods,[27] a subset of atoms is periodically

randomly assigned with new velocities based on a Maxwell-Boltzmann distribution for the target temperature. The Andersen thermostat and Langevin thermostat are the examples of this method.[27] The Nosé-Hoover thermostat uses an extended system method and is one of the more popular thermostats for MD simulation.[28, 29] An artificial variable associated with a fictional heat bath mass are added to the equations of motion. Thus, a heat bath is considered to be an integral part of the system for this method.

Pressure control algorithms, also known as barostats, are essential tools in molecular dynamics simulations. They are used to maintain a constant pressure within the system or to apply an external stress. This control is typically achieved by dynamically altering the size and shape of the unit cell and correspondingly rescaling all atomic coordinates throughout the simulation, a process similar to that of temperature control. The Berendsen barostat,[30] for instance, is an algorithm that applies a scaling factor to gently drive the pressure towards a target value, allowing for the rapid stabilization of a system's pressure. However, it's important to note that, while effective, the Berendsen barostat does not generate a true NPT ensemble, as it doesn't accurately represent the correct pressure fluctuations. On the other hand, the Parrinello-Rahman method,[31] an extension of the Andersen method,[32] is another widely employed barostat in molecular dynamics simulations. This method allows for changes not only in the size but also in the shape of the simulation cell, accommodating external stresses and providing a more accurate representation of an NPT ensemble. Its widespread use can be attributed to its ability to more precisely reflect the pressure conditions of real-world experimental scenarios.

1.3 Force fields and force field parameters

As discussed above, to determine the energy and forces required during a simulation one uses simple equations and parameters. Indeed, the quality of a simulation depends on the accuracy

of the energy calculation. Here, we discuss these equations in more detail. Potential energy functions using internal coordinates of molecules were used as early as the 1930 ~1940s, well before MM approaches were developed.[33] This type of approach was used to describe the vibrational frequencies of small molecules including ethanol and ethyl halides. However, the potential energy functions were specific to the molecule during these early stages.[34] More recently, classical MD simulations have become popular computational methods to investigate biological systems including proteins, lipids, nucleic acids, carbohydrates, and also small organic molecules. In 1969, Levitt and Lifson proposed simple empirical expressions for determining the potential energy and forces from a set of atomic coordinates, namely a force field (FF). Modern conventional FFs commonly used in MD simulation were established by Lifson's group.[35]

Classical FFs typically contain a series of bonded interactions that represent any bonds, bond angles, and dihedral angles in the system, together with a series of non-bonded terms that includes both electrostatics and van der Waals interactions. The Lennard-Jones form is often used to describe the van der Waals term, while Coulomb's law is usually applied to represent the electrostatic interactions.[36] For the classical MD simulations, the electrons are not described explicitly due to the size range of molecule implicated and the time scale. The systems are considered to be in the electronic ground state. Atoms are represented by spheres located in space with an associated mass, partial atomic charge and van der Waals parameters.[37]

The potential energy is assumed to be additive as indicated in Equation 1.6.

$$V(r) = V_{bond}(r) + V_{angle}(r) + V_{dihedral}(r) + V_{improper}(r) + V_{elec}(r) + V_{vdW}(r) \qquad 1.6$$

where V(r) is the potential energy function for the system and r is the coordinate of all atoms in the system. The V_{bond}, V_{angle}, $V_{dihedral}$, and $V_{improper}$ terms describe the bond stretch, angle bending, torsion or dihedral angle, and improper dihedral angle bending motions in molecules.

The V_{elec} term represents electrostatic terms, and V_{vdW} represents van der Waal terms. These two terms correspond to non-bonded terms.[38] While the general form of the equations used to describe these different interactions has remained the same, the parameters associated with a particular FF have been consistently improved since this general FF form was established.[39]

1.3.1 Bonded terms in force fields

The bonded terms in a force field include covalently bonded interactions between 2 ~4 atoms. A harmonic functional form is usually adopted to describe bond stretching,

$$V_{bond} = \frac{1}{2} k_b (r - r_0)^2 \tag{1.7}$$

where k_b is the force constant, r_0 is the equilibrium bond length, and r is the bond length. Gravitational forces, which are negligible compared to other forces at the molecular scale, are not considered in MD simulations. Furthermore, there are no Quantum Mechanical (QM) vibrational energy levels involved in classical MD simulations. Bond dissociation is also not allowed due to application of the harmonic functional form. More complicated bond stretching terms, such as morse potential, have been used for the inclusion of anharmonicity. However, it is more computationally demanding and does not affect the dynamics of biomolecules. Thus, the simple harmonic bond potential is most commonly used. It should be noted that the use of the simple harmonic bond potential effectively constrains bond lengths, preventing them from changing significantly during the simulation. These constraints reduce the complexity of the system and allow for more efficient computation.

A harmonic potential is also implemented to describe the 3-body bond-angle vibrational potential energy. For example, for the angle formed by three atoms A-B-C, as shown in Figure 1.5, the bond angle potential takes the form,

$$V_{angle} = \frac{1}{2} k_a (\theta - \theta_0)^2 \quad 1.8$$

This is similar in form to the bond stretching equation where the parameters k_a and θ_0 are defining the stiffness and equilibrium geometry of each angle.

The energy for rotation around bonds are expressed in terms of the dihedral angle of four atoms connected in series and often takes the form,

$$V_{dihedral} = k_\phi [1 + \cos(n\phi - \delta)] \quad 1.9$$

where k_ϕ represents force constant that controls the size of the maxima and minima, n represents the periodicity which determines the number of maxima and minima in a full rotation (360 degrees), and the phase shift is represented as δ, which allows one to shift the position of the maxima and minima. An alternative "improper" dihedral is also used to ensure planarity of atoms due to hybridization and conjugation effects. This is usually enforced with a simple harmonic potential,

$$V_{improper} = \frac{1}{2} k_\phi (\phi - \phi_0)^2 \quad 1.10$$

In contrast to the proper dihedrals, the improper dihedral angle can be expressed by using three atoms which are all attached to one central atom as illustrated in Figure 1.5.

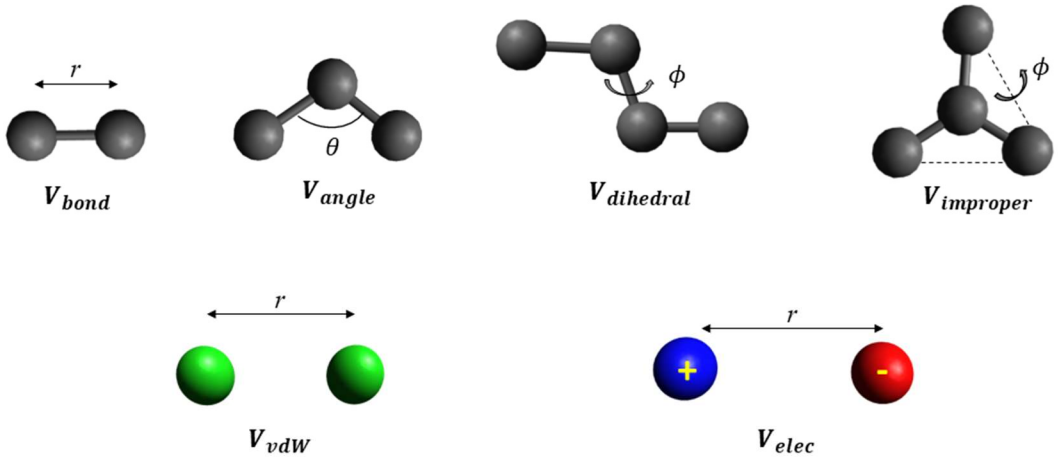


Figure 1.5 Illustration of the interaction energy types in a traditional potential energy function.

1.3.2 Nonbonded terms in force fields

Nonbonded interactions refer to the through space or intermolecular interactions. Nonbonded interactions occur either between atoms in separate molecules or between atoms separated by three or more bonds within the same molecule. One of the equations commonly used to model VDWs interaction is the Lennard-Jones (LJ) equation which takes the form,

$$V_{vdW} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad 1.11$$

where r is distance between two interacting atoms. At short distances the interaction is repulsive, while at long distances the (dispersion) interactions are always attractive. The ε parameter represents the well depth and determines how strongly the two particles attract each other. The σ parameter corresponds to the distance at which the intermolecular potential between the two particles is zero. Hence, σ controls how close two nonbonding particles can get and is commonly indicated as the van der Waals radius. The LJ parameters depend on the two interacting atom types (C, N, etc). Combination rules are applied to obtain the LJ parameters between two different atom types given the parameters between two atoms of the same type. Different Force Fields often use different types of combination rules. Generally, the Lorentz-Berthelot combination rule employ the arithmetic mean for σ, and the geometric mean for ε as indicated in Equation 1.12,

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad 1.12$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \quad 1.13$$

where *i* and *j* are atom types.

The electrostatic potential between two charged atoms is expressed as

$$V_{elec} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r} \quad 1.14$$

where q_i and q_j are the parameters that represent the effective charges on atoms i and j separated by a distance r. The vacuum permittivity and relative permittivity (dielectric constant) are represented as ε_0 and ε_r respectively. The effective charges correspond to partial atomic charges, with non-integer values, which are assigned to each atom to represent the overall charge distribution of a molecule. The sum of the partial charges in a molecule is required to be equal to the molecule's net formal charge. This type of FF is a simple, but very effective approach, to investigate various systems. Partial charges of the atoms are fixed over the simulation periods to reduce computational cost even though explicit polarization effects may be important in some systems.[39] The application of fixed charge FFs has proved a successful way to study biological systems, and they remain the most commonly applied FF in use today.

1.3.3 Traditional FFs for biomolecules

MD simulations have been used to study the overall structure, dynamics and thermodynamics of biomolecular systems for many years. The overall quality of these types of simulations have been gradually improved over time aided by increasing computing power and software development. Nevertheless, the FF remains one of the most important factors determining the quality of a simulation.[40, 41]

The FFs models most commonly used to simulate biomolecules are either all atom, united atom (UA), or coarse grained (CG) models.[42] Every atom in a system, including all hydrogens, is explicitly included for all atom FFs, whereas only the heavy atoms and the polar hydrogen atoms are explicitly parameterized in UA FFs. Several atoms are usually grouped together as a single "bead" for use in CG FF. Of course, as the number of explicit atoms included in a simulation decreases the simulation becomes more efficient, but also more approximate. The size of the system and the sampling time are therefore a major factor in the choice of a particular FF. The

study of large macromolecular assemblies, or systems requiring long simulation times, may require a CG FF, while UA approaches offer a moderate increase in the sampling of conformational space.[43]

The AMBER, CHARMM, and OPLS models are among the most widely utilized all-atom FFs,[39, 44, 45] while GROMOS and KBFF represent some of the most prevalent UA force fields.[46, 47] The MARTINI force field is a particularly popular CG model.[48] There exist multiple versions of each force field, continuously optimized over time for various types of molecules; including proteins, DNA, RNA, and small organic compounds. Another categorization of force fields are starting to appear in the literature. Traditional classical force fields do not explicitly account for polarization, resulting in these being referred to as non-polarizable or fixed charge force fields. However, recent advancements in computational power have enabled the development of polarizable force fields, which do provide a more precise description of charge redistribution due to changes in the local environment. Examples of such polarizable force fields include POL3, a model specifically designed for water,[49] and AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications), a comprehensive model designed for a wide range of biomolecules.[50] These polarizable models capture dynamic changes in the electron clouds around atoms in response to their environment. However, the increased level of detail comes at the cost of significantly higher computational demand compared to fixed charge models. Despite these advancements, classical fixed charge force fields continue to dominate the field due to their lower computational cost, proving sufficiently accurate for many biomolecular simulation studies. Nonetheless, the choice of force field greatly depends on the specific requirements of the study and the balance between computational feasibility and accuracy.

Most biomolecular FFs available today have been established using QM calculation to provide many of the required parameters, and the effective partial atomic charges for non-polarizable models, in particular. The FF parameters are considered to be additive in this model. It is also assumed that the parameters for small molecules, amides for example, can be transferred and used for larger molecules, such as polypeptides and proteins, where the same functional groups appear.

These FFs have been very useful for the simulation of globular proteins and short peptides where one is primarily interested in the properties of folded protein structures.[51] However, recent comparison with experiment have uncovered that these FFs have major problems reproducing the properties of intrinsically disordered proteins (IDP),[52] and denatured state ensembles of globular proteins. Traditional FFs have generally provided over-collapsed structures. The over-collapse problem generally arises from poor solvation, which is a result of the over estimation of solute-solute interactions and under estimation of solute-solvent interactions.[53, 54] This is directly related to the corresponding small molecule FFs which also show excessive solute aggregation in condensed phase simulations. To improve the over-collapse problem, FFs have been recently re-evaluated and optimized. Recent modifications in force field development have included altering the traditional ε combination rules between the water oxygen and protein atoms. This has led to the reparameterization of protein atoms in force fields such as Amber ff99SB/GAFF+TIP4P-Ew, and Amber ff03ws. The effectiveness of these modifications in improving the modelling of protein-water interactions has been assessed in subsequent tests by Mercadantes *et al*.[53] Additional strategies for FF improvement include the development of the new TIP4P-D water model, refinement via a99SB-disp, and the CHARMM36m model which scales water interactions.[55-57] Although, FFs have undoubtedly been improved with these many efforts, it is acknowledged

problems still exist. For example, the recent improvements in FFs generally lead to better results for IDP simulations, but also results in increased instability for structured globular proteins for which the folded state may no longer by stable.[57]

The above issues illustrate the difficulties in obtaining a good balance between solute-solute and solute-solvent interactions. If the solute-solute interactions are too strong then folded proteins will be (too) stable and IDPs will over-collapse. On the contrary, if the solute-solvent interactions are too strong then IDPs will not collapse but folded proteins may become unstable. The problem is to improve this balance. However, it is difficult to quantify the solute-solute and solute-solvent interactions to determine their relative strengths. One way to do this has been developed by the Smith group.[47] Here, one obtains simulation data regarding the solute-solute and solute-solvent distributions in solution that can be compared to the corresponding experimental quantities. The approach used to achieve this goal involves the Kirkwood-Buff theory of solutions.

1.4 Kirkwood-Buff (KB) theories

The Kirkwood-Buff (KB) theory of solutions was first introduced in 1951. Kirkwood and Buff established new relationships between thermodynamic quantities and molecular distribution functions for multicomponent system defined in the μVT or Grand Canonical ensemble (GCE).[58] In this way KB theory can be considered to provide a link between macroscopic thermodynamic properties – such as the compressibility, the partial molar volumes, and various types of chemical potential derivatives – and the underlying microscopic "structure" of the solution. The relationships are exact and the results can be applied to study any multicomponent solution containing any types of molecules as long as the mixture is miscible. Recently, KB theory has been used to help improve models for computer simulation by the Smith group and others.[59, 60]

The relationship between the thermodynamic properties and the molecular distributions is achieved through the KB integrals (KBI) defined as following,

$$G_{ij} = \int_0^\infty \left[g_{ij}(r) - 1\right] 4\pi r^2 \, dr \tag{1.15}$$

where $g_{ij}(r)$ is the radial distribution function (RDF) defined in μVT ensemble system for the two species i and j, separated by a distance, r. The KB integral, G_{ij}, quantifies the deviation in the distribution of j molecules around a central molecule i from that of a random distribution. Therefore, the KBIs can provide quantitative information concerning the affinity between species i and j. If the KBI between i and j is positive then the net interaction between the two species is favorable. On the other hand, the interaction between two species can be considered unfavorable if they display a negative KBI. The units of the KBIs are volume and are commonly provided in cm^3/mol or L/mol depending on their magnitude.

In the μVT ensemble system, the number of particles fluctuates since its conjugate variable, the chemical potential, μ, is fixed. Consequently, one way to understand the KBIs, defined in the μVT ensemble, is to consider two closely related open systems. The KBIs can be best understood after multiplication by a number density, $\rho_j = N_j/V$, to give an "excess" coordination number as follows,

$$N_{ij} \equiv \rho_j G_{ij} = 4\pi\rho_j \int_0^R g_{ij}(r)\, r^2 dr - \rho_j V = \langle N_j \rangle_i - \langle N_j \rangle_0 \tag{1.16}$$

where V is the (spherical) volume of the μVT system out to a radius, R, and the angular brackets represent a μVT ensemble average. Therefore, N_{ij} quantifies the change in the number of j molecules surrounding a central i molecule $\left(\langle N_j \rangle_i\right)$ from the number of j molecules found in the same range of space before insertion of the i molecule $\left(\langle N_j \rangle_0\right)$. The coordination number can change because the i molecule occupies space, and the interactions between other molecules can

be changed both directly and indirectly due to the presence of the *i* molecule. In the theory the volume is extended to infinity to obtain the numerical values for the KBIs. From simulation one can determine the KBIs as a function of integration distance, *R*. An example of this is provided in Figure 1.6.

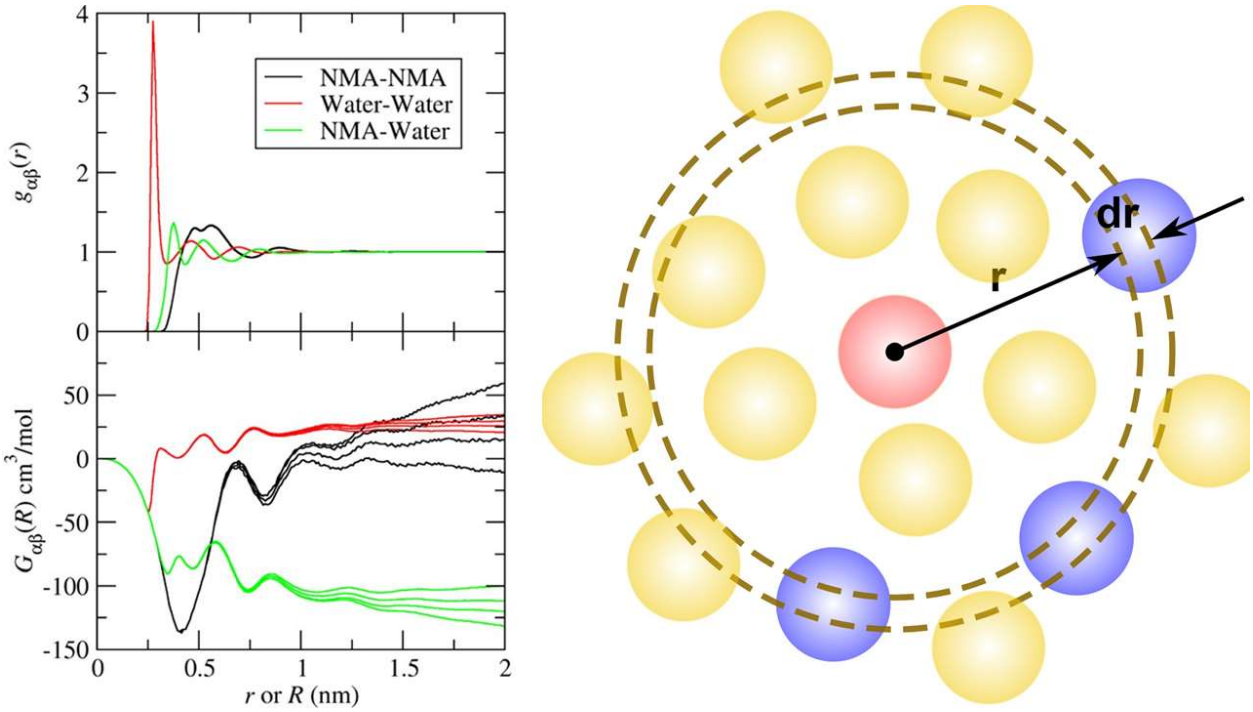


Figure 1.6 The top left panel of the figure presents RDFs derived from simulations where the mole fraction of N-methylacetamide (xNMA) is 0.1. The bottom left, one can observe the depiction of the distance-dependent KBIs, also corresponding to an xNMA of 0.1. The illustrations of the RDFs and the KBIs have been reprinted with permission from Ploetz, E. A.; Karunaweera, S.; Bentenitis, N.; Chen, F.; Dai, S.; Gee, M. B.; Jiao, Y. F.; Kang, Y. S.; Kariyawasam, N. L.; Naleem, N.; et al. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. Journal of Chemical Theory and Computation 2021, 17 (5), 2964-2990. DOI: 10.1021/acs.jctc.1c00075. Copyright (2021) American Chemical Society.[47] The diagram on the right adopted from Wikipedia represents RDF.[61]

The KBIs can also be expressed in terms of the molecule number fluctuations in the GCE as given by, Kirkwood and Buff.[62]

$$\rho_i\left(\delta_{ij} + \rho_j G_{ij}\right) = \frac{\langle \delta N_i \delta N_j \rangle}{V} \qquad 1.17$$

where $\delta_{ij} = 0$ when $i \neq j$, and $\delta_{ij} = 1$ when $i = j$, and $\delta N_i = N_i - \langle N_i \rangle$ represents a fluctuation in the number of i molecules. In turn, the fluctuations are related to chemical potential derivatives in the GCE through,

$$\frac{\langle \delta N_i \delta N_j \rangle}{V} = \frac{1}{V}\left(\frac{\partial \langle N_j \rangle}{\partial \beta \mu_i}\right)_{\beta,V,\{\beta\mu\}'} \tag{1.18}$$

where $\beta = 1/kT$. Consequently, the chemical potential derivatives in the GCE are related to the corresponding KBIs. The major accomplishment of KB theory was to manipulate the above derivatives, valid for open systems, to alternative thermodynamic derivatives valid for equivalent closed systems more commonly representative of experimental systems.

$$\mu_i = \mu_i^o + RTln(\gamma_i \rho_i) \tag{1.19}$$

In Equation 1.19, the chemical potential μ_i is expressed in terms of the standard chemical potential μ_i^o, the activity coefficient, γ_i and the number density ρ_i. This relation becomes essential when considering real solutions, as the activity coefficient γ_i accounts for deviations from ideal behavior that often occur in such systems. Furthermore, this equation connects the gap between the molecular level properties represented by the Kirkwood-Buff integrals and the macroscopic thermodynamic properties. Following this, the chemical potential will play a key role in understanding how concentration changes in a μVT ensemble, as will be demonstrated.

For a simple binary mixture, KB theory can be simply derived without complex matrix manipulations as we now outline. Temperature remains constant throughout. The number densities in the μVT ensemble only depend on the chemical potentials at constant temperature. Hence, one can write the following differential,

$$d\rho_2 = \frac{1}{V}\left(\frac{\partial \langle N_2 \rangle}{\partial \beta \mu_1}\right)_{\beta,V,\{\beta\mu\}'} d\beta\mu_1 + \frac{1}{V}\left(\frac{\partial \langle N_2 \rangle}{\partial \beta \mu_2}\right)_{\beta,V,\{\beta\mu\}'} d\beta\mu_2 + \cdots \tag{1.20}$$

for a series of components 1, 2, etc. This can be simplified using Equations 1.18 and 1.19 to give,

$$dln\rho_2 = \rho_1 G_{21} d\beta\mu_1 + (1 + \rho_2 G_{22}) d\beta\mu_2 + \cdots \quad 1.21$$

We will also require the Gibbs-Duhem (GD) equation at constant T,

$$dP = \rho_1 d\mu_1 + \rho_2 d\mu_2 + \cdots \quad 1.22$$

For a two component solution consisting of a solvent 1 and a solute 2 there will be three unique KBI values, G_{11}, G_{22}, and $G_{12} = G_{21}$ that characterize the mixture.

When T and P are constant Equation 1.22 provides a link between changes in the two chemical potentials that can be used to eliminate the solvent chemical potential from Equation 1.21. Taking a derivative with respect to $ln\rho_2$ then provides the following equation,

$$\left(\frac{\partial \beta\mu_2}{\partial ln\rho_2}\right)_{P,T} = \frac{1}{1 + \rho_2(G_{22} - G_{12})} \quad 1.23$$

upon rearrangement. Consequently, the change in the solute chemical potential as a function of solute concentration is given by two KBIs, G_{22} and G_{12}, and the solute number density (molarity), where the former characterizes the solute-solute distribution the later characterizes the solute-solvent distribution. Changes in the solute molar activity coefficient with solute concentration can be obtained by additional modification of Equation 1.23 as follows,

$$\left(\frac{\partial \beta\mu_2}{\partial ln\rho_2}\right)_{P,T} = 1 + \left(\frac{\partial ln\gamma_2}{\partial ln\rho_2}\right)_{P,T} \quad 1.24$$

and therefore,

$$\left(\frac{\partial ln\gamma_2}{\partial ln\rho_2}\right)_{P,T} = -\frac{\rho_2(G_{22} - G_{12})}{1 + \rho_2(G_{22} - G_{12})} \quad 1.25$$

Thus, the equation demonstrates the relationship between γ_2 and concentration ρ_2, where γ_2 decreases when $G_{22} > G_{12}$, i.e. solute self association is greater that solute solvation, and vice versa.

If we take the derivative of Equation 1.21 with respect to pressure with T and composition constant, then the isothermal compressibility κ_T of the solution mixture can be obtained as follows,

$$RT\kappa_T = RT\left(\frac{\partial ln\rho_2}{\partial P}\right)_{T,\{N\}} = \rho_1 G_{12}\bar{V}_1 + (1+\rho_2 G_{22})\bar{V}_2 \quad 1.26$$

where $\bar{V}_1$ is partial molar volume (PMV) of the solvent and $\bar{V}_2$ represents PMV of the solute. Under the same conditions the GD equation provides,

$$1 = \rho_1\bar{V}_1 + \rho_2\bar{V}_2 \quad 1.27$$

Using this in Equation 1.26 provides,

$$\bar{V}_2 = \frac{RT\kappa_T - G_{12}}{1+\rho_2(G_{22}-G_{12})} \quad 1.28$$

and,

$$\bar{V}_1 = \frac{RT\kappa_T - G_{12}}{1+\rho_1(G_{11}-G_{12})} \quad 1.29$$

Hence, one can write,

$$\frac{\bar{V}_2}{\bar{V}_1} = \frac{1+\rho_1 G_{11} - \rho_1 G_{21}}{1+\rho_2 G_{22} - \rho_2 G_{12}} \quad 1.30$$

Solving for the solute PMV using Equation 1.27 provides,

$$\bar{V}_2 = \frac{1+\rho_1 G_{11} - \rho_1 G_{21}}{\rho_2(1+\rho_1 G_{11} - \rho_1 G_{21}) + \rho_1(1+\rho_2 G_{22} - \rho_2 G_{12})} \quad 1.31$$

and,

$$\bar{V}_1 = \frac{1+\rho_2 G_{22} - \rho_2 G_{21}}{\rho_2(1+\rho_1 G_{11} - \rho_1 G_{21}) + \rho_1(1+\rho_2 G_{22} - \rho_2 G_{12})} \quad 1.32$$

for the solvent PMV. The denominator is normally abbreviated and simplified as η,

$$\eta = [\rho_2(1+\rho_1 G_{11} - \rho_1 G_{21}) + \rho_1(1+\rho_2 G_{22} - \rho_2 G_{12})] \quad 1.33$$

Finally, using Equations 1.28 and 1.29 in Equation 1.26 provides κ_T according to,

$$RT\kappa_T = \frac{(1+\rho_1 G_{11})(1+\rho_2 G_{22}) - \rho_1\rho_2 G_{12}^2}{\rho_2(1+\rho_1 G_{11} - \rho_1 G_{21}) + \rho_1(1+\rho_2 G_{22} - \rho_2 G_{12})} \quad 1.34$$

Hence, we have demonstrated that the three KBIs in a binary solution can be related to the three thermodynamic quantities: derivatives of chemical potentials, the partial molar volumes, and the isothermal compressibility.

Although KB theory was established a while ago, the application of the theory in real life problems came some 20 years after it was published. There was very little information at that time about the nature of the distribution functions that form a central part of KB theory. The first significant breakthrough for KB theory was 1972 when the KB theory was recognized to be useful for understanding certain properties of water and aqueous solutions.[63] A major breakthrough occurred in 1978 with the KB inversion approach of Ben-Naim.[64] Here, instead of using the KBIs to describe the corresponding thermodynamic properties, he instead used the experimental thermodynamic properties to extract the experimental KBIs. This then provides a variety of information regarding the mixing behavior of solutions. For a binary solution mixture the KBIs can be extracted from the experimental data as follows,

$$\delta_{ij} + \rho_j G_{ij} = \rho_j RT\kappa_T + \frac{1}{x_1 x_2}\frac{x_j}{x_i}\frac{(1-\rho_i\bar{V}_i)(1-\rho_j\bar{V}_j)}{\left(\partial\beta\mu_i/\partial x_j\right)_{T,P}} \quad 1.35$$

where x_j represents the mole fraction of species j.

When studying solution mixtures it is often convenient to compare with properties expected for an ideal mixture.[65] For an ideal binary solution on the mole fraction scale the ideal KBIs are given by,

$$G_{ij} = RT\kappa_T - V_i^0 - V_j^0 + \rho_i (V_i^0)^2 + \rho_j \left(V_j^0\right)^2 \quad 1.36$$

$$G_{11} + G_{22} - 2G_{12} = 0 \quad 1.37$$

where the zero superscript indicates pure liquid values. Hence, the KBIs provide valuable quantitative information concerning the behavior of both ideal and real solution mixtures.

1.5 KBFF+KBFF20

The KB theory is an exact theory of solution mixtures which has no approximation and no free parameters except for the fact that the system is in the thermodynamic limit. KB theory allows direct application to systems in the condensed phase. The ability to quantify the balance between solute-solute and solute-solvent distributions in the condensed phase provides an opportunity to evaluate and improve force fields for the simulation of condensed phases. Indeed, several models for small solute molecules have been provided for better balance between the solute-solute and solute-solvent mixture interactions.[47, 66, 67] The Kirkwood–Buff-derived force fields (KBFF) have been developed in an effort to better understand and model solution mixture interactions. In these studies water is usually the solvent employed,[47] although methanol has also been used for low solubility solutes.[66] The main focus of the model development is an improvement of the effective partial atomic charges used to mimic the polarity of the solutes, as it is difficult to establish accurate effective charge distribution for use in the condensed phase due to significant polarization effects, especially in polar solvent such as water. Traditionally, gas-phase QM calculations are used to obtain partial atomic charges for the CHARMM FFs,[39] while an enhanced dipole moment approach has been utilized in the AMBER FFs.[44] However, these methods do not accurately describe polarization effects on the solute (and solvent). Polarizable FFs can provide an explicit description of polarization in solution, which appears to be an improvement, but the computational cost is very expensive. Thus, fixed-charge or non-polarizable FFs remain the most common FFs to study biological systems. Hence, there is still the desire to develop non-polarizable FFs with improved effective partial atomic charges.

Recently, the KBFF20 model for peptides and proteins was published which has focused on the development of a FF from a series of small solute molecules representative of the backbone and side-chain functional groups found in amino acids.[47, 59] In addition, the development of new ϕ/ψ and χ potentials, not amenable to study through the KBIs, were tested to reproduce the conformational properties of polypeptide chains.[59] The use of KB theory, and the KBIs in particular, appears to be a general, accurate, and viable approach for the development of new FFs and the improvement of existing FFs.

1.6 Amino acid

More than 300 amino acids (AA) are found in nature even though only 20 of them (α-amino acid) are used as building blocks of peptides and proteins. AAs are organic substances including both an amino and acid group in close proximity. All αAAs are chiral molecules and present optical activity except for glycine. All AAs have a primary amino group and a carboxyl group connected to the α-carbon atom except for proline. Alternatively, the amino group links to the β-carbon atom for some amino acids such as taurine and β-alanine. The configuration of the AAs are labelled as either L- or D- isomers, which are defined with reference to glyceraldehydes. AAs play a pivotal role in biological systems, not only serving as the building blocks of proteins, but also acting as intermediates in metabolism, neurotransmitters, and even as osmolytes.[68] These osmolytes can help maintain cell volume and integrity under various stress conditions. The behavior of individual AAs in solution differs significantly from their behavior in the polymeric form, such as when incorporated into peptides or proteins. In aqueous solutions, AAs typically exist in their zwitterionic form, featuring both a positively charged amino group ($-NH_3^+$) and a negatively charged carboxyl group ($-COO^-$). This dual nature contributes to their versatile roles in biological systems.

The solvation of biological molecules is an important aspect to understanding their structure, dynamics and functionality. Understanding the details of the solute-solute and solute-solvent interactions is central to their physicochemical properties. The properties of aqueous AA solutions are often studied by simulation to help understand protein solvation behavior due to its simplicity of structure. Miller *et al*.[69] simulated the osmotic coefficient of AAs with various FFs and different water models and concluded that the simulation of proteins have the common problem of overestimating the strength of solute-solute interactions. The results from the paper demonstrate how the sensitivity of osmotic coefficients to the force field parameters assigned to zwitterions led to a reparameterization of amino acid interactions. Specific attention was given to the interactions of sodium ions with the side chain carboxylate and the backbone carboxylate groups of Glutamate (Glu). This attention to detail resulted in a notable enhancement of the simulation accuracy, evidenced by improved osmotic coefficients and diminished Na-Glu clustering. These findings underscore the importance of accurately representing interactions with both side chain and backbone carboxylate groups in such simulations. Another study by Qiu *et al*.[70] suggested that scaling the LJ parameter between AA atoms and water oxygen improves the simulation results for the hydration of IDPs. Many studies have been performed to help understand protein solution properties. However, since AAs themselves are important biological molecules, and their properties differ from that when incorporated in a polypeptide chain, different parameters need to be developed for the simulation study of zwitterionic AA, and the FF parameters need to be tested to reproduce properties more specific to AAs themselves.

1.7 Sodium dodecyl sulfate (SDS): membrane-mimetic environment

SDS is an anionic alkyl sulfate amphiphilic molecule containing a long hydrophobic aliphatic twelve carbon chain tail and a hydrophilic sulfate head group as shown in Figure 1.6.[71]

SDS is a surfactant which means it is a surface active agent. Thus, SDS is not highly soluble in either polar or organic solvents because of its amphiphilic properties. The hydrophilic sulfate head group would prefer to be surrounded by a polar solvent, whereas the hydrophobic tail would prefer to be in contact with an organic solvent. Thus, SDS molecules often prefer regions between two immiscible phases, such as the boundary between a liquid and a gaseous phase, or between two liquids, or between a liquid and a solid.[72]

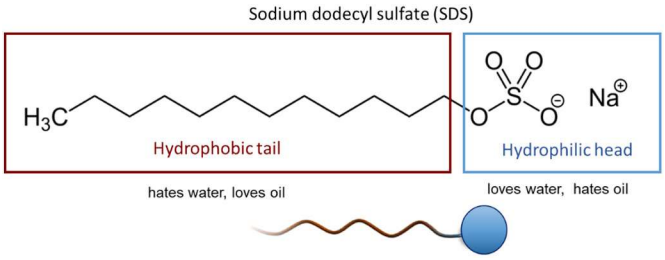


Figure 1.7 Amphiphilic chemical structure of SDS. The figure is reprinted with permission by Wolowicz, A.; Staszak, K. Study of surface properties of aqueous solutions of sodium dodecyl sulfate in the presence of hydrochloric acid and heavy metal ions. Journal of Molecular Liquids 2020, 299. DOI: 10.1016/j.molliq.2019.112170. This is an open access article distributed under the terms of the Creative Commons CC-BY license (2020).[71]

The presence of SDS adsorbed at an interface leads to a reduction in the surface tension between the phases.[73] The surface tension is an important parameter for a number of physical phenomena: such as adsorption, wetting, catalysis, distillation and much more, with direct relevance for the behavior of industrial products in coatings, food, detergents, cosmetics and so on.[74] Surface tension is defined as the free energy required to create a unit area between two phases. Many surfactants play a critical role in reducing the surface tension of water at the air/water interface from about 70 mN m^{-1} to about 25-40 mN m^{-1}.[75] When micromolar concentrations of strong surfactants are mixed with water, the air/water interface is occupied by surfactant monomers, where the hydrophilic head orientates towards water and the hydrophobic chains towards air.[76] This behavior leads to a decrease in the surface tension due to the increase in surfactant at the interface, i.e. surface adsorption. After the surfactant concentration is high enough

that the surface is fully saturated with surfactant, known as the maximum packing condition, then aggregation proceeds in the bulk solution leading to the formation of spheroidal aggregates, known as micelles.[77] The concentration at which aggregation arises is called the critical micelle concentration, generally known as the CMC ,which is typically determined by calculation of the inflection point in the surface tension as a function of SDS concentration.

Various characteristics of the structure of surface adsorbed and micelle forming surfactants have been investigated, including shape of charged surface of micelles, the thickness of water/hydrocarbon interfaces, and the configuration of SDS head and tail groups at interfaces etc.[74] This structural information has been examined mainly with scattering techniques which provide information concerning the size, shape and charge of micelles. Generally, air is considered as the hydrophobic phase, but the oil/water interface has also been studied.[78]

SDS is often used as a membrane memetic to study interaction with biological molecules such as proteins due to its micelle formation property. In particular, several high-resolution X-ray crystallography and solution NMR peptide structures have relied heavily on the use of detergents because membrane proteins may be solubilized without denaturation into a nonbiological state.[79] In addition, conformation changes of IDP at different interfaces (lipid/water) have been observed in the presence of SDS as lipid bilayer mimicking environment.[80] Clearly, SDS has interesting effects on the structure of peptides and proteins. These effects should be amenable to simulation studies if a reliable FF for SDS exists or can be developed, and one that is consistent with the FFs used to study peptides and proteins.

1.8 Amylin: Intrinsically disordered protein (IDP)

Intrinsically disordered proteins (IDP) do not have a stable structure in isolation. One of the more important IDPs is Human Islet Amyloid Polypeptide (hIAPP, or amylin) which is a

peptide hormone with 37 residues.[81] The most potent form has a disulfide bond between cysteine residues C2 and C7, and the C-terminus is amidated. Amylin is co-secreted with insulin from pancreatic β-cells. Under normal physiological condition,[82] hIAPP is associated with appetite suppression and involved in the maintenance of glycemic levels. In pathological environments, such as type II diabetes, it is found as insoluble fibrillar aggregates which are known as amyloid fibrils. Cytotoxicity of amylin is due to its interaction with cell membrane.[83] Despite some controversy, it appears that intermediate states, or oligomeric species, tend to be cytotoxic, not the mature amyloid fibrils. Interestingly, the amylin peptides of rodents do not exhibit amyloid formation, whereas amylin from many species, such as human, cat, non-human primate, do exhibit formation of amyloid fibrils.[84] When amylin interacts with lipids or SDS micelles, it forms α-helical like structures according to NMR studies.[85] However, it is known that the amyloid state involves β-sheet structures.[85] Consequently, it is thought that fibril formation must require conversion from a helical to a β-sheet type of structure, catalyzed by surface interactions, which can then aggregate.[86] Unfortunately, no experimental data concerning the three dimensional structure of amyloid fibrils formed from full length amylin are available. However, a small fragment (residues 22 to 27) of the amylin sequence, NFGAIL, has been shown to form fibrils morphologically similar to those formed by full-length amylin.[87]

It is difficult to study fibril formation with full length amylin both experimentally and computationally for different reasons. Many computational studies have been hampered by FF issues observed in simulations of IDPs as mentioned in previous chapter.[53] In addition to the FF problems, the low experimental concentrations of peptide required to achieve fibril formation, and the long times required for fibril formation, are difficult to realize in simulation. Thus, a different approach is necessary. It is clear that there are conformation changes in amylin at different

environments, and that several alternative amylin sequences do not display amyloid formation. As fibril formation is known to be initiated at the interface between a membrane and water by forming α-helix conformations,[88] it appears informative to investigate the role of different interfaces, such as vacuum/water and SDS/water, on the conformation of amylin and amylin mutants.

1.9 Organization of the dissertation

In this thesis, we will discuss the development of new force field (FF) parameters using KB theory to study the thermodynamics and behavior of AAs and SDS. This will be followed by tests of the new SDS force field, as well as further simulation studies to elucidate the behavior of Amylin and Amylin mutants at different interfaces.

In Chapter 2 we discuss the parameterization of force field parameters for models of zwitterionic amino acids in aqueous solution based on Kirkwood-Buff theory. Although the recent update of KBFF20 has provided all the parameters for the terminal nitrogen and oxygen charge of protein chains, these parameters may not be fully optimized for zwitterionic amino acid due to close presence of both charged groups. Thus, the parameters are optimized to reproduce the experimental Kirkwood-Buff integrals calculated from the thermodynamic properties of solution mixtures. Additionally, the diffusion coefficients, the dielectric constants, the surface tension are obtained from simulation to help validate the FF parameters.

In Chapter 3, we present the development of force field parameters for SDS through the parameterization of the sulfate ion and methyl/ethyl sulfate ions mixtures in water. An existing parameter set for the sulfate ion suffered from aggregation issues, which necessitated the creation of new parameters. Specifically, the non-bonded parameters, including the charge on the oxygen atom and the Lennard-Jones (LJ) parameters determining the size of the terminal oxygen, were optimized to address the aggregation problems in aqueous solution. This optimization process also

involved applying the KB theory. To facilitate the development of the SDS force field, we leveraged the parameters previously developed for the sulfate ion to describe the SDS head group. This allowed us to examine the transferability of the parameters from the sulfate ion to the methyl/ethyl sulfate ions for the non-bonded interactions. These parameters were rigorously tested during the development process to ensure their applicability and accuracy. By addressing the aggregation issues and refining the non-bonded parameters, we established a robust force field for SDS, which is crucial for accurately simulating its behavior and interactions in various environments.

In Chapter 4, we investigate the conformational behavior of peptides in aqueous solution, both in the absence and presence of Sodium Dodecyl Sulfate (SDS). Peptides have the ability to adopt diverse conformations influenced by their sequence and surrounding environment. By comparing the conformational behavior of peptides under different conditions, specifically in the absence and presence of SDS, we aim to build confidence in the accuracy of the SDS FF, and also gain insights into the stability and propensity of peptides to adopt specific secondary structures, such as helices, sheets, or random coils. This analysis allows for a deeper understanding of the relationship between environmental factors and conformational preferences.

In Chapter 5, we investigate the conformational changes of amylin, an IDP, at various interfaces: including vacuum/water, SDS/water, and lipid/water interfaces. Despite amylin being recognized as an IDP, there have been observations of fibrilization induced by the lipid membrane. Additionally, previous studies have reported a structural transition, such as the formation of α-helices in the presence of SDS micelles, as indicated by NMR structures. To explore the influence of different environments, we employ newly developed force fields and conduct simulations of amylin at diverse interfaces, aiming to observe and elucidate the effects of these distinct

environments on its conformation and behavior. By studying amylin at these interfaces, we hope to enhance our understanding of its structural dynamics and the interplay between amylin and various environments.

In Chapter 6, our primary focus is to conduct a mutation study of amylin, specifically targeting the amyloidogenic region within the hIAPP sequence (NFGAIL mutations). Building upon the findings from Chapter 5, where we examined the conformational changes of wild type hIAPP at various interfaces. In this chapter we will investigate mutated forms of hIAPP. Using molecular dynamics simulations, we will investigate the behavior of pramlintide, rat/mice amylin, and deamidated amylin at the air/water interface, as well as in the presence and absence of SDS micelles. These mutations have been studied for their distinct behaviors and conformational structures at different interfaces compared to hIAPP. Through these simulations, our aim is to understand how these alterations impact the conformational dynamics and interactions of amylin at various interfaces.

1.10 Reference

(1) Bolcer, J. D.; Hermann, R. B. *The Development of Computational Chemistry in the United States.*; 1994.

(2) Pritchard, H. O. Computational chemistry in the 1950s. *Journal of Molecular Graphics & Modelling* **2001**, *19* (6), 623-627.

(3) Jensen, F. Computational Chemistry: The Exciting Opportunities and the Boring Details. *Israel Journal of Chemistry* **2022**, *62* (1-2). DOI: 10.1002/ijch.202100027.

(4) Vanommeslaeghe, K.; Guvench, O.; MacKerell, A. D. Molecular Mechanics. *Current Pharmaceutical Design* **2014**, *20* (20), 3281-3292. DOI: 10.2174/13816128113199990600.

(5) Alder, B. J.; Wainwright, T. E. Phase Transition For A Hard Sphere System. *Journal of Chemical Physics* **1957**, *27* (5), 1208-1209. DOI: 10.1063/1.1743957.

(6) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267* (5612), 585-590. DOI: 10.1038/267585a0.

(7) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Accounts of Chemical Research* **2017**, *50* (3), 539-543. DOI: 10.1021/acs.accounts.6b00532.

(8) Malola, S.; Hakkinen, H. Prospects and challenges for computer simulations of monolayer-protected metal clusters. *Nature Communications* **2021**, *12* (1), Editorial Material. DOI: 10.1038/s41467-021-22545-x.

(9) Zhang, X.; Zheng, P.; Ma, Y.; Jiang, Y.; Li, H. Atomic-scale understanding of oxidation mechanisms of materials by computational approaches: A review. *Materials & Design* **2022**, *217*, Review. DOI: 10.1016/j.matdes.2022.110605.

(10) Kohn, W.; Sham, L. J. Sefl-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140* (4A).

(11) Kanungo, B.; Rufus, N. D.; Gavini, V. Efficient All-Electron Time-Dependent Density Functional Theory Calculations Using an Enriched Finite Element Basis. *Journal of chemical theory and computation* **2023**. DOI: 10.1021/acs.jctc.2c01052.

(12) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophysical Journal* **2007**, *92* (12), 4289-4303. DOI: 10.1529/biophysj.106.094425.

(13) Tozzini, V. Coarse-grained models for proteins. *Current Opinion in Structural Biology* **2005**, *15* (2), 144-150. DOI: 10.1016/j.sbi.2005.02.005.

(14) Frankel, D. *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier, 2002.

(15) Leach, A. R. *Molecular Modelling: principles and applications*; Henry Ling LTd, 2001.

(16) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford Science Publications, 1987.

(17) George, A.; Mondal, S.; Purnaprajna, M.; Athri, P. Review of Electrostatic Force Calculation Methods and Their Acceleration in Molecular Dynamics Packages Using Graphics Processors. *Acs Omega* **2022**, *7* (37), 32877-32896. DOI: 10.1021/acsomega.2c03189.

(18) Arnold, A.; Fahrenberger, F.; Holm, C.; Lenz, O.; Bolten, M.; Dachsel, H.; Halver, R.; Kabadshow, I.; Gahler, F.; Heber, F.; et al. Comparison of scalable fast methods for long-range interactions. *Physical Review E* **2013**, *88* (6). DOI: 10.1103/PhysRevE.88.063308.

(19) Liu, Y. L.; Hu, C. J.; Zhao, C. C. Efficient parallel implementation of Ewald summation in molecular dynamics simulations on multi-core platforms. *Computer Physics Communications* **2011**, *182* (5), 1111-1119. DOI: 10.1016/j.cpc.2011.01.007.

(20) Yang, S. C.; Wang, Y. L.; Jiao, G. S.; Qian, H. J.; Lu, Z. Y. Accelerating electrostatic interaction calculations with graphical processing units based on new developments of ewald method using non-uniform fast fourier transform. *Journal of Computational Chemistry* **2016**, *37* (3), 378-387. DOI: 10.1002/jcc.24250.

(21) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008**, *4* (3), 435-447. DOI: 10.1021/ct700301q.

(22) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26* (16), 1701-1718. DOI: 10.1002/jcc.20291.

(23) Fennell, C. J.; Gezelter, J. D. Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics. *Journal of Chemical Physics* **2006**, *124* (23). DOI: 10.1063/1.2206581.

(24) Sagui, C.; Darden, T. A. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Annual Review of Biophysics and Biomolecular Structure* **1999**, *28*, 155-179. DOI: 10.1146/annurev.biophys.28.1.155.

(25) Zheng, L.; Alhossary, A. A.; Kwoh, C.-K.; Mu, Y. Molecular Dynamics and Simulation. In *Encyclopedia of Bioinformatics and Computational Biology*, Academic Press, 2019; pp 550-566.

(26) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **1984**, *81* (8), 3684-3690. DOI: 10.1063/1.448118.

(27) Goga, N.; Rzepiela, A. J.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C. Efficient Algorithms for Langevin and DPD Dynamics. *Journal of Chemical Theory and Computation* **2012**, *8* (10), 3637-3649. DOI: 10.1021/ct3000876.

(28) Nose, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics* **1984**, *52* (2), 255-268. DOI: 10.1080/00268978400101201.

(29) Hoover, W. G. Canonical Dynamics -Equilibrium Phase-Space Distributions. *Physical Review A* **1985**, *31* (3), 1695-1697. DOI: 10.1103/PhysRevA.31.1695.

(30) Berendsen, H. J. C. Transport properties computed by linear response through weak coupling to a bath. *Computer Simulations in Material Science. Meyer, M., Pontikis, V. eds. Kluwer.* **1991**, 139–155.

(31) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - A New Molecular-Dynamics Method. **1981**, *52* (12), 7182-7190. DOI: 10.1063/1.328693.

(32) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72* (2384).

(33) Kraka, E.; Zou, W. L.; Tao, Y. W. Decoding chemical information from vibrational spectroscopy data: Local vibrational mode theory. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2020**, *10* (5). DOI: 10.1002/wcms.1480.

(34) Zhu, X.; Lopes, P. E. M.; MacKerell, A. D. Recent developments and applications of the CHARMM force fields. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2012**, *2* (1), 167-185. DOI: 10.1002/wcms.74.

(35) Carvalho, A. T. P.; Barrozo, A.; Doron, D.; Kilshtain, A. V.; Major, D. T.; Kamerlin, S. C. L. Challenges in computational studies of enzyme structure, function and dynamics. *Journal of Molecular Graphics & Modelling* **2014**, *54*, 62-79. DOI: 10.1016/j.jmgm.2014.09.003.

(36) Dauber-Osguthorpe, P.; Hagler, A. T. Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? *Journal of Computer-Aided Molecular Design* **2019**, *33* (2), 133-203. DOI: 10.1007/s10822-018-0111-4.

(37) Guvench, O.; MacKerell, A. D. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In *Molecular Modeling of Proteins. Methods Molecular Biology.*, Kukol, A. Ed.; Vol. 443; Humana Press., 2008.

(38) Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *Journal of Chemical Information and Modeling* **2018**, *58* (3), 565-578, Review. DOI: 10.1021/acs.jcim.8b00042.

(39) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690. DOI: 10.1002/jcc.21367.

(40) van Gunsteren, W. F.; Daura, X.; Hansen, N.; Mark, A. E.; Oostenbrink, C.; Riniker, S.; Smith, L. J. Validation of Molecular Simulation: An Overview of Issues. *Angewandte Chemie-International Edition* **2018**, *57* (4), 884-902. DOI: 10.1002/anie.201702945.

(41) Steinhauser, M. O.; Hiermaier, S. A Review of Computational Methods in Materials Science: Examples from Shock-Wave and Polymer Physics. *International Journal of Molecular Sciences* **2009**, *10* (12), 5135-5216. DOI: 10.3390/ijms10125135.

(42) Chen, C. X.; Depa, P.; Maranas, J. K.; Sakai, V. G. Comparison of explicit atom, united atom, and coarse-grained simulations of poly (methyl methacrylate). *Journal of Chemical Physics* **2008**, *128* (12). DOI: 10.1063/1.2833545.

(43) Qi, Y. F.; Cheng, X.; Han, W.; Jo, S.; Schulten, K.; Im, W. CHARMM-GUI PACE CG Builder for Solution, Micelle, and Bilayer Coarse-Grained Simulations. *Journal of Chemical Information and Modeling* **2014**, *54* (3), 1003-1009. DOI: 10.1021/ci500007n.

(44) He, X. B.; Man, V. H.; Yang, W.; Lee, T. S.; Wang, J. M. A fast and high-quality charge model for the next generation general AMBER force field. *Journal of Chemical Physics* **2020**, *153* (11). DOI: 10.1063/5.0019056.

(45) Loschwitz, J.; Olubiyi, O. O.; Hub, J. S.; Strodel, B.; Poojari, C. S. Computer simulations of protein-membrane systems. *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly* **2020**, *170*, 273-403. DOI: 10.1016/bs.pmbts.2020.01.001.

(46) Oliveira, M. P.; Goncalves, Y. M. H.; Gheta, S. K. O.; Rieder, S. R.; Horta, B. A. C.; Hunenberger, P. H. Comparison of the United- and All-Atom Representations of (Halo)alkanes Based on Two Condensed-Phase Force Fields Optimized against the Same Experimental Data Set. *Journal of Chemical Theory and Computation* **2022**, *18* (11), 6757-6778. DOI: 10.1021/acs.jctc.2c00524.

(47) Ploetz, E. A.; Karunaweera, S.; Bentenitis, N.; Chen, F.; Dai, S.; Gee, M. B.; Jiao, Y. F.; Kang, Y. S.; Kariyawasam, N. L.; Naleem, N.; et al. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 2964-2990. DOI: 10.1021/acs.jctc.1c00075.

(48) Beu, T. A.; Ailenei, A. E.; Costinas, R. I. Martini Force Field for Protonated Polyethyleneimine. *Journal of Computational Chemistry* **2020**, *41* (4), 349-361. DOI: 10.1002/jcc.26110.

(49) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theoretical Chemistry Accounts* **2009**, *124* (1-2), 11-28. DOI: 10.1007/s00214-009-0617-x.

(50) Shi, Y.; Xia, Z.; Zhang, J. J.; Best, R.; Wu, C. J.; Ponder, J. W.; Ren, P. Y. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* **2013**, *9* (9), 4046-4063. DOI: 10.1021/ct4003702.

(51) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996**, *118* (45), 11225-11236. DOI: 10.1021/ja9621760.

(52) Levine, Z. A.; Shea, J. E. Simulations disordered proteins and systems with conformational heterogeneity. *Current Opinion in Structural Biology* **2017**, *43*, 95-103. DOI: 10.1016/j.sbi.2016.11.006.

(53) Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Grater, F. Kirkwood-Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. *Journal of Physical Chemistry B* **2015**, *119* (25), 7975-7984. DOI: 10.1021/acs.jpcb.5b03440.

(54) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. Developing force fields from the microscopic structure of solutions. *Fluid Phase Equilibria* **2010**, *290* (1-2), 43-47. DOI: 10.1016/j.fluid.2009.11.023.

(55) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *Journal of Physical Chemistry B* **2015**, *119* (16), 5113-5123. DOI: 10.1021/jp508971m.

(56) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115* (21), E4758-E4766. DOI: 10.1073/pnas.1800690115.

(57) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmueller, H.; MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14* (1), 71-73. DOI: 10.1038/nmeth.4067.

(58) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. *Journal of Chemical Physics* **1951**, *19* (6), 774-777. DOI: 10.1063/1.1748352.

(59) Ploetz, E. A.; Karunaweera, S.; Smith, P. E. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Applications of KBFF20. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 2991-3009. DOI: 10.1021/acs.jctc.1c00076.

(60) Lbadaoui-Darvas, M.; Takahama, S. Water Activity from Equilibrium Molecular Dynamics Simulations and Kirkwood-Buff Theory. *Journal of Physical Chemistry B* **2019**, *123* (50), 10757-10768. DOI: 10.1021/acs.jpcb.9b06735.

(61) Wikipedia. *Radial Distribution Function*. 2023. https://en.wikipedia.org/wiki/Radial_distribution_function (accessed 2023 August 1st 2023).

(62) Kirkwood, J. G.; Buff, F. P. THE STATISTICAL MECHANICAL THEORY OF SOLUTIONS .1. *Journal of Chemical Physics* **1951**, *19* (6), 774-777. DOI: 10.1063/1.1748352.

(63) Ben-Naim, A. *Molecular Theory of Solution*; Oxford University Press, Inc, 2006.

(64) Ben-Naim, A. Inversion of the Kirkwood-Buff theory of solutions: Application to the water-ethanol system. **1977**, *67* (11), 4884-4890.

(65) Atkins, P. J.; De Paula, J.; Keeler, J. *Physical Chemistry*; Oxford University Press, 2014.

(66) Ploetz, E. A.; Smith, P. E. A Kirkwood-Buff force field for the aromatic amino acids. *Physical Chemistry Chemical Physics* **2011**, *13* (40), 18154-18167. DOI: 10.1039/c1cp21883b.

(67) Naleem, N.; Bentenitis, N.; Smith, P. E. A Kirkwood-Buff derived force field for alkaline earth halide salts. *Journal of Chemical Physics* **2018**, *148* (22). DOI: 10.1063/1.5019454.

(68) Schweikhard, E. S.; Ziegler, C. M. Chapter One - Amino Acid Secondary Transporters: Toward a Common Transport Mechanism. In *Current Topics in Membrane*, Vol. 70; Academic Press, 2012; pp 1-28.

(69) Miller, M. S.; Lay, W. K.; Elcock, A. H. Osmotic Pressure Simulations of Amino Acids and Peptides Highlight Potential Routes to Protein Force Field Parameterization. *Journal of Physical Chemistry B* **2016**, *120* (33), 8217-8229. DOI: 10.1021/acs.jpcb.6b01902.

(70) Qiu, Y. J.; Shan, W. J.; Zhang, H. Y. Force Field Benchmark of Amino Acids. 3. Hydration with Scaled Lennard-Jones Interactions. *Journal of Chemical Information and Modeling* **2021**, *61* (7), 3571-3582. DOI: 10.1021/acs.jcim.1c00339.

(71) Wolowicz, A.; Staszak, K. Study of surface properties of aqueous solutions of sodium dodecyl sulfate in the presence of hydrochloric acid and heavy metal ions. *Journal of Molecular Liquids* **2020**, *299*. DOI: 10.1016/j.molliq.2019.112170.

(72) Aguirre-Ramirez, M.; Silva-Jimenez, H.; Banat, I. M.; De Rienzo, M. A. D. Surfactants: physicochemical interactions with biological macromolecules. *Biotechnology Letters* **2021**, *43* (3), 523-535. DOI: 10.1007/s10529-020-03054-1.

(73) Sharma, T.; Sangwai, J. S. Silica nanofluids in polyacrylamide with and without surfactant: Viscosity, surface tension, and interfacial tension with liquid paraffin. *Journal of Petroleum Science and Engineering* **2017**, *152*, 575-585. DOI: 10.1016/j.petrol.2017.01.039.

(74) Rosen, M. J.; KunJappu, J. T. *Surfactant and Interfacial Phenomena*; John Wiley & Sons, Inc., 2012.

(75) Wnetrzak, A.; Chachaj-Brekiesz, A.; Kobierski, J.; Karwowska, K.; Petelska, A. D.; Dynarowicz-Latka, P. Unusual Behavior of the Bipolar Molecule 25-Hydroxycholesterol at the Air/Water Interface-Langmuir Monolayer Approach Complemented with Theoretical Calculations. *Journal of Physical Chemistry B* **2020**, *124* (6), 1104-1114. DOI: 10.1021/acs.jpcb.9b10938.

(76) Das, D.; Dash, U.; Meher, J.; Misra, P. K. Improving stability of concentrated coal-water slurry using mixture of a natural and synthetic surfactants. *Fuel Processing Technology* **2013**, *113*, 41-51. DOI: 10.1016/j.fuproc.2013.02.021.

(77) Zhu, Y. K.; Free, M. L.; Woollam, R.; Durnie, W. A review of surfactants as corrosion inhibitors and associated modeling. *Progress in Materials Science* **2017**, *90*, 159-223. DOI: 10.1016/j.pmatsci.2017.07.006.

(78) Staples, E.; Penfold, J.; Tucker, I. Adsorption of mixed surfactants at the oil-water interface. *Journal of Physical Chemistry B* **2000**, *104* (3), 606-614. DOI: 10.1021/jp992955d.

(79) Tulumello, D. V.; Deber, C. M. SDS Micelles as a Membrane-Mimetic Environment for Transmembrane Segments. *Biochemistry* **2009**, *48* (51), 12096-12103. DOI: 10.1021/bi9013819.

(80) Bénarouche, A.; Habchi, J.; Cagna, A.; Maniti, O.; Girard-Egrot, A.; Cavalier, J.-F.; Longhi, S.; Carrière, F. Interfacial Properties of NTAIL, an Intrinsically Disordered Protein. *Biophysical Journal* **2017**, *113* (12), 2723-2735.

(81) Scollo, F.; La Rosa, C. Amyloidogenic Intrinsically Disordered Proteins: New Insights into Their Self-Assembly and Their Interaction with Membranes. *Life-Basel* **2020**, *10* (8). DOI: 10.3390/life10080144.

(82) Bower, R. L.; Hay, D. L. Amylin structure-function relationships and receptor pharmacology: implications for amylin mimetic drug development. *British Journal of Pharmacology* **2016**, *173* (12), 1883-1898. DOI: 10.1111/bph.13496.

(83) Ahmad, E.; Ahmad, A.; Singh, S.; Arshad, M.; Khan, A. H.; Khan, R. H. A mechanistic approach for islet amyloid polypeptide aggregation to develop anti-amyloidogenic agents for type-2 diabetes. *Biochimie* **2011**, *93* (5), 793-805. DOI: 10.1016/j.biochi.2010.12.012.

(84) Akter, R.; Cao, P.; Noor, H.; Ridgway, Z.; Tu, L. H.; Wang, H.; Wong, A. G.; Zhang, X. X.; Abedini, A.; Schmidt, A. M.; et al. Islet Amyloid Polypeptide: Structure, Function, and Pathophysiology. *Journal of Diabetes Research* **2016**, *2016*. DOI: 10.1155/2016/2798269.

(85) Nanga, R. P. R.; Brender, J. R.; Vivekanandan, S.; Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochimica Et Biophysica Acta-Biomembranes* **2011**, *1808* (10), 2337-2342. DOI: 10.1016/j.bbamem.2011.06.012.

(86) Buchanan, L. E.; Dunkelberger, E. B.; Tran, H. Q.; Cheng, P. N.; Chiu, C. C.; Cao, P.; Raleigh, D. P.; de Pablo, J. J.; Nowick, J. S.; Zanni, M. T. Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient beta-sheet. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (48), 19285-19290. DOI: 10.1073/pnas.1314481110.

(87) Madine, J.; Jack, E.; Stockley, P. G.; Radford, S. E.; Serpell, L. C.; Middleton, D. A. Structural Insights into the Polymorphism of Amyloid-Like Fibrils Formed by Region 20-29 of Amylin Revealed by Solid-State NMR and X-ray Fiber Diffraction. *Journal of the American Chemical Society* **2008**, *130* (45), 14990-15001. DOI: 10.1021/ja802483d.

(88) Patil, S. M.; Xu, S.; Sheftic, S. R.; Alexandrescu, A. T. Dynamic alpha-Helix Structure of Micelle-bound Human Amylin. *Journal of Biological Chemistry* **2009**, *284* (18), 11982-11991. DOI: 10.1074/jbc.M809085200.

Chapter 2 – Kirkwood-Buff Derived Force Field for Zwitterionic α-Amino Acids

2.1 Introduction

While the primary function of the twenty standard proteinogenic α-amino acids (AAs) is as building blocks of peptides and proteins, the monomers themselves also have functions, such as being intermediates of metabolism,[1, 2] and acting as neurotransmitters.[3] Thus, it is desirable that a protein force field includes a reliable set of parameters for simulations of zwitterionic (the dominant form at neutral pH) amino acids in aqueous solution. Recently, we published peptide/protein parameters for the Kirkwood-Buff (KB) derived Force Field (KBFF), KBFF20. However, the parameters were not meant to be applied to zwitterionic α-amino acid simulations.[4, 5] Herein, these additional parameters are provided and validated based upon simulations of twelve of the twenty standard α-amino acids that display a significantly high solubility.

The KBFF models have been developed to reproduce experimental Kirkwood-Buff integrals (KBIs).[6-12] KB theory provides a two-way bridge between macroscopic information (select thermodynamic properties) and microscopic information (the pair distributions of components in the solution). From the microscopic perspective, the Kirkwood-Buff integrals (KBIs) are defined by[13]

$$G_{ij} = G_{ji} = 4\pi \int_0^\infty [g_{ij}^{\mu VT}(r) - 1] r^2 dr \qquad 2.1$$

where G_{ij} is the KBI between species i and j, $g_{ij}(r)$ is the radial distribution function (RDF) defined in the open system for species i and j, and r is typically chosen to be the distance between the centers of mass (COM) of molecules i and j. If $G_{ij} > 0$, i and j have a net affinity for each other.

The process of converting experimental bulk thermodynamic data for solution mixtures into KBIs (to compare to those obtained from simulation) is well established.[14] In closed systems, the thermodynamic properties used to calculate the KBIs are the solution activity coefficients, partial molar volumes, and compressibilities.[7, 14, 15] For mixtures under ambient conditions, the thermodynamic property the KBIs depend upon the most is the activity.[16] [13] For closed binary systems, the concentration dependence of the solute activity coefficient can be expressed in terms of the KBIs *via,*[14]

$$\left(\frac{\partial \ln \gamma_s^c}{\partial \ln \rho_s}\right)_{T,p} = -\frac{\rho_s(G_{ss} - G_{ws})}{1 + \rho_s(G_{ss} - G_{ws})} \qquad 2.2$$

where γ_i^c is the molarity (c) based activity coefficient of component i, ρ_i is the number density of i, T is the absolute temperature, and p the pressure. Component w is the solvent and component s is the solute. A force field that reproduces the changes in solute activity with solute concentration will provide a correct description of the microscopic structure of solution mixtures.[17] More importantly, the KBIs are a sensitive measure of the molecular distributions arising from the relative strengths of the solute-solute, solute-solvent, and solvent-solvent interactions.[8, 18]

KB theory is an exact theory of solutions that contains no free parameters. No approximations or assumptions are required besides typical statistical mechanical assumptions, *i.e.* ergodicity and that the system is in the thermodynamic limit. The theory can be applied to any miscible solution with any number of molecular components of any size and complexity. The theory does not assume pairwise additive interactions, although the KBFF models are pairwise additive.

In osmotic systems, changes in the osmotic pressure, Π, with composition also provide access to G_{ss} *via*,[13, 19]

$$\frac{1}{RT}\left(\frac{\partial \Pi}{\partial \rho_s}\right)_{\mu_w} = \frac{1}{1+\rho_s G_{ss}} \qquad 2.3$$

If a simulated G_{ss} is positive and larger than experiment, it may be interpreted as meaning that the change in the osmotic pressure as the concentration of s is increased is too small. Likewise, from the study of binary closed systems, this too large G_{ss} value would mean that the derivative of the natural log of the solute activity coefficient with composition was decreasing too much as the solute concentration increased. That is, both types of analyses (of osmotic pressure or activity coefficients) are informative regarding the interactions between the solute molecules in solution. An advantage of using the closed system approach is that the G_{ws} and G_{ww} are also provided, which are much more statistically reliable as $\rho_s \rightarrow 0$.

The raw experimental data reported in the literature for aqueous amino acid systems are typically the osmotic coefficients, from which the authors often derive and also report activity coefficients. Since Roux's 2010 publication of a MD simulation method to directly calculate the osmotic pressure from analyzing the force a semipermeable wall exerts on a non-permeable solute,[20] several similar studies have been performed on amino acid solutions. Based upon these studies, it has become apparent that protein force fields often lack a reliable parameter set for aqueous zwitterionic amino acids that is able to quantitatively reproduce the experimental degree of solute self-association.

For example, in 2016, Miller *et al.*[21, 22] performed osmotic simulations of sixteen zwitterionic amino acids (*aq*), one zwitterionic amino acid (*aq*) mixture, Glu + Lys (*aq*), and four dipeptides (GG, AA, AG, GA) (*aq*) using the AMBER ff99SB-ILDN + (TIP3P or TIP4P-Ew) force fields. Nine of these systems were also simulated with the TIP4P-D water model instead and with an unpublished version of KBFF+SPC/E, taken from a previous thesis.[23] When using the TIP3P water model, AMBER ff99SB-ILDN gave too much aggregation for all systems. In

contrast, with the TIP4P-Ew water model, it gave good results for the aliphatic AAs but too much AA aggregation for Thr and Ser. Five neutral AAs and the four dipeptides were also tested with the TIP4P-D water model and with the unpublished KBFF+SPC/E force field. With these latter two FFs, AA aggregation was generally reduced except for KBFF+SPC/E simulation of zwitterionic Gly, which was too aggregating. For the dipeptide systems, the authors rederived the AMBER charged termini partial charges to give better agreement with experiment for AMBER ff99SB-ILDN+TIP3P. In a second osmotic simulation study, Miller *et al*.[21, 22] examined the ability of the AMBER ff99SB-ILDN+TIP3P, AMBER ff99SB-ILDN+TIP4P-Ew, CHARMM36+TIP3P, GROMOS54a7+SPC, and OPLSAA+TIP4P force fields to reproduce the experimental osmotic coefficients for seven neutral amino acids in water and for five small molecules in water. Besides GROMOS54a7+SPC and AMBER ff99SB-ILDN+TIP4P-Ew, the force fields produced osmotic coefficients that were typically lower than experiment, which implies overly favorable solute−solute interactions. The authors showed that the adjustment of various LJ parameters, with differing approaches taken for the different force fields, greatly improved the agreement with experiment.

Although the above paragraph highlights examples of how some non-KBFF FFs have had certain parameters adjusted to fix the balance of interactions in selected cases, KBFF is the only united atom FF we are aware of that has used this type of approach as the motivation for the entire FF development and the basis for the parameters for each of the small molecule FF building blocks (unless the experimental data was unavailable for a given building block). The use of KB theory for FF development has several advantages over the more common FF development approaches. Namely, while most FFs rely on gas phase data to develop their condensed phase FFs, KB theory

provides a way of making use of condensed phase experimental mixture data directly in the parameterization process.

Additionally, computer simulation can relatively easily obtain KBIs, because they are calculated from the easily and frequently analyzed RDFs (or coordination numbers). Therefore, the species' distributions in solution and the distribution variation with composition can be quantitatively analyzed. The simulated KBIs can even be broken down into distance-dependent effects, while any information about the system at these molecular length scales is typically unavailable from experiments. Specifically, the Smith group has used KBIs mostly for checking the electrostatic terms, because electrostatic forces are long ranged and constitute a major part of intermolecular interactions, yet are typically the least well-known parameters in the condensed phase.[7] KBIs are known be sensitive to the input parameters of the force field, especially (typically) to the charge distributions. KBIs are also very sensitive to the van der Waals parameters, but these parameters are typically considered more well-established through comparison with the enthalpy of vaporization or the reproduction of crystal lattice parameters, *etc*.

Although direct transfer of the KBFF20 peptide parameters to zwitterionic α-amino acids were reasonable and did not lead to excessive aggregation, the results were not optimal. Specifically, application of KBFF20 to zwitterionic Gly gave a too unfavorable Gly-Gly interaction in aqueous solution (results not shown). For β-Alanine, the direct transfer gave reasonable parameters.[5] These findings suggested that the proximity of the NH_3^+ and CO_2^- in α-amino acids, where they are separated by only one carbon, was the root cause for parameter adjustment. Previously, parameters for Betaine, DL-Alanine, NH_4Cl, NH_4Br, $N(CH_3)_4Cl$, and $N(CH_3)_4Br$ have appeared in theses by the Smith group,[24] but not in the peer-reviewed literature. Aqueous simulations of all twenty AAs have also appeared in a thesis,[23] and publications of KBFF

Glycine have appeared.[19 23, 25] As noted above, other research groups have also made use of these KBFF AA parameters.[21, 22, 26] The present work serves as the first complete KBFF models for the non-Gly zwitterionic AAs, and future KBFF Gly simulations should use the parameters described in this work. Herein, we followed the same approach as earlier investigations into this issue by Gee,[27] where it was first discovered that the charge distribution used for the termini of peptides/proteins required adjustment before use on zwitterionic α-amino acids. The reason we are revisiting the issue that was previously addressed in the thesis of Gee is because of a significant change made to the OT and NT atom types in KBFF20.[5]

In the present study, eight amino acids with neutral side chains, namely, Ala, Gly, Cys, Pro, Ser, Thr, His and Val, and three amino acid salt solutions with positively charged side chain (using Cl^- counterions), namely, ArgH, LysH, and HisH, and one negatively charged side chains (using Na^+ counterions), namely, Glu were studied to optimize the relevant charge distributions. The remaining AAs were not studied due to their low solubility in water. Optimization of the CO_2^- charge was achieved by reproducing the experimental KBIs. Partial molar volumes were also calculated from the KBIs and compared with experiment. Additional properties of aqueous zwitterionic α-amino acids – including the self-diffusion coefficients, relative permittivities, and surface tension changes – were analyzed to further validate the KBFF models.

2.2 Methods

2.2.1 Kirkwood-Buff analysis of experimental data

Kirkwood-Buff inversion is the procedure established by Ben-Naim to extract the KBIs from the bulk experimental thermodynamic data, and comprehensive details about this established procedure have been provided elsewhere.[9, 10, 12, 15, 17, 28-30] We will adopt the notation that w is the solvent (water) and s is the solute (amino acid or amino acid salt for the charged systems). For the

neutral amino acids, this is identical to the usual nomenclature that the solvent is denoted as 1 and the solute (amino acid) as 2; this latter nomenclature will also be used.

The KB inversion procedure involves the analysis of the experimental chemical potential (μ_i) derivatives with respect to composition, the partial molar volumes ($\overline{V_i}$), and the solution isothermal compressibility (κ_T), all as a function of composition. Details of the experimental data used to obtain the experimental KBIs are listed in Table 2.1. By using this data, the three KB integrals (G_{ww}, G_{ss}, and $G_{ws} = G_{sw}$) can be obtained from the following relationships for a two-component mixture[16, 31]

$$\begin{aligned} 1 + \rho_w G_{ww} &= \rho_w RT\kappa_T + \rho_w \phi_s \bar{V}_s/\mu_{ss} \\ 1 + \rho_s G_{ss} &= \rho_s RT\kappa_T + \phi_w^2/\mu_{ss} \\ \rho_s G_{ws} &= \rho_s RT\kappa_T + \phi_w\phi_s/\mu_{ss} \\ \mu_{ss} &= \frac{1}{RT}\left(\frac{\partial \mu_s}{\partial \ln m_s}\right)_{T,P} \end{aligned} \qquad 2.4$$

where R is the gas constant, T is the absolute temperature, and $\phi_i = \rho_i \overline{V_i}$ is the volume fraction of species i.

Amino Acids with Neutral Side Chains

The solution mass density (d) in g/cm^3 and solute activity coefficient on the molal scale (γ_s^m) as a function of amino acid molality (m_s) for the neutral amino acids were obtained from the literature and fitted to the following polynomial equations,

$$\begin{aligned} d &= 0.9970 + a_1 m_s + a_2 m_s^2 + a_3 m_s^3 \\ \ln\gamma_s^m &= b_1 m_s + b_2 m_s^2 + b_3 m_s^3 \end{aligned} \qquad 2.5$$

where 0.9970 g/cm^3 is the mass density of pure water at 298K[32] and the a's and b's are fitting parameters with no physical meaning. Partial molar volumes and activity derivatives for use in the KB inversion equations were then calculated from the above analytic fitted equations. The experimental analysis is relatively insensitive to the exact value of κ_T for most solutions under

ambient conditions.[6] Consequently, κ_T was treated as ideal according to a simple mixing approximation, $\kappa_T \approx \phi_w \kappa^o_{T,w} + \phi_s \kappa^o_{T,s}$, where one only requires the pure (^{o}) values. $\kappa^o_{T,w}$ was taken to be 4.5×10^{-5} bar^{-1} for pure water, and $\kappa^o_{T,s}$ was taken to be zero for the pure amino acid crystals.[7] Under this commonly used approximation, the compressibility functions as a constraint equation for the three KB integrals. Based on testing this approximation on systems for which experimental compressibility data exists, the absence of the true compressibility deviations from ideality does not significantly affect the accuracy of the KBI results.[5]

Table 2.1 Experimental thermodynamic data sources used for Kirkwood-Buff inversion procedure

Solute	Density		V_s° (cm^3/mol)	Activity Coefficient	
	m_s range	Ref.		m_s range	Ref.
Gly	0.000-3.540	[33]	46.616[34]	0.100-3.500	[35] (solid)
				0.100-3.000	[36](dot)
				0.101-3.299	[37](dash)
				0.227-1.140	[38](dot-dash)
Ala	0.000-1.800	[39](DL)	63.555[40](DL)	0.200-1.900	[41] (DL,solid)
				0.224-1.056	[38](DL,dot)
				0.100-1.900	[42] (DL,dash)
				0.101-1.387	[37](L,dot-dash)
Cys	0.000-0.692	[43](L)	84.080[44](DL)	0.300-1.000	[45](L)
Pro	0.250-5.321	[46](L)	l84.100[47](L)	0.200-7.300	[48] (DL)
Ser	0.175-3.694	[49](L, *)	68.462[50](DL)	0.100-4.200	[51] (L,solid)
				0.200-0.500	[48] (L,dot)
				0.054-3.992	[37](DL,dash)
Thr	0.005-0.787	[52](L)	81.533[53] (L)	0.200-2.000	[48] (DL)
His	0.022-0.201	[54](L)	102.075[55](DL)	0.100-0.250	[36](L)
Val			91.500[56] (L)	0.200-0.650	[41] (DL)
HisH			123.812[57] (L)	0.010-0.600	[36](L,solid)
				0.100-0.700	[58] (**, dot)
ArgH			139.295[59] (L)	0.100-5.000	[58] (**)
LysH			132.259[60] (DL)	0.100-3.500	[58] (**)
Glu			92.694[61] (L)	0.100-4.000	[62] (L)

s is the solute (amino acid or amino acid salt). m_s is the molality (mol/kg) of s and V_s° is the volume of pure s. DL indicates DL-AA mixture, and L indicates L-AA for the experimental data. Pure water volume used 18.069 cm^3/mol. Solid, dot, dash, or dot-dash refers to the line style used to display the KBIs based on the different experimental activity datasets used in Figures 2.2-2.3. * density data is reported at 10MPa. ** stereochemistry of AA is not reported. The experimental density data for thr and pro are not available at the same concentration as activity data thus density data are extrapolated to extract experimental KBI. Gly, Ala, Cys, Ser, and His have similar range of data compare to activity data. Val and all charged residues (HisH, ArgH, LysH, Glu) do not have density data.

Amino Acids with Charged Side Chains

Chloride was used as the counterion for the three systems consisting of net positively charged amino acids, and no additional salt was added. For these salt systems, the experimental data was fitted to the following equation,

$$\ln\gamma_2^m = -\frac{A\sqrt{m_s}}{1+c_1\sqrt{m_s}} - \ln(1-c_2 m_s) + c_3 m_s \qquad 2.6$$

where $\gamma_2^m = \gamma_\pm^m$ is the mean molal activity coefficient of the salt, A=0.5115ln(10) is the Debye-Hückel-Onsager parameter at 298K for aqueous 1:1 electrolytes,[63] m_s is the molality of the salt, and the *c*'s are fitting parameters without any specific physical meaning. The first term on the right-hand side of the above equation is a Debye-Hückel term for a 1:1 salt. Density data for the three studied amino acids with charged side chains in solution were not available to as high of a concentration as the literature activity data. Thus, the partial molar volumes in these systems were held constant according to the values in the pure crystals for all compositions and the value of pure water.[64] A set of KBIs from the activity derivatives provides G_{ss}, G_{sw}, and G_{ww}, which represent salt (s) and water (w) molecules when the KB inversion approach is applied.

We assume that amino acid salts dissociate completely into the cation (+) and anion (-), and this is what we generally observe. Thus, we make no attempt to calculate RDFs between "salt molecules." Instead the amino acid salt solution is treated as a binary system of indistinguishable ions (2) and water (1), making use of $\nu N_s = N_2 = N_+ + N_-$, where ν is the total number of ions generated by the dissolution of the salt.[14] Thus, the three KBIs that are presented for the salt systems corresponding to *i–j* equal to indistinguishable ion–indistinguishable ion (22), water–indistinguishable ion (12 or 21), and water–water (11). The constraint of electroneutrality within small volumes of the solution surrounding each species proves additional relationships between

the KBIs for salt molecules or indistinguishable ions and the KBIs for individual cations and anions, but those KBIs are not shown in this work.[7] Details of the procedure can be found elsewhere.[14]

2.2.2 Simulation

All molecular dynamics simulations were performed using modifications to the KBFF20 models (http://kbff.chem.k-state.edu) together with the SPC/E water model as implemented in the GROMACS 2016.4 package[65] using the leap-frog algorithm.[66] Concentration of AAs simulation are varied based on individual solubility. Gly, LysH (3m), Cys, Ala (1m), Ser, Pro, ArgH, Glu (4m), Thr (2m), HisH (0.6m), Val (0.5m), and His (0.25m) were simulated. His was simulated with tautomer mixture. All simulations were performed at 298 K and a pressure of 1 bar. A time-step of 2 fs was used, and the bond lengths (and H-O-H angle for water) were constrained using the Lincs[67] (solutes) and Settle[68] (water) algorithms. The particle mesh Ewald (PME) technique was used to calculate electrostatic[69] (0.12 nm grid spacing) and LJ[69] interactions with a cutoff distance of 1.0 nm for both the real space electrostatic and LJ contributions. We note that geometric combination rules are used in the reciprocal part of the LJ-PME calculation. Because some combination rules are broken in KBFF20,[4] this does introduce very small errors in the forces and energies. The Verlet cut-off scheme was used, with the 'Verlet-buffer-tolerance' parameter set to the Gromacs default of 0.005 kJ/mol/ps/atom. Random initial configurations of molecules in a cubic box (L=10 nm) were generated for all the systems. A test simulation of a racemic mixture of DL-Ser compared to pure L-Ser showed no statistically significant differences in the KBIs; thus, all simulation results presented correspond to pure L-stereoisomers. This is consistent with the simulation-based observations of Miller *et al.* and experimental-based observations of Robinson, both based upon studies of Ala.[21] The steepest descent method was used to perform energy

minimization. The systems were then equilibrated for 100 ps in the *NpT* ensemble. The temperature was fixed at 298K unless otherwise specified using the Berendsen thermostat with the time constant for coupling group $\tau_t = 0.1$ ps. The pressure coupling with a reference pressure of 1 bar was performed applying the Berendsen[70] barostat with time constant for pressure coupling $\tau_P = 0.5$ps. Production simulations were then performed for 100 ns in the *NpT* ensemble The temperature was fixed at 298K using the Nosé-Hoover[71, 72] thermostat with the time constant for coupling group τ_t=0.5 ps. The pressure coupling with a reference pressure of 1 bar was achieved using the Parrinello-Rahman[73, 74] barostat with time constant for pressure coupling $\tau_P = 2.5$ps. The simulation data between 10 to 100 ns were used for calculating ensemble averages. Error bars were calculated as standard deviation from the mean. Configurations were saved every 2 ps for analysis. Self-diffusion constants were calculated by using the mean square fluctuation approach of molecules using 'gmx msd'. Relative permittivities were obtained from the dipole moment fluctuations of the system using 'gmx dipoles'.[75] For salt systems, 'gmx dipoles' subtracts off the net charge on each molecule.

2.2.3 Kirkwood-Buff analysis of the simulation data

Radial distribution functions were obtained from the simulations for each system and composition. The pair RDFs correspond to the 22, 21, and 11 distributions relative to a random distribution after averaging over all other 2 and 1 molecules at that particular composition. The indistinguishable ion treatment for salts provides ion-ion, ion-water, and water-water RDFs and disregards the identity of the ion.[27] KB integrals were determined from the simulation data in closed systems (*NpT* ensemble) even though the KBIs are defined for an open system in the grand canonical ensembles (μVT), according to

$$G_{ij} \approx 4\pi \int_0^R \left[g_{ij}^{NpT}(r) - 1\right] r^2 dr \qquad 2.7$$

where R represents a correlation region within which the solution composition differs from the bulk composition. All RDFs are assumed to be unity beyond R, such that the G_{ij} are converged. For comparison with experiment, a numerical value for the simulated G_{ij} was determined by averaging the integral over a short-range of distances corresponding to approximately the width of one solvation shell.

The KBI values can be used to provide expressions for the three associated thermodynamic properties of a binary solution mixture, as established.[13] In this work, only the $\overline{V_i}$ were calculated in this way. The relevant expressions are,[16]

$$\bar{V}_w = \frac{1 + \rho_s(G_{ss} - G_{ws})}{\eta} \qquad 2.8$$

$$\bar{V}_s = \frac{1 + \rho_w(G_{ww} - G_{ws})}{\eta}$$

where $\eta = \rho_w + \rho_s + \rho_w\rho_s(G_{ww} + G_{ss} - 2G_{ws})$.

2.2.4 Parameter development and analysis

KBFF20 model of zwitterionic AA are represented in Figure 2.1

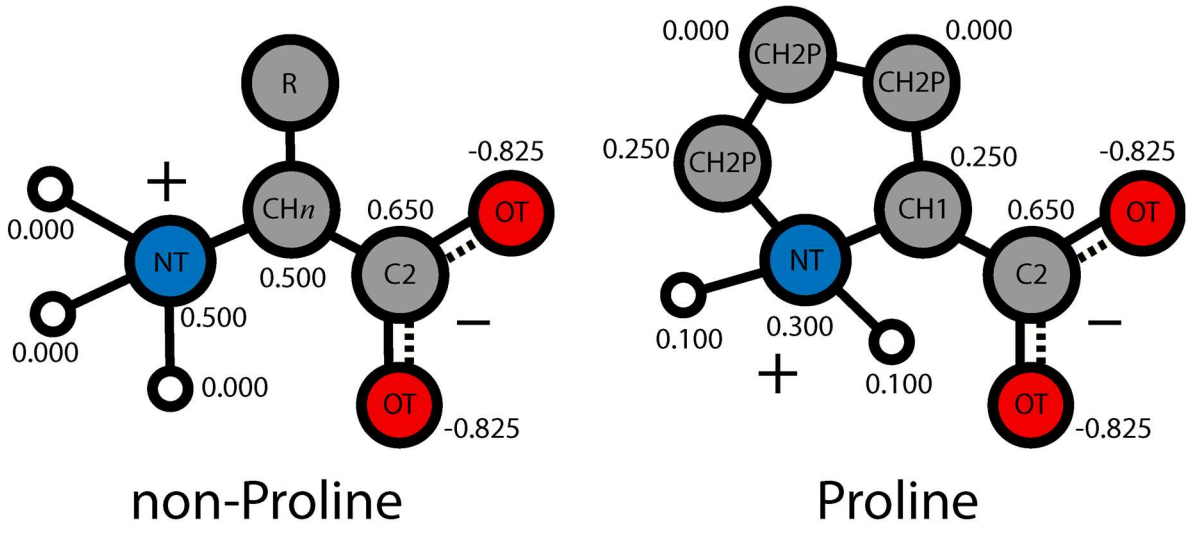


Figure 2.1 Atom types and charges used in building blocks of KBFF20 zwitterionic amino acids. Beads with no charges are only included to orient the reader. CHn refers to atom types with n=1 or 2.

The solutions studied here are composed of zwitterionic amino acids and water molecules. Counter Cl^- ions are additionally included for the charged amino acid salt solutions. All KBFF models in this study contain a simple, classical, non-polarizable, pair-wise additive description for each molecule. As in KBFF20, bond and angle parameters were adopted from GROMOS53a6[76] force field with permission. Dihedral parameters are also the same as would be expected for KBFF20 peptides and proteins.[4] The intermolecular interactions between each pair of atoms *i* and *j* were described by a Lennard-Jones 6-12 plus Coulomb potential, V_{ij}, given by

$$V_{ij} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad 2.9$$

where σ_{ij} and ε_{ij} are the Lennard-Jones diameter and the interaction strength parameters, respectively, which were taken from KBFF20.[4] Geometric combination rules were used for both the σ_{ij} and ε_{ij} parameters. q_i is the partial charge on atom *i*, and all other parameters have their usual meaning. The exclusions and 1-4 scalings are the same as in KBFF20.[4]

The effective partial atomic charges in aqueous solution for the Coulomb potential were parameterized to reproduce the KBIs of binary mixtures. For zwitterionic amino acids, the N-terminus is positively charged and the C-terminus is negatively charged with atom types NT and OT, respectively. Herein, we followed the same approach as the initial Smith group investigations into this issue by Gee,[27] where it was found that a unique charge distribution was required to model zwitterionic AAs, and that the KBIs were more sensitive to the CO_2^- charge distribution than to the NH_3^+ charge distribution. Therefore, following Gee, we focused only on reparametrizing the CO_2^- charge distribution for the present study, with the exception of the charges on the protonated NH_2^+ of Pro, which were also parameterized. The resulting charge distributions are shown in Figure 2.1.

The simulated KBIs were compared with those extracted from the experimental data for the same system over a range of solution compositions. As discussed in the introduction, the simulated KBIs are usually obtained from integration of the center of mass (COM) based radial distribution functions (RDFs), which can be obtained from the simulations. In this work, instead of using the COM, a single atom's coordinates were chosen for each molecule, namely the α carbon of the amino acid residue and the water oxygen. In the thermodynamic limit, the simulated KBI is independent of the choice of the origin.[14]

The surface tension, γ, was calculated from our simulations as,[77]

$$\gamma = \frac{1}{2} L_z \left[\langle P_{zz} \rangle - \frac{1}{2} \left(\langle P_{xx} \rangle + \langle P_{yy} \rangle \right) \right] \qquad 2.10$$

In equation 2.10 L_z is the size of the box along the direction z, perpendicular to the vacuum/water interface. The factor 1/2 outside the bracket considers the fact that there are two interfaces in the system, and P_{xx}, P_{yy}, and P_{zz} are components of the pressure tensor along the x, y, and z direction.

2.3 Results and Discussion

The main goal for the force field development of this study is to reproduce the experimental KBIs for aqueous AA salt solutions as a function of solute concentration. The three unique KBIs for each binary mixture and concentration were calculated to understand and quantify the molecular distributions. The experimental G_{22} for each solution is compared to the simulated results in Figure 2.2. For the neutral AAs, 22 is the AA-AA KBI. For the charged AAs, 22 is the indistinguishable ion-indistinguishable ion KBI. The Gly, Ala, Ser, and the HisH systems have multiple experimental KBI determinations, based on the use of different activity data sets from the literature, which was always combined with the same density data set. The AA salts displayed higher simulated and experimental G_{22} values at lower solute concentrations, indicating high self-

association, due to the known Debye-Huckel behavior of electrolytes. Most of the simulated G_{22} values agree with the experimental values, when taking into account the simulated error bars and the difference in the experimental KBIs when using different literature activity data sets. However, the ArgH simulated G_{22} value is slightly underestimated. The Cys, His, and HisH solutions have positive G_{22} values, indicating there is a strong solute-solute interaction. The G_{22} values for Gly and Ser are positive at lower concentration and become negative at higher concentrations except for one of the experimental determinations for Gly. The KBFF simulations reproduce this trend.

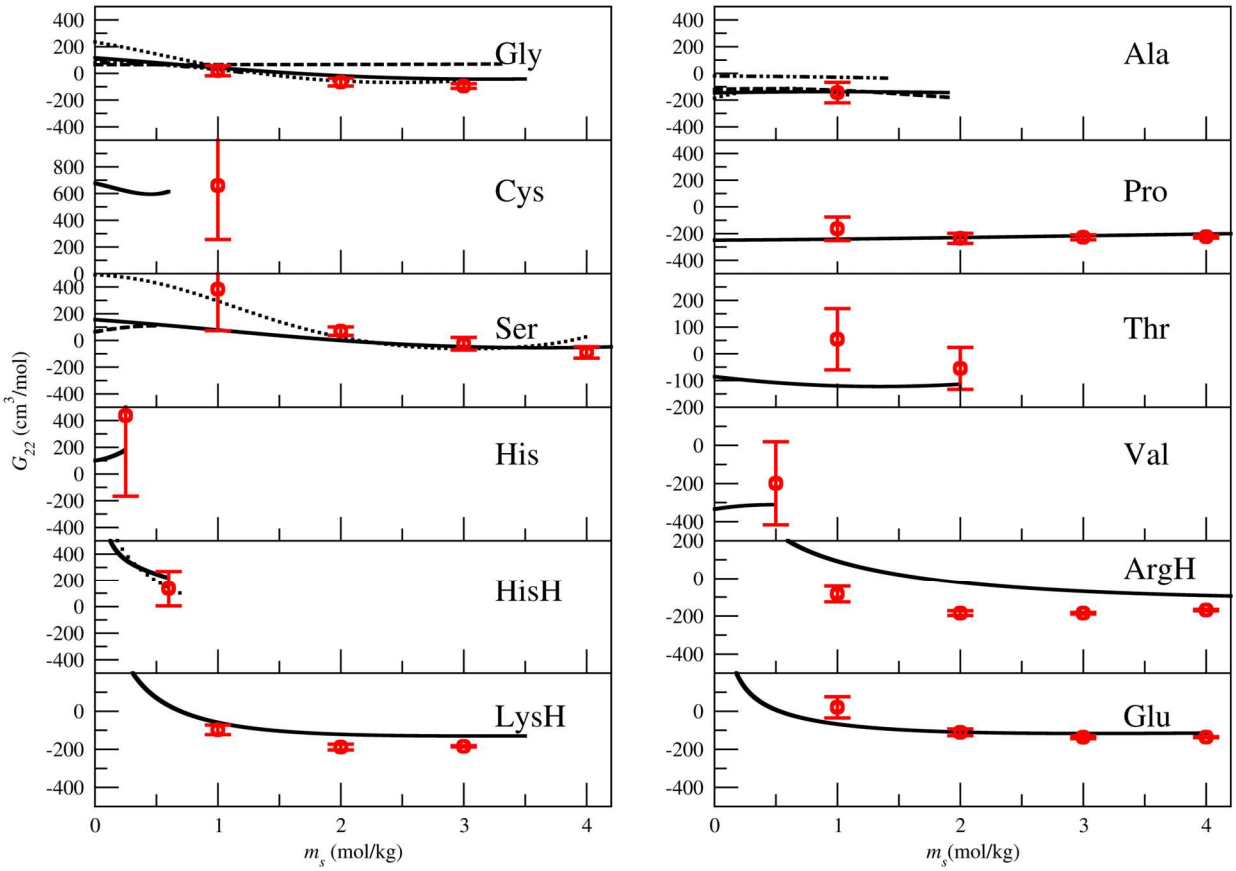


Figure 2.2 G_{22} (cm^3/mol) as a function of solute or salt molality (mol/kg). Experimentally derived values (black solid, dotted, and dashed curves) are compared to simulated values (red points with error bars). For the neutral amino acids, 2 is the AA and the x-axis is the molality. For the charged AAs, 2 is the indistinguishable ion KBI while the x-axis concentration is that of the AA salt.

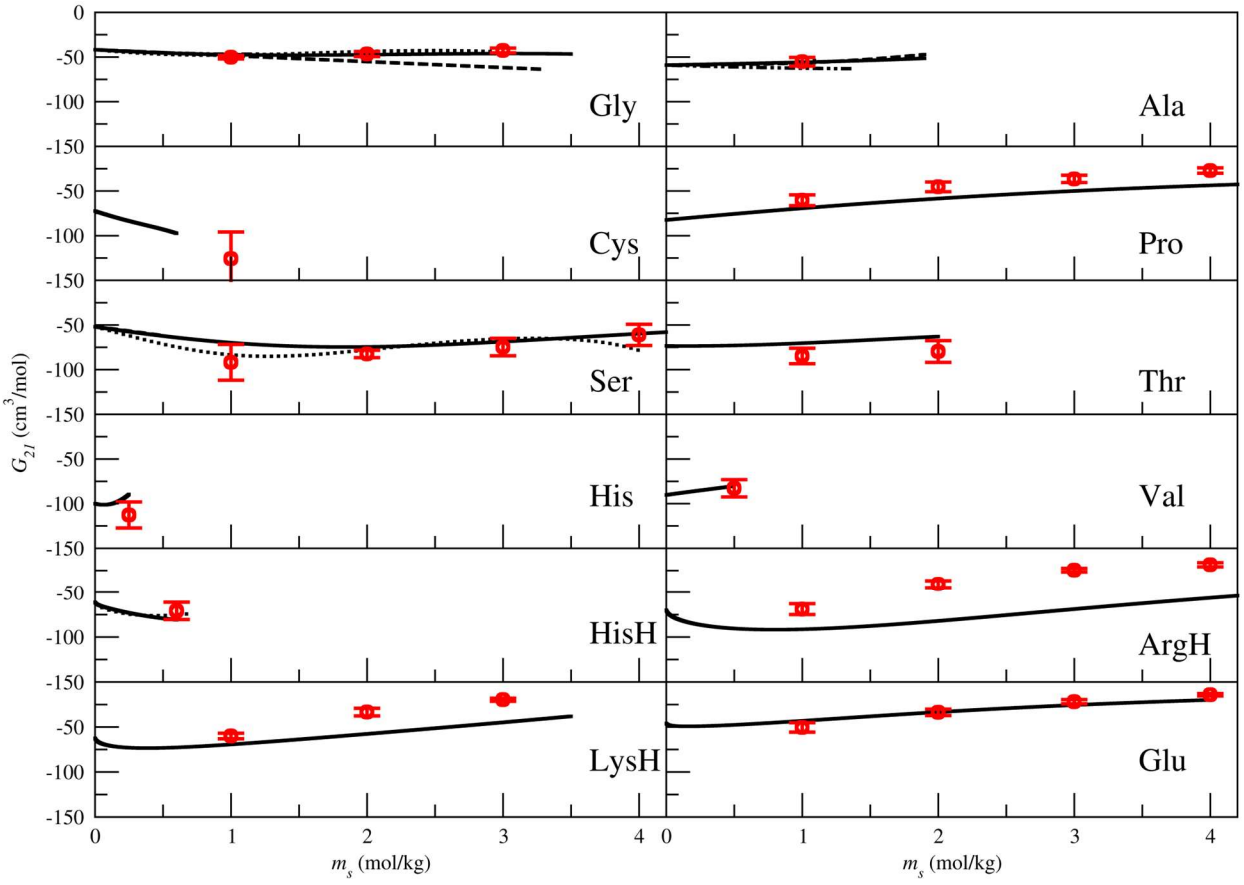


Figure 2.3 G_{12} (cm^3/mol) as a function of solute or salt molality (mol/kg). Experimentally derived values (black solid, dotted, and dashed curves) are compared to simulated values (red points with error bars). For the neutral amino acids, 2 is the AA and the *x*-axis is the AA molality. For the charged AAs, 2 is the indistinguishable ion KBI while the *x*-axis concentration is that of the AA salt.

The experimental and simulated $G_{12} = G_{21}$ (*i.e.*, neutral AA-water or indistinguishable ion-water) KBIs are shown in Figure 2.3. Above 1 *m*, the experimental G_{12} increases as the solute concentration increases for all systems studied and the same trends are exhibited by the simulated values. The simulated G_{12} values for Pro and the ArgH and LysH salt solutions are slightly overestimated compared to experiment, indicating too much solute-solvent mixing.

The simulated and experimental partial molar volumes (PMV) of the solutes and water as a function of solute concentration are displayed in Figures 2.4 and 2.5. The experimental solute PMVs for Gly and Ala aqueous solutions slightly increased as the solute concentration increased, while the experimental solute PMVs for Cys and His solutions decreased as the solute

concentration increased. Experimental solute PMVs for Ser and Thr solutions increased then decreased within the range of concentrations analyzed. The Pro PMV remained nearly constant. Experimental solute PMVs for all AA salt solutions were simply set (fixed) according to the experimental pure crystal densities of the AA salts because the experimental concentration dependent volumetric data was not available in the literature. All the simulated solute PMVs increased monotonically as the solute concentration increased and are in good agreement with experiment, except for the simulated Pro PMV, which is too low. Also, the solute PMV increases as the number of atoms in the solute molecule (and molecular weight) increases for both simulated and experimental values, as expected. The experimental water PMV for Ser and Thr solutions slightly decreased then increased within the range of concentrations studied, while the other non-salt solutions' experimental water PMVs remained nearly constant. For the AA salt solutions, the experimental water PMVs were held constant at the pure water value due to the same reason as noted for the salt PMVs. The simulated water PMV values were in good agreement with experiment except for Ser at 4 *m*.

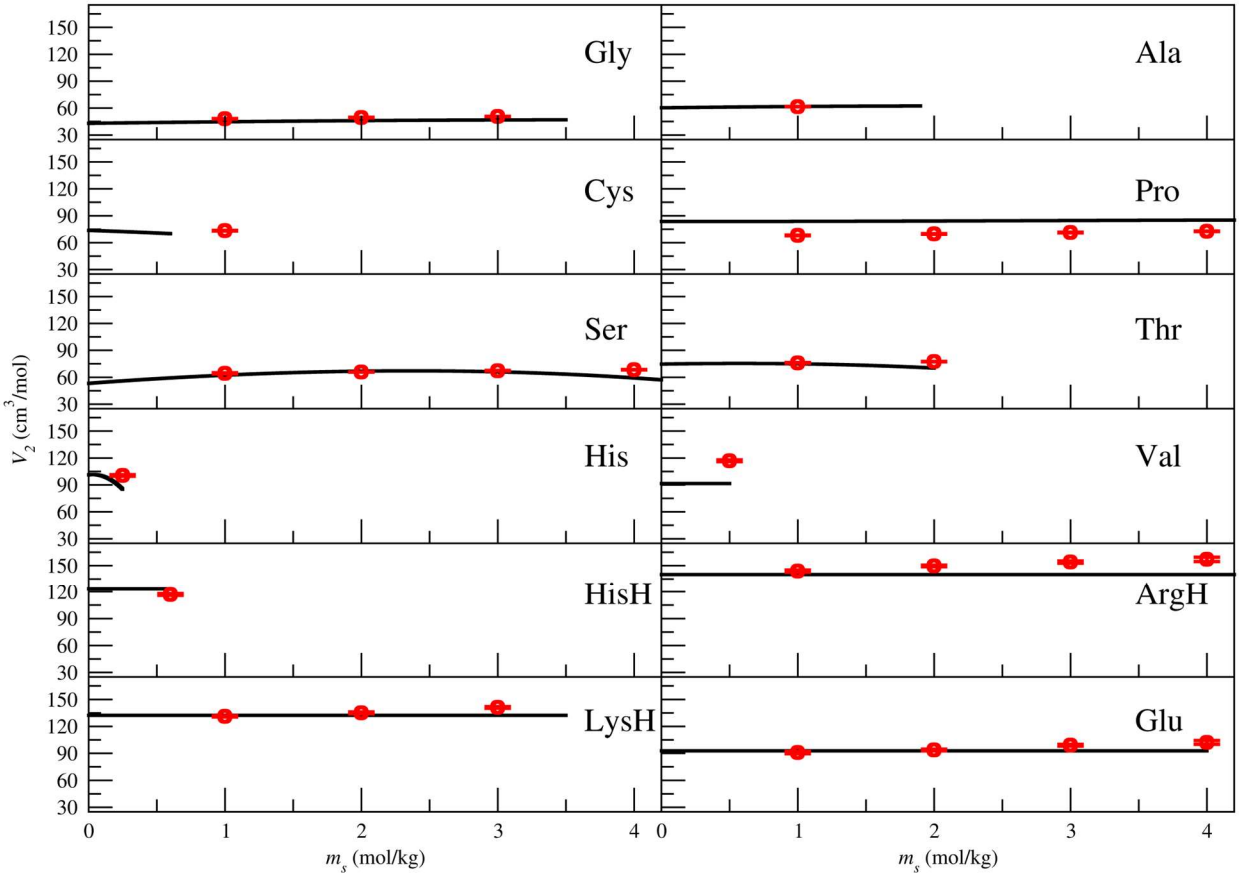


Figure 2.4 Partial molar volume of the solute (cm^3/mol) as a function of the solute or salt molality (mol/kg). Experimentally derived values (black curves) are compared to simulated values (red points with error bars). For the charged AAs, the volume and the x-axis concentration is that of the AA salt.

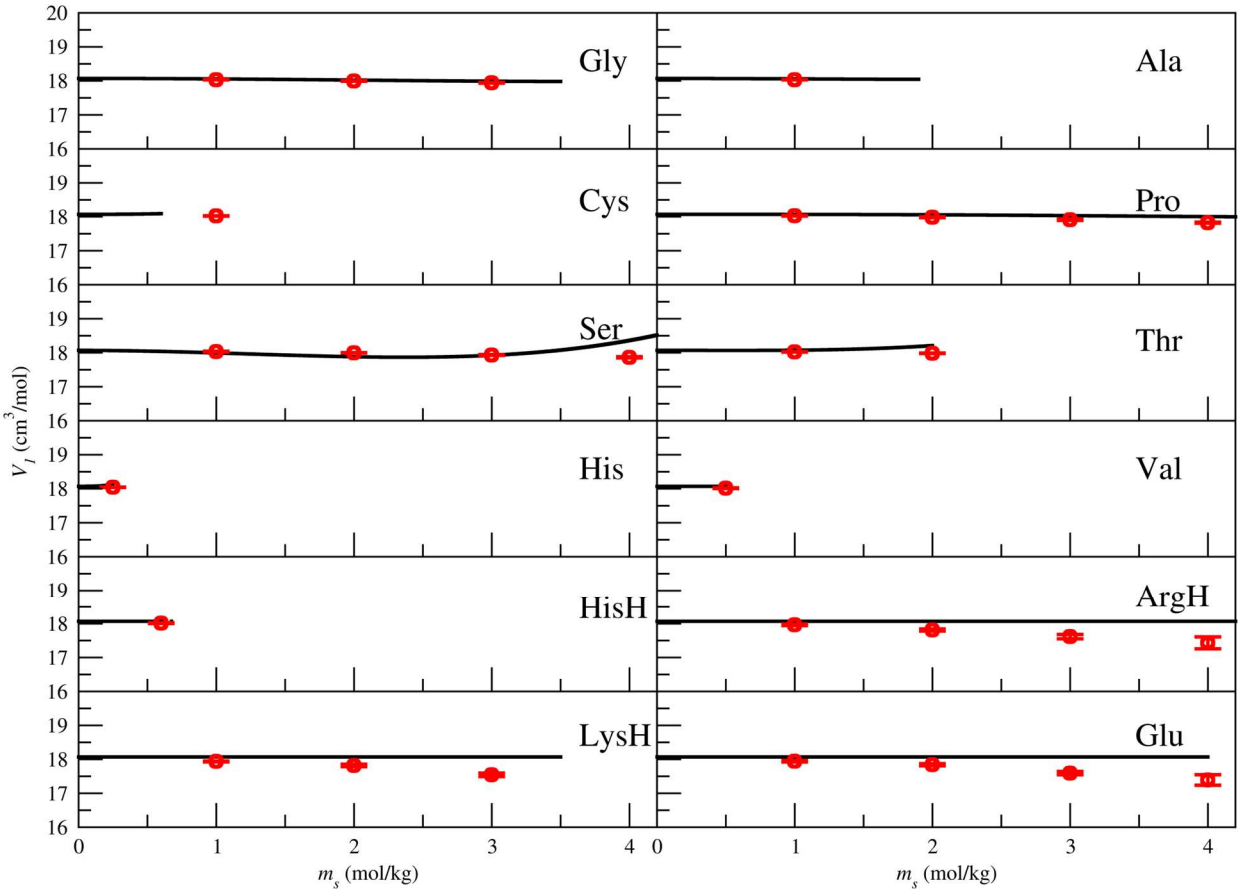


Figure 2.5 Partial molar volume of water (cm^3/mol) as a function of the solute or salt molality (mol/kg). Experimentally derived values (black curves) are compared to simulated values (red points with error bars).

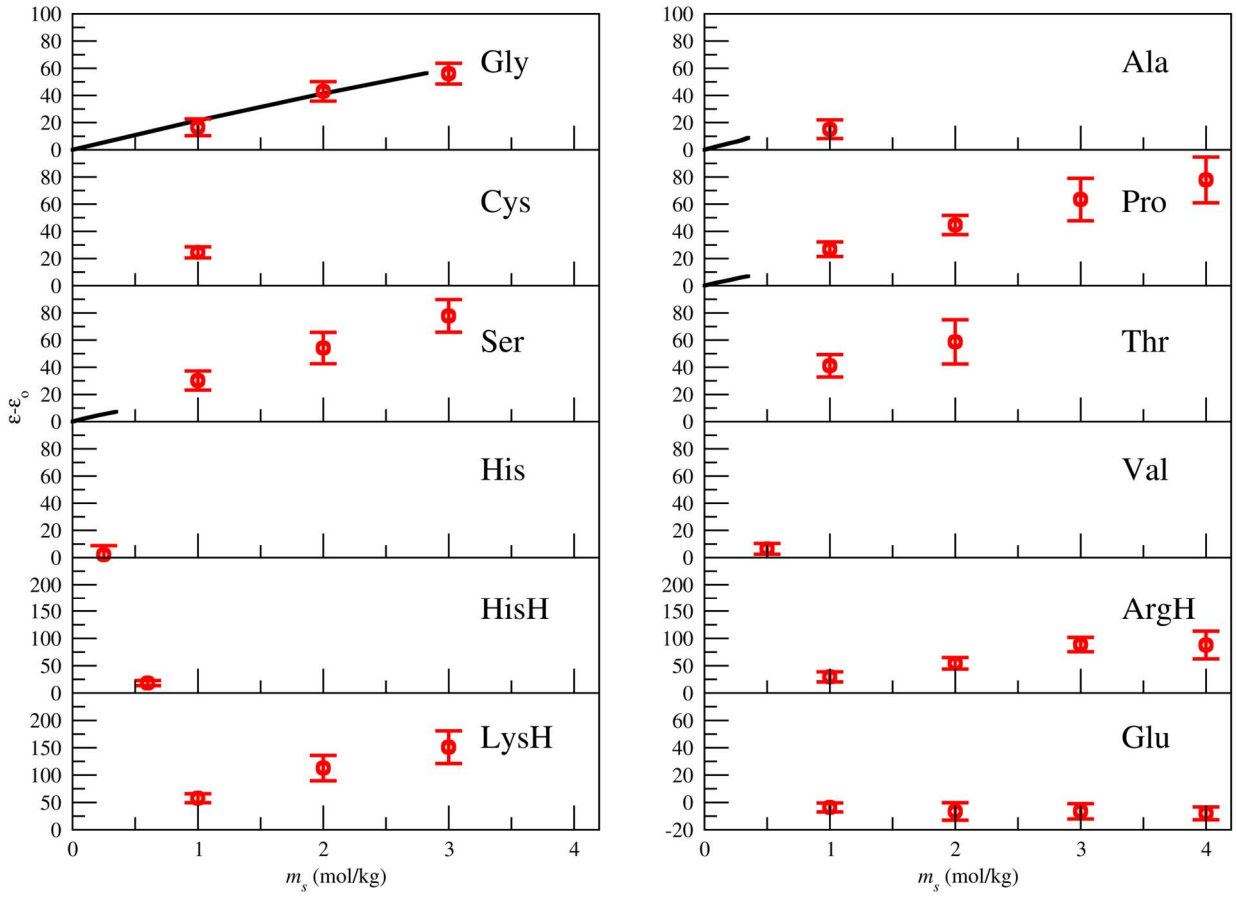


Figure 2.6 Dielectric increment (ε-ε_0) as a function of solute or salt molality (mol/kg). Curves (black) in Gly,[78] Ala,[79] Pro[79], and Ser[79] were obtained from the experimental dielectric constant data while dots with error bars (red) correspond to data obtained from simulation using the KBFF models. For the charged amino acids, the x-axis concentration is that of the AA salt.

The dielectric increments ($\varepsilon - \varepsilon_0$) of AA salts solutions, calculated from the dipole moment fluctuations, are displayed in Figure 2.6. Here, ε is the relative permittivity of the solution, and ε_0 is the relative permittivity of pure water. The value of $\varepsilon_0 = 72.1$ was obtained from a simulation of pure water using the SPC/E model. This value is slightly low compared to the experimental value of 78.30.[80] Hence, a comparison between the experimental and simulated values for the absolute permittivity is inappropriate with this water model. Only four (Gly, Ala, Ser, and Pro) of the studied AAs' experimental relative permittivity values were available in the literature. The experimental relative permittivity for all those solutions increased as a function of molality. This

experimental trend is reproduced by the current KBFF models for all systems. This indicates that the modification of the water structure and dynamics by the AA is correctly reproduced.

The simulated AA self-diffusion constants calculated using the mean square fluctuation approach are displayed in Figure 2.7 as a function of solute molality. Only the Gly and Ala experimental self-diffusion constants are available, and they display an essentially linear decrease with increasing solute molality. All the simulated AA self-diffusion constants decreased with solute concentration. Also, all simulated AA self-diffusion constants decreased as the number of atoms in the solute (and molecular weight) increased. For the systems simulated at multiple compositions, the values were extrapolated to infinite dilution for comparison to the experimental infinite dilution values, as shown in Table 2.2. Overall, the simulated diffusion coefficient values are slightly larger than the experimental values for Gly and Ala at all finite concentrations, indicating that the simulated Gly and Ala are moving faster than they should experimentally. Likewise, all the experimental AA self-diffusion coefficients at infinite dilution are smaller than the extrapolated simulated values except for one of the experimental values for ArgH. Corrections were made in this study to enhance the accuracy of the simulation data. Firstly, adjustments were made to correct for the SPC/E water model, known to produce self-diffusion values lower than the experimental results. Secondly, the system size correction method proposed by Hummer *et al*.[81] was employed. This adjustment, based on hydrodynamic arguments, helps counter the system-size dependency observed in the calculation of translational diffusion coefficients under periodic boundary conditions. The correction compensates for the finite size of the simulated systems, which otherwise leads to an overestimation of self-diffusion coefficients.

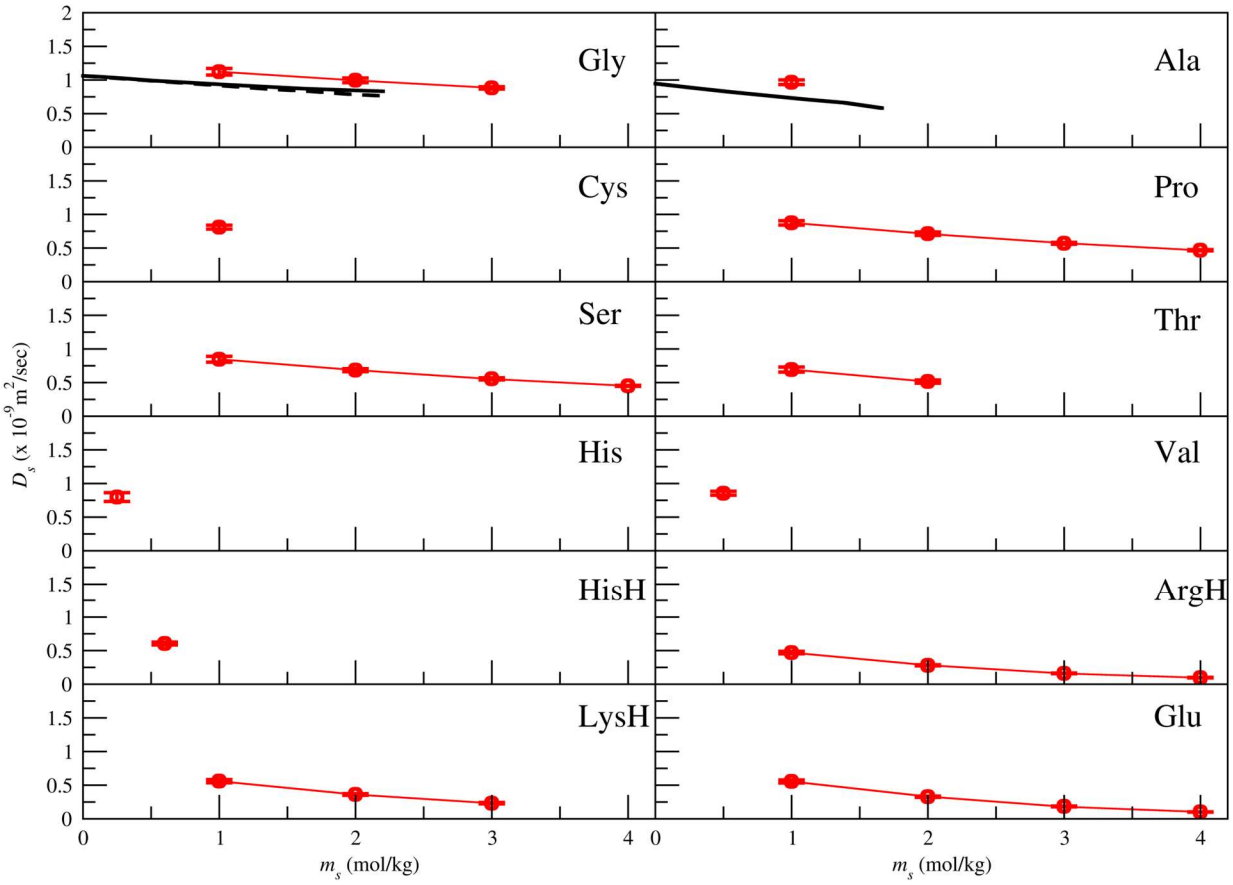


Figure 2.7 Solute self-diffusion constant ($\times 10^{-9}$ m^2/s) as a function of solute or salt molality. The black lines in Gly[82, 83] and Ala[83] are obtained from the experimental solute self-diffusion constant data and the red dots with error bars are obtained from simulation performed using the KBFF model (not corrected for finite size effect or viscosity). For the charged AAs, the x-axis concentration is that of the AA salt.

Table 2.2 Amino acid self-diffusion constant at infinite dilution D_{AA}^{∞}($\times 10^{-9} m^2/s$) at 298 K

Amino Acid	Simulated Data (before correction)	Simulated Data (after correction)	Experimental Data
Ala	-	-	0.95[84], 0.91[85], 0.92[86], 0.92[87]
ArgH	0.72	0.66	0.66[84], 0.73[86], 0.68[87]
Cys	-	-	0.86[87]
Gly	1.27	1.10	1.06[82, 88]
HisH	-	-	0.70[87]
LysH	0.82	0.74	0.66[87]
Pro	1.07	0.94	0.91[84]
Ser	1.04	0.92	0.92[86], 0.89[87]
Thr	0.87	0.78	-
Glu	0.67	0.62	-

Select site-site RDFs obtained from the Gly solution simulations are displayed in Figure 2.8. Interactions between the N of NH_3^+ (NT) and the C of CO_2^- (with OT atom types) are investigated, as well as interactions of the N and the C with the water oxygen (OW). The solute concentration does not have a large effect on the results. No strong interactions between the NT and the OT are observed (as inferred by the lack of a strong N-C first solvation shell), which is consistent with the non-aggregating Gly KBIs. While these same site-site RDFs are available in all the systems studied, they looked so similar to those shown for Gly that they were not replotted for all AA systems. This indicates that the termini are well solvated and there is no strong N-C attraction in all the AA systems. For the other AA systems, additional select site-site RDFs

obtained from the 1 *m* AA/AA salt solution simulations are displayed in Figure 2.9. Side chain-side chain and side chain interactions with the N of NH_3^+ (or NH_2^+ for Pro) and the C of CO_2^- are investigated. Thr and Cys have strong intermolecular side chain-side chain interactions. Ser, Thr, Cys, and His show strong interactions between NT and side chain, while no residues show a strong interaction between the side chain and the C-terminus.

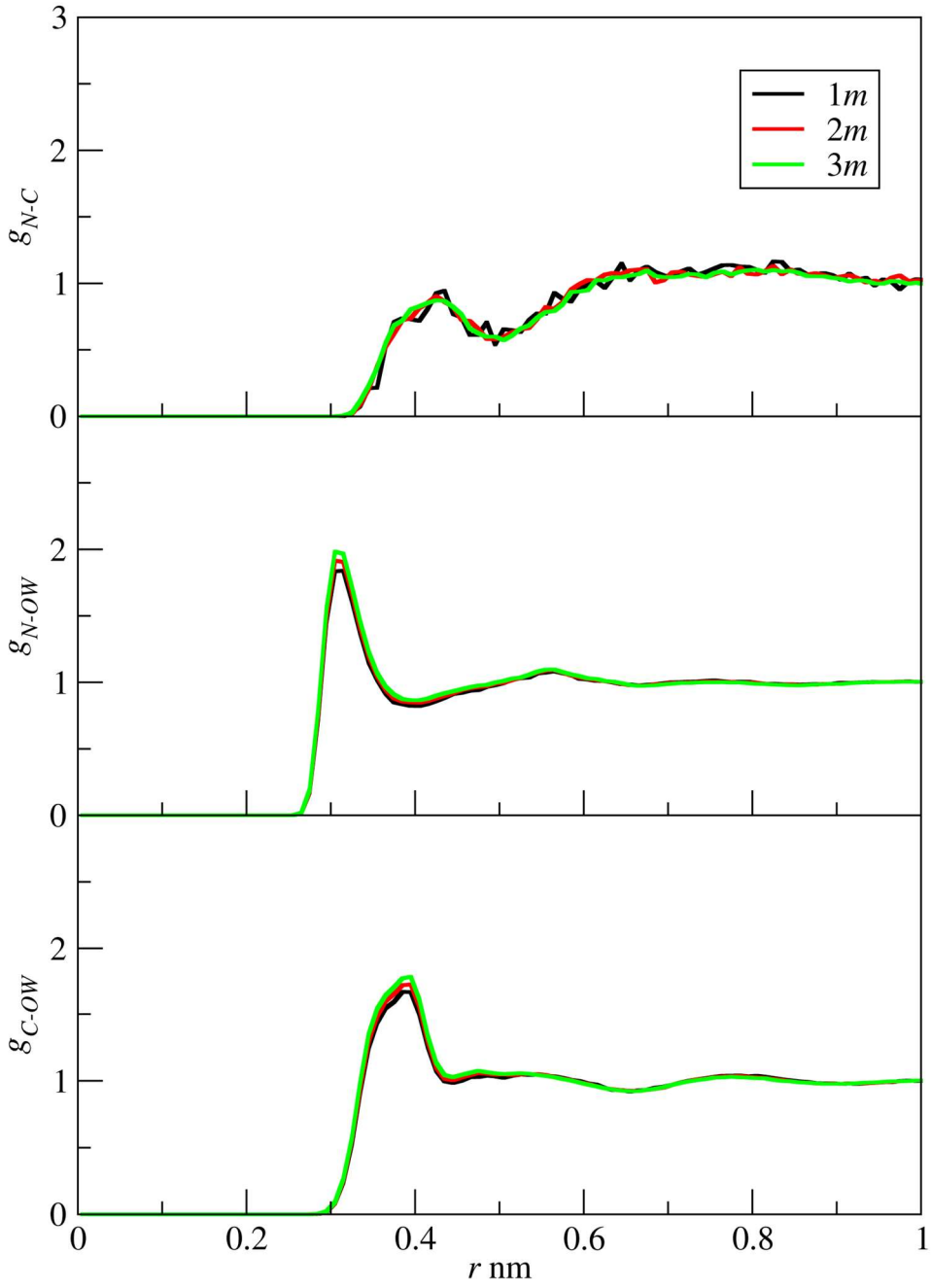


Figure 2.8 Select site-site radial distribution function obtained from simulation of Gly solutions. Interactions between the N of NH_3^+ (NT) and the C of CO_2^- (with OT atom types) are investigated as well as interactions of the N and the C with the water oxygen (OW).

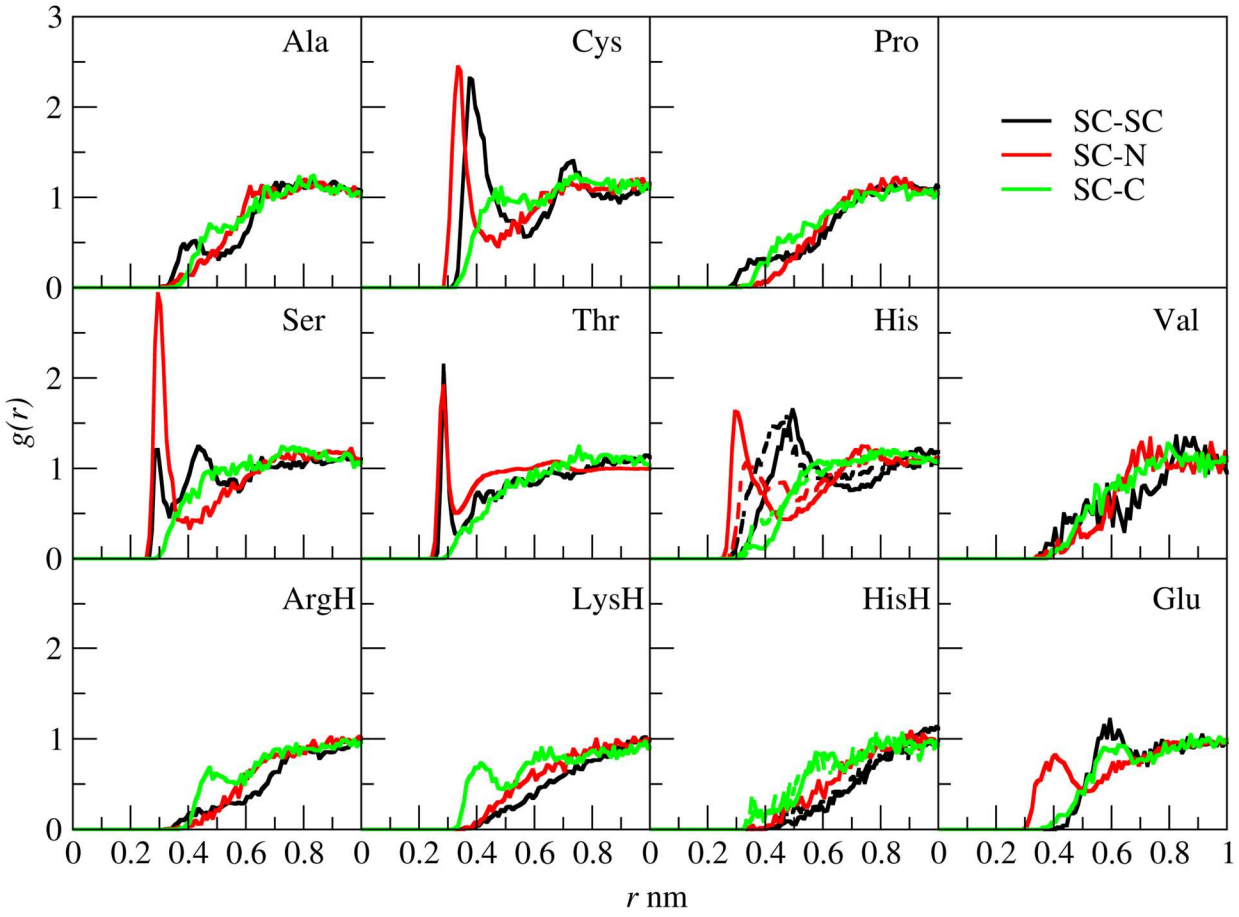


Figure 2.9 Select site-site radial distribution function obtained from simulation of 1m AA/AA salts solution. SC-SC represents side chain-side chain interaction, SC-N represents side chain-NT interaction, and SC-C represent side chain-carboxylate carbon interaction.

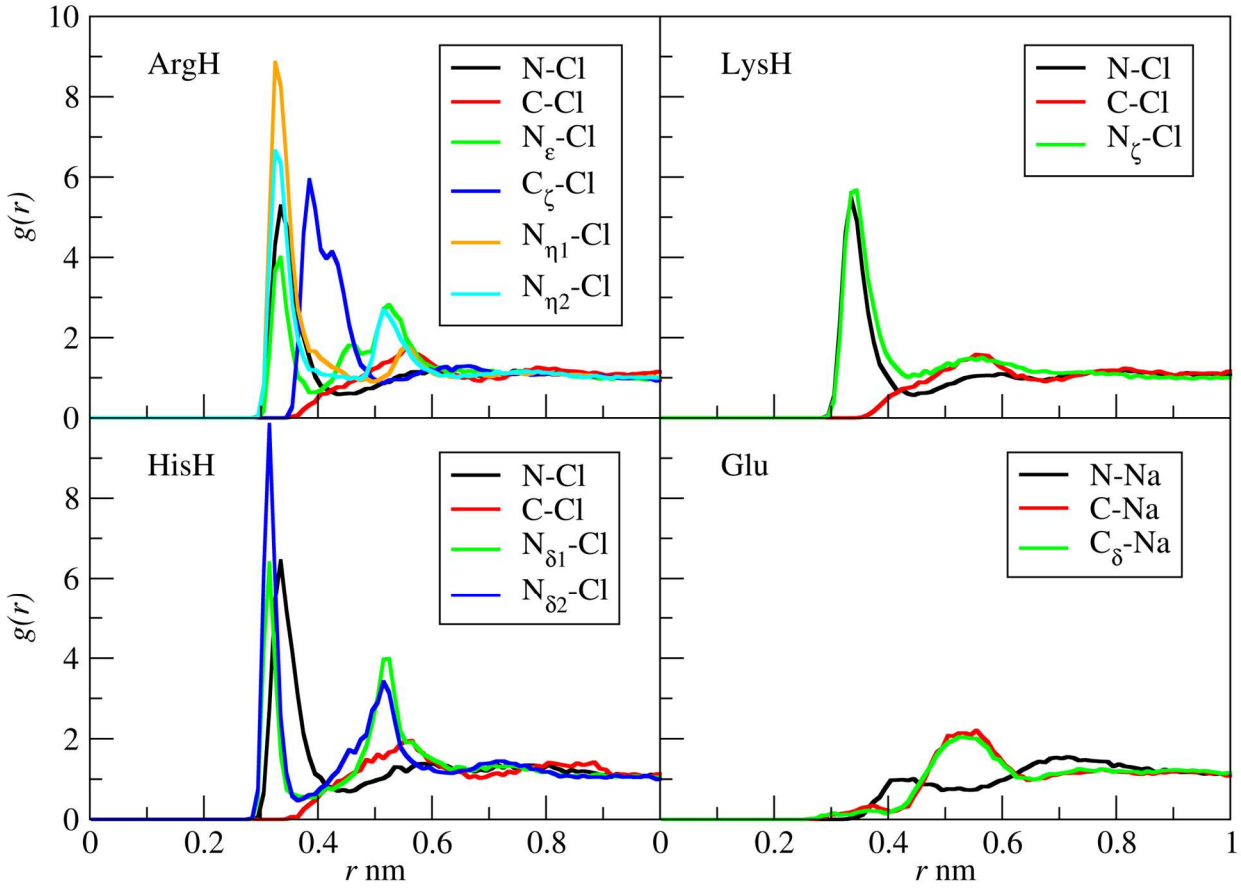


Figure 2.10 Select site-site radial distribution function obtained from simulations of 1m AA salt solutions. N is the N-terminal (amino) nitrogen and C is the C-terminal (carboxylate)carbon.

For the charged AAs, select site-site RDFs to the counterion are displayed in Figure 2.10. As expected, a strong interaction is observed between the positively charged amino nitrogen (NT atom type) and the Cl^- counterion for all three charged AAs. However, for ArgH ($N_{\eta 1}$ and $N_{\eta 2}$) and HisH ($N_{\varepsilon 2}$), the charged side chain nitrogens had a stronger interaction with the Cl^- counterion than did the N-terminal amino (NT) nitrogen. The NT and the side chain nitrogen for LysH have nearly equal strong contact peaks with the Cl^- counterion. No strong interaction exists between the negatively charged C-terminus and the Na^+ ion.

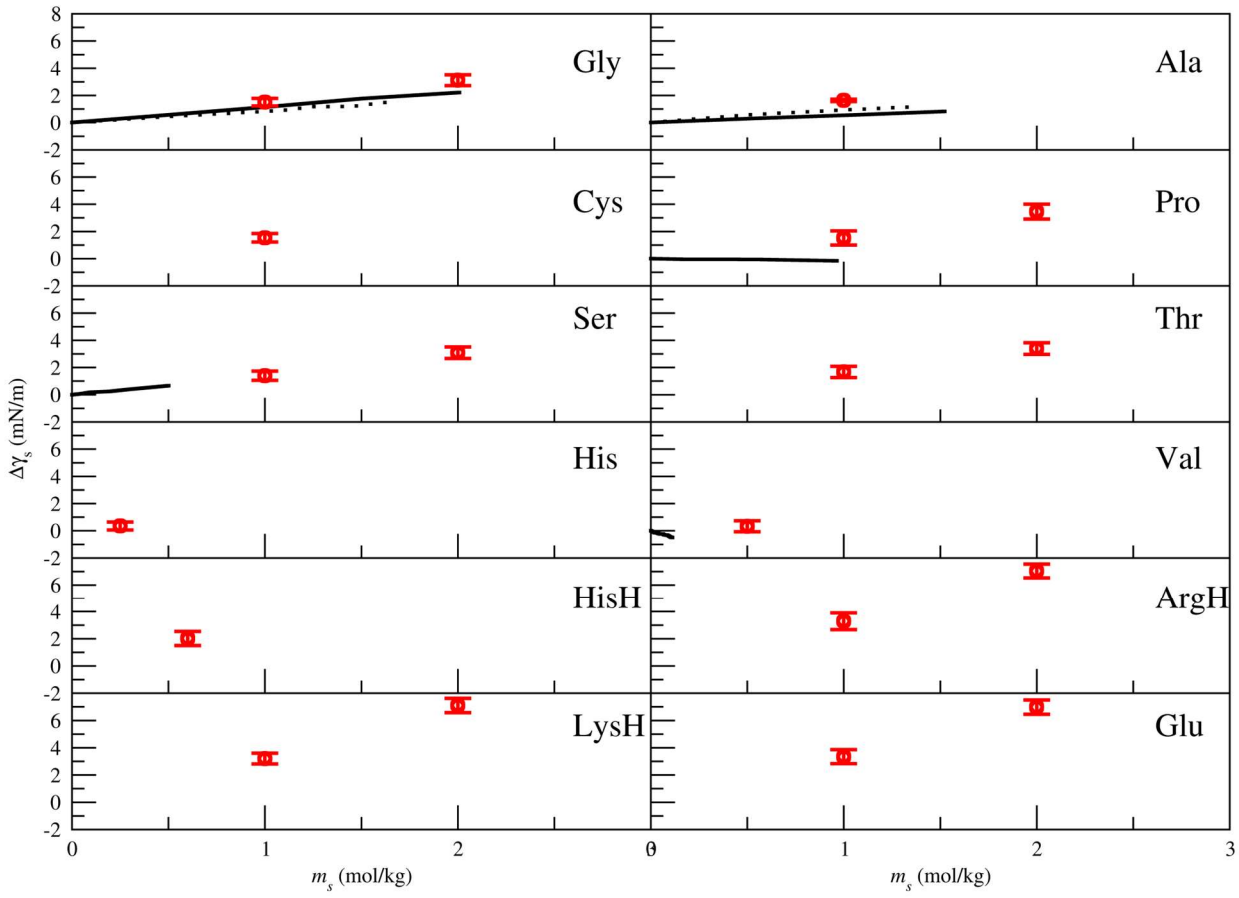


Figure 2.11 Surface Tension as function of solute or salt molality (mol/kg). Curves (black solid and dotted line) in Gly,[89, 90] Ala,[89, 90] Pro,[91] Ser,[91] and Val[92] were obtained from experimental surface tension while dots with error bar (red) correspond to data obtained from simulation using the KBFF20 models. For charged AA, the x-axis concentration is that of the AA salts at 298K.

Only five experimental values of surface tension data are available (Gly, Ala, Ser, Pro, and Val) and are displayed in Figure 2.11, together with simulated values. Simulation values of Gly, Ala, and Ser have similar trend as experimental values which increase with concentration. This indicates that the polar AAs are excluded from the surface in both the experiments and simulations. The nonpolar AAs, Pro and Val, are surface absorbed according to the experimental data. Unfortunately, our simulation results for the surface tension of pro solutions increase with concentration, indicating surface exclusion. However, this may not be because of new parameters

for NT and OT developed here, but because of the parameters for side chains. Further study may be needed to clarify this aspect.

2.4 Conclusions

A new force field is presented for the simulation of zwitterionic α-amino acids. The parameters were optimized to reproduce the KBIs for a range of soluble AAs in aqueous solution. Good agreement with the experimental data was possible for all AAs with the possible exception of Arg, for which future changes to the side chain parameters may be required. The new models reproduce the known diffusion constants and dielectric decrements. Surface tension changes were also well reproduced for all AAs except Pro. Again, this could signal a possible need for side chain parameter changes. The new models are fully compatible with the KBFF20 models for peptides and proteins.

2.5 Reference

(1) Okumoto, S.; Funck, D.; Trovato, M.; Forlani, G. Editorial: Amino Acids of the Glutamate Family: Functions beyond Primary Metabolism. *Frontiers in Plant Science* **2016**, *7*. DOI: 10.3389/fpls.2016.00318.

(2) Wu, G. Y. Amino acids: metabolism, functions, and nutrition. *Amino Acids* **2009**, *37* (1), 1-17. DOI: 10.1007/s00726-009-0269-0.

(3) Dalangin, R.; Kim, A.; Campbell, R. E. The Role of Amino Acids in Neurotransmission and Fluorescent Tools for Their Detection. *International Journal of Molecular Sciences* **2020**, *21* (17). DOI: 10.3390/ijms21176197.

(4) Ploetz, E. A.; Karunaweera, S.; Smith, P. E. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Applications of KBFF20. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 2991-3009. DOI: 10.1021/acs.jctc.1c00076.

(5) Ploetz, E. A.; Karunaweera, S.; Bentenitis, N.; Chen, F.; Dai, S.; Gee, M. B.; Jiao, Y. F.; Kang, Y. S.; Kariyawasam, N. L.; Naleem, N.; et al. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 2964-2990. DOI: 10.1021/acs.jctc.1c00075.

(6) Bentenitis, N.; Cox, N. R.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Thiols, Sulfides, and Disulfides. *Journal of Physical Chemistry B* **2009**, *113* (36), 12306-12315. DOI: 10.1021/jp904806f.

(7) Gee, M. B.; Cox, N. R.; Jiao, Y. F.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Aqueous Alkali Halides. *Journal of Chemical Theory and Computation* **2011**, *7* (5), 1369-1380. DOI: 10.1021/ct100517z.

(8) Ploetz, E. A.; Smith, P. E. A Kirkwood-Buff force field for the aromatic amino acids. *Physical Chemistry Chemical Physics* **2011**, *13* (40), 18154-18167. DOI: 10.1039/c1cp21883b.

(9) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. *Journal of Chemical Physics* **2004**, *121* (5), 2180-2186. DOI: 10.1063/1.1768938.

(10) Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006**, *27* (13), 1477-1485. DOI: 10.1002/jcc.20441.

(11) Naleem, N.; Bentenitis, N.; Smith, P. E. A Kirkwood-Buff derived force field for alkaline earth halide salts. *Journal of Chemical Physics* **2018**, *148* (22). DOI: 10.1063/1.5019454.

(12) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B* **2003**, *107* (16), 3891-3898. DOI: 10.1021/jp022049s.

(13) Kirkwood, J. G.; Buff, F. P. THE STATISTICAL MECHANICAL THEORY OF SOLUTIONS .1. *Journal of Chemical Physics* **1951**, *19* (6), 774-777. DOI: 10.1063/1.1748352.

(14) Smith, P. E. *Fluctuation Theory of Solution: Application in Chemistry, Chemical Engineering, and Biophysics*; CRC Press LLC, 2013.

(15) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005**, *109* (31), 15080-15086. DOI: 10.1021/jp051773i.

(16) Ben-Naim, A. *Molecular Theory of Solution*; Oxford University Press, 2006.

(17) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. Developing force fields from the microscopic structure of solutions. *Fluid Phase Equilibria* **2010**, *290* (1-2), 43-47. DOI: 10.1016/j.fluid.2009.11.023.

(18) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. Kirkwood-Buff integrals for ideal solutions. *Journal of Chemical Physics* **2010**, *132* (16). DOI: 10.1063/1.3398466.

(19) Karunaweera, S.; Gee, M. B.; Weerasinghe, S.; Smith, P. E. Theory and Simulation of Multicomponent Osmotic Systems. *Journal of Chemical Theory and Computation* **2012**, *8* (10), 3493-3503. DOI: 10.1021/ct300079v.

(20) Luo, Y.; Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *Journal of Physical Chemistry Letters* **2010**, *1* (1), 183-189. DOI: 10.1021/jz900079w.

(21) Miller, M. S.; Lay, W. K.; Elcock, A. H. Osmotic Pressure Simulations of Amino Acids and Peptides Highlight Potential Routes to Protein Force Field Parameterization. *Journal of Physical Chemistry B* **2016**, *120* (33), 8217-8229. DOI: 10.1021/acs.jpcb.6b01902.

(22) Miller, M. S.; Lay, W. K.; Li, S. X.; Hacker, W. C.; An, J. D.; Ren, J. L.; Elcock, A. H. Reparametrization of Protein Force Field Nonbonded Interactions Guided by Osmotic Coefficient Measurements from Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2017**, *13* (4), 1812-1826. DOI: 10.1021/acs.jctc.6b01059.

(23) Karunaweera, S. Theory and simulation of molecular interactions in biological systems. Kansas State University, Manhattan, Kansas, U.S.A., 2017.

(24) Gee, M. B. Computer Simulation and Theory of Amino Acid Interaction in Solution. Kansas State University, Manhattan, Kansas, 2010.

(25) Ploetz, E. A.; Weerasinghe, S.; Kang, M.; Smith, P. E. Accurate Force Fields for Molecular Simulation. In *Fluctuation Theory of Solutions: Application in Chemistry, Chemical Engineering, and Biophysics*, CRC press, 2013.

(26) Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Grater, F. Kirkwood-Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. *Journal of Physical Chemistry B* **2015**, *119* (25), 7975-7984. DOI: 10.1021/acs.jpcb.5b03440.

(27) Gee, M. B. COMPUTER SIMULATION AND THEORY OF AMINO ACID INTERACTIONS IN SOLUTION. KANSAS STATE UNIVERSITY, Manhattan, Kansas, 2010.

(28) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for sodium chloride in water. *Journal of Chemical Physics* **2003**, *119* (21), 11342-11349. DOI: 10.1063/1.1622372.

(29) Weerasinghe, S.; Smith, P. E. Kirkwood-Buff derived force field for mixtures of acetone and water. *Journal of Chemical Physics* **2003**, *118* (23), 10663-10670. DOI: 10.1063/1.1574773.

(30) Matteoli, E.; Lepori, L. SOLUTE SOLUTE INTERACTIONS IN WATER .2. AN ANALYSIS THROUGH THE KIRKWOOD-BUFF INTEGRALS FOR 14 ORGANIC SOLUTES. *Journal of Chemical Physics* **1984**, *80* (6), 2856-2863. DOI: 10.1063/1.447034.

(31) Ploetz, E. A.; Smith, P. E. Local fluctuations in solution mixtures. *Journal of Chemical Physics* **2011**, *135* (4). DOI: 10.1063/1.3615718.

(32) Haynes, W. M.; Lide, D. R.; Bruno, T. J. *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data*; CRC Press, 2016.

(33) Grineva, O. V.; Belyaeva, E. V. Structure of water-glycine solutions in saturated and near-saturated regions according to compressibility data. *Journal of Structural Chemistry* **2011**, *52* (6), 1139-1143. DOI: 10.1134/s0022476611060199.

(34) Aree, T.; Buergi, H.-B.; Chernyshov, D.; Tornroos, K. W. Dynamics and Thermodynamics of Crystalline Polymorphs. 3. gamma-Glycine, Analysis of Variable-Temperature Atomic Displacement Parameters, and Comparison of Polymorph Stabilities. *Journal of Physical Chemistry A* **2014**, *118* (43), 9951-9959. DOI: 10.1021/jp506659c.

(35) Smith, P. K.; Smith, E. R. B. The Activity of Glycine in Aqueous Solution at Twenty-Five Degrees. *journal of Biological Chemistry* **1937**, *117*, 209-216.

(36) Tsurko, E. N.; Neueder, R.; Kunz, W. Water activity and osmotic coefficients in solutions of glycine, glutamic acid, histidine and their salts at 298.15 K and 310.15 K. *Journal of Solution Chemistry* **2007**, *36* (5), 651-672. DOI: 10.1007/s10953-007-9131-8.

(37) Kuramochi, H.; Noritomi, H.; Hoshino, D.; Nagahama, K. Measurements of vapor pressures of aqueous amino acid solutions and determination of activity coefficients of amino acids. *Journal of Chemical and Engineering Data* **1997**, *42* (3), 470-474. DOI: 10.1021/je960113r.

(38) Romero, C. M.; Gonzalez, M. E. Osmotic and activity coefficients of glycine, DL-alpha-alanine and DL-alpha-aminobutyric acid in aqueous solutions at temperatures between 288.15 and 303.15 K. *Fluid Phase Equilibria* **2006**, *250* (1-2), 99-104, Article. DOI: 10.1016/j.fluid.2006.10.012.

(39) Rodriguez, H.; Soto, A.; Arce, A.; Khoshkbarchi, M. K. Apparent molar volume, isentropic compressibility, refractive index, and viscosity of DL-alanine in aqueous NaCl solutions. *Journal of Solution Chemistry* **2003**, *32* (1), 53-63. DOI: 10.1023/a:1022640715229.

(40) Iwama, S.; Horiguchi, M.; Sato, H.; Uchida, Y.; Takahashi, H.; Tsue, H.; Tamura, R. Observation of the Preferential Enrichment Phenomenon for Essential alpha-Amino Acids with a Racemic Crystal Structure. *Crystal Growth & Design* **2010**, *10* (6), 2668-2675. DOI: 10.1021/cg100152c.

(41) Smith, P. K.; Smith, E. R. B. Thermodynamic Properties of Solutions of Amino Acids and Related Substance. II. The Activity of Aliphatic Amino Acids in Aqueous Solution at Twenty-Five Degrees. *Journal of Biological Chemistry* **1937**, *121*, 607-613.

(42) Robinson, R. A. The Vapor Pressure of Aqueous Solutions of Alanine. *Journal of Biological Chemistry* **1952**, *199* (1).

(43) Kumar, H.; Verma, A.; Chadha, C. Synthesis and thermodynamics studies of ionic liquid 1-methyl-3-pentylimidazolium bromide (C(5)mim Br) with amino acids (L-cysteine and N-acetyl-L-cysteine) at different temperatures. *Journal of Chemical Thermodynamics* **2017**, *111*, 238-249. DOI: 10.1016/j.jct.2017.03.040.

(44) Luger, P.; Weber, M. DL-cysteine at 298 K. *Acta Crystallographica Section C-Crystal Structure Communications* **1999**, *55*, 1882-1885. DOI: 10.1107/s0108270199009208.

(45) Held, C.; Cameretti, L. F.; Sadowski, G. Measuring and Modeling Activity Coefficients in Aqueous Amino-Acid Solutions. *Industrial & Engineering Chemistry Research* **2011**, *50* (1), 131-141, Article. DOI: 10.1021/ie100088c.

(46) Ninni, L.; Meirelles, A. J. A. Water activity, pH and density of aqueous amino acids solutions. *Biotechnology Progress* **2001**, *17* (4), 703-711, Article. DOI: 10.1021/bp0100427.

(47) Seijas, L. E.; Delgado, G. E.; Mora, A. J.; Fitch, A. N.; Brunelli, M. On the crystal structures and hydrogen bond patterns in proline pseudopolymorphs. *Powder Diffraction* **2010**, *25* (3), 235-240. DOI: 10.1154/1.3478557.

(48) Smith, P. K.; Smith, E. R. B. Thermodynamic Properties of Solutions of Amino Acids and Related Substances. V. The Activities of Some Hydroxy- and N-methylamino Acids and Proline in Aqueous Solution at Twenty-Five Degrees. *The Journal of biological chemistry* **1940**, *132*, 57-64.

(49) Marriott, R. A. The volumetric properties of aqueous solutions of glycylglycine and L-serine at elevated temperatures and pressures. *Journal of Chemical Thermodynamics* **2001**, *33* (8), 959-982. DOI: 10.1006/jcht.2001.0823.

(50) Dittrich, B.; Hubschle, C. B.; Messerschmidt, M.; Kalinowski, R.; Girnt, D.; Luger, P. The invariom model and its application: refinement of D,L-serine at different temperatures and resolution. *Acta Crystallographica a-Foundation and Advances* **2005**, *61*, 314-320. DOI: 10.1107/s0108767305005039.

(51) Hutchens, J. O.; Granito, S. M.; Figlio, K. M. An Isopiestic Comparison Method for Activities - Activities of L-Serine and L-Arginine. Hydrochloirde. *Journal of Biological Chemistry* **1963**, *238* (4).

(52) Kumar, H.; Singla, M.; Jindal, R. Studies of interionic interactions of L-serine/L-threonine in aqueous trilithium citrate solutions using density and speeds of sound measurements at different temperatures. *Journal of Molecular Liquids* **2015**, *208*, 170-182. DOI: 10.1016/j.molliq.2015.04.040.

(53) Shoemaker, D. P.; Donohue, J.; Schomaker, V.; Corey, R. B. The Crystal Structure of LS-Threonine. *Journal of the American Chemical Society* **1950**, *72* (6), 2328-2349.

(54) Yuan, Q.; Li, Z. F.; Wang, B. H. Partial molar volumes of L-alanine, DL-serine, DL-threonine, L-histidine, glycine, and glycylglycine in water, NaCl, and DMSO aqueous solutions at T=298.15 K. *Journal of Chemical Thermodynamics* **2006**, *38* (1), 20-33. DOI: 10.1016/j.jct.2005.03.015.

(55) Edington, P.; Harding, M. M. Crystal-Structure of DL-Histidine. *Acta Crystallographica Section B-Structural Science* **1974**, *B 30* (JAN15), 204-206. DOI: 10.1107/s0567740874002470.

(56) Mallikarjunan, M.; Rao, S. T. Crystal Structure of DL-Valine. *Acta Crystallographica Section B-Structural Crystallography and Crystal Chemistry* **1969**, *B 25*. DOI: 10.1107/s0567740869002147.

(57) Fuess, H.; Hohlwein, D.; Mason, S. A. Neutron-Diffraction Study of L-Histidine Hydrochloride Monohydrate. *Acta Crystallographica Section B-Structural Science* **1977**, *33* (MAR15), 654-659. DOI: 10.1107/s0567740877004415.

(58) Bonner, O. D. Osmotic and Activity-Coefficients of Some Amino-Acids and Their Hydrochloride Salts at 298.15-K. *Journal of Chemical and Engineering Data* **1982**, *27* (4), 422-423. DOI: 10.1021/je00030a016.

(59) Dow, J.; Jensen, L. H.; Mazumdar, S. K.; Srinivas.R; Ramachan.Gn. Refinement of Structure of Arginine Hydrochloride Monohydrate. *Acta Crystallographica Section B-Structural Crystallography and Crystal Chemistry* **1970**, *B 26* (NOV15). DOI: 10.1107/s0567740870004697.

(60) Bhaduri, D.; Saha, N. N. Crystal and Molecular-Structure of DL-Lysine Hydrochloride. *Journal of Crystal and Molecular Structure* **1979**, *9* (6), 311-316. DOI: 10.1007/bf01200523.

(61) Nagashima, N.; Sano, C.; Kawakita, T.; Iitaka, Y. Crystal-Structure of Sodium DL-Glutamate Dihydrate. *Analytical Sciences* **1992**, *8* (2), 267-269. DOI: 10.2116/analsci.8.267.

(62) Bonner, O. D. Osmotic and Activity-Coefficients of Sodium and Potassium Glutamate at 298.15 K. *Journal of Chemical and Engineering Data* **1981**, *26* (2), 147-148. DOI: 10.1021/je00024a014.

(63) Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*; Butterworth Scientific Publications, 1955.

(64) Cox, R. A.; McCartney, M. J.; Culkin, F. Pure Water for Relative Density Standard. *Deep-Sea Research* **1968**, *15*. DOI: 10.1016/0011-7471(68)90008-9.

(65) Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B.; GROMACS; team, d. *GROMACS User Manual version 2016*; 2018.

(66) Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet High-Resolution Computer Models of a Plasma. *Journal of Computational Physics* **1974**, *14* (2), 148-158. DOI: 10.1016/0021-9991(74)90010-2.

(67) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463-1472. DOI: 10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.

(68) Miyamoto, S.; Kollman, P. A. SETTLE - An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992**, *13* (8), 952-962. DOI: 10.1002/jcc.540130805.

(69) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large System. *Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397.

(70) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **1984**, *81* (8), 3684-3690.

(71) Nose, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics* **1984**, *52* (2), 255-268. DOI: 10.1080/00268978400101201.

(72) Hoover, W. G. Canonical Dynamics -Equilibrium Phase-Space Distributions. *Physical Review A* **1985**, *31* (3), 1695-1697. DOI: 10.1103/PhysRevA.31.1695.

(73) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - A New Molecular-Dynamics Method. **1981**, *52* (12), 7182-7190. DOI: 10.1063/1.328693.

(74) Nose, S.; Klein, M. L. Constant Pressure Molecular-Dynamics for Molecular-Systems. *Molecular Physics* **1983**, *50* (5), 1055-1076. DOI: 10.1080/00268978300102851.

(75) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *Journal of Chemical Physics* **1998**, *108* (24), 10220-10230. DOI: 10.1063/1.476482.

(76) van Gunsteren, W. F.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mar, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The Gromos96 Manual and User Guide*; Hochschulverlag AG an der ETH Zurich and BIOMOS b.v., 1996.

(77) Senapati, S.; Berkowitz, M. L. Computer simulation study of the interface width of the liquid/liquid interface. *Physical Review Letters* **2001**, *87* (17). DOI: 10.1103/PhysRevLett.87.176101.

(78) Wyman, J., Jr. Dielectric Constants of Polar Solutions. *J. Am. Chem. Soc* **1934**, *56* (3).

(79) Zhuo, K. L.; Chen, Y. J.; Kang, L.; Xu, S. J.; Wang, J. J. Dielectric Constants for Binary Amino Acid-Water Solutions from (278.15 to 313.15) K. *Journal of Chemical and Engineering Data* **2009**, *54* (1), 137-141. DOI: 10.1021/je800644b.

(80) Malmberg, C. G.; Maryott, A. A. Dielectric Constant of Water from 0-Degree-C to 100-Degree-C. *Journal of Research of the National Bureau of Standards* **1956**, *56* (1), 1-8. DOI: 10.6028/jres.056.001.

(81) Yeh, I. C.; Hummer, G. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *Journal of Physical Chemistry B* **2004**, *108* (40), 15873-15879, Article. DOI: 10.1021/jp0477147.

(82) Wang, J. H. Tracer-Diffusion in Liquids .5. Self-Diffusion Isoelectric Glycine in Aqueous Glycine Solutions. *Journal of the American Chemical Society* **1953**, *75* (11), 2777-2778. DOI: 10.1021/ja01107a518.

(83) Altunina, L.; Bezrukov, O.; Smirnova, N.; Moskvicheva, I.; Fokanov, V. *Self-diffusion in aqueous solutions of amino acids, peptides, and proteins.*; Springer US, 1971.

(84) Polson, A. On The Diffusion Constants of The Amino Acids. *Biochem. J.* **1937**, *31* (10), 1903-1912.

(85) Longsworth, L. G. *Electrochemistry in biology and medicine*; John Wiley & Sons Inc., 1955.

(86) Ma, Y. G.; Zhu, C. Y.; Ma, P. S.; Yu, K. T. Studies on the diffusion coefficients of amino acids in aqueous solutions. *Journal of Chemical and Engineering Data* **2005**, *50* (4), 1192-1196. DOI: 10.1021/je049582g.

(87) Germann, M. W.; Turner, T.; Allison, S. A. Translational diffusion constants of the amino acids: Measurement by NMR and their use in modeling the transport of peptides. *Journal of Physical Chemistry A* **2007**, *111* (8), 1452-1455. DOI: 10.1021/jp068217o.

(88) Altunina, L.; Bezrukov, O.; Smirnova, N.; Moskvicheva, I.; Fokanov, V. *Self-diffusion in aqueous solutions of amino acids, peptides, and proteins.*; Springer, 1971.

(89) Pappenhbimer, J. R.; Lepie, M. P.; Wyman, J. J. The Surface Tension of Aqueous Solution of Dipolar ions. *J Am. Chem. Soc.* **1936**, *58* (10).

(90) J.W., B. The Capillary properties of alpha-amino-acids. *Trans. Faraday Soc.* **1939**, *35*, 1293-1298.

(91) Raza, M. A.; Hallett, P. D.; Liu, X.; He, M.; Afzal, W. Surface Tension of Aqueous Solutions of Small-Chain Amino and Organic Acids. *Journal of Chemical and Engineering Data* **2019**, *64* (12), 5049-5056. DOI: 10.1021/acs.jced.9b00026.

(92) Matubayasi, N.; Miyamoto, H.; Namihira, J.; Yano, K.; Tanaka, T. Thermodynamic quantities of surface formation of aqueous electrolyte solutions - V. Aqueous solutions of

aliphatic amino acids. *Journal of Colloid and Interface Science* **2002**, *250* (2), 431-437. DOI: 10.1006/jcis.2002.8307.

Chapter 3 – Sulfate, Alkyl Sulfate, and SDS Force Fields

3.1 Introduction

SDS is an amphiphilic molecule with a long hydrophobic aliphatic tail and a ionic hydrophilic sulfate head group that can reduce surface tension between phases due to its amphiphilic properties.[1] This reduction in surface tension is a critical parameter for various physical phenomena, and surfactants play a vital role in reducing the surface tension of water.[2, 3] Micelle formation occurs when the surfactant achieves maximum packing, and SDS is often used as a membrane mimetic to study interactions with biological molecules such as proteins.[4] Conformational changes of intrinsically disordered proteins at different interfaces have been studied in the presence of SDS as a lipid bilayer-mimicking environment.[5, 6] The general process of developing a force field is iterative, with the parameters of the potential energy function being adjusted until the force field accurately reproduces the properties of the molecules being studied. Once the force field has been developed, it can be used to simulate the behavior of larger systems and to study a wide range of chemical and biological phenomena. Developing FF parameters based on KBFF models requires experimental activity coefficient and density data,[7] which are currently unavailable for SDS. To overcome this limitation, we use a similar strategy for developing peptide parameters,[7] by separating the molecule into small functional groups such as the sulfate head groups and the hydrophobic tail. Since experimental data for sodium sulfate activities are available, we can extract the required experimental KBIs and use them to develop parameters. These parameters can then be tested for transferability to alkyl sulfates and an expanded parameter development for dihedral angles. When all parameters are developed from a simple model system, they are employed to simulate SDS and validated with experimental properties such as surface tension.

3.1.1 Sulfate ion

The sulfate anion is a polyatomic oxoanion with empirical formula SO_4^{2-} that has diverse usage including biological systems. Inorganic sulfate is an essential electrolyte which is required by all organisms for life. Sulfate is required for proper cell growth and the development of many organisms.[8] Important biological processes, such as biosynthesis and detoxification, are managed by sulfation of various endogenous and exogenous molecules.[9] Sulfate is also critical for cell matrix synthesis and for the maintenance of cell membranes.[10] Sulfate ions are to the left of the Hofmeister series, as shown in Figure 3.1,[9] and are therefore categorized as kosmotrope ions (water structure making), and have a propensity to stabilize the native (folded) structure of proteins, and also lead to salting-out behavior.

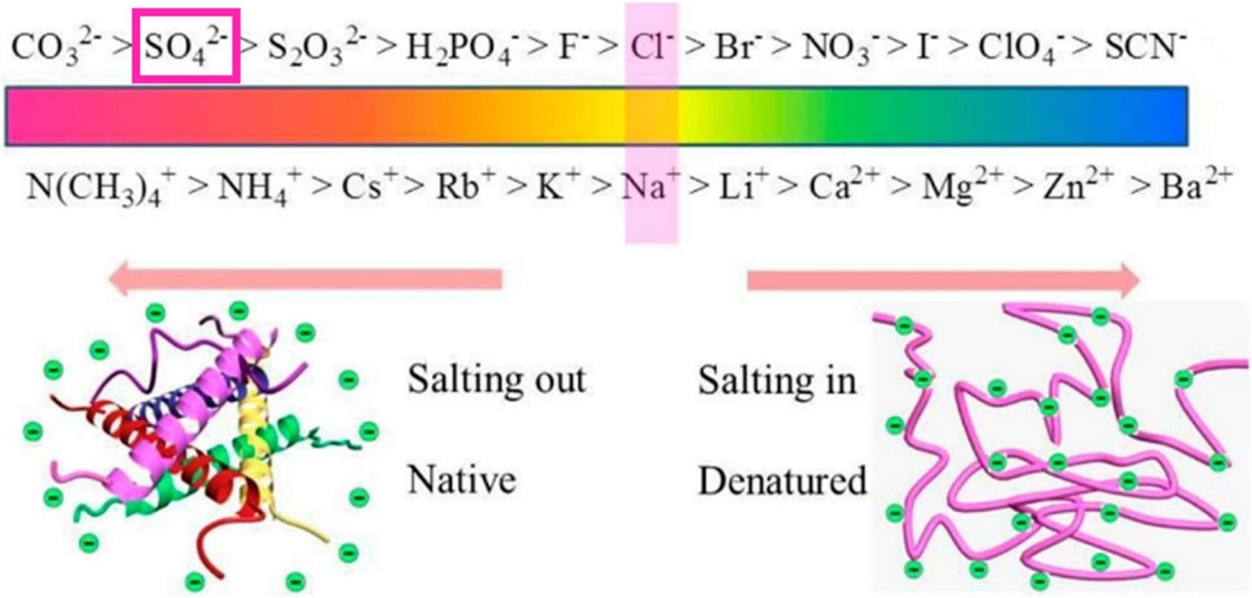


Figure 3.1 Ranking of ions in the Hofmeister Series. The figure is reprinted with permission by Kang, B. B.; Tang, H. C.; Zhao, Z. D.; Song, S. S. Hofmeister Series: Insights of Ion Specificity from Amphiphilic Assembly and Interface Property. Acs Omega 2020, 5 (12), 6229-6239. DOI: 10.1021/acsomega.0c00237. Copyright American Chemical Society.(2020)[9]

It is known to be a challenge to describe salts containing divalent oxo anions with non-polarizable force fields. Several simulations of oxoanion containing salts have appeared in the

literature. Generally, significant ion aggregation has been reported for these simulations. Excessive ion aggregation is commonly observed in aqueous sodium sulfate simulations, including our own unpublished data.[11-13] This aggregation behavior can be reduced with polarizable models.[14, 15] However, due to the limitations of polarizable models, such as system size and simulation duration, it would be advantageous to improve the aggregation problem with non-polarizable models. To achieve this goal, it is important to find a good balance between solvation and ion pairing when developing the FFs parameters.

Xu *et al.*'s X-ray diffraction studies show that Na^+ and SO_4^{2-} form contact ion pairs in aqueous sodium sulfate solutions, particularly under high concentrations and temperatures. This formation is likely to affect the hydration structure around these ions, possibly increasing the number of water molecules associated with each ion. Specifically, the study identified seven to twelve water molecules forming a hydration shell around SO_4^{2-} and 6 in the first hydration shell of Na^+ (with 8−12 for the second shell).[16] A recent simulation study using a nonpolarizable FF resolved the sulfate ion aggregation problem, but in a somewhat complex way. Mamatkulov *et al.*[17] optimized the sulfate FF parameters using a mechanically flexible model for the sulfate anion. Intramolecular interactions in the sulfate anion are usually excluded in sulfate anion FFs. However, they included the O–O electrostatic interaction within the same sulfate anion, after scaling by a factor $\lambda_q = 0.833$. They were then able to reproduce the experimental dielectric spectra of aqueous $MgSO_4$ and Na_2SO_4 solutions with these models. In another study, Zeron *et al.*[18] have explored scaling the charges on ions to account for the electronic contribution to the permittivity. However, this leads to a net charge for the sulfate anion of -1.7, which is unusual and difficult to incorporate into other FFs.

3.1.2 Alkyl sulfate

Short chain organosulfate salts such as sodium/potassium methyl/ethyl sulfate are widely used for pharmaceutical formulations and are a source of ionic liquids, which are good alternatives organic solvent.[19] SDS is a longer chain alkyl sulfate that is the most widely used anionic surfactant. For example, SDS is used in the SDS-PAGE electrophoresis method that allows for protein separation based on mass of proteins.[20] Other uses of SDS involve the structural characterization of membrane proteins, particularly by crystallization and solution NMR techniques.[21] SDS is used as a lipid mimetic in these characterizations.

SDS is an amphiphilic molecule containing a long hydrophobic aliphatic tail and a hydrophilic sulfate head group. Due to its amphiphilic properties, SDS can be adsorbed at the interface and reduce surface tension between phases. The surface tension is an important parameter for various physical phenomena, and surfactants play a critical role in reducing the surface tension of water. Micelle formation occurs when surfactant achieves maximum packing. Conformational changes of intrinsically disordered proteins at different interfaces have been performed in the presence of SDS as a lipid bilayer-mimicking environment role in biological system.

SDS forms micelles in aqueous solution. The first critical micelle concentration (CMC) for SDS is around 0.008 M,[22, 23] and average aggregation number at this concentration is about 60 SDS per micelle in aqueous solution at 25^{o}C.[24, 25] Recently, MD simulation studies of SDS in aqueous solutions have been performed extensively.[26, 27] As mentioned previously, the quality of a simulation relies on the FF. Thus, the development of good FF parameter is important. The CHARMM, AMBER, GROMOS, and OPLS FFs have been applied to develop UA or all atom FF parameters for SDS.[28, 29] Generally, the parameters are adopted from quantum mechanical calculations of small molecules, and the parameters are then transferred for use in larger molecules.

The process of parameter validation usually compares simulation results to experimental data. However, SDS is a challenging molecule to study as there is limited experimental data available. Surface tension data is sensitive to the force field parameters and experimental data for SDS is available. The surface tension of SDS slowly decreases with an increase of concentration, drops sharply near the CMC, and then stays fixed above the CMC concentration.[30] However, validation with this process is difficult for SDS because the low CMC requires very large simulation boxes. Consequently, in their simulations Cheng *et al*.[31] calculated the surface tension of SDS at higher concentrations than the CMC. Since it is difficult to perform simulations around the CMC concentration, we will instead compare with experimental surface tension changes as a function of surface coverage in order to validate the SDS FF.

3.2 Methods

3.2.1 General process of SDS parameter development with KBFF20 model

The detailed KB theory is described in Chapter 1. The experimental activity coefficients and densities for sodium/potassium sulfate and alkyl sulfates are available in the literature.[32-47] A similar strategy as described in Chapter 2, using the Kirkwood Buff inversion procedure, was applied to extract the experimental KBI. Therefore, to obtain experimental KBI values we used the experimental composition dependent activity and density data. The experimental activity derivatives were fitted to the Pitzer equation for the salt solution, as described previously in Naleem *et al*.,[48] and the densities were fitted to a simple polynomial expression as shown in Chapter 2. For a single electrolyte, $M_{\nu+}X_{\nu-}$, the Pitzer equation is expressed as following,

$$ln\gamma_{\pm} = |z_+ z_-| f^{\gamma} + m\left(\frac{2\nu_+\nu_-}{\nu}\right)B^{\gamma} + m^2\left(\frac{2(\nu_+\nu_-)^{3/2}}{\nu}\right)C^{\gamma}$$

$$f^{\gamma} = -A_{\phi}\left[\frac{\sqrt{I/m^0}}{1+b\sqrt{I/m^0}} + \frac{2}{b}\ln\left(1+b\sqrt{I/m^0}\right)\right] \qquad 3.1$$

$$B^{\gamma} = 2\beta^{(0)} + \frac{2\beta^{(1)}}{\alpha^2(I/m^0)}\left[1 - exp\left(-\alpha\sqrt{I/m^0}\right)\left(1 + \alpha\sqrt{I/m^0} - (1/2)\,\alpha^2 - (I/m^0)\right)\right]$$

$$C^{\gamma} = (3/2)C^{\phi}$$

$$A_{\phi} = \frac{1}{3}(2\pi N_A \rho_A)^{1/2}\left[\frac{e^2}{4\pi\varepsilon_0 DkT}\right]^{3/2}$$

where *I* is a molality scale ionic strength, Z is the charge number of ions, and the $\nu = \nu_+ + \nu_-$ is the number of ions dissociated in one unit of electrolyte formula. The constants $\beta^{(0)}$, $\beta^{(1)}$, and C^{ϕ} are Pitzer equation fitting parameters. Also, α = 2 $kg^{1/2}\,mol^{-1/2}$, b = 1.2 $kg^{1/2}\,mol^{-1/2}$, and A_{ϕ} is the Debye-Huckel term for the osmotic coefficient. Lastly, N_A, ρ_A, D, k, and ε_0, are the Avogadro number, density of the solution mixture, dielectric constant, Boltzmann constant, and the vacuum permittivity, respectively.

3.2.2 Parameter development

3.2.2.1 Sulfate ion

To develop a sulfate FF, one requires effective partial charges (q) used in the electrostatic interactions, together with the LJ interaction size (σ) and the interaction strength (ε) for sulfur and oxygen. However, SO_4^{2-} is a polyatomic anion that also requires bond and angle parameters. The equilibrium bond length and bond angle were adopted from KBFF20.[7] Hence, the main focus of the optimization will be the non-bonded terms such as the effective charges and LJ parameters for the SO_4^{2-} ion.

Our previous unpublished sulfate FF parameters were used as initial parameters.[49] Then, the partial charges and oxygen σ parameters were adjusted to obtain reasonable agreement with the experimental KBI values. Most of the parameter optimization simulations were performed for aqueous Na_2SO_4 solutions using the KBFF sodium cation parameters.[50] This was followed by

simulations of aqueous K_2SO_4 solutions using the KBFF potassium cation parameters.[50] The final sulfate parameters are provided in Table 3.1.

In addition, we also investigated the role of the O-S-O bond angle force constant in the simulations. The initial value of 520 kJ/mol led to instability in many of the simulations. This instability persisted through different versions of the GROMACS software package. Hence, we attempted to optimize the force constant by comparing simulated and experimental vibrational spectra with different angle force constant (k_ϕ) ranging between 1200 ~ 5200 kJ/mol.

3.2.2.2 Alkyl sulfate ion

All bonded terms are adopted from KBFF20,[7] but the dihedral angle (C-O-S-O and C-C-O-S) parameters for rotation around the O-S and C-O bonds were parameterized in order to reproduce quantum mechanical calculations.[51] Simulations were performed for aqueous solutions of the following alkyl sulfates: potassium methyl sulfate, sodium methyl sulfate, potassium ethyl sulfate, and sodium ethyl sulfate. Initial effective charge set was taken from Ferro *et al.*[52] Sulfur and oxygen charges were adjusted, along with the remaining alkyl group charges, to ensure that the overall charge of the molecule was -1. Once the best effective charge distribution is selected for the methyl sulfate ion, see Table 3.2, the transferability of the LJ oxygen size parameter from SO_4^{2-} to methyl sulfate (atom type OT) was investigated. All parameters were tested for the transferability between sodium and potassium ion salts for methyl sulfate. Lastly, we combined all the alkyl sulfate and KBFF20 hydrocarbon parameters together to generate the SDS FF. Additional bond, angle and dihedral parameters were taken from KBFF20. The final parameters are summarized in Tables 3.2 and 3.3.

3.2.3 Simulation methods

3.2.3.1 Simulation for KBI analysis

All the molecular dynamics simulations were performed with the GROMACS 2016 software pakage.[53] The SPC/E water model was used in the simulations.[54] All molecules were modeled using the KBFF force field for sodium sulfate (Na_2SO_4), potassium sulfate (K_2SO_4), sodium methyl sulfate (CH_3OSO_3Na), potassium methyl sulfate (CH_3OSO_3K), sodium ethyl sulfate ($C_2H_5OSO_3Na$), and potassium ethyl sulfate ($CH_2H_5OSO_3K$) systems. All simulations were performed in the isothermal isobaric ensemble (NpT) at 298K temperature and 1 atm pressure. The temperature and pressure were weakly coupled to a bath with relaxation times of 0.1 ps and 0.5 ps, respectively. A time step of 2 fs was used to integrate the equations of motion. All the non-water bonds were constrained with LINCS algorithm,[55] and the water bonds were constrained with the SETTLE algorithm.[56] To calculate electrostatic interactions, the particle mesh Ewald technique[57] was used with a 1 nm cut off distance for real space calculations, as well as 1.0 nm cutoff was used for the van der Waals interactions. A random initial configuration of ions and waters in a cubic box with length of 10 nm was generated to provide a known concentration of aqueous salt solution using a custom written Fortran code. An energy minimization using the steepest descent method followed by 100 ps of equilibration was performed before the production run. The production run of 100 ns was used to calculate ensemble averages. In the production run the configurations were saved for analysis every 2 ps.

3.2.3.2 Simulation for SDS surface tension measurement

The GROMACS 2016 software package[53] was used to integrate the equations of motion using the leapfrog algorithm with a time step of 2 fs for the production. The simulations for surface tension measurement performed in the canonical ensemble in which the number of particles (N),

the box volume (V), and the temperature (T) were kept constant (NVT). Temperature was kept constant using the Nose−Hoover thermostat,[58, 59] with a relaxation time of 0.1ps. All simulations were conducted at T = 298 K. To calculate electrostatic interactions, the particle mesh Ewald technique[57] was used with a 1 nm cut-off distance for real space calculations. A 1.0 nm cut-off was also used for the van der Waals interactions. Periodic boundary conditions were employed in all three dimensions. The box size is 10 × 10 × 30 nm^3. Water molecules occupied a volume of 10×10×10 nm^3 in the center of the simulation box, surrounded in each z direction by a similar volume of empty space (vacuum) as shown in Figure 3.2. Several simulations were performed in which the two 10×10 nm^2 interfaces between vacuum and water were covered by either 25%, 50%, 75%, or 100% of SDS molecules. Based on the estimates from Kawai *et al*.[60] and Tajima *et al*.[61], the average area occupied by an SDS head group was approximated as 0.5 nm². Therefore, 100 SDS molecules for 25% coverage, 200 SDS molecules for 50% coverage, 300 SDS molecules for 75% coverage, and 400 SDS molecules for 100% coverage were distributed over each 10×10 nm² surface area. Each production simulation was performed for 100 ns.

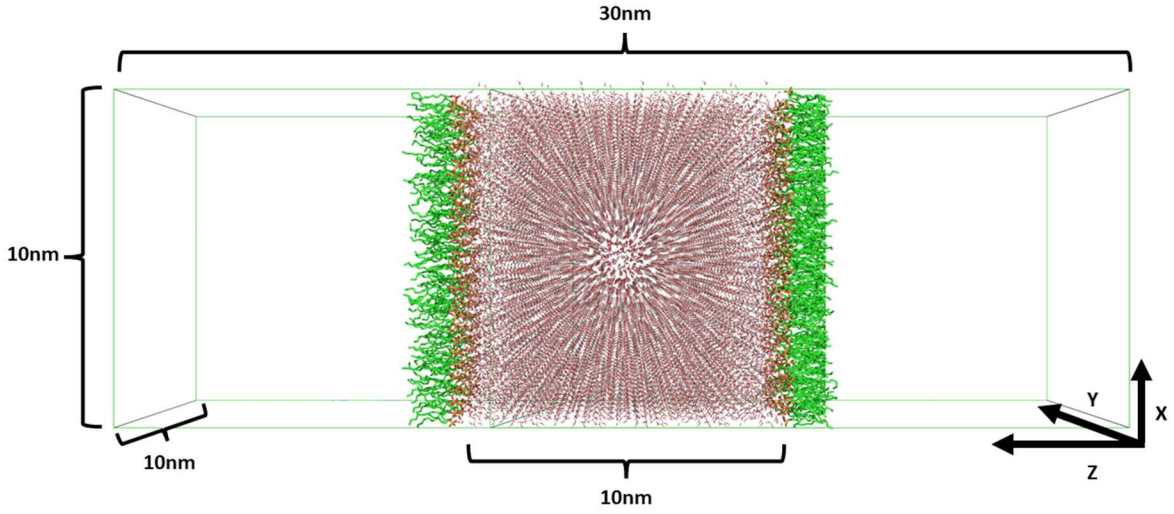


Figure 3.2 Schematic representation of the simulation box used to determine interfacial surface tensions.

Table 3.1 Final LJ Nonbonded Parameters for SO_4^{2-}

Atoms	Atom Type	σ (nm)	ε (kJ/mol)	Ref
S	S3	0.3580	1.6545	[7]
O	OT4	0.4200	0.6047	This work
O-O	OW-OT4	0.3647	0.6272	This work

Table 3.2 Final L-J Nonbonded Parameters for alkyl sulfates

Atoms	Atom Type	σ (nm)	ε (kJ/mol)	Ref
S	S3	0.3580	1.6545	[7]
O	OT	0.3850	0.6047	[7]
CH_2	CH_2	0.4070	0.4105	[7]
CH_3	CH_3	0.3748	0.8672	[7]
O-O	OW-OT	0.3329	0.6272	[7]

Table 3.3 Dihedral Force Field Parameters

Atoms	Force $k_\phi = (kJmol^{-1})$	Multiplicity n	Ref
O-S-O-CH_2	1.167	3	62, 63
S-O-CH_2-$CH_{2(3)}$	-12.500	1	64
	0.500	2	64
	4.000	3	64
	-0.500	4	64
O-CH_2-CH_2-$CH_{2(3)}$	-0.750	1	51
	0.500	2	51
	8.500	3	51
CH_2-CH_2-CH_2-CH_3	-2.510	1	7
	-0.250	2	7
	5.250	3	7

All phase shifts (δ) are zero. The C-C-C-C parameters taken from Ref. 7 correspond to a Lysine side chain in KBFF20.[7] All other references correspond to the target data used for the fitting process. The O-S-O-C and S-O-C-C dihedral parameters were generated by Elizabeth Ploetz for this work. The O-C-C-C dihedral parameters were generated previously by Shu Dai.[65]

3.3 Results and discussion

3.3.1 Sodium sulfate (Na_2SO_4) and potassium sulfate (K_2SO_4)

The SO_4^{2-} ion is tetrahedral in shape (Figure 3.3) with overall net charge of -2, which represents a relatively high charge density. In an effort to model the properties of the sulfate oxo anion, we first tried several different charge distributions consistent with an overall charge of -2. However, the KBIs determined from the simulations of aqueous Na_2SO_4 and K_2SO_4 solutions proved to be insensitive to the quite large changes in the charge distribution. Furthermore, all the

distributions led to excessive aggregation of the ions. Thus, we decided to keep our initial charge distribution, which was taken from McCammon *et al*.,[66] and proceeded to adjust the LJ σ parameter for oxygen (original atom type OT). The aggregation problem was resolved by increasing the size of the sulfate oxygen to 0.4200 nm.

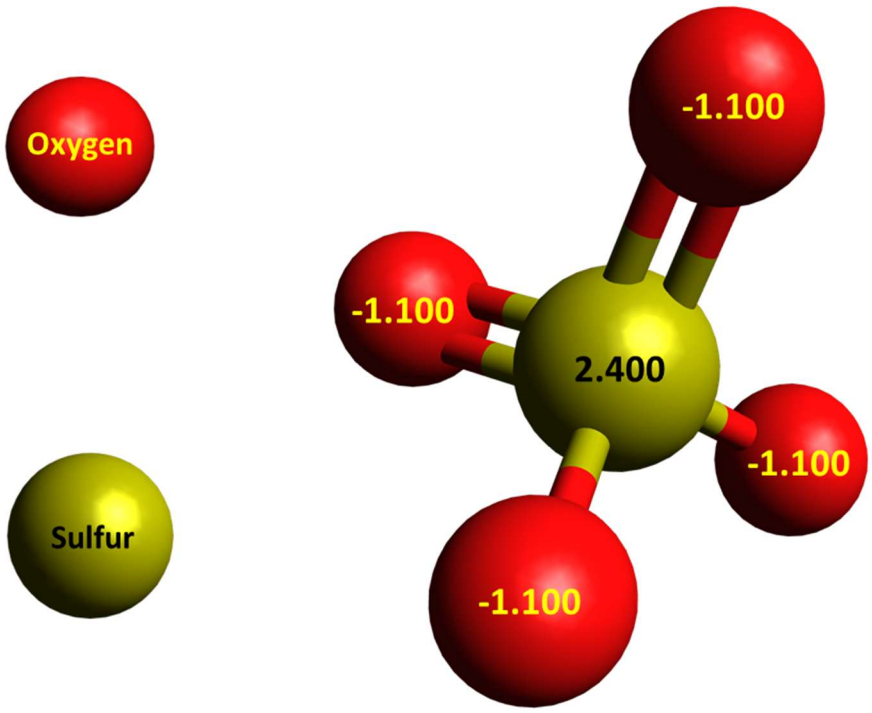


Figure 3.3 Effective charge distribution of Sulfate anion where red represents oxygen (O) and yellow represents sulfur (S) atoms.

The center of mass (COM) radial distribution functions for Na_2SO_4 and K_2SO_4 solutions are shown in Figure 3.4. All the rdfs converged to unity beyond 1 nm. The distributions appear to be insensitive to the concentration for Na_2SO_4 solutions. The ion-ion rdfs (g_{22}) for K_2SO_4 display a larger excluded volume due to larger size of potassium ions compared to sodium. The ion-solvent (g_{12}) distributions corresponding to the first coordination shell of Na_2SO_4 display a higher intensity compared to K_2SO_4, indicating that sodium ions are more strongly solvated than potassium ions. The water-water distributions are not significantly affected by changing the cation from sodium to potassium.

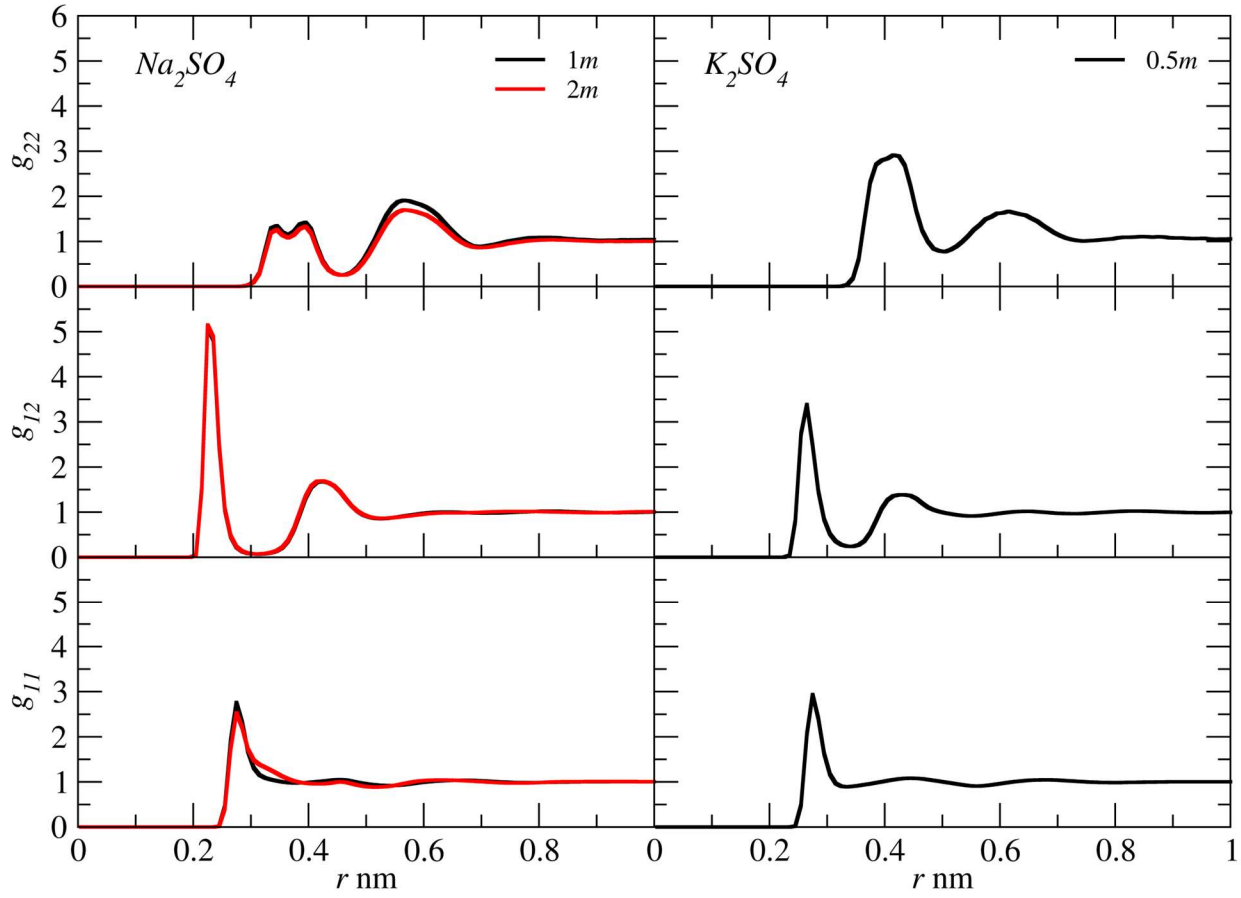


Figure 3.4 Radial distribution function for aqueous Na_2SO_4 (left) and K_2SO_4 (right) solutions.

The KB integrals obtained from the simulations of Na_2SO_4 and K_2SO_4 solution mixtures are compared to the experimental values in Figure 3.5. The Na_2SO_4 KBI value for ion-ion interaction (G_{22}) at 1m is within the error bar, while the 2m value is slightly underestimated compared to experimental. Overall, the simulated KBIs are in good agreement for Na_2SO_4. The experimental solubility of K_2SO_4 is ~0.78m,[67] therefore we were restricted to simulations at 0.5m. The simulated G_{22} value is slightly overestimated compared to some of the experimental KBIs, obtained with different experimental datasets, but can be considered acceptable in comparison to the range of experimental KBI values. The experimental ion-ion KBIs are all positive indicating a preference for ion association, over ion solvation, which is reproduced by the simulations. The simulated ion-solvent KBI ($G_{12} = G_{21}$) values for Na_2SO_4 solutions increased with increasing ion

concentration, whereas the experimental KBIs decreased with increasing concentration. The reasons for this are unclear at present. The K_2SO_4 G_{12} values are also slightly lower than the experimental values. The solvent-solvent KBI (G_{11}) values for both Na_2SO_4 and K_2SO_4 solutions do not change much with increasing salt concentration. The simulated G_{11} values were slightly overestimated for both Na_2SO_4 and K_2SO_4 solutions. In comparison with the results for other salt systems,[50] the agreement with experiment provided here can be considered acceptable.

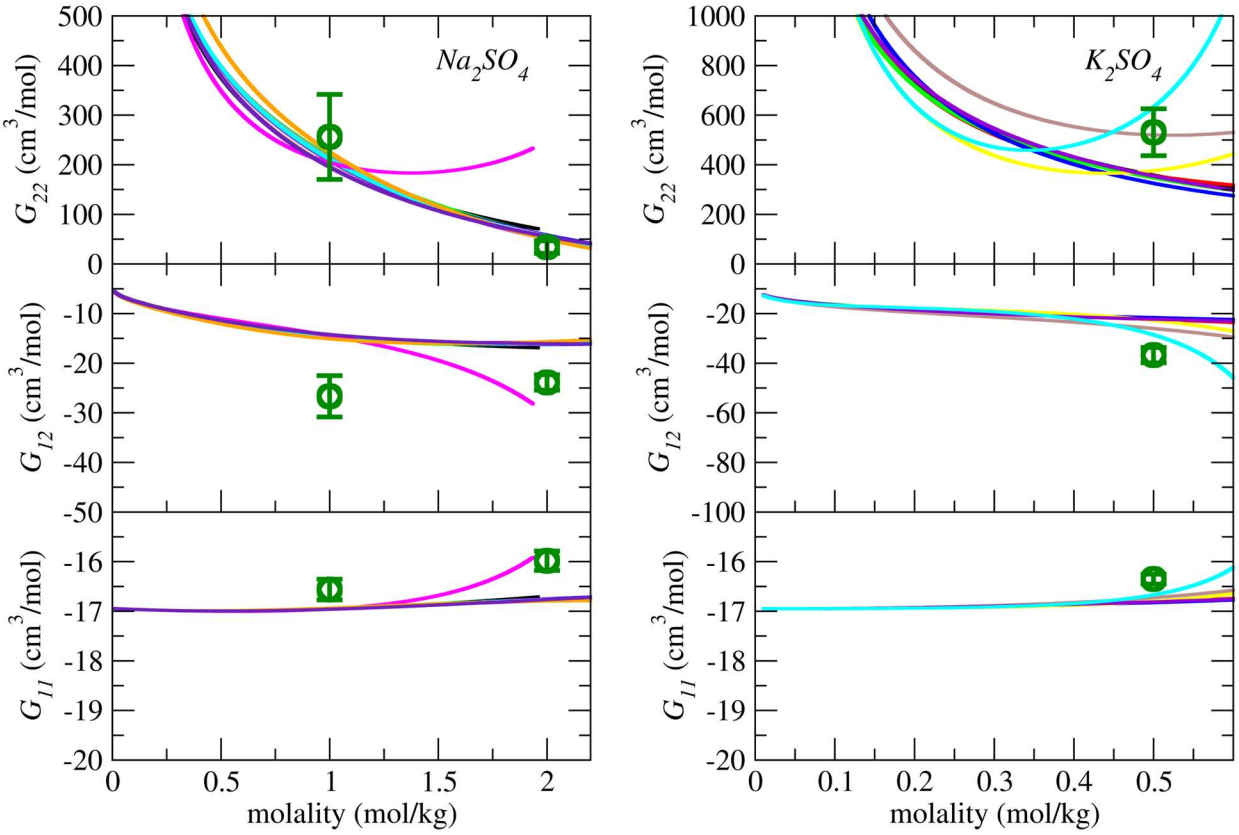


Figure 3.5 KB integrals for Na_2SO_4 (left) and K_2SO_4 (right) as function of concentration. All solid lines are experimental KBIs and deep green dot with error bars are KBIs from MD simulation.

The simulated and experimental partial molar volumes are displayed in Figure 3.6. The experimental solute partial molar volumes (PMV_2) of both Na_2SO_4 and K_2SO_4 solutions increased with concentration, and this trend was reproduced by the simulations. However, the simulated PMV_2 values were higher than experiment. This may be related to the much larger sulfate oxygen

parameters used here to avoid the aggregation of ions. The partial molar volume of solvent (PMV_1) does not change with concentration of both salts, and the experimental and simulation data are in good agreement.

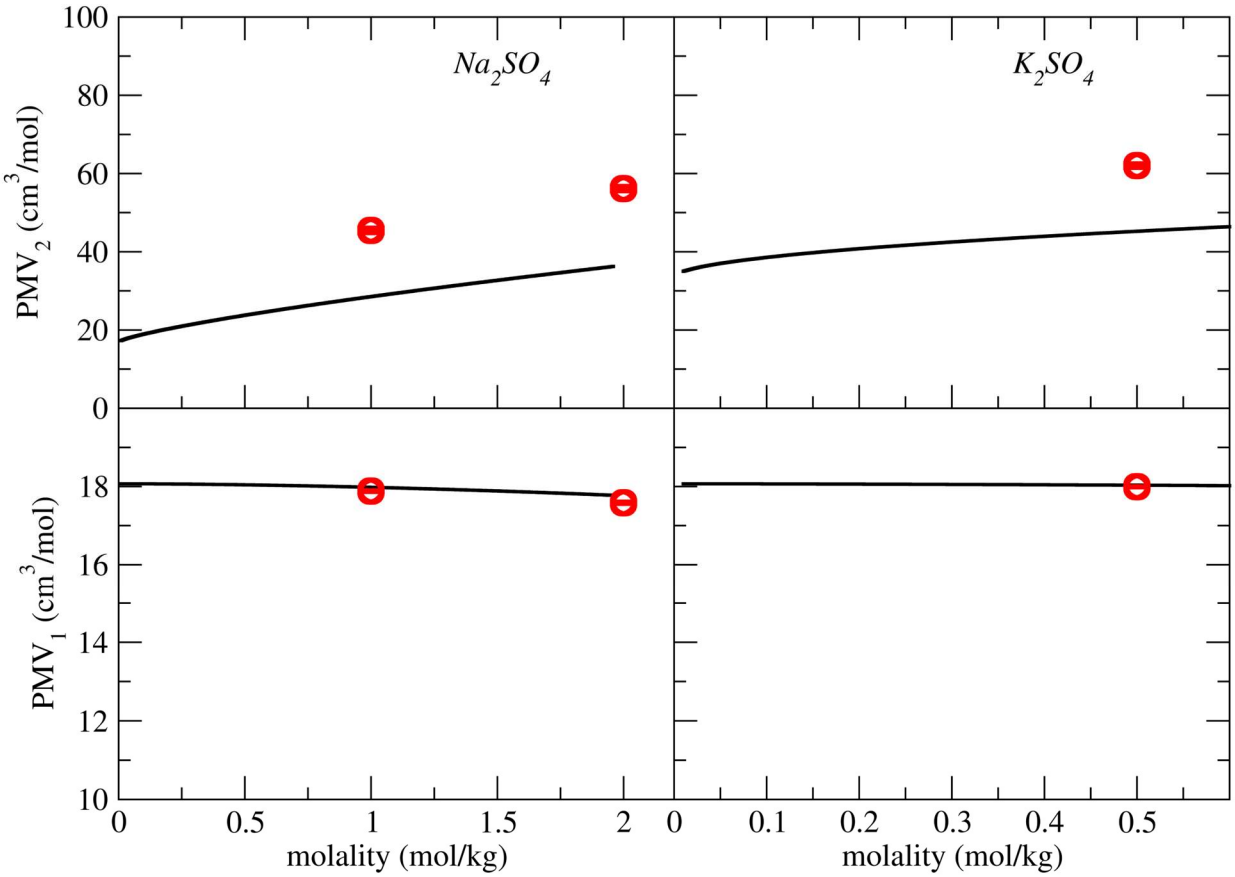


Figure 3.6 Solute (top) and solvent (bottom) partial molar volumes for Na_2SO_4 (left) and K_2SO_4 (right) solutions as function of concentration for. All solid lines are experimental data, while red circles and error bars correspond to simulated data.

Selected site-site RDFs obtained from the Na_2SO_4 and K_2SO_4 solution simulations are displayed in Figure 3.7. Interactions between the *Na/K* of Na_2SO_4/K_2SO_4 and the O of SO_4^{2-} (with OT atom types) are investigated, as well as interactions of OT with the water oxygen (OW). The solute concentration does not have a large effect on the results for Na_2SO_4 solutions. There are strong interactions between the Na and the OT based on the intensity of first shell peak. The corresponding K_2SO_4 first shell peak has an even higher intensity, which agrees with the simulated

KBI results (positive G_{22} values for both salts, and larger values for K_2SO_4 compared to Na_2SO_4). The first shell coordination peak for the interaction of OT with OW and S with OW in the present simulations appears at a further distance compared to AIMD results for the solvated sulfate anion.[14] This is probably the consequence of the increased oxygen atom size used here. Nevertheless, the same features are observed in both cases.

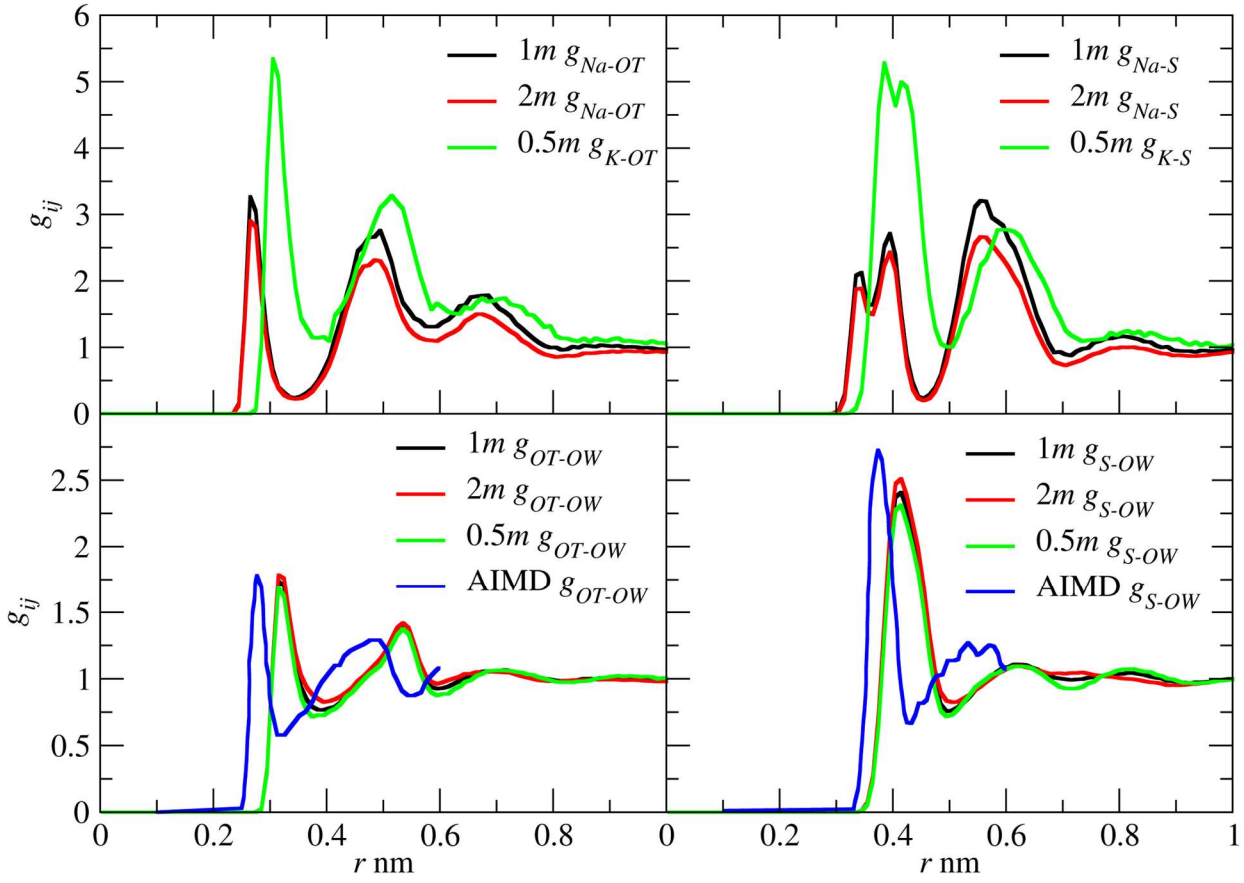


Figure 3.7. Radial distribution function for site-site interactions. Black solid line – 1m Na_2SO_4. Red solid line - 2m Na_2SO_4. Green solid line – 0.5m K_2SO_4. Blue solid line – *ab initio* MD simulation results.[14]

The stability of the simulations was sensitive to the force constant used for the O-S-O bending mode. To investigate this further we ran several simulations with different force constant values. Vibrational spectra were then obtained from the simulations and compared to the experimental Raman spectra of hydrated sodium sulfate,[68] as shown in Figure 3.8. The Raman

spectrum displays split peaks at 439, 451 and 613, 619, 642 cm^{-1} corresponding to the O-S-O vibrational mode. The simulation with a force constant of k_θ = 1200 kJ/mol displayed a broad peak at similar wavenumber to experiment. Hence, this value was used for all simulations. Interestingly, as the force constant increased the simulated sulfate vibrational mode gradually split into two different modes. However, we did not investigate this issue any further. Although there were significant differences in the vibrational frequencies for different k_θ values, there was no difference in the KBI results (data not shown).

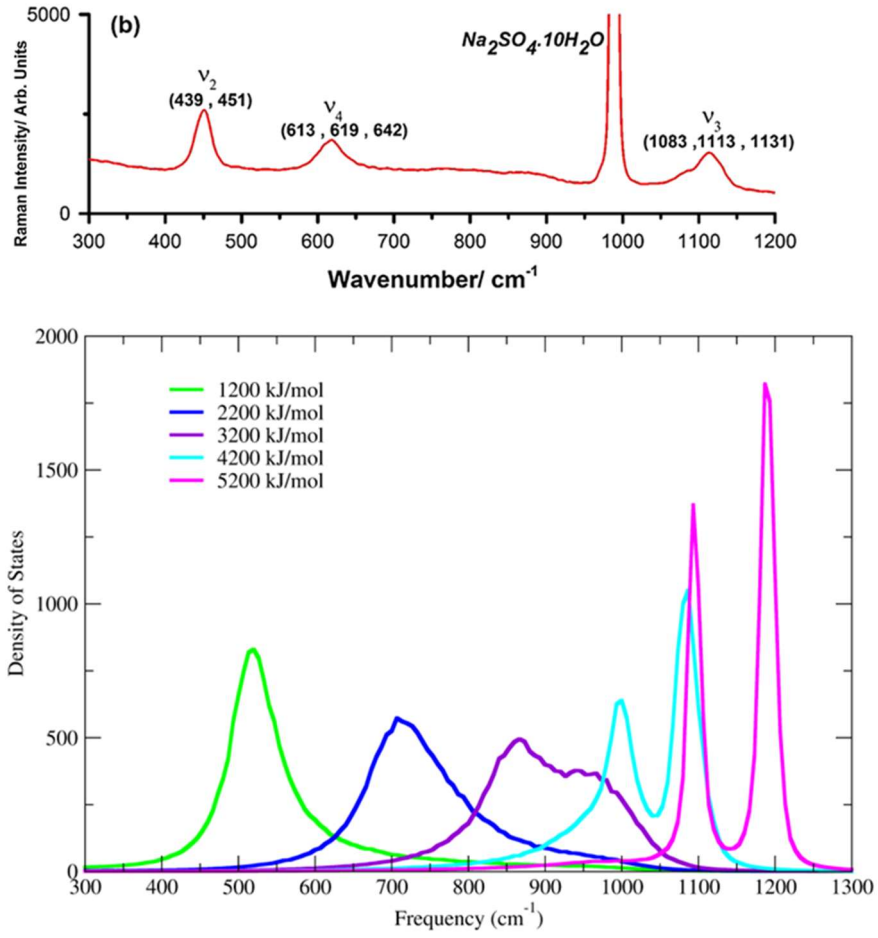


Figure 3.8. Raman (top) and simulated (bottom) spectra of Na_2SO_4 (top) solutions obtained with different O-S-O angle force constants. Experimental spectra (top) was reprinted with permission from Ben Mabrouk, K.; Kauffmann, T. H.; Aroui, H.; Fontana, M. D. Raman study of cation effect on sulfate vibration modes in solid state and in aqueous solutions. Journal of Raman Spectroscopy 2013, 44 (11), 1603-1608. DOI: 10.1002/jrs.4374.Copyright John Wiley & Sons, Inc.(2013).[68]

3.3.2 Methyl sulfate

The overall net charge of an alkyl sulfate anion is -1, which is not as high a charge density as the sulfate ion. We started with an initial charge distribution adopted from Ferro *et al.*[52] and redistributed as final charged as shown in Figure 3.9. Effective partial charge of methyl sulfate charge distribution is not as sensitive as sulfate ion but still optimized from initial charge. Final optimized charge distribution is displayed in Figure 3.9. Once we optimized the effective charge distribution, we tested the LJ σ parameter obtained for OT in the sulfate ion, but decided to keep parameters developed for a terminally charged oxygen in KBFF20 for use in methyl sulfate, as summarized in Table 3.2.

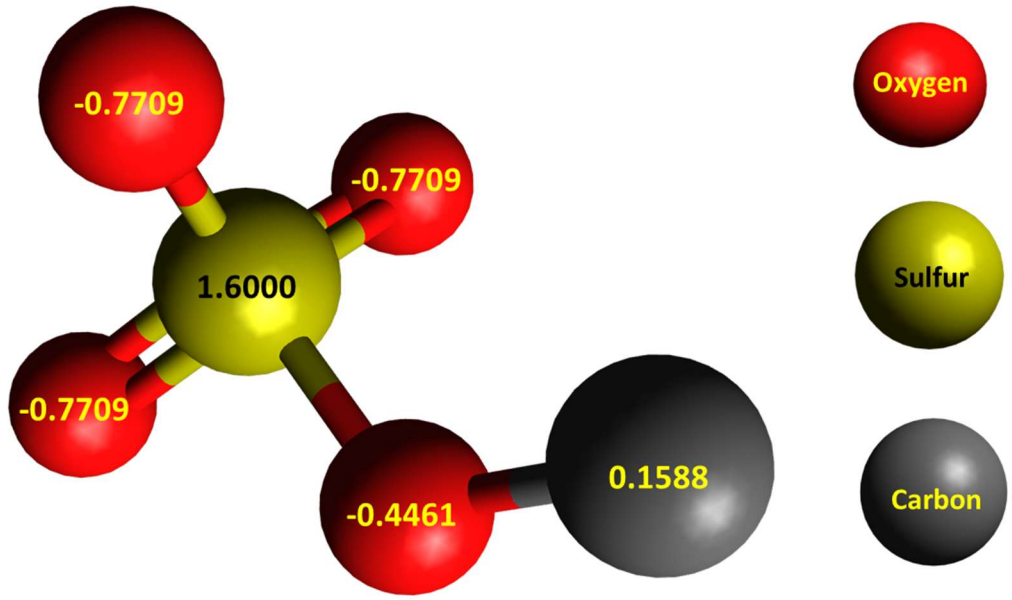


Figure 3.9 Final effective charge distribution for the methyl sulfate ion.

The RDFs obtained from simulations of sodium and potassium methylsulfate mixtures are provided in Figure 3.10. As expected, the simulations demonstrate less intense peaks compared to those for the corresponding sulfate ion in Figure 3.5. However, the g_{21} RDFs first shell peak for methyl sulfate is of a similar intensity to sulfate ion in Figure 3.5. Sodium methyl sulfate g_{21} has a higher first peak intensity than potassium methyl sulfate g_{12}. The solvent RDFs g_{11} for both sodium

and potassium methyl sulfate are very similar to each other. Simulations have been performed up to 4m for both sodium potassium sulfate solutions. However, concentration does not affect the solute or solvent distributions for either of the methyl sulfate solutions as shown Figure 3.10.

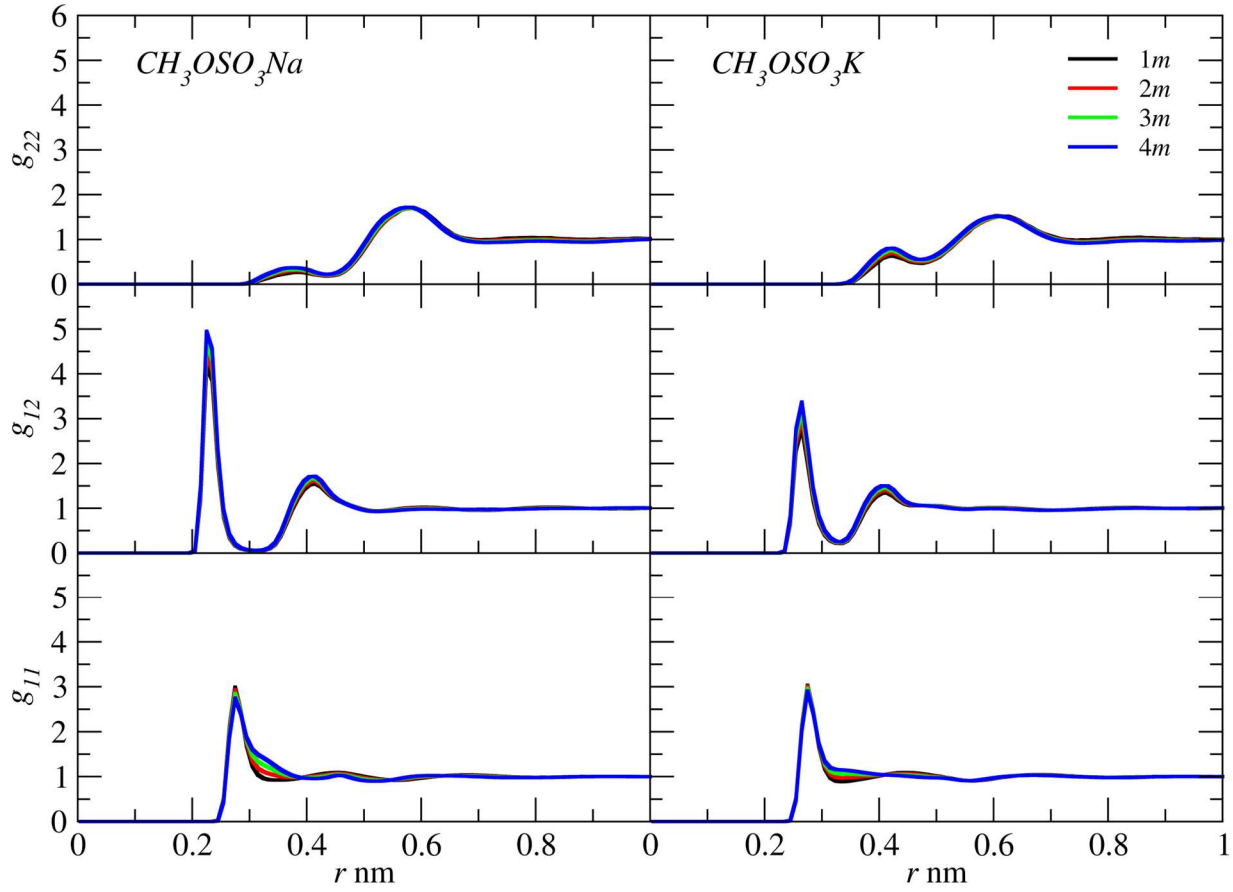


Figure 3.10 RDFs for sodium methyl sulfate (*CH_3OSO_3Na*, right) and potassium methyl sulfate (*CH_3OSO_3K*, left) aqueous solutions.

The negative G22 KBIs for the methyl sulfates at higher concentrations indicates that the alkyl sulfates are more hydrated than the corresponding sulfate salts (G22>0, Figure 3.5), as shown in Figure 3.11. The simulated G22 values for sodium methyl sulfate reproduced the experimental KBI values very well, but for potassium methyl sulfate solutions the simulated G22 values are slightly underestimated. Unfortunately, we could not find a charge distribution that improved the KBIs for the potassium salts while maintaining good agreement for the sodium salts. The simulated

G12=G21 values for both sodium and potassium methyl sulfate increase with increases in concentration. The 1m and 2m sodium methyl sulfate G12 values are lower than experiment, but the 3m and 4m simulated G12 values are in good agreement. The 1m simulation G12 value for potassium methyl sulfate is similar to the experimental value, but then slowly deviates from experiment for 2m, 3m, and 4m solutions, which provide higher than experimental values. In summary, the agreement with experiment for the KBIs is good for sodium methyl sulfate, while the results for potassium methyl sulfate are not as good but considered acceptable.

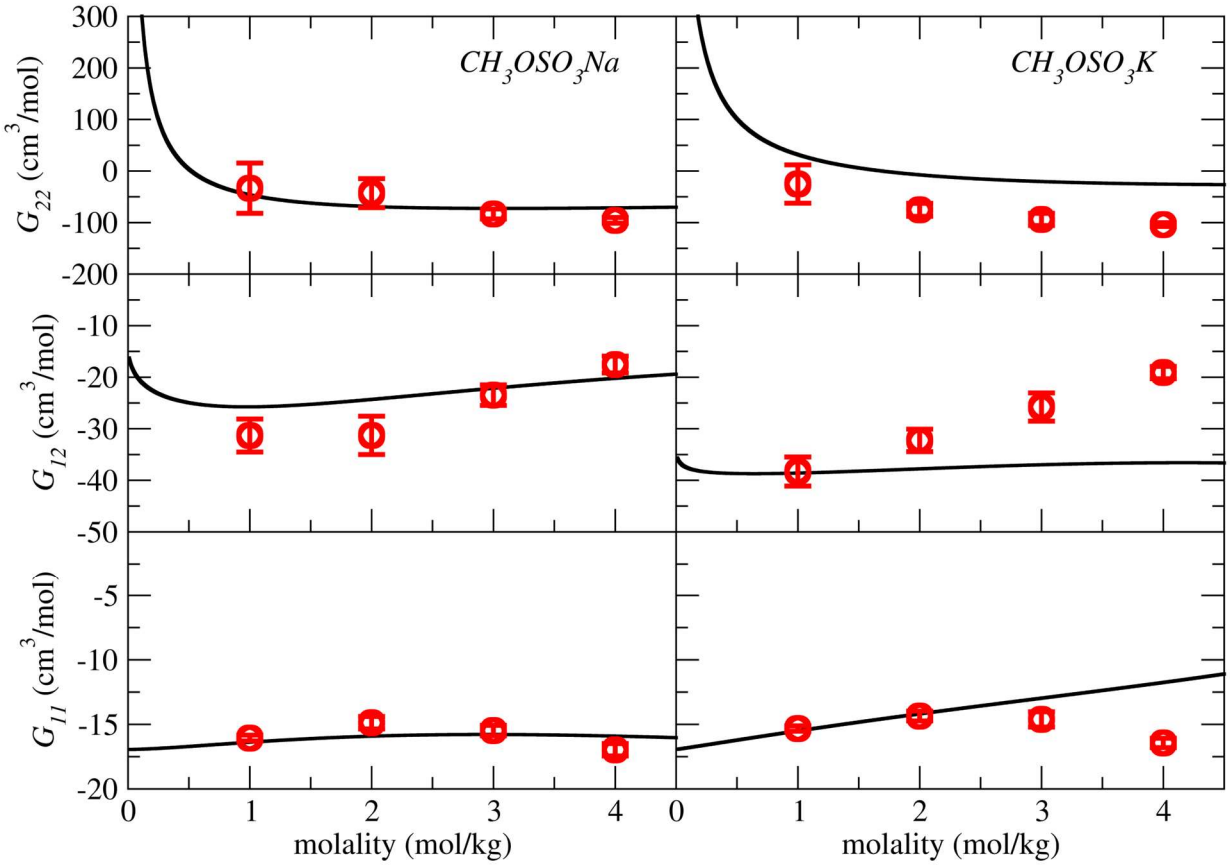


Figure 3.11 KB integrals for *CH₃OSO₃Na* (left) and *CH₃OSO₃K* (right) solutions as function of concentration. Black solid lines are experimental KBIs and red circles with error bars are simulated KBIs.

The PMVs for both solution mixtures are displayed in Figure 3.12. The solute PMVs were slightly overestimated in the simulations. The deviation from experiment was smaller than for the

sulfate salts due to the use of a smaller oxygen size in the alkyl sulfates. It should be noted that the experimental PMVs for the potassium methylsulfate solutions are only approximate as composition dependent density data was unavailable. Hence, the PMV of potassium methyl sulfate was held fixed according to the experimental crystal densities of the pure potassium sulfate salt. Therefore, the experimental PMV_2 remains constant for that reason. The simulated water PMV values were in good agreement with experiment.

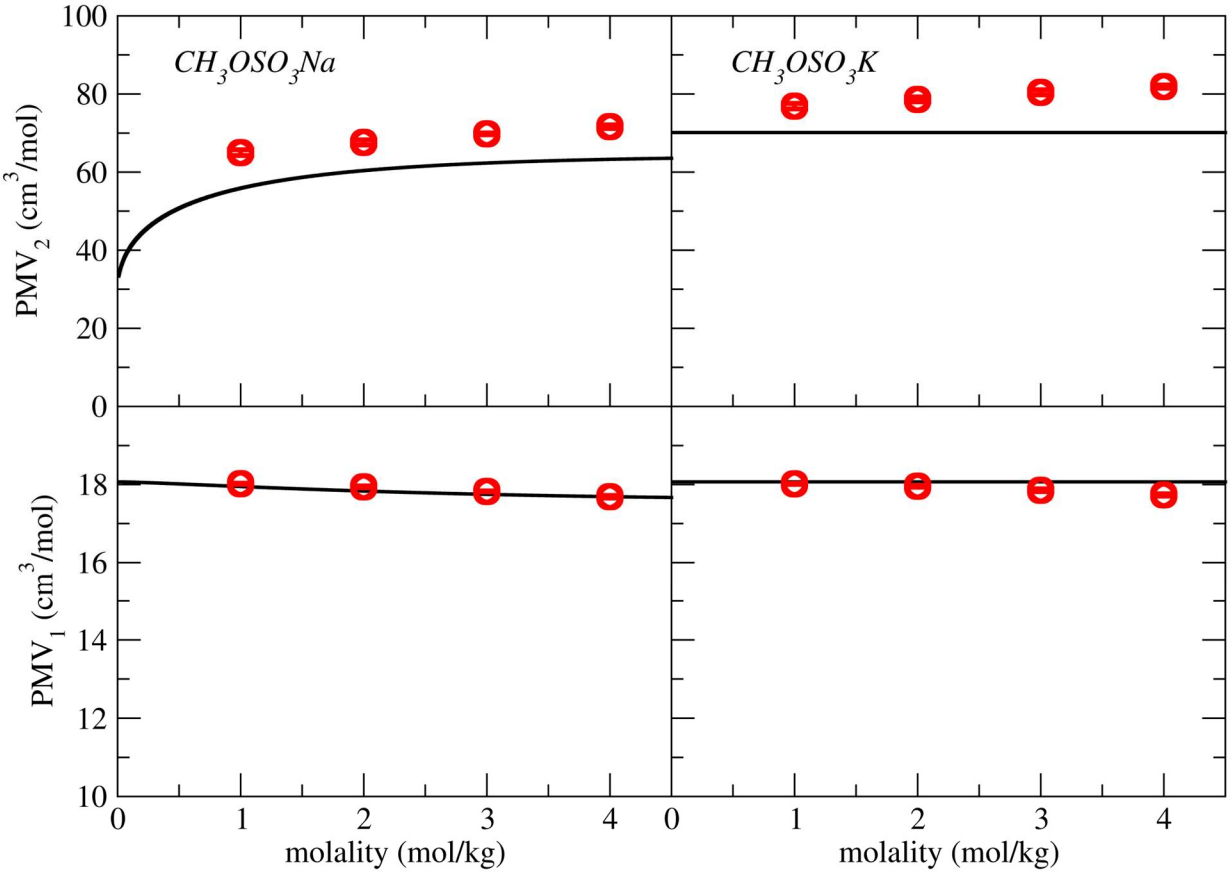


Figure 3.12 PMVs for *CH_3OSO_3Na* (left) and *CH_3OSO_3K* (right) as function of concentration. Black solid lines are experimental PMVs and red dots with error bars are PMVs from MD simulation.

3.3.3 Ethyl sulfate

Transferability of non-bonded parameters developed for methyl sulfate were tested for use with sodium and potassium ethyl sulfate. Since ethyl sulfate is also an alkyl sulfate, we simply

adopted the same methyl sulfate parameters. The final charge was assigned as Figure 3.13. All charges determined for methyl sulfate were transferred to ethyl sulfate, and the additional methyl group hydrophobic tail charge was set to zero.

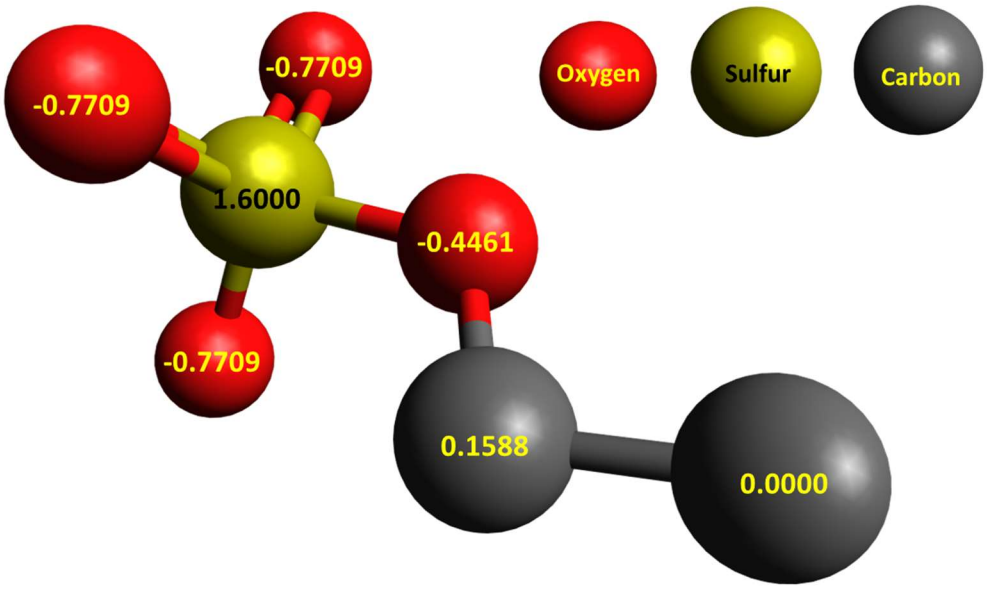


Figure 3.13 Final effective charge distribution for ethyl sulfate.

All the RDFs for ethyl sulfate solutions demonstrated very similar results to those of the methyl sulfates, as shown in Figure 3.14. All RDFs were insensitive to concentration just like the sulfate and methyl sulfate solutions.

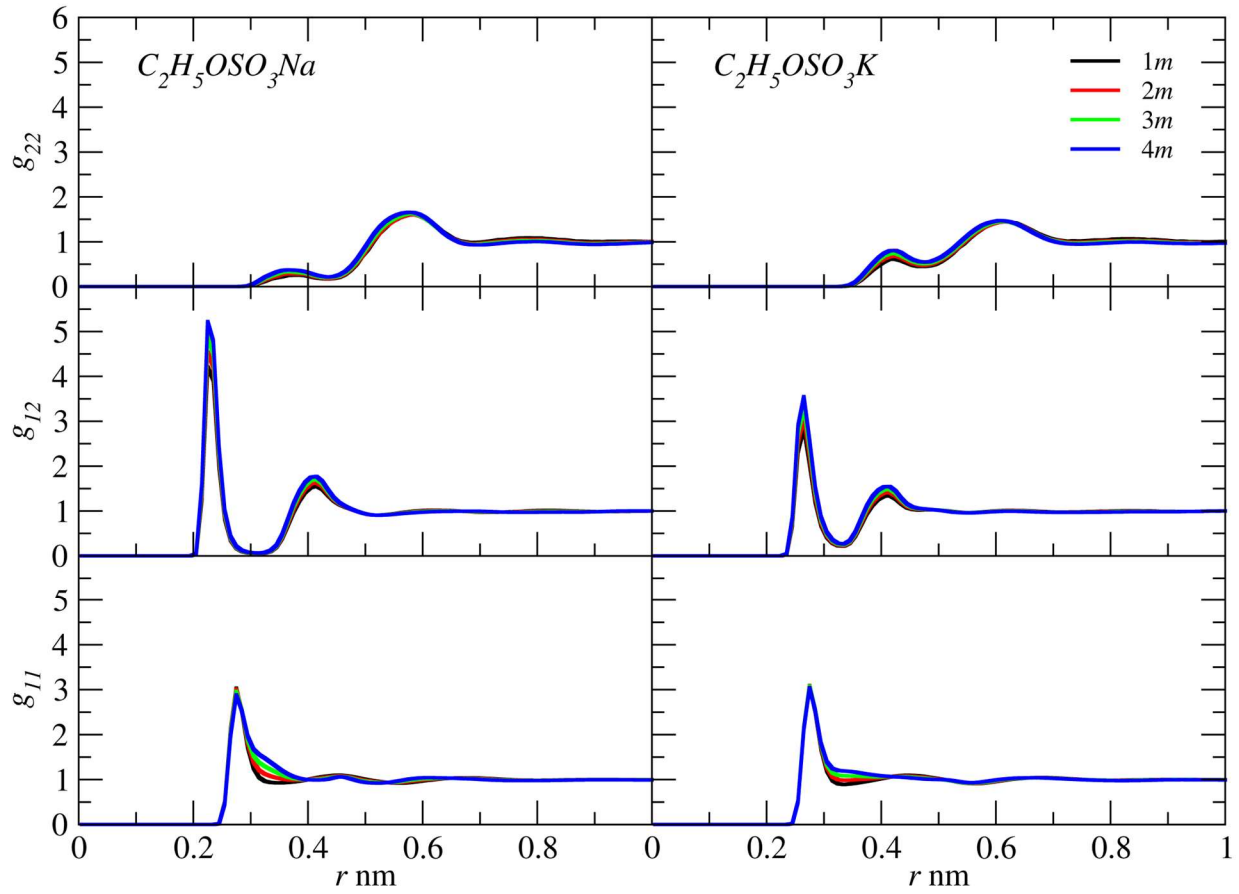


Figure 3.14 RDFs for $C_2H_5OSO_3Na$ (left) and $C_2H_5OSO_3K$ (right) aqueous solutions as a function of distance.

The G_{22} value was overestimated for lower concentrations (1m, 2m) of sodium ethyl sulfate, but higher concentrations are well reproduced, as shown in Figure 3.15. Whereas the potassium ethyl sulfate results are good for the 1m concentration, they become slightly more underestimated for higher concentrations. The overall trend for both ethyl sulfates involved a small decrease in G_{22} with increasing concentration. The simulated G_{12}=G_{21} values increased with increasing concentration for both sodium and potassium ethyl sulfates. Both G_{11} values increased with increasing concentration for all ethyl sulfate solutions and show good agreement between experiment and simulation results.

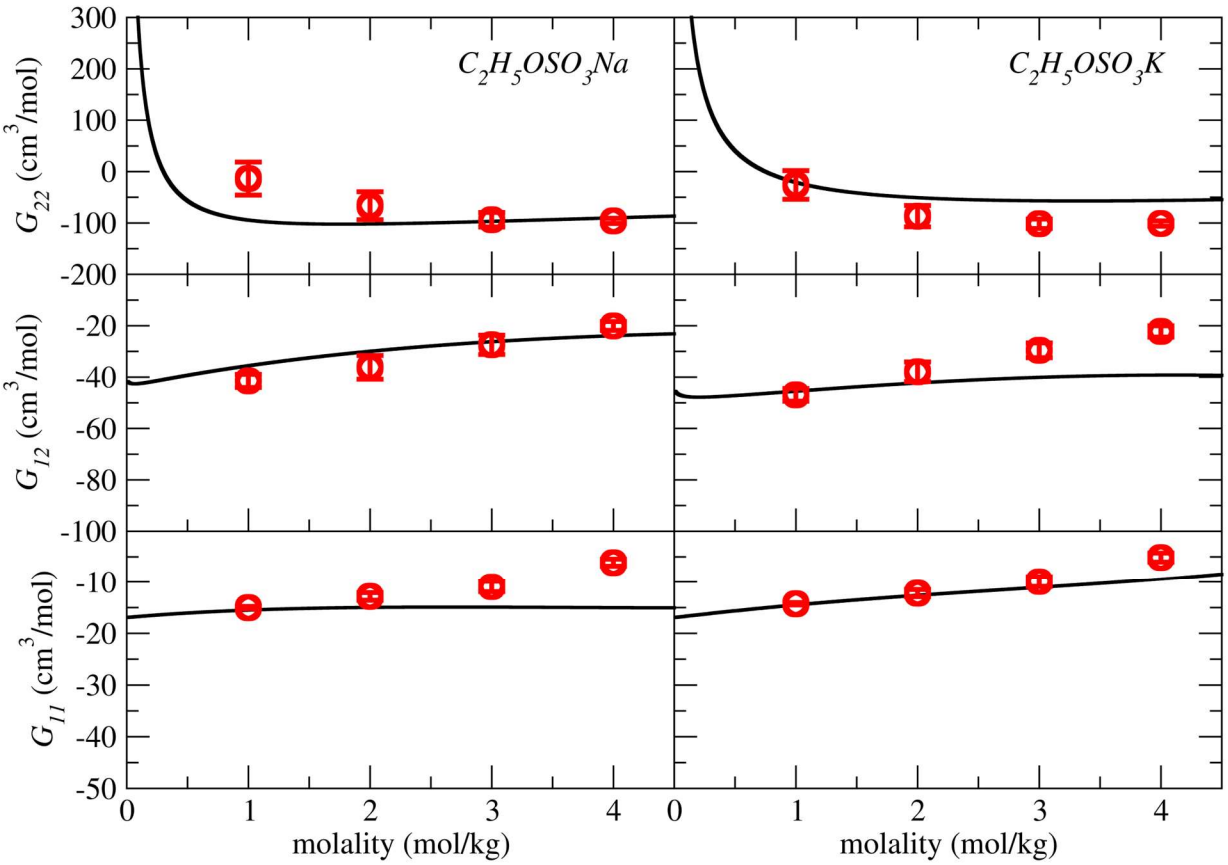


Figure 3.15 KBIs for sodium ethyl sulfate (left) and potassium ethyl sulfate as function of concentrations. Black solid line is experimental KBIs and red circles with error bars are simulated KBIs.

The PMVs of ethyl sulfate solutions were held fixed according to the experimental pure water and pure crystal densities of potassium ethyl sulfate salts, because the experimental concentration dependent volumetric data was not available in the literature. These estimates and the simulated PMVs are displayed in Figure 3.16. The results were reasonable.

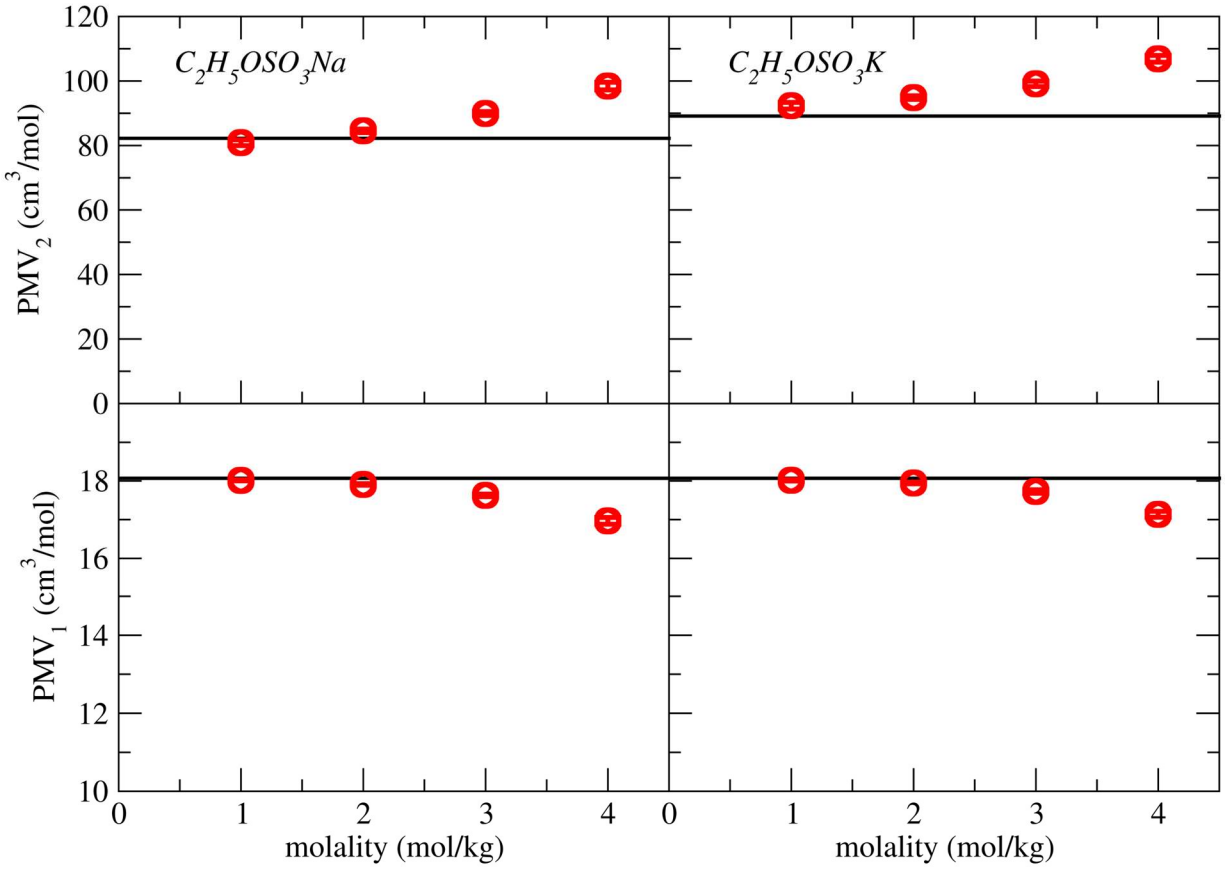


Figure 3.16 Partial molar volumes of sodium ethyl sulfate (left) and potassium ethyl sulfate (right) solutions. Black solid lines are PMVs from experiment and red circle with error bars are simulated PMVs.

The methyl sulfate anion has one rotatable bond between O-S, while the ethyl sulfate anion has two rotatable bonds between O-S and C-O. The dihedral potentials for rotation around these bonds needed to be determined. These intramolecular potentials do not affect the simulated intermolecular (KBI) results. Our target data was gas phase quantum mechanical calculations performed for the C-O-S-O dihedral angle. This is a three-fold potential with energy minima at ±60 and 180 degrees, and an energy barrier between 8.4-11.3 kJ/mol.[62, 63] Our final results for barrier height were ~ 11.3kJ/mol for both sodium and potassium methyl sulfate, and ~8.3kJ/mol for both sodium and potassium ethyl sulfate, as shown in Figure 3.17. The concentration of the solute does not affect the results.

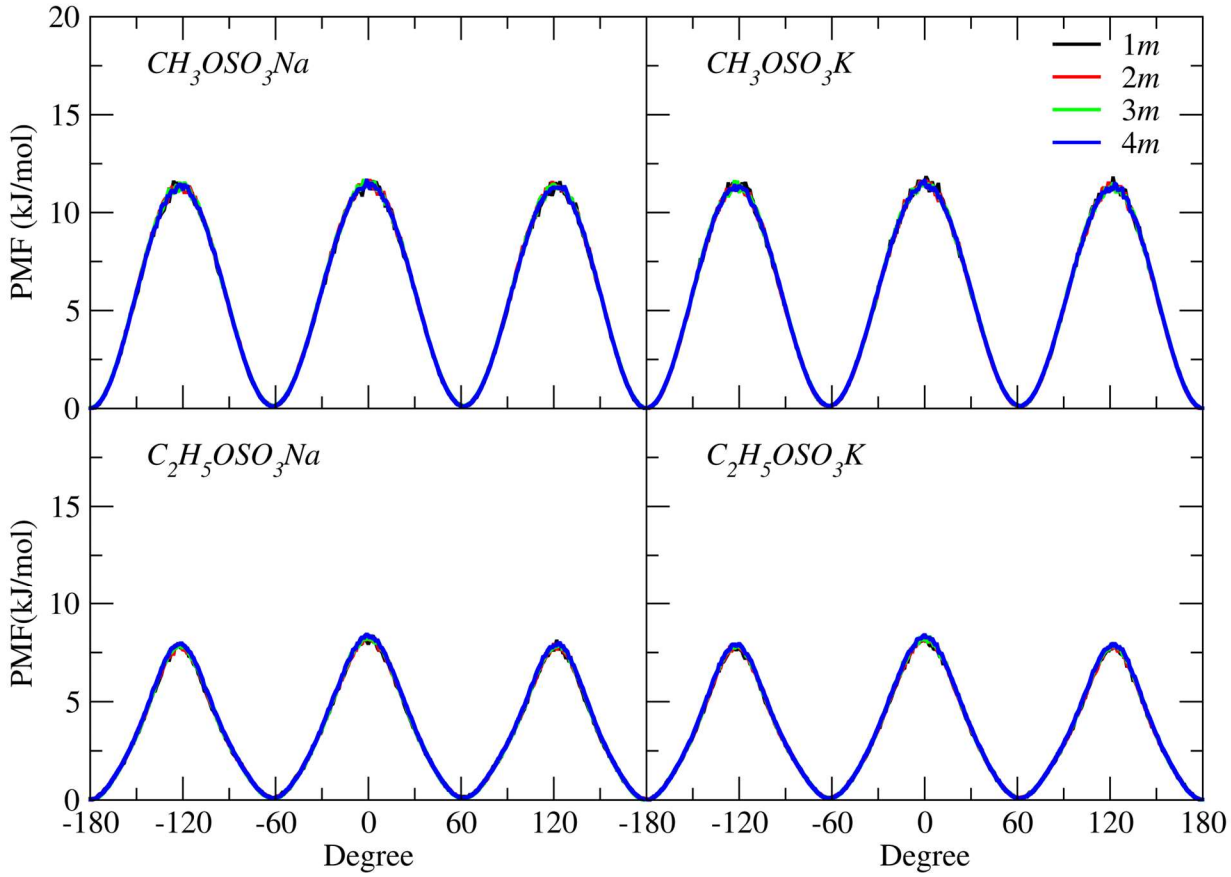


Figure 3.17 Simulation results for the potential of mean force for rotation around the C-O-S-O dihedral for sodium/potassium methyl/ethyl sulfate solutions at 298K.

The C-C-O-S dihedral potential for ethylsulfates was also optimized to match a DFT calculation for the gas phase.[64] The results are shown in Figure 3.18. Good agreement with the QM surface was obtained except for the region around ±90 degrees. A summary of the final dihedral parameters is provided in Table 3.3.

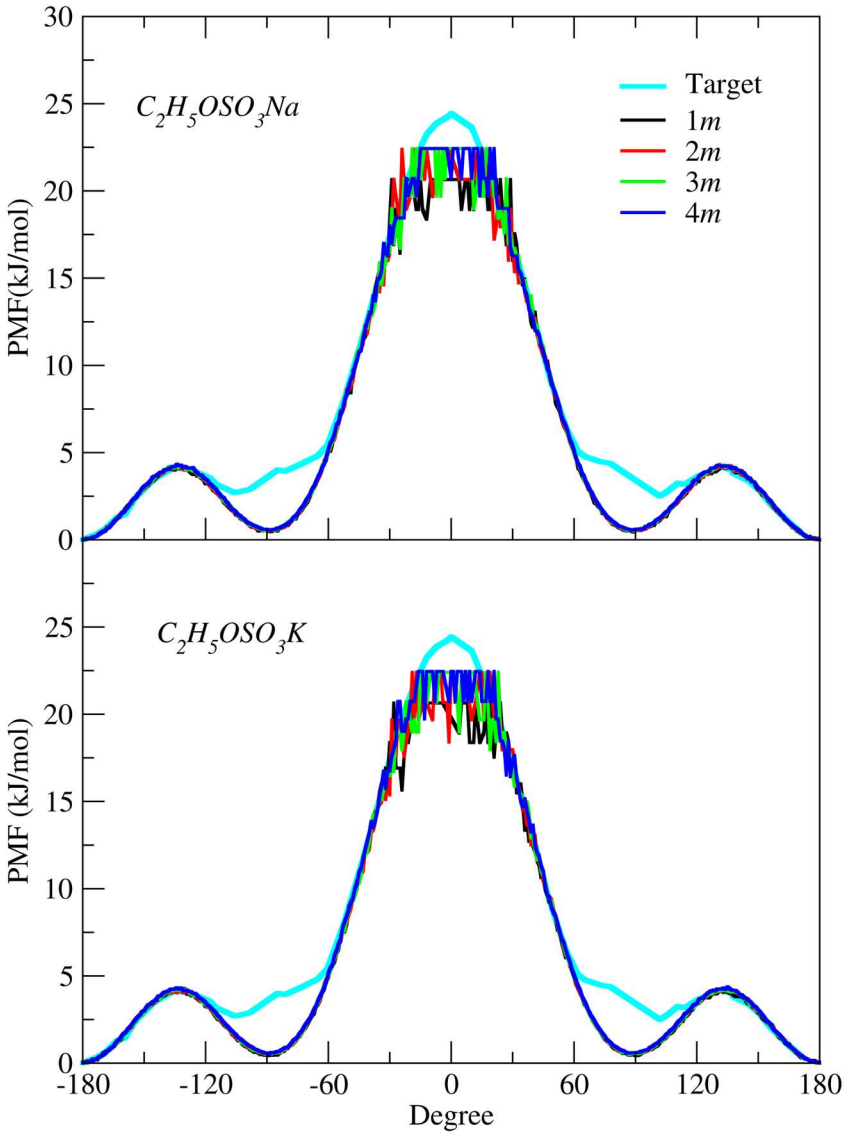


Figure 3.18 Simulation results for the C-C-O-S dihedral potential for sodium/potassium ethyl sulfate solutions at 298K.

3.3.4 SDS

All the KBIs results for methyl sulfate solutions were very similar to the results for ethyl sulfate solutions, thus we decided to transfer all the non-bonded parameters form ethyl sulfate to form a parameter set for SDS. Additional bond, angle and dihedral parameters were taken from KBFF20.[7] The dihedral angle parameters are summarized in Table 3.3.

Testing a model for SDS is difficult. The SDS micelle aggregation number is reported as 60,[23] with an experimental CMC concentration of 8mM. Hence, it is almost impossible to achieve the CMC concentration with 60 SDS due to the limitation in the size of simulation boxes. As an alternative, we performed two simulations of SDS to ensure we observed reasonable behavior. First, a micelle was preassembled with 60 SDS and simulated 100 ns in a 10 nm cubic box. Second, 60 randomly placed SDS were placed in the same size box (~105 mM) and simulated for 1 μs. The initial and final configurations are provided in Figure 3.19. The preassembled micelle maintained the same micelle size and shape. The randomly placed SDS molecules aggregated and formed micelles that gradually increased in size over the length of the simulation. This suggests the current force field is accurate enough to favor micelle formation even if one cannot determine a simulated value for the CMC.

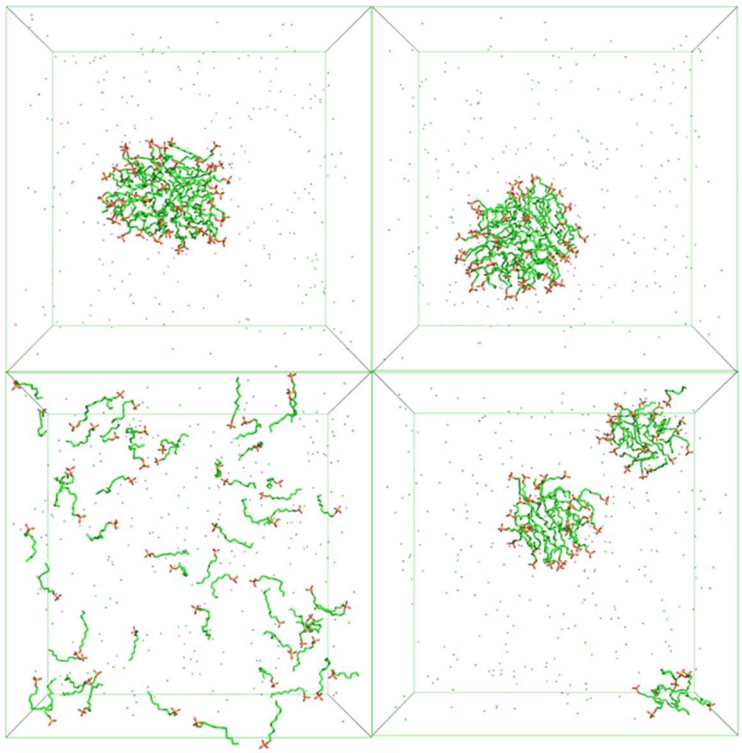


Figure 3.19 Initial (top left) and final (top right) configurations for a preassembled micelle with 60 SDS simulated for 100 ns. Initial (bottom left) and final (bottom right) configurations after 400ns of simulation starting from 60 randomly placed SDS molecules.

There is no experimental data available to calculate KBIs for SDS due to its low solubility. However, there is surface tension data for the vapor-liquid interface. Thus, we have also attempted to validate our parameters by comparing experimental and simulated surface tensions as shown in Figure 3.20. Surfactant action is caused in part by a dramatic reduction in surface tension. Experimentally, the surface tension declines only slightly when the occupancy of the air/water interface increases from 0 to 60% of the maximum, followed by a sudden drop in surface tension after about 60% surface coverage.[69] The simulated surface tension results underestimated the surface tension changes, but the trend and sudden drop are well reproduced, especially considering possible issues determining the exact surface coverage. Menger *et al.*[69] established the surface excess (Γ), which is equivalent to the interfacial concentration units of mol/cm² as shown in equation 3.3,

$$\Gamma = -\frac{d\gamma/d(lnc)}{nRT} \qquad 3.2$$

Here, γ is the surface tension and c is the bulk surfactant concentration. The parameter *n* which represents the stoichiometric coefficient, accounts for the change in the number of moles during the reaction or process being described. The percent surface coverage was obtained from Equation 3.3, and full saturation was reported to occur at 4.8×10^{-10} mol/cm².[69] Kawai *et al.*[60] used infrared external reflection spectroscopy to measure surface coverage, and their result was 0.48 nm² per SDS. This result is similar to the values obtained by Tajima *et al.*,[61] which was 0.52 nm² per SDS. Kawai and Tajima's results estimated a slightly bigger surface area per SDS compared to Menger's (~0.35nm^2). In our simulations, we used Kawai and Tajima's value to prepare the system, which may lead to a difference in the results for the relationship between surface coverage and surface tension. Also, the SPC/E water model provides a lower surface tension compared to pure water,[70, 71] which may also affect the difference between the simulation and experimental results.

The surface tension, γ, was calculated from our simulations as,

$$\gamma = \frac{1}{2} L_z \left[\langle P_{zz} \rangle - \frac{1}{2} \left(\langle P_{xx} \rangle + \langle P_{yy} \rangle \right) \right] \qquad 3.3$$

where L_z is the size of the box along the direction z, perpendicular to the vacuum/water interface. The factor 1/2 outside the bracket considers the fact that there are two interfaces in the system. The quantities P_{xx}, P_{yy}, and P_{zz} are components of the pressure tensor along the x, y, and z direction.

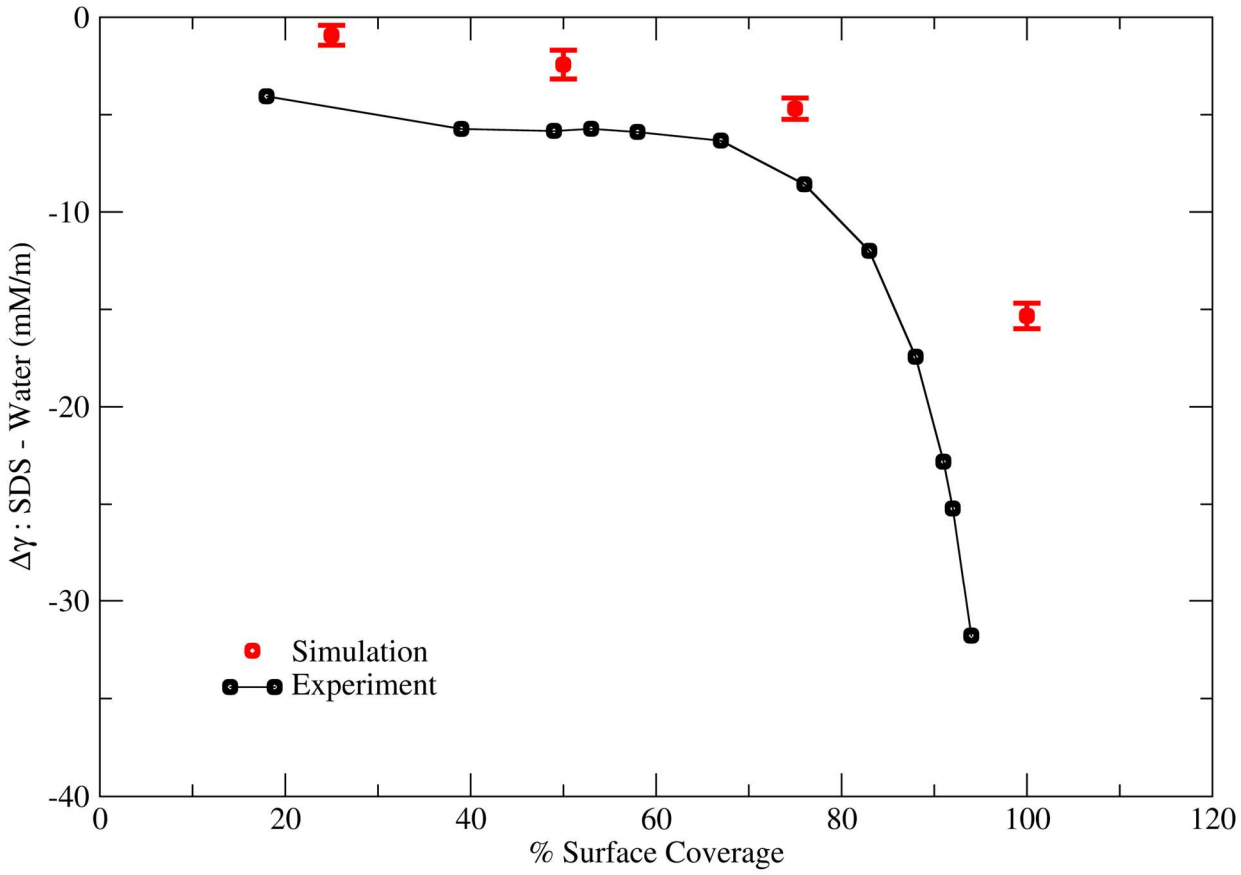


Figure 3.20 Relationship between surface tension changes and surface coverage.

Finally, we preassembled a micelle with 80 SDS molecules and simulated it for 100 ns. The micelle remained intact throughout the simulation. We then simulated a micelle with 80 SDS molecules and an additional 10 randomly placed SDS molecules. During this simulation, the random SDS molecules either aggregated with themselves or became part of the existing micelle. This type of behavior is expected when the concentration of SDS is higher than its critical micelle

concentration, as multiple micelle formation or elongated shapes have been observed at higher SDS concentrations.[29, 72, 73] Our simulation results, shown in Figure 3.21, are therefore consistent with this experimental trend.

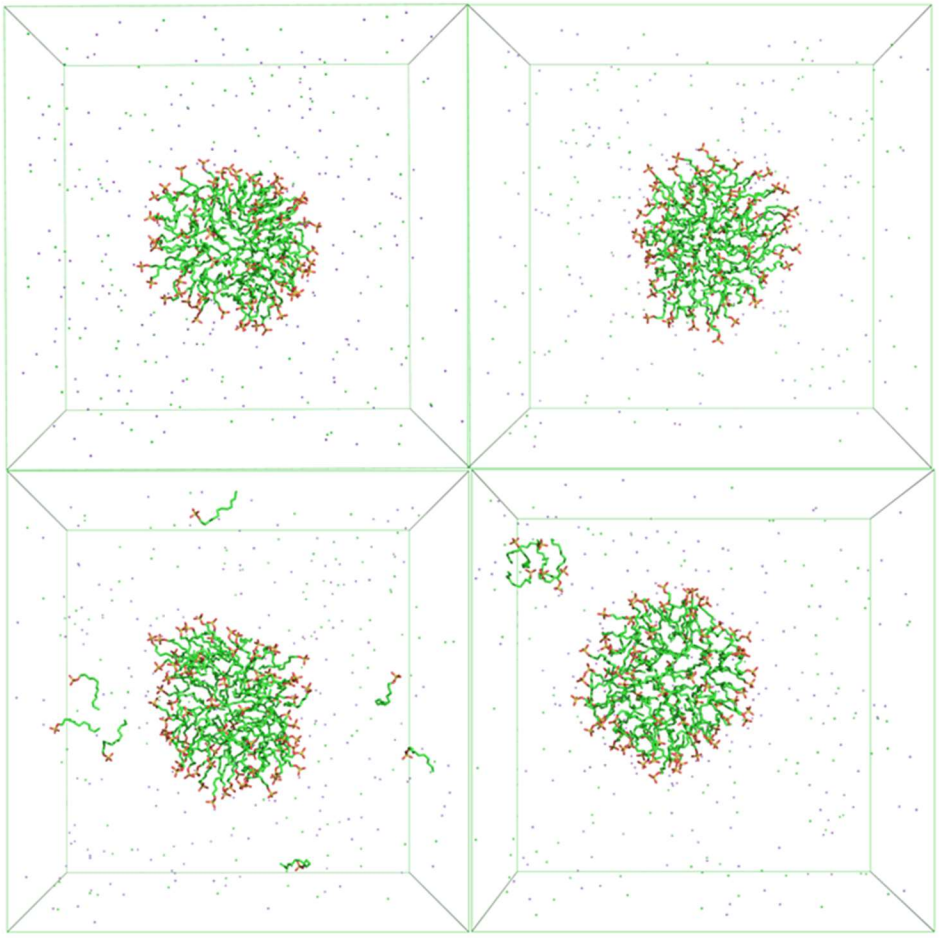


Figure 3.21 The initial configuration of a micelle with 80 SDS molecules (top left). The top right panel displays the micelle configuration after a 100 ns simulation. The bottom left panel shows the initial configuration of a micelle with 80 SDS molecules and an additional 10 randomly placed SDS molecules. Finally, the bottom right panel displays the configuration of the micelle with 80 SDS molecules and 10 random SDS molecules after a 100 ns simulation.

3.4 Conclusions

Sulfate, alkylsulfate and SDS force field parameters have been developed. The parameters reproduce the KBIs for solutions of sodium and potassium sulfate, methylsulfate and ethylsulfate salts. The resulting SDS force field is shown to form stable micelles, even though a CMC value

was not determined. The surface aggregation of SDS was investigated via simulation of surface tension changes with surface coverage. The model of SDS provided here gave qualitative agreement with experiment, which was considered acceptable given the difficulties establishing the exact degree of surface coverage.

3.5 Reference

(1) West, A. Intermolecular Forces and Solvation. In *Interface Science and Technology*, Ball, V. Ed.; Elsevier, 2018; pp 49-130.

(2) Ilyas, S. U.; Ridha, S.; Kareem, F. A. A. Dispersion stability and surface tension of SDS-Stabilized saline nanofluids with graphene nanoplatelets. *Colloids and Surfaces a-Physicochemical and Engineering Aspects* **2020**, *592*. DOI: 10.1016/j.colsurfa.2020.124584.

(3) Saien, J.; Fadaei, V. The study of interfacial tension of kerosene-water under influence of CTAB surfactant and different size silica nanoparticles. *Journal of Molecular Liquids* **2018**, *255*, 439-446. DOI: 10.1016/j.molliq.2018.01.120.

(4) Tulumello, D. V.; Deber, C. M. SDS Micelles as a Membrane-Mimetic Environment for Transmembrane Segments. *Biochemistry* **2009**, *48* (51), 12096-12103. DOI: 10.1021/bi9013819.

(5) Nanga, R. P. R.; Brender, J. R.; Vivekanandan, S.; Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochimica Et Biophysica Acta-Biomembranes* **2011**, *1808* (10), 2337-2342. DOI: 10.1016/j.bbamem.2011.06.012.

(6) Bénarouche, A.; Habchi, J.; Cagna, A.; Maniti, O.; Girard-Egrot, A.; Cavalier, J.-F.; Longhi, S.; Carrière, F. Interfacial Properties of NTAIL, an Intrinsically Disordered Protein. *Biophysical Journal* **2017**, *113* (12), 2723-2735.

(7) Ploetz, E. A.; Karunaweera, S.; Bentenitis, N.; Chen, F.; Dai, S.; Gee, M. B.; Jiao, Y. F.; Kang, Y. S.; Kariyawasam, N. L.; Naleem, N.; et al. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 2964-2990. DOI: 10.1021/acs.jctc.1c00075.

(8) Priyamvada, S.; Saksena, S.; Alrefai, W. A.; Dudeja, P. K. Intestinal Anion Absorption. In *Physiology of the Gastrointestinal Tract*, 6 ed.; Said, H. M. Ed.; Academic Press, 2018; pp 1317-1362.

(9) Kang, B. B.; Tang, H. C.; Zhao, Z. D.; Song, S. S. Hofmeister Series: Insights of Ion Specificity from Amphiphilic Assembly and Interface Property. *Acs Omega* **2020**, *5* (12), 6229-6239. DOI: 10.1021/acsomega.0c00237.

(10) Markovich, D. Physiological roles and regulation of mammalian sulfate transporters. *Physiological Reviews* **2001**, *81* (4), 1499-1533. DOI: 10.1152/physrev.2001.81.4.1499.

(11) Cerutti, D. S.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P. Simulations of a Protein Crystal: Explicit Treatment of Crystallization Conditions Links Theory and Experiment in the Streptavidin-Biotin Complex. *Biochemistry* **2008**, *47* (46), 12065-12077. DOI: 10.1021/bi800894u.

(12) Mason, P. E.; Dempsey, C. E.; Neilson, G. W.; Brady, J. W. Nanometer-scale ion aggregates in aqueous electrolyte solutions: Guanidinium sulfate and guanidinium thiocyanate. *Journal of Physical Chemistry B* **2005**, *109* (50), 24185-24196. DOI: 10.1021/jp052799c.

(13) Chialvo, A. A.; Simonson, J. M. SOLVATION AND ION PAIR ASSOCIATION IN AQUEOUS METAL SULFATES: INTERPRETATION OF NDIS RAW DATA BY ISOBARIC-ISOTHERMAL MOLECULAR DYNAMICS SIMULATION. *Collection of Czechoslovak Chemical Communications* **2010**, *75* (4), 405-424. DOI: 10.1135/cccc2009535.

(14) Pegado, L.; Marsalek, O.; Jungwirth, P.; Wernersson, E. Solvation and ion-pairing properties of the aqueous sulfate anion: explicit versus effective electronic polarization. *Physical Chemistry Chemical Physics* **2012**, *14* (29), 10248-10257. DOI: 10.1039/c2cp40711f.

(15) Wernersson, E.; Jungwirth, P. Effect of Water Polarizability on the Properties of Solutions of Polyvalent Ions: Simulations of Aqueous Sodium Sulfate with Different Force Fields. *Journal of Chemical Theory and Computation* **2010**, *6* (10), 3233-3240. DOI: 10.1021/ct100465g.

(16) Xu, J. X.; Fang, Y.; Fang, C. H. Structure of aqueous sodium sulfate solutions derived from X-ray diffraction. *Chinese Science Bulletin* **2009**, *54* (12), 2022-2027. DOI: 10.1007/s11434-009-0232-1.

(17) Mamatkulov, S. I.; Rinne, K. F.; Buchner, R.; Netz, R. R.; Bonthuis, D. J. Water-separated ion pairs cause the slow dielectric mode of magnesium sulfate solutions. *Journal of Chemical Physics* **2018**, *148* (22). DOI: 10.1063/1.5000385.

(18) Zeron, I. M.; Abascal, J. L. F.; Vega, C. A force field of Li+, Na+, K+, Mg2+, Ca2+, Cl-, and SO42-in aqueous solution based on the TIP4P/2005 water model and scaled charges for the ions. *Journal of Chemical Physics* **2019**, *151* (13). DOI: 10.1063/1.5121392.

(19) Zhang, Q. G.; Wang, N. N.; Yu, Z. W. The Hydrogen Bonding Interactions between the Ionic Liquid 1-Ethyl-3-Methylimidazolium Ethyl Sulfate and Water. *Journal of Physical Chemistry B* **2010**, *114* (14), 4747-4754. DOI: 10.1021/jp1009498.

(20) Walker, J. M. SDS Polyacrylamide Gel Electrophoresis of Proteins. In *The Protein Protocols Handbook.* , 2 ed.; Walker, J. M. Ed.; Humana Press., 2002.

(21) Kosol, S.; Schrank, E.; Krajacic, M. B.; Wagner, G. E.; Meyer, N. H.; Gobl, C.; Rechberger, G. N.; Zangger, K.; Novak, P. Probing the Interactions of Macrolide Antibiotics with Membrane-Mimetics by NMR Spectroscopy. *Journal of Medicinal Chemistry* **2012**, *55* (11), 5632-5636. DOI: 10.1021/jm300647f.

(22) Mazer, N. A.; Benedek, G. B.; Carey, M. C. An investigation of the micellar phase of sodium dodecyl sulfate in aqueous solution sodium chloride solution using quasielastic scattering spectroscopy. *Journal of Physical Chemistry* **1976**, *80* (10), 1705-1805.

(23) Reiss-Husson, F.; Luzzati, L. The structure of the micellar solution of some amphiphilic compounds in pure water as determined by absolute small-angle x-ray scattering techniques. *Journal of Physical Chemistry* **1964**, *68* (12), 3054-3511.

(24) Bezzobotnoy, V. Y.; Borbely, S.; Cser, L.; Farago, B.; Gladkih, I. A.; Ostanevich, Y. M.; Vass, S. Temperature and Concentration-Dependence of Properties of Sodium Dodecyl-Sulfate Micelles Determined from Small-Angle Neutron-Scattering Experiments. *Journal of Physical Chemistry* **1988**, *92* (20), 5738-5743. DOI: 10.1021/j100331a038.

(25) Itri, R.; Amaral, L. Q. Distance Distribution Function of Sodium Dodecyl-Sulfate Micelles by X-ray Scattering. *Journal of Physical Chemistry* **1991**, *95* (1), 423-427. DOI: 10.1021/j100154a074.

(26) Volkov, N. A.; Shchekin, A. K.; Tuzov, N. V.; Lebedeva, T. S.; Kazantseva, M. A. Molecular modeling of ionic aggregates at several concentrations of SDS in aqueous solution. *Journal of Molecular Liquids* **2017**, *236*, 414-421. DOI: 10.1016/j.molliq.2017.04.018.

(27) Ruiz-Morales, Y.; Romero-Martinez, A. Coarse-Grain Molecular Dynamics Simulations To Investigate the Bulk Viscosity and Critical Micelle Concentration of the Ionic Surfactant Sodium Dodecyl Sulfate (SDS) in Aqueous Solution. *Journal of Physical Chemistry B* **2018**, *122* (14), 3931-3943. DOI: 10.1021/acs.jpcb.7b10770.

(28) Chun, B. J.; Choi, J. I.; Jang, S. S. Molecular dynamics simulation study of sodium dodecyl sulfate micelle: Water penetration and sodium dodecyl sulfate dissociation. *Colloids and*

Surfaces a-Physicochemical and Engineering Aspects **2015**, *474*, 36-43. DOI: 10.1016/j.colsurfa.2015.03.002.

(29) Tang, X. M.; Koenig, P. H.; Larson, R. G. Molecular Dynamics Simulations of Sodium Dodecyl Sulfate Micelles in Water-The Effect of the Force Field. *Journal of Physical Chemistry B* **2014**, *118* (14), 3864-3880. DOI: 10.1021/jp410689m.

(30) Mysels, K. J. Surface-Tension of Solution of Pure Sodium Dodecyl-Surfate. *Langmuir* **1986**, *2* (4), 423-428. DOI: 10.1021/la00070a008.

(31) Cheng, T.; Chen, Q.; Li, F.; Sun, H. A. Classic Force Field for Predicting Surface Tension and Interfacial Properties of Sodium Dodecyl Sulfate. *Journal of Physical Chemistry B* **2010**, *114* (43), 13736-13744. DOI: 10.1021/jp107002x.

(32) El Guendouzi, M.; Aboufaris, M. Comparative Study of Sodium Phosphate and Sodium Sulfate in Aqueous Solutions at (298.15 to 353.15) K. *Journal of Chemical and Engineering Data* **2015**, *60* (8), 2308-2319. DOI: 10.1021/acs.jced.5b00203.

(33) El Guendouzi, M.; Mounir, A.; Dinane, A. Water activity, osmotic and activity coefficients of aqueous solutions of Li2SO4, Na2SO4, K2SO4, (NH4)(2)SO4, MgSO4, MnSO4, NiSO4, CuSO4, and ZnSO4 at T=298.15 K. *Journal of Chemical Thermodynamics* **2003**, *35* (2), 209-220.

(34) Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*; Butterworth Scientific Publications, 1955.

(35) Harned, H. N.; Owen, B. B. *The Physical Chemistry of Electrolyte Solutions*; Reinhold Publishing Corporation, 1958.

(36) Rard, J. A.; Clegg, S. L.; Palmer, D. A. Isopiestic determination of the osmotic coefficients of Na2SO4(aq) at 25 and 50 degrees C, and representation with ion-interaction (Pitzer) and mole

fraction thermodynamic models. *Journal of Solution Chemistry* **2000**, *29* (1), 1-49. DOI: 10.1023/a:1005182316790.

(37) Rard, J. A.; Miller, D. G. Isopiestic determination of the Osmotic coefficients of aqueous sodium sulfate, magnesium su. *Journal of Chemical & Engineering Data* **1981**, *26* (1), 33-38.

(38) Holmes, H. F.; Mesmer, R. E. Thermodynamics of Aqueous-Solutions of the Alkali-Metal Sulfates. *Journal of Solution Chemistry* **1986**, *15* (6), 495-518. DOI: 10.1007/bf00644892.

(39) Goldberg, R. N. Evaluated Activity and Osmotic Osmotic Coefficient for Aqueous-Solutions - 36 Uni-Bivalent Electrolytes. *Journal of Physical and Chemical Reference Data* **1981**, *10* (3), 671-764. DOI: 10.1063/1.555646.

(40) Robinson, R. A.; Wilson, J. M.; Stokes, R. H. The Activity Coefficients of Lithium, Sodium and Potassium Sulfate and Sodium Thiosulfate at 25° from Isopiestic Vapor Pressure Measurements. *Journal of the American Chemical Society* **1941**, *63* (4), 1011-1013.

(41) Pearce, J. N.; Erkstrom, H. C. Vapor Pressures and Partial Molal Volumes of Aqueous Solutions of the Alkali Sulfates at 25°C. *Journal of American Chemical Society* **1937**, *59* (12), 2689-2691.

(42) Sirbu, R. Thermodynamic studies of mixed aqueous solutions of binary electrolytes of Na2SO4 and K2SO4. *Thermochimica Acta* **2000**, *352*, 1-10. DOI: 10.1016/s0040-6031(99)00429-3.

(43) Zaytsev, I. D.; Aseyev, G. G. *Properties of Aqueous Solutions of Electrolytes;*; CRC Press, 1992.

(44) Koda, S.; Nomura, H. AQUEOUS-SOLUTIONS OF SODIUM METHYLSULFATE BY RAMAN-SCATTERING, NMR, ULTRASOUND, AND DENSITY-MEASUREMENTS. *Journal of Solution Chemistry* **1985**, *14* (5), 355-366. DOI: 10.1007/bf00646171.

(45) Rodil, E.; Vera, J. H. Measurement and correlation of the activity coefficients of individual ions in aqueous electrolyte solutions of Na2SO4 and K2SO4. *Canadian Journal of Chemical Engineering* **2001**, *79* (5), 771-776. DOI: 10.1002/cjce.5450790511.

(46) Gampe, T.; Libus, Z. Osmotic and Activity-Coefficients of C2H5SO4NA(aq) and C2H5SO4K(aq) at the Temperature 298.15-K. *Journal of Chemical Thermodynamics* **1994**, *26* (3), 261-269. DOI: 10.1016/0021-9614(94)90004-3.

(47) Gampe, T.; Libus, Z. Osmotic and activity coefficients of CH3SO4Na(aq) and CH3SO4K(aq) at 25 degrees C. *Journal of Solution Chemistry* **1999**, *28* (7), 837-847. DOI: 10.1023/a:1021780130601.

(48) Naleem, N.; Bentenitis, N.; Smith, P. E. A Kirkwood-Buff derived force field for alkaline earth halide salts. *Journal of Chemical Physics* **2018**, *148* (22). DOI: 10.1063/1.5019454.

(49) Nawavi, M. Molecular dynamics simulations of aqueous ion solutions. Kansas State University, Manhattan, KS, U.S.A., 2017.

(50) Gee, M. B.; Cox, N. R.; Jiao, Y. F.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Aqueous Alkali Halides. *Journal of Chemical Theory and Computation* **2011**, *7* (5), 1369-1380. DOI: 10.1021/ct100517z.

(51) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996**, *118* (45), 11225-11236. DOI: 10.1021/ja9621760.

(52) Ferro, D. R.; Pumilia, P.; Ragazzi, M. An improved force field for conformational analysis of sulfated polysaccharides. *Journal of Computational Chemistry* **1997**, *18* (3), 351-367. DOI: 10.1002/(sici)1096-987x(199702)18:3<351::aid-jcc6>3.3.co;2-r.

(53) Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B.; team, G. d. GROMACS User Manual version 2016. 2016.

(54) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. THE MISSING TERM IN EFFECTIVE PAIR POTENTIALS. *Journal of Physical Chemistry* **1987**, *91* (24), 6269-6271. DOI: 10.1021/j100308a038.

(55) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463-1472. DOI: 10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.

(56) Miyamoto, S.; Kollman, P. A. SETTLE - An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992**, *13* (8), 952-962. DOI: 10.1002/jcc.540130805.

(57) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large System. *Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397.

(58) Nose, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics* **1984**, *52* (2), 255-268. DOI: 10.1080/00268978400101201.

(59) Hoover, W. G. Canonical Dynamics -Equilibrium Phase-Space Distributions. *Physical Review A* **1985**, *31* (3), 1695-1697. DOI: 10.1103/PhysRevA.31.1695.

(60) Kawai, T.; Kamio, H.; Kondo, T.; Kon-No, K. Effects of concentration and temperature on SDS monolayers at the air-solution interface studied by infrared external reflection spectroscopy. *Journal of Physical Chemistry B* **2005**, *109* (10), 4497-4500. DOI: 10.1021/jp046858i.

(61) Tajima, K.; Muramats.M; Sasaki, T. Radiotracer Studies on Adsorption of Surface Active Substance at Aqueous Surface .1. Accurate Measurement of Adsorption of Tritiated Sodium

Dodecylsulfate. *Bulletin of the Chemical Society of Japan* **1970**, *43* (7). DOI: 10.1246/bcsj.43.1991.

(62) Huige, C. J. M.; Altona, C. Force-Field Parameters for Sulfates and Sulfamates Based on Ab-initio Calculations - Extensions of AMBER and CHARMM Fields. *Journal of Computational Chemistry* **1995**, *16* (1), 56-79. DOI: 10.1002/jcc.540160106.

(63) Whitfield, D. M.; Tang, T. H. Binding-Properties of Carbohydrate O-Sulfate Esters Based on Ab-initio 6-31+G** Calculations on Methyl and Ethyl Sulfate Anions. *Journal of the American Chemical Society* **1993**, *115* (21), 9648-9654. DOI: 10.1021/ja00074a033.

(64) Kelkar, M. S.; Shi, W.; Maginn, E. J. Determining the Accuracy of Classical Force Fields for Ionic Liquids: Atomistic Simulation of the Thermodynamic and Transport Properties of 1-Ethyl-3-methylimidazolium Ethylsulfate (emim EtSO4) and Its Mixtures with Water. *Industrial & Engineering Chemistry Research* **2008**, *47* (23), 9115-9126. DOI: 10.1021/ie800843u.

(65) Dai, S. Properties of biologically relevant solution mixtures by theory and simulation. Kansas State University, Manhattan Kansas, 2014.

(66) Cannon, W. R.; Pettitt, B. M.; McCammon, J. A. Sulfate Anion in Water - Model Structural, Thermodynamic, and Dynamic Properties. *Journal of Physical Chemistry* **1994**, *98* (24), 6225-6230. DOI: 10.1021/j100075a027.

(67) Zhou, H. Y.; Zeng, D. W.; Han, H. J.; Dong, O. Y.; Li, D. D.; Yao, Y. Solubility Isotherm of the System Li2SO4-K2SO4-MgSO4-H2O at 273.15 K. *Journal of Chemical and Engineering Data* **2013**, *58* (6), 1692-1696. DOI: 10.1021/je4001125.

(68) Ben Mabrouk, K.; Kauffmann, T. H.; Aroui, H.; Fontana, M. D. Raman study of cation effect on sulfate vibration modes in solid state and in aqueous solutions. *Journal of Raman Spectroscopy* **2013**, *44* (11), 1603-1608. DOI: 10.1002/jrs.4374.

(69) Menger, F. M.; Rizvi, S. A. A. Relationship between Surface Tension and Surface Coverage. *Langmuir* **2011**, *27* (23), 13975-13977. DOI: 10.1021/la203009m.

(70) Vargaftik, N. B.; Volkov, B. N.; Voljak, L. D. International Tables of the Surface-Tension of Water. *Journal of Physical and Chemical Reference Data* **1983**, *12* (3), 817-820. DOI: 10.1063/1.555688.

(71) Chen, F.; Smith, P. E. Simulated surface tensions of common water models. *Journal of Chemical Physics* **2007**, *126* (22). DOI: 10.1063/1.2745718.

(72) Hayashi, S.; Ikeda, S. Micelle Size and Shape of Sodium Dodecyl Sulfate In Concentrated NaCl Solution. *J. Phys. Chem.* **1980**, *80* (7), 744-751.

(73) Benrraou, M.; Bales, B. L.; Zana, R. Effect of the nature of the counterion on the properties of anionic surfactants. 1. Cmc, ionization degree at the cmc and aggregation number of micelles of sodium, cesium, tetramethylammonium, tetraethylammonium, tetrapropylammonium, and tetrabutylammonium dodecyl sulfates. *Journal of Physical Chemistry B* **2003**, *107* (48), 13432-13440. DOI: 10.1021/jp021714u.

Chapter 4 – Peptide Interactions with SDS

4.1 Introduction

The most well-known biological application of sodium dodecyl sulfate (SDS) is the denaturing of proteins for SDS-PAGE methods. SDS-PAGE is a commonly used technique for separating proteins based on their molecular weight.[1, 2] In this technique, proteins are denatured and uniformly charged by the anionic detergent SDS, and then separated by electrophoresis on a polyacrylamide gel. The proteins are then visualized by staining with a dye or by transferring them to a membrane for western blotting.[3]

Another popular biological application is the use of SDS micelles as lipid mimetics to study the structure of proteins by techniques such as nuclear magnetic resonance (NMR), X-ray crystallography, and cryo-electron microscopy (CryoEM).[4] In these techniques, one can investigate proteins that are often difficult to study in their native state because they require a specific environment, or the formation of a complex with other molecules, to become structurally stable. SDS micelles provide a stable environment for some proteins by mimicking the hydrophobic core of a lipid bilayer and can thereby help to maintain the native structure of the protein.[5] For example, SDS micelles are often used to study the structure and dynamics of membrane proteins. The SDS micelles provide a hydrophobic environment that mimics the membrane, and the interactions between the protein and the SDS can then be probed using NMR spectroscopy.[6] Similarly, in X-ray crystallography and CryoEM, SDS micelles are often used as a tool for protein structure determination. The SDS micelles can help to stabilize the protein and keep it in a soluble, monodisperse, form that is suitable for crystallography or CryoEM.[7] In general, SDS micelles are a useful tool for studying the structure and function of proteins, particularly membrane proteins, in a controlled environment.

The presence of SDS micelles in solution can induce changes in the secondary structure of proteins, including the loss of α-helical structure and an increase in β-sheet structure.[8] This has been observed in many different proteins such as alpha-lactalbuim, apomyoglobin, and ribonuclease A.[9-11] The presence of high concentrations of SDS can lead to the unfolding and denaturation of proteins, resulting in the loss of protein activity and function. Lysozyme undergoes denaturation and unfolding at high concentration of SDS, has been demonstrated by circular dichroism spectroscopy, fluorescence spectroscopy, and X-ray crystallography.[12, 13] Green fluorescent protein (GFP) also undergoes the unfolding process in the presence of a high concentration of SDS.[14] At lower concentrations of SDS proteins can bind to the surface of SDS micelles, which can then alter their conformation and stability. This effect has been observed in many different proteins, such as Cytochrome c, hemoglobin, and many membrane proteins.[15-17] The interactions of proteins with SDS micelles have been reported to lead to protein aggregation or precipitation in some cases, particularly if the protein concentration is high or if the protein has a high hydrophobicity. Amyloid β and α-Synuclein have displayed aggregation in the presence of SDS micelles.[18, 19]

Several studies have used molecular dynamics (MD) simulations to investigate the effect of SDS on protein folding. It is beneficial to combine accurate simulation studies with experimental measurements to fully understand the atomistic detail and key conformational features related to these interactions. MD simulations have provided valuable insights into the mechanism of protein unfolding during SDS-PAGE and have helped to explain the observed experimental results.[20] At high temperature, protein unfolding has been simulated in the presence of SDS. Winogradoff *et al*. reported MD simulation studies to investigate the mechanism of protein unfolding in the presence of SDS at high temperature.[20] Spontaneous unfolding in the presence of

SDS at the boiling point of water was observed on the time scale of several microseconds in this study. Another MD study of Aβ40 MD with SDS micelles at room temperature demonstrated that Aβ40 shows increased helix structure in the presence of SDS.[21] Studying the conformational changes of peptides in the presence of SDS can provide valuable insights into the behavior of proteins and peptides under physiological and experimental conditions, and can have important implications for the development of peptide-based therapeutics or many other biological applications.

In the previous chapter we developed KBFF parameters for aqueous SDS solutions. These were tested to ensure the stability of micelle structures and the changes in surface tension induced by SDS at the vapor-liquid interface. More rigorous testing is still required to establish the accuracy of the force field parameters. In this chapter, we examined 5 small peptide NMR structures in the presence of SDS micelle, as obtained from PDB databank and summarized in Table 4.1, by MD simulation. The peptides were treated with our recently published force field for peptides and proteins, KBFF20.[22] MD simulations have been performed for each peptide in the absence and presence of SDS to investigate the effect of SDS on the peptides. This provides a test of both the existing KBFF20 and new SDS force fields.

Table 4.1 Selected peptide information

PDB ID	1DN3	2JNI	2OJM	2RMY	6WPB
Type	NMR	NMR	NMR	NMR	NMR
Temp (K)	293	288	318	298	298
pH	6.0	3.3	4.12	7.4	6.5
Ionic Strength	0	0	0	0	Undefined
Pressure	Ambient	Ambient	Ambient	Ambient	1atm
Residue Number	15	21	22	34	15
Structure	helix	β-sheet	helix	helix	helix
Sequence	QAPAYKKAAK KLAES1209	RWCVYAYVRI RGVLVRYRRC W	FFHHIFRGIV HVGKTIHRLV TG	MAEMGSKGVT AGKIASNVQK KLTRAQEKVL QKLY	GILDAIKAIA KAAG-NH2

4.2 Methods

4.2.1 Simulation methods

4.2.1.1 System preparation: Peptide without SDS

The systems were prepared using the GROMACS 2016 software package.[23] 5 peptides were retrieved from the Protein Data Bank which are summarized in Table 4.1. The peptides were simulated using the KBFF20 force field.[22] Each system was solvated with SPC/E water molecules in a 10 x 10 x 10 nm cubic box. The system was neutralized by adding an appropriate amount of counter ions (Na^+ or Cl^-) and an additional physiological concentration (0.15M) of NaCl was added.

4.2.1.2 System preparation: Peptide with SDS

The same 5 peptides in Table 4.1 were placed in individual box size 10 x 10 x 10 nm^3 in the presence of SDS. Three different initial SDS placements were investigated as the SDS

molecules do not appear in the PDB coordinates. These included an intact SDS micelle, two split half SDS micelles, and a collection of random SDS molecules in solution, as indicated in Figure 4.1. Each box was then solvated with SPC/E water and an appropriate number of counter ion were added (Na^+ or Cl^-) to neutralize the system. Additional Na^+ and Cl^- ions are added for the simulations under physiological conditions (0.15M).

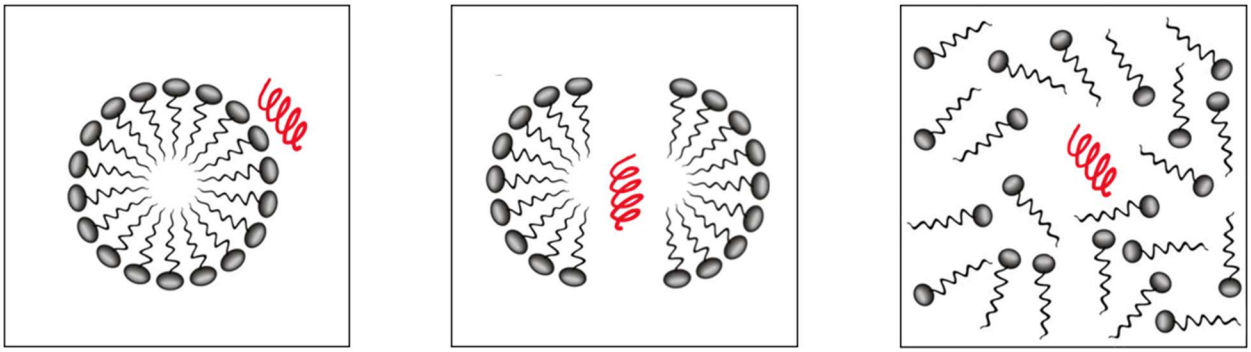


Figure 4.1 Schematic diagram of the different initial setups of SDS and peptide in the simulation box cells. Grey polar head with hydrophobic tail represents SDS and red helix represents peptide.

All simulations in this study were performed in the isothermal isobaric ensemble (NpT) at 298K temperature and 1 atm pressure. The temperature and pressure were weakly coupled to a bath with relaxation times of 0.1 ps and 0.5 ps, respectively. A time step of 2 fs was used to integrate the equations of motion. Peptide and SDS bonds in the simulations were constrained with LINCS algorithm,[24] while the water bonds were constrained with SETTLE.[25] To calculate the electrostatic interactions, the particle mesh Ewald technique was used with a 1.0 nm cutoff distance for real space calculations. A 1.0 nm cutoff was also used for the van der Waals interactions.[26] An energy minimization using the steepest descent method followed by 100ps of equilibration was performed before the production run. The production run of 500ns was used to calculate ensemble averages for peptide without SDS. Configurations were saved for analysis every 1 ns. For the production simulations for the peptides in the presence of SDS, Cα restraints were applied for the

first 100ns of the simulation. Additional 500ns simulations were then performed without Cα constraints. All the configurations were saved for analysis every 1 ns as well.

The previous peptides were simulated starting from their (folded) PDB structures. We also performed simulations in the presence of SDS starting from unfolded structures. Unfolded random initial structures of each peptide were obtained from the simulations of the peptides in the absence of SDS, which then unfolded. These unfolded configurations were placed with 60 random SDS and simulated for 1-2 μs without Cα restraints.

4.3 Results and discussion

We have investigated the structure of 5 small peptides in the presence or absence of SDS. All obtained NMR structures were α-helix, except for 2JNI which has an antiparallel β-hairpin structure with a disulfide bond connecting the ends of the hairpin. The structures of the 5 peptides obtained from PDB databanks are the structures in the presence of SDS micelles.

4.3.1 1DN3

Montserret *et al*. investigated the structure of a hydrophilic, charged, amphipathic α-helical peptide derived from the sequence of the 56-70 segment of human platelet factor 4 (PF4), referred to as PF4(56−70).[27] This peptide's sequence is QAPLYKKIIKKLLES and has several hydrophobic and hydrophilic residues. The previous experimental studies have focused on the basic version of the peptide (QAPAYKKAAKKLAES), with the PDB identifier 1DN3, examining its conformation both in the absence and presence of Sodium dodecyl sulfate (SDS), using Circular Dichroism (CD) and Nuclear Magnetic Resonance (NMR).[27] The results revealed that in water and in the presence of neutral, zwitterionic, or cationic detergents, the peptide remained unstructured. Interestingly, however, the addition of SDS at a neutral pH induced the peptide to adopt an α-helical conformation.[27]

To further analyze these findings, simulations of 1DN3 were performed mirroring the experimental conditions both in the absence and presence of SDS. The findings from these simulations were consistent with the experimental data as shown in Figure 4.2. When simulated without SDS, the peptide's helical structure was initially maintained, but then gradually lost its helicity over the first 100ns of the simulation. Indeed, between 400 and 500ns, the peptide completely unfolded in the absence of SDS, as visualized by the black lines in Figure 4.2. However, in simulations conducted with SDS, the peptide's α-helical structure remained stable throughout the entire 500ns simulation period, as indicated by the red lines in Figure 4.2. Further analysis, using the end-to-end distance and radius of gyration measurements, corroborated these observations, as depicted in Figure 4.3. These measurements showed a greater dispersion for the peptide in the absence of SDS, contrasting with the more consistent results observed in the presence of SDS. This underscores the significant role of SDS in promoting and stabilizing the α-helical conformation of the peptide. Furthermore, the results were reproducible between sets of three independent simulations indicating that sufficient sampling was achieved for this system.

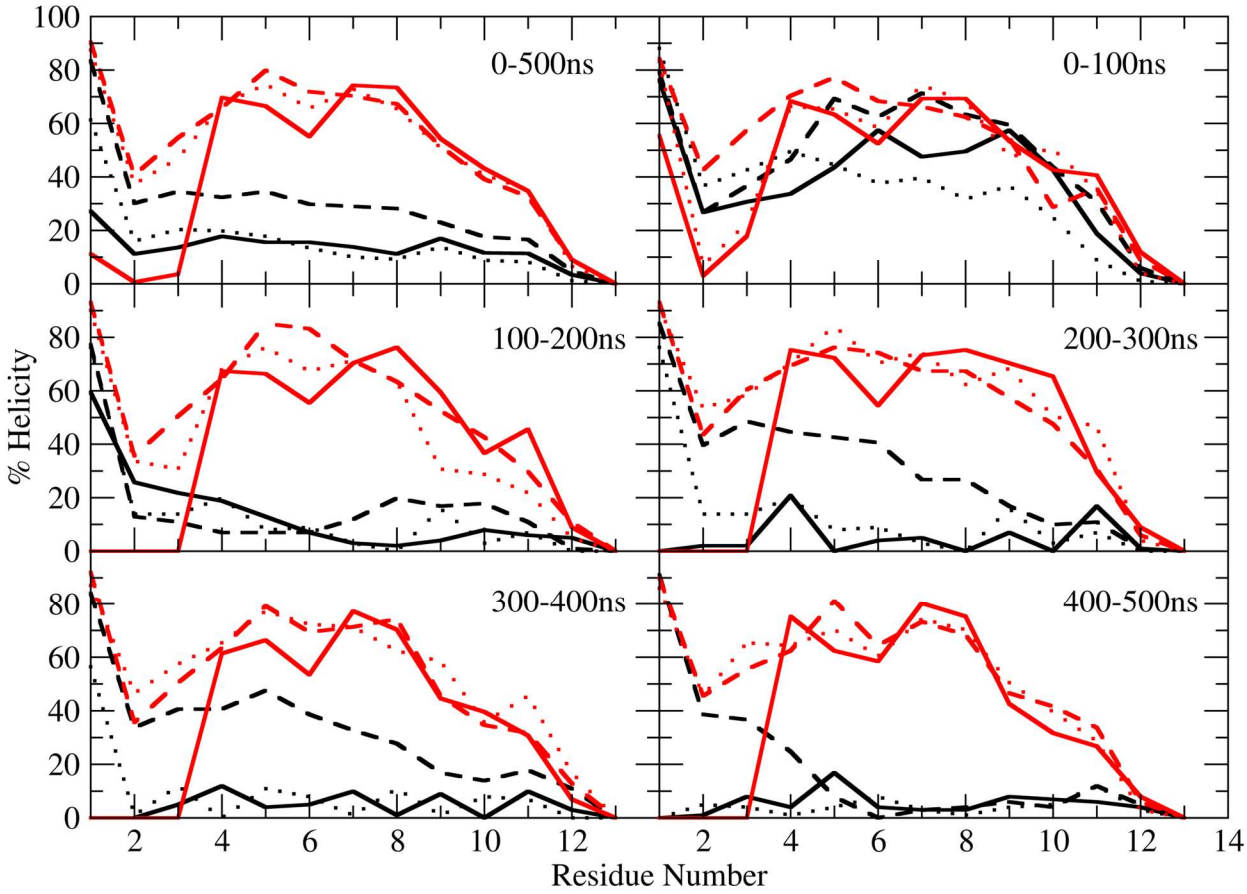


Figure 4.2 1DN3 % helicity in the absence or presence of SDS. Solid, dotted and dash Black, lines represent 1DN3 in absence of SDS. Red solid dotted and dash line represents 1DN3 with SDS in various starting forms.

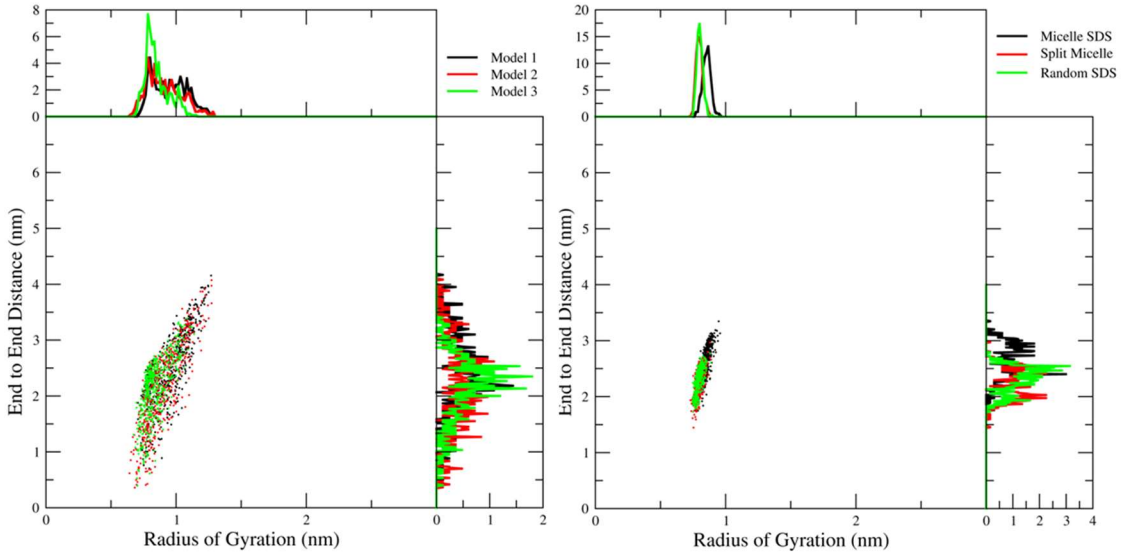


Figure 4.3 Radius of Gyration and end to end distance analyses results for 1DN3 in absence (left graph) and presence (right graph) of SDS.

4.3.2 2OJM

Piscidin 1 (Pis-1, WT PDB ID: 2OJM) is a peptide of notable cytotoxic properties, featuring a cationic α-helical structure.[28] This peptide was initially isolated from the mast cells of hybrid striped bass and, interestingly, it exhibits no selectivity between bacterial and mammalian cells. However, an exploration of the peptide's tertiary structure in the presence of SDS micelles, as revealed by NMR spectroscopy, indicated an α-helical structure stretching from the second phenylalanine (F2) to threonine at position 21 (T21).[28] Further insights into the secondary structure of Pis-1 and its analogues in membrane-like environments were gleaned through the analysis of their Circular Dichroism (CD) spectra in various membrane mimetic settings.[28] While Pis-1 and its analogues demonstrated random structures in an aqueous solution, they underwent conformational changes to adopt α-helical configurations in a 50% Trifluoroethanol (TFE)/water solution, SDS micelles, and Dodecylphosphocholine (DPC) micelles.

In our simulations, the wild-type Pis-1 peptide (PDB ID: 2OJM) was shown to maintain its helical conformation when SDS was present, regardless of the initial state of the SDS as shown in Figure 4.4. However, without the SDS, the peptide gradually lost its helicity over the course of the simulation. In the absence of SDS (all black lines in Figure 4.4), both N and C-termini of the peptide began to unfold, resulting in complete unfolding after 400ns. Conversely, when simulated with an SDS micelle, the peptide did lose helicity for residue 13 to the C-terminus during the 300-500ns period (red solid line in Figure 4.4). Interestingly, however, both when simulated with split micelle SDS and random SDS, Pis-1 maintained its helical configuration throughout the simulation. This partial loss of helicity could be attributed to the initial positioning of the peptide. The peptide, in the presence of random or split micelle SDS, had more leeway to adjust its conformation as the SDS slowly formed a micelle. Contrarily, when interacting with a

preassembled SDS micelle, the peptide may have found it challenging to revert to its original helical structure once a conformational change occurred due to its close interaction with the SDS and a lack of sampling.

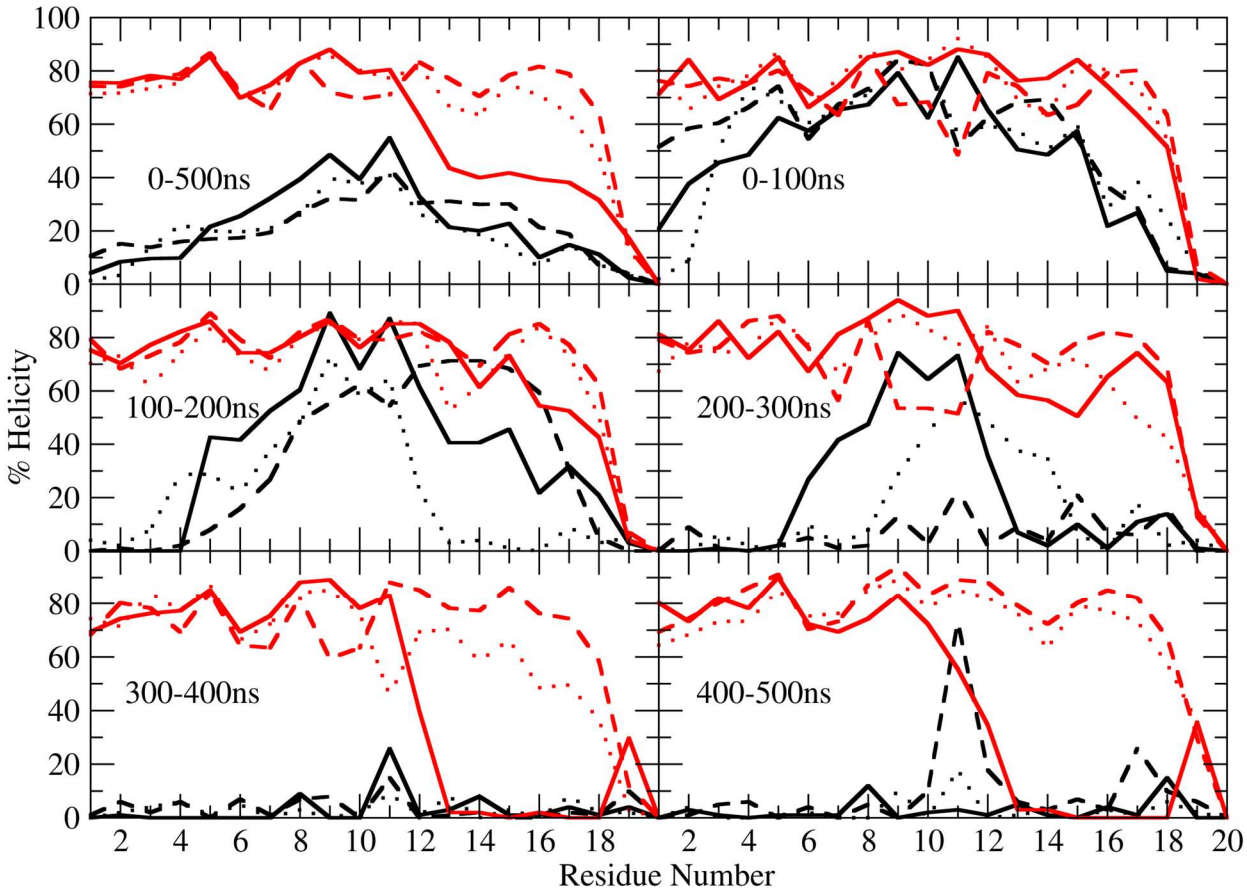


Figure 4.4 2OJM % helicity in the absence and presence of SDS. Solid, dotted and dash black lines represent 3 different configurations of 2OJM in absence of SDS. Red solid dotted and dash line represents 2OJM with SDS micelle, split SDS micelle and random SDS each.

The results of the radius of gyration and end-to-end distance analyses are presented in Figure 4.5. The radius of gyration for 2OJM remained steady at the peak value of 1.2nm in the presence of SDS. However, without the SDS, the results spread over a wider range, from 1.0 to 2.0 nm. The end-to-end distance analysis echoed the trends observed in the radius of gyration analysis. Without SDS, the range of distances for 2OJM stretched from 0-6nm, while a narrower range of 3-4nm was evident for 2OJM in the presence of SDS.

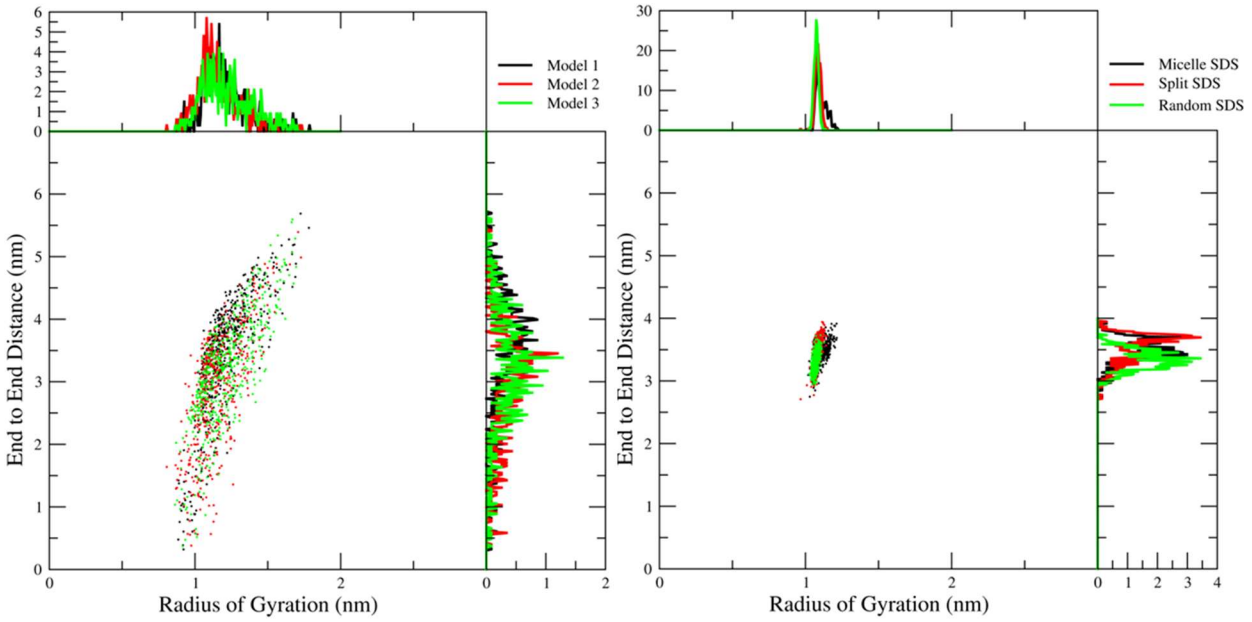


Figure 4.5 Radius of gyration and end to end distance analyses results for 2OJM in absence (left graph) and presence (right graph) of SDS.

4.3.3 2RMY

Bin/Amphiphysin/Rvs-homology (BAR) domains, as represented by PDB ID: 2RMY, play a pivotal role in generating and sensing membrane curvature due to their ability to bind negatively charged membranes through their positively charged concave surfaces.[29] Detailed insights into the behavior of these domains were obtained from NMR structure and MD simulations of 2RMY, performed in SDS and dodecylphosphocholine micelles. NMR studies have revealed that the binding of the N-BAR peptide to micelles induces significant changes in both the backbone and side-chain resonances, as compared to the random coil chemical shifts dominating in aqueous solution.[29] This shift confirms the interaction of N-BAR with the micellar environment, inducing a defined secondary structure, as also observed by far-UV CD spectra.[29] The well-ordered residues, 8–34 in SDS and 10–34 in DPC micelles, as highlighted by a heavy atom root mean-square

deviation < 1.1 Å, form an amphipathic helix with negatively charged side chains on the convex side and hydrophobic side chains on the concave side.

This experimental data again align with our MD simulation results, where an initially disordered N-terminal region is observed in the absence of SDS (as illustrated in Figure 4.6). In the presence of SDS, 2RMY maintains its helicity throughout the simulation, as shown by the red lines in Figure 4.6. Conversely, without SDS, 2RMY gradually loses its helicity from the N-terminus to residue 14, as depicted by the black lines in Figure 4.6.

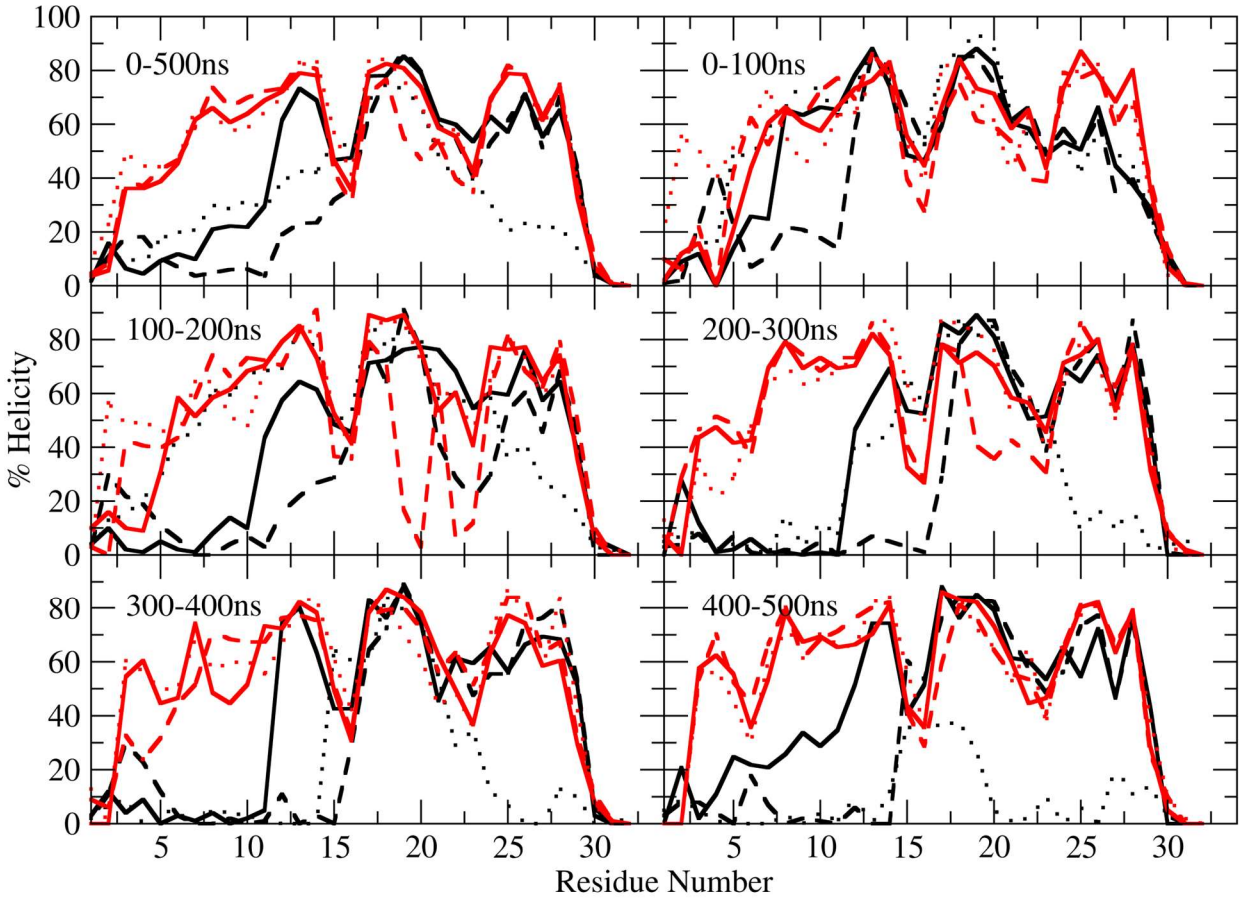


Figure 4.6 2RMY % helicity in the absence and presence of SDS. Solid, dotted and dash black lines represent 3 different configurations of 2RMY in the absence of SDS. Red solid, dotted and dash line represent 2RMY with SDS micelle, split SDS micelle, random SDS each.

Furthermore, we employed radius of gyration and end-to-end distance calculations to analyze 2RMY. The results, as shown in Figure 4.7, demonstrated wider ranges of both measures

in the absence of SDS, implying potential greater structural heterogeneity or flexibility in the peptide when SDS is absent. This illustrates that SDS plays a critical role in stabilizing the helical structure of the peptide.

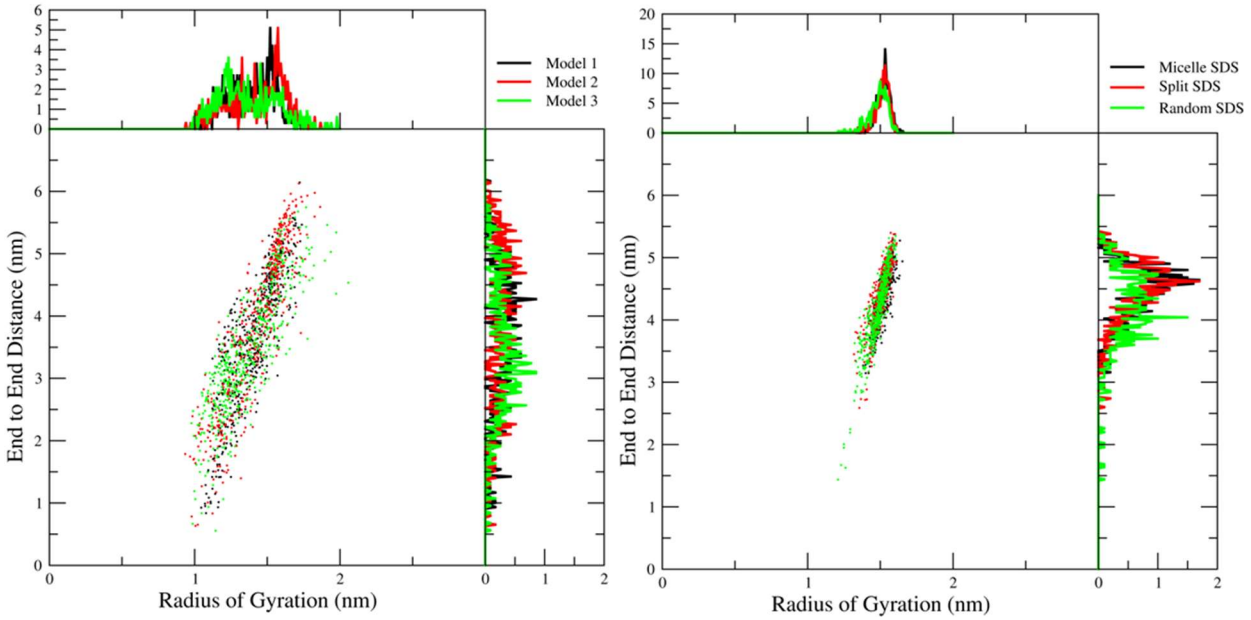


Figure 4.7 Radius of gyration and end to end distance analyses results for 2RMY in absence (left graph) and presence (right graph) of SDS.

4.3.4 6WPB

Gomes *et al.*,[30] characterized HSP1-NH2 using Circular Dichroism (CD) spectroscopy, noting that in an aqueous solution, the peptide showed a disordered conformation. However, in the presence of either DPC or SDS micelles, even at low detergent concentrations, they observed negative maxima characteristic of helical structures. Interestingly, they found that the helical content of HSP1-NH2 increased in the presence of anionic micelles in a SDS concentration-dependent manner, unlike in the presence of DPC zwitterionic micelles, where no significant change was observed. Further experimental insights into the three-dimensional structures of HSP1-NH2 were provided using solution NMR methodologies. In the presence of deuterium labeled SDS-d25 and DPC-d38 micelles, the peptide showed substantial amphipathic characteristics.[30] Of

particular interest was a slight bending of the helix near the D4 residue in the presence of DPC-d38 micelles, causing the G1 and L3 residues to orient towards the hydrophobic face.

To assess and compare to these experimental findings, we conducted molecular dynamics simulations. In the absence of SDS, our simulations showed that the peptide 6WPB progressively lost its helicity, ultimately completely unfolding after 400ns, as shown in Figure 4.8 (black lines). However, when SDS was present, the peptide maintained its helical conformation throughout the entire simulation period, as shown by the red lines in Figure 4.8.

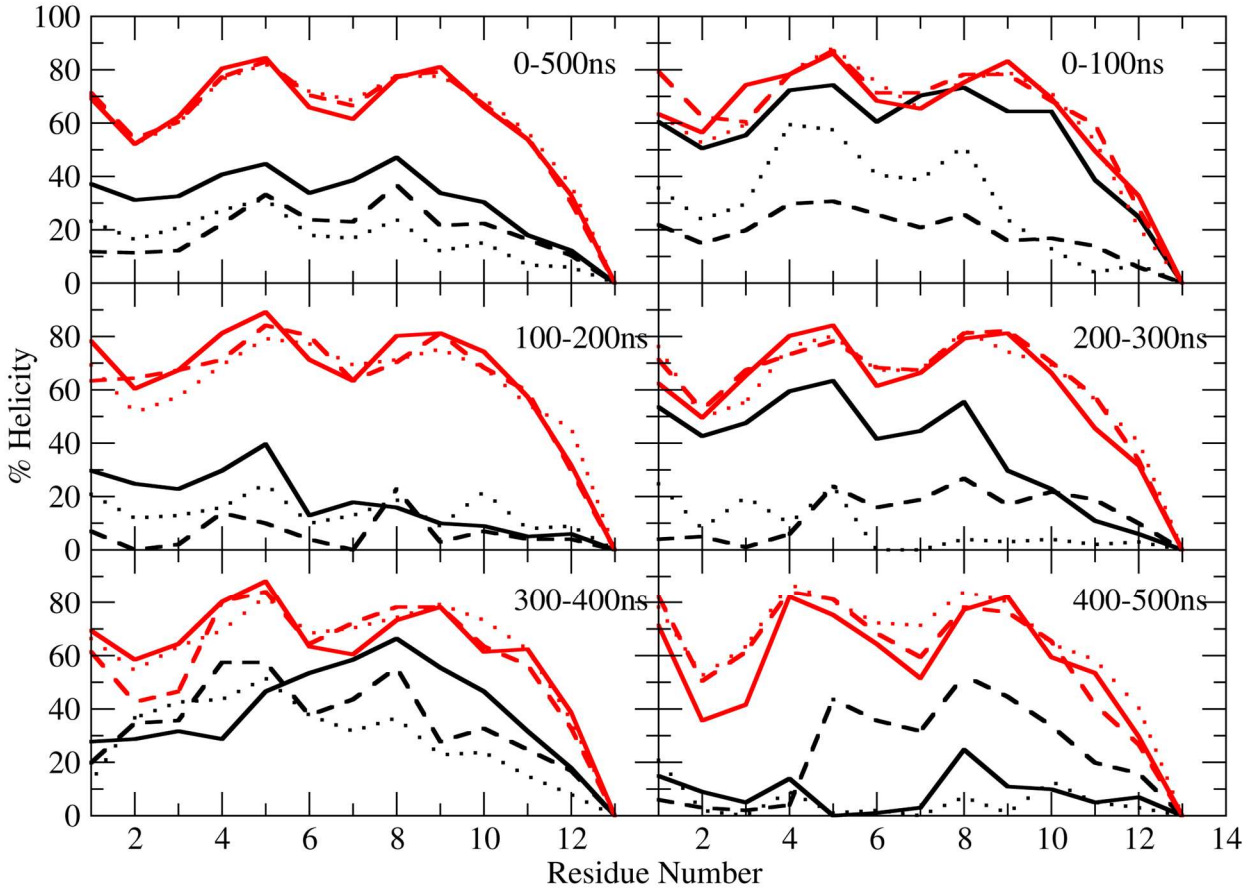


Figure 4.8 6WPB % helicity in the absence and presence of SDS. Solid, dotted and dash black lines represent 3 different configurations of 6WPB in the absence of SDS. Red solid dotted and dash line represent 6WPB with SDS micelle, split SDS micelle, random SDS each.

These observations are further substantiated by the results of the radius of gyration and end-to-end distance analyses (see Figure 4.9). In the absence of SDS, 6WPB showed a less compact and more varied structure. However, in the presence of SDS, 6WPB's structure appeared more compact, with a narrower range in the histogram (right graph in Figure 4.9). This aligns with the percent helicity results, where 6WPB maintains a consistent helicity in the presence of SDS but loses helicity over time in its absence. The end-to-end distance analysis, which measures the distance between the N- and C-termini of the peptide, showed a larger distance for 6WPB in the absence of SDS. This indirectly suggests that without SDS, 6WPB loses helicity and adopts a more extended structure. In contrast, a more compact structure is maintained in the presence of SDS.

Collectively, these analyses provide an in-depth perspective on the conformational dynamics of the 6WPB peptide, highlighting the significant role of SDS in maintaining the peptide's helical structure and compact conformation.

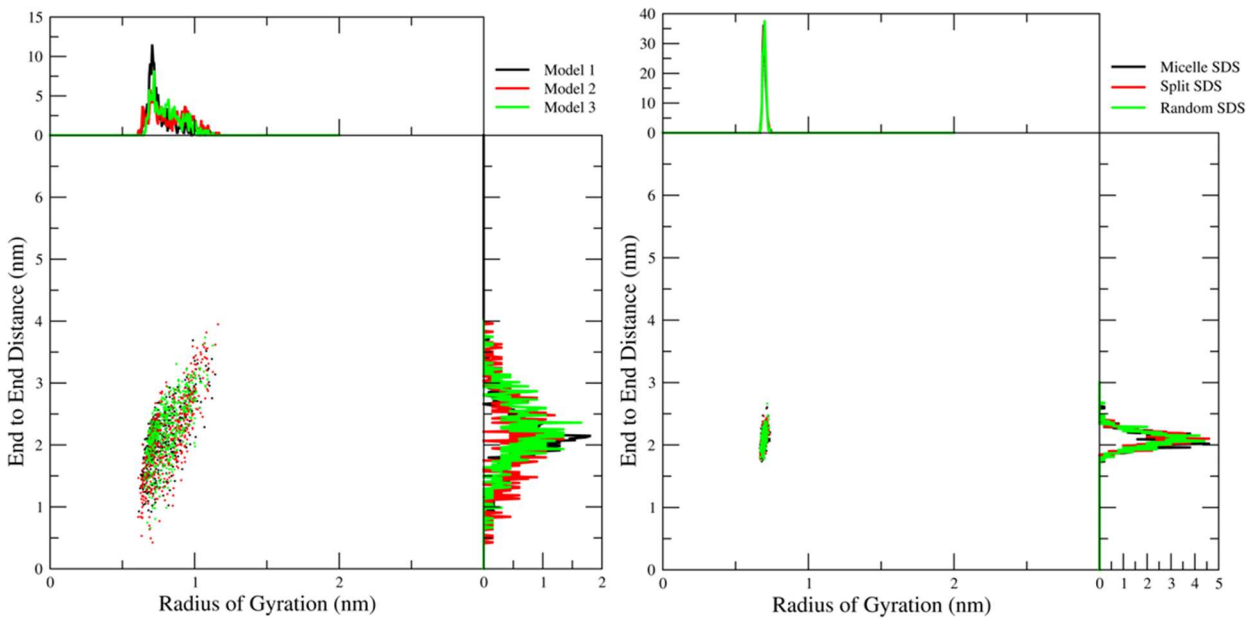


Figure 4.9 Radius of gyration and end to end distance analyses results for 6WPB in absence (left graph) and presence (right graph) of SDS.

4.3.5 2JNI

Arenicin-2 (PDB ID: 2JNI) is a cationic antimicrobial peptide, characterized by its unique beta-sheet structure known as a "β-hairpin".[31] This structure is established by two antiparallel β-strands that run in opposite directions and is stabilized by one disulfide bond and nine hydrogen bonds. The amphipathicity, or segregation of polar and hydrophobic side chains on a peptide surface, is a critical attribute for the membrane activity of cationic antimicrobial peptides (AMPs) such as arenicin-2. The researchers hypothesized that if biological membranes are targeted by arenicin-2, its spatial structure should undergo changes upon contact with lipid bilayers, becoming more amphipathic. Ovchinnikova *et al.*, utilized circular dichroism (CD) spectra to explore possible changes in arenicin-2 conformation when interacting with lipid bilayers (see Figure

4.10).[31] In an aqueous solution, the CD spectrum of arenicin-2 exhibited four bands, an unusual characteristic for a beta-structural peptide. This peculiar feature could be attributed to distortions in the beta-sheet structure and CD signals originating from aromatic residues and the disulfide bond. The spectral behavior of arenicin-2 changes significantly in the presence of detergent micelles (SDS and DPC). Specifically, the CD spectra in the presence of these micelles are not similar to each other, but they deviate significantly from the spectrum of arenicin-2 in aqueous solution, resembling more closely the canonical spectrum of a β-structure.

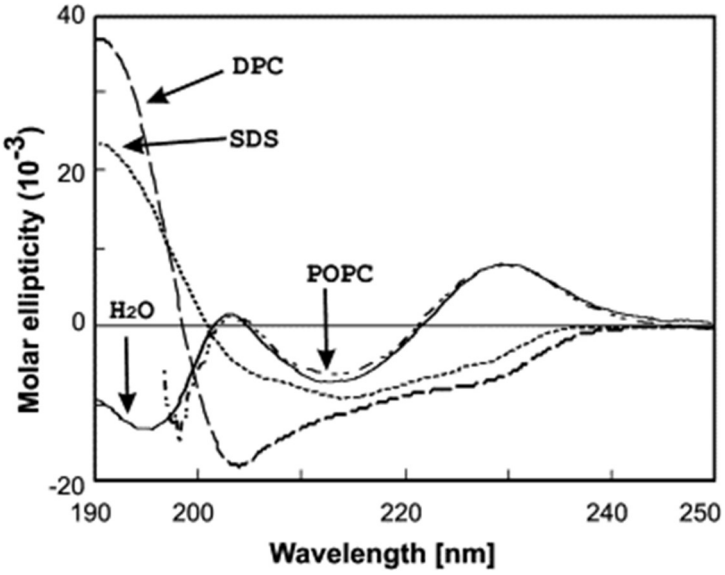


Figure 4.10 CD spectra of 2JNI in different environments. The figure is reprinted with permission from Ovchinnikova, T. V.; Shenkarev, Z. V.; Nadezhdin, K. D.; Balandin, S. V.; Zhmak, M. N.; Kudelina, I. A.; Finkina, E. I.; Kokryakov, V. N.; Arseniev, A. S. Recombinant expression, synthesis, purification, and solution structure of arenicin. Biochemical and Biophysical Research Communications 2007, 360 (1), 156-162. DOI: 10.1016/j.bbrc.2007.06.029. Copyright Elsevier Inc. (2007). [31]

We examined the conformational changes of arenicin-2 through simulation studies. Within this context, arenicin-2 (2JNI) stands out as the only peptide featuring an antiparallel β-sheet structure for study in this chapter. The simulations revealed that 2JNI maintains its antiparallel-β conformation both in the absence and presence of SDS throughout the simulation period. The radius of gyration and end-to-end distance analyses for 2JNI, displayed in Figure 4.11, indicated a

very tight range, suggestive of the peptide's ability to preserve its initial conformation. Nonetheless, a temporary loss of hydrogen bonds observed at certain time intervals (70 to 100 ns and 350 to 400 ns of model 1 simulation) may hint at a transient disruption of the antiparallel-β structure in the absence of SDS.

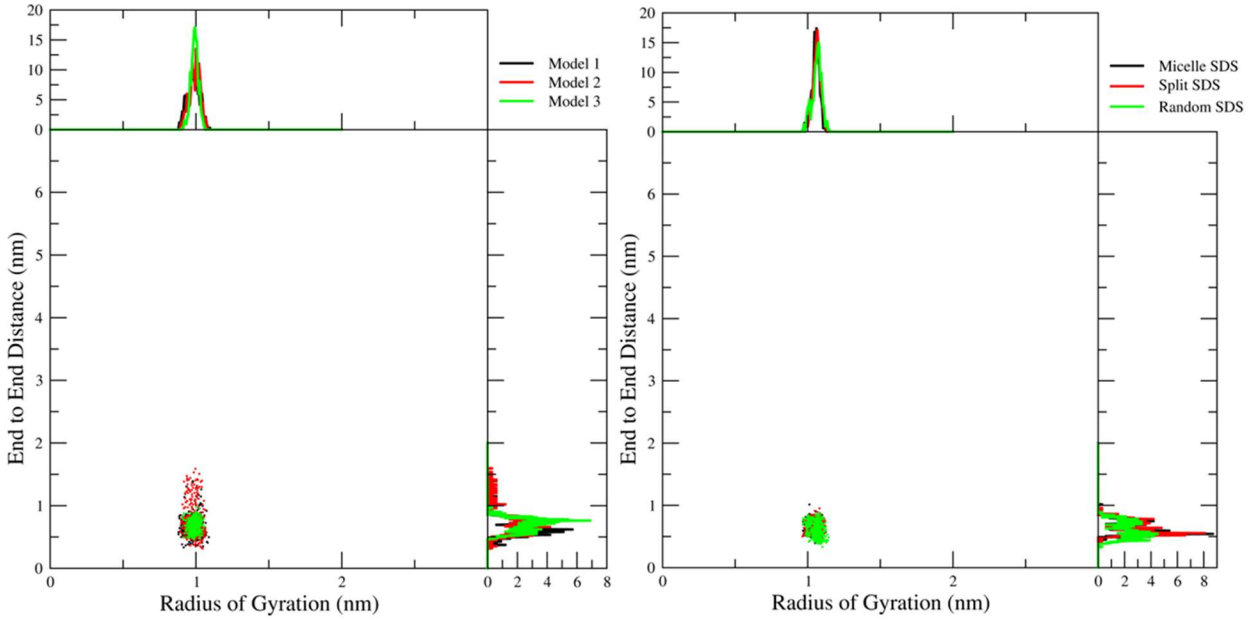


Figure 4.11 Radius of gyration and end to end distance analyses results for 2JNI in absence (left graph) and presence (right graph) of SDS.

Although 2JNI generally forms an antiparallel-β structure both in the absence and presence of SDS, some distortions akin to those reported in experimental results could be observed in the absence of SDS. An overlay of the 500 ns 2JNI conformation in the absence and presence of SDS, shown in Figure 4.12, reveals such distortion or change in curvature of the antiparallel-β structure. In another simulation experiment, a randomly selected 2JNI configuration was exposed to random SDS over 2 μs. The peptide was found to prefer staying close to SDS, but it did not revert to antiparallel-β structures. This behavior could be due to either the duration of the simulation or the sampling of higher energy conformations during the simulation. A potent interaction with SDS residues could slow the process of beta-sheet formation, requiring a more extended simulation period.

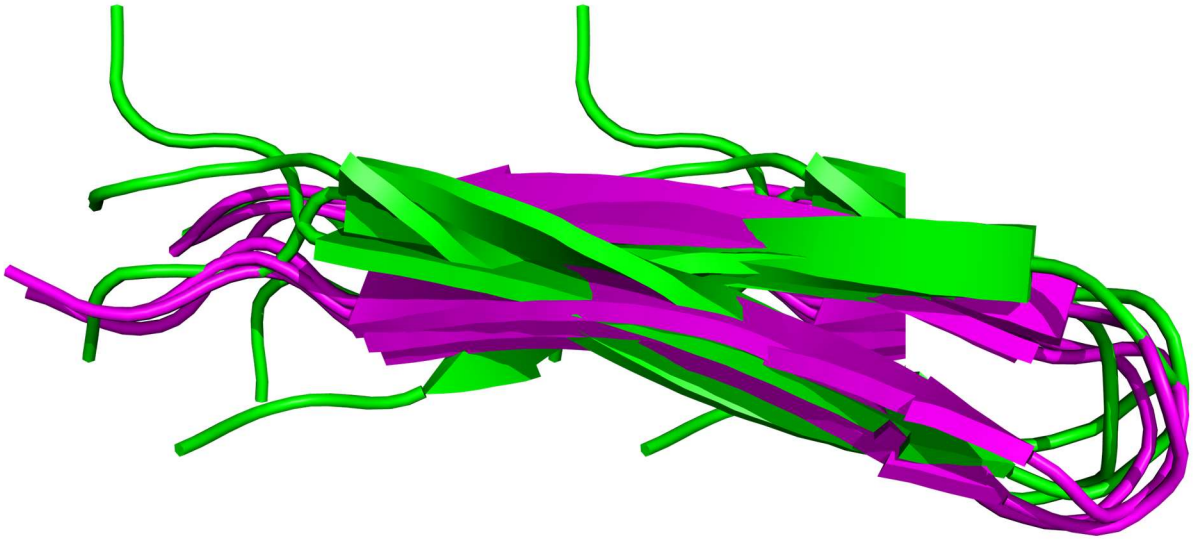


Figure 4.12 Overlay of the conformation at 500ns from 6 different simulations. All green represents 2JNI in absence of SDS and all magenta represent 2JNI in presence of SDS.

4.3.6 Peptide structure in the absence or presence of SDS

The initial and final configurations after 500 ns of simulation for each peptide, both in the absence and presence of SDS, are presented in Figure 4.13. The peptides 1DN3, 2OJM, and 6WPB underwent complete unfolding after 500 ns of simulation. However, 2RMY only partially unfolded during the simulation conducted in the absence of SDS. As illustrated by the overlaid configurations in Figure 4.14 (depicted in orange), all the peptides adopting a helical structure (1DN3, 2OJM, 2RMY, and 6WPB) successfully preserved their helical form throughout the 500 ns simulation. The peptide 2JNI, which assumes a beta-sheet hairpin structure, maintained its configuration both in the absence and presence of SDS over the same duration of the simulation. It's worth noting that, in the absence of SDS, 2JNI exhibited more twisting compared to its structure in the presence of SDS. Overall, these results are in line with the percent helicity analyses for each helix-forming peptide.

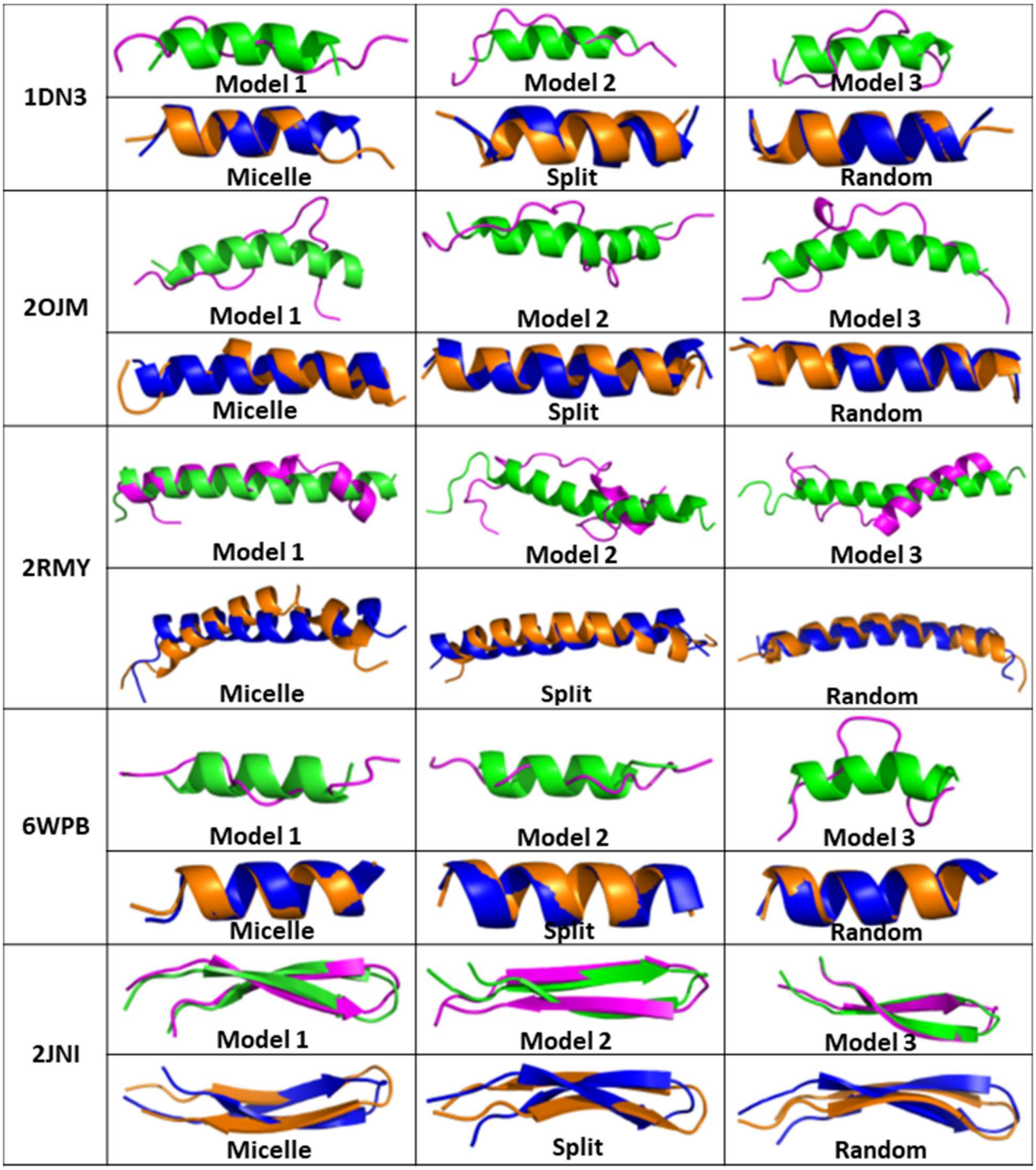


Figure 4.13 Initial and final configuration overlayed for each peptide. Green color -0ns in the absence of SDS. Magenta color -500ns in absence of SDS. Blue color -0ns in the presence of SDS. Orange color -500ns in the presence of SDS

4.3.7 Peptide folding in the presence of SDS

All simulations in the presence of SDS have been performed starting from the appropriate folded PDB structure. All the helical peptides remained helical after 500 ns of simulation. However, it is a more demanding test to attempt folding these peptides from an initial random structure in the presence of SDS. Thus, to support our results we chose unfolded structures, obtained from the helix forming peptide simulations in absence of SDS, and used these peptide configurations with random SDS and simulated up to 2 μs. Percent helicities for the simulation of each peptide in the presence of random SDS molecules are shown in Figure 4.14. The 6WPB peptide formed a stable helix after 100ns of simulation and the helical structure was maintained throughout the simulation period. Thus, this simulation was stopped after 1μs. The 1DN3 and 2RMY formed a complete helix between 1.5-2 μs of simulation However, the 2OJM peptide only recovered the C-terminus half of the helix, which may be due to strong interaction with SDS. Further analysis may be needed with enhanced sampling to fully fold and/or determine exact free energy differences between folded and unfolded states.

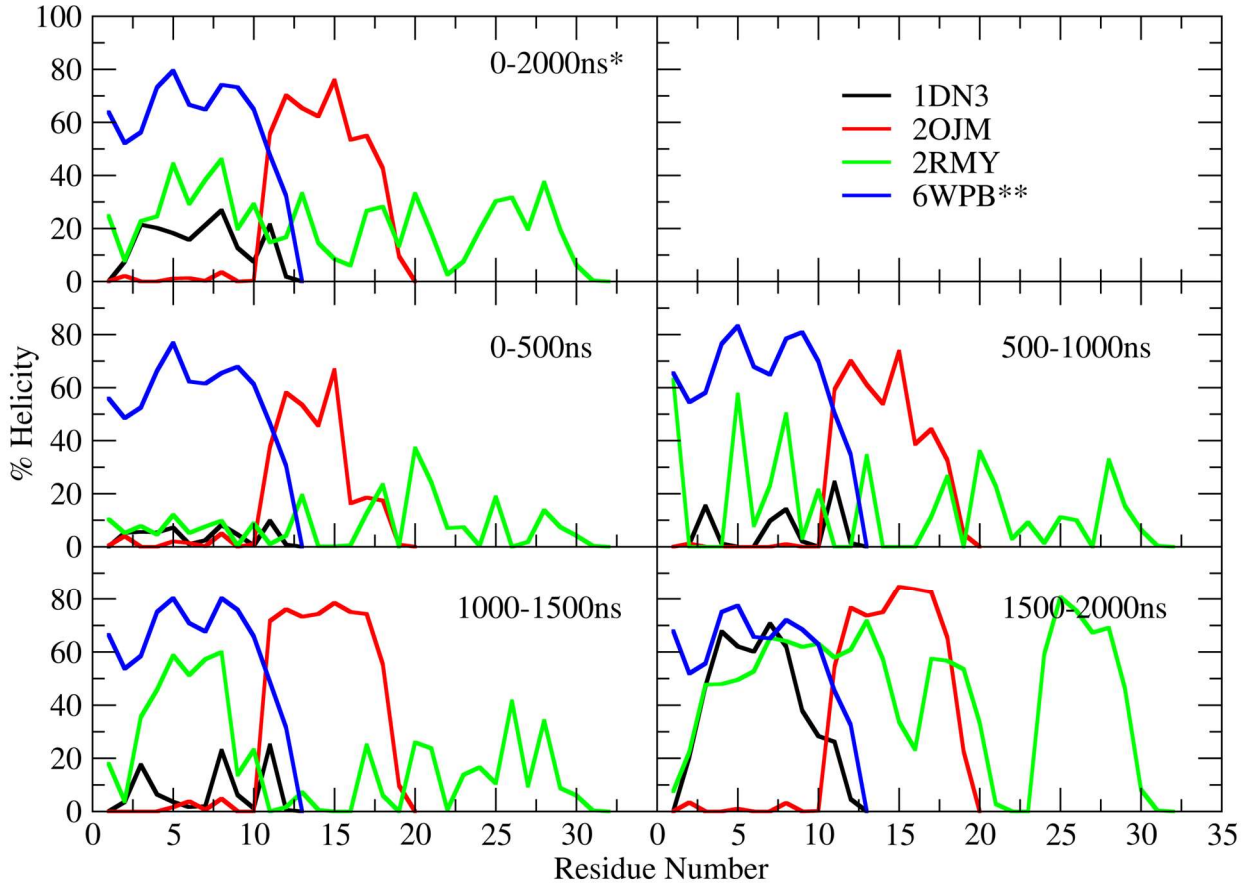


Figure 4.14 % helicity of initially unfolded peptides in the presence of random SDS. The 1DN3, 2OJM, 2RMY average helicity in top left graph is average of 2 μs simulation. The 6WPB average helicity is only for 1 μs**.**

4.4 Conclusions

A series of peptide simulations in absence and presence of SDS have been performed. All 4 peptides underwent an unfolding process in absence of SDS during the simulation period. All simulations the presence of SDS maintained the helix form during the simulation period. All unfolded peptides in the presence of randomly positioned SDS molecules were able to recover their helix form at the end of simulation. The peptide 2JNI, which is the only antiparallel β-sheet hairpin structure, maintained its structure in both absence and presence of SDS throughout the simulation time (500ns). The results provide a test of both the new SDS model generated in the previous chapter and the KBFF20 force field in general. It is encouraging for both models that

helical structure was not only maintained in the presence of SDS but could be formed from random initial structures. Furthermore, in light of the overstabilization of folded structures associated with other traditional force fields, it is clear that KBFF20 also reproduces the instability of these helices in pure water.

4.5 Supporting Information

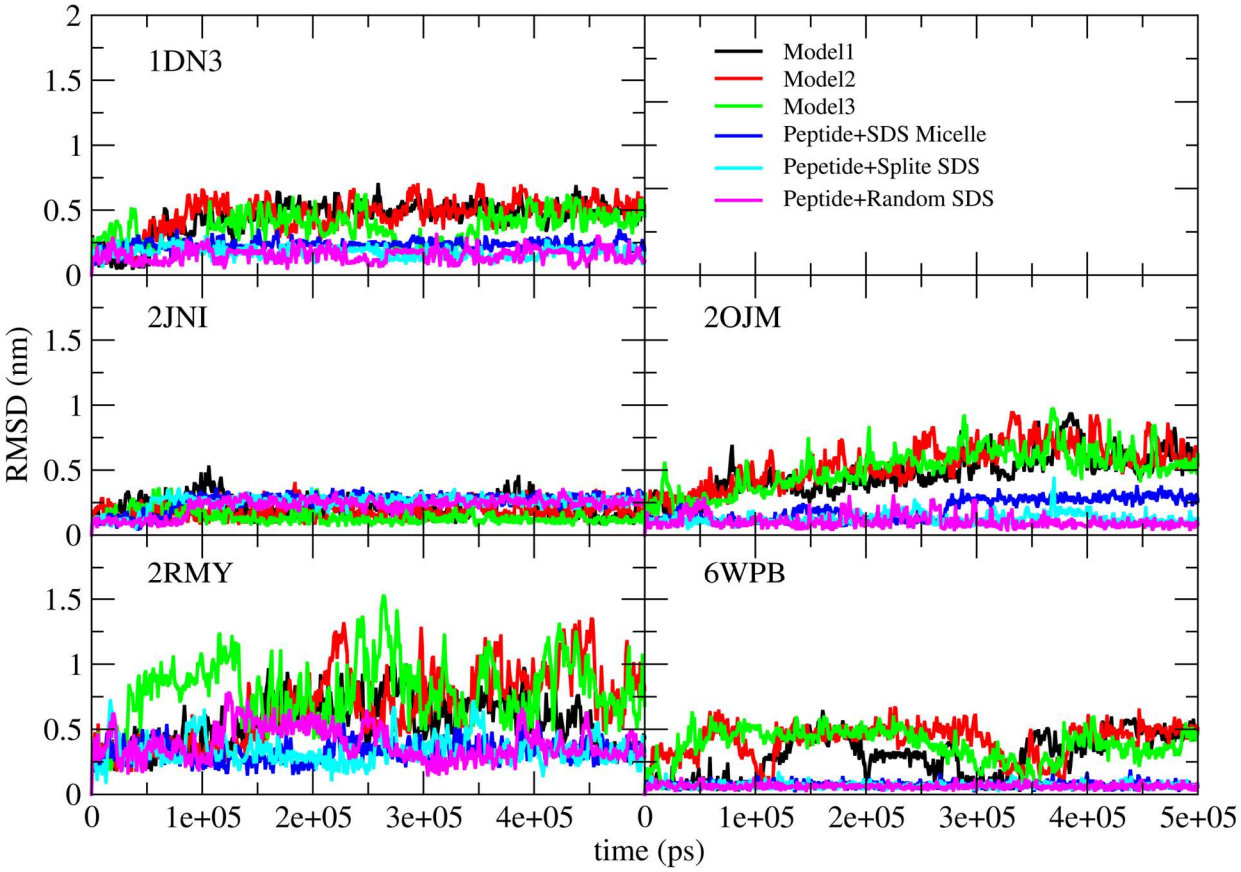


Figure S 4.1 RMSD is for 5 peptides in different simulations. Model 1, 2, 3 are simulated in absence of SDS.

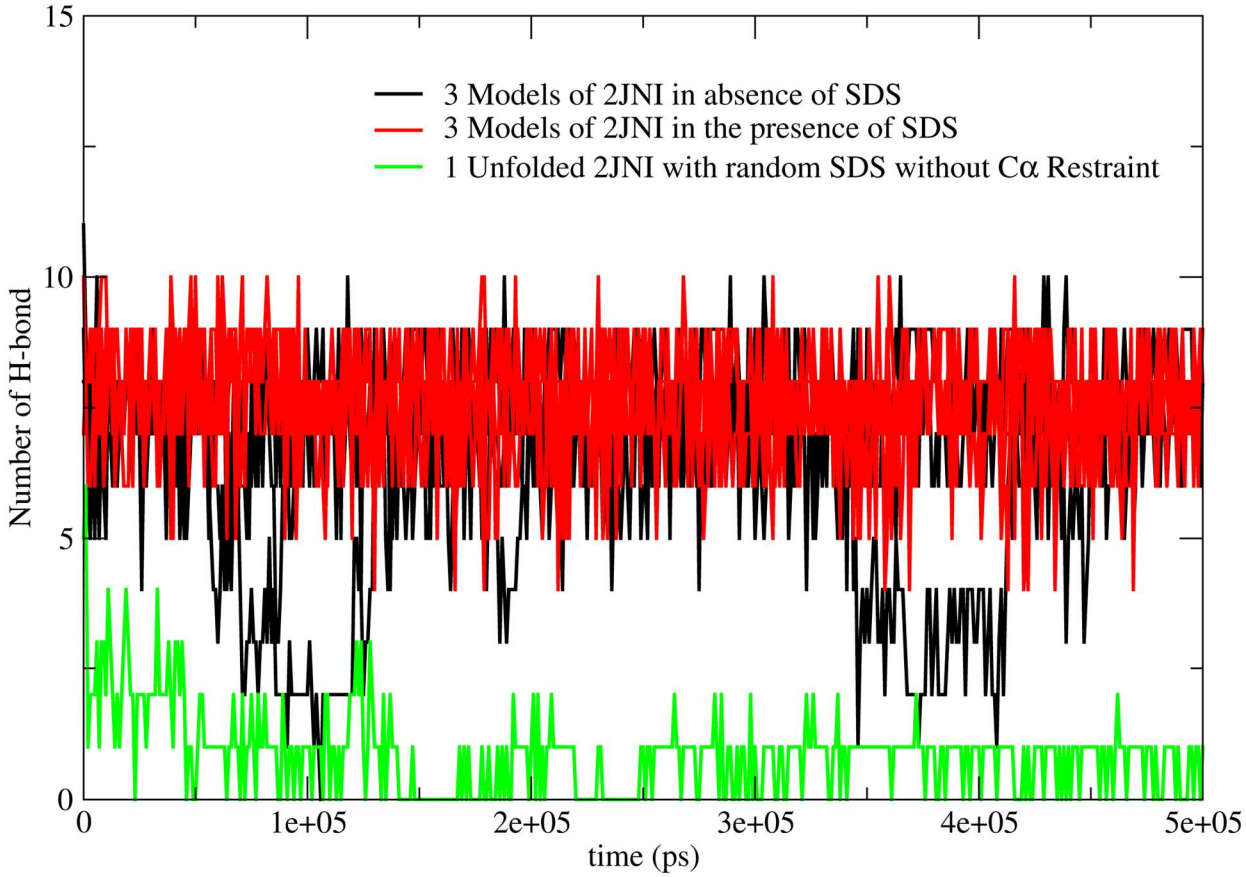


Figure S 4.2 Number of hydrogen bond in 2JNI.

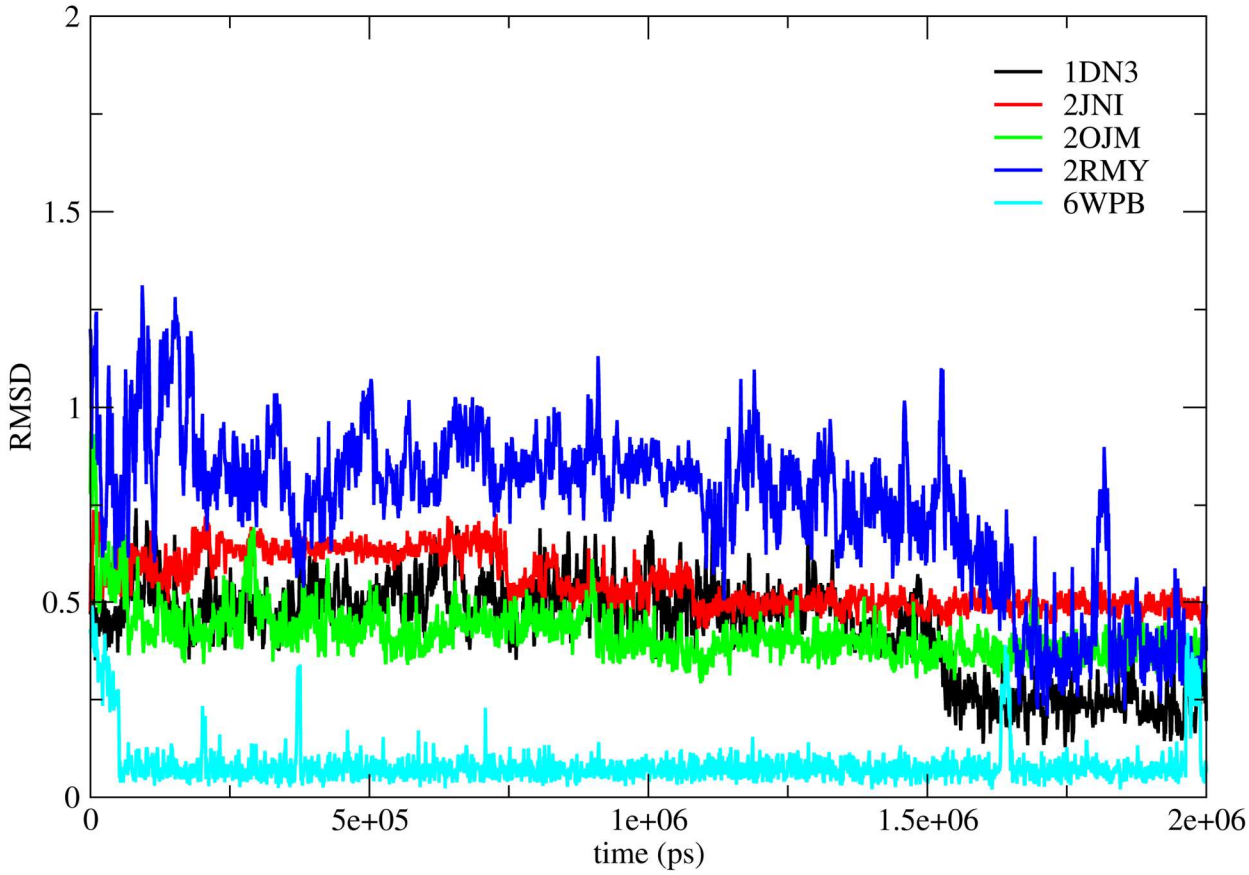


Figure S 4.3 RMSD for unfolded peptide in the presence of random SDS.

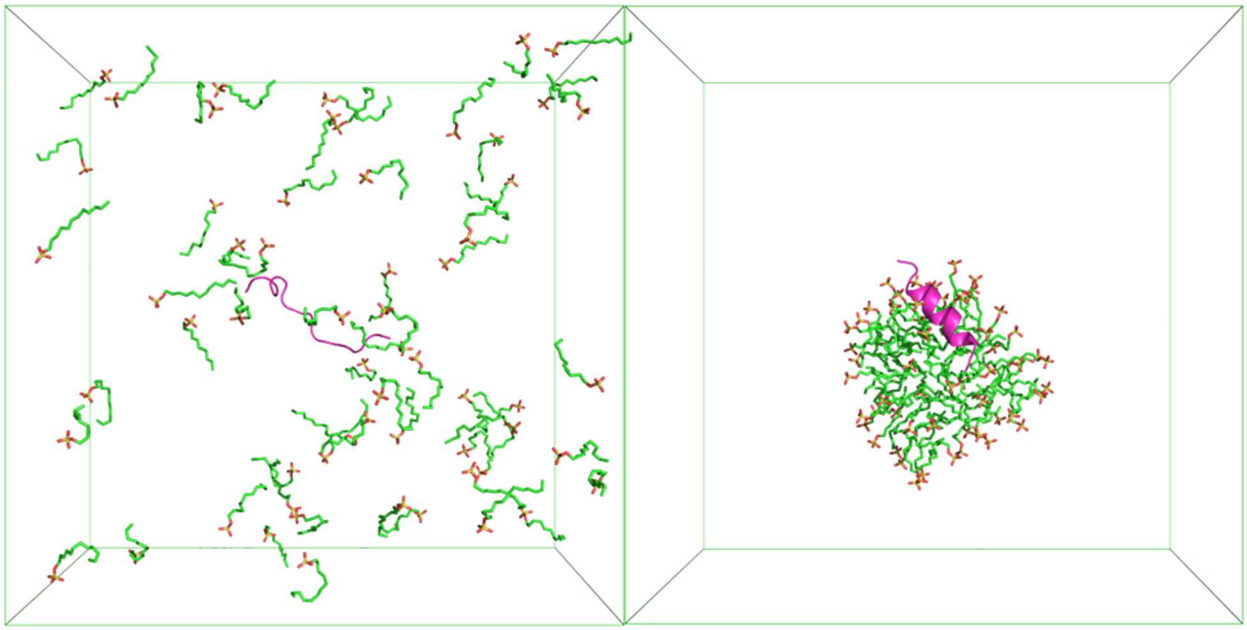


Figure S 4.4 Unfolded 1DN3 in the presence of random SDS simulation. Initial conformation (0ns, left box) and final conformation (2μs, right box).

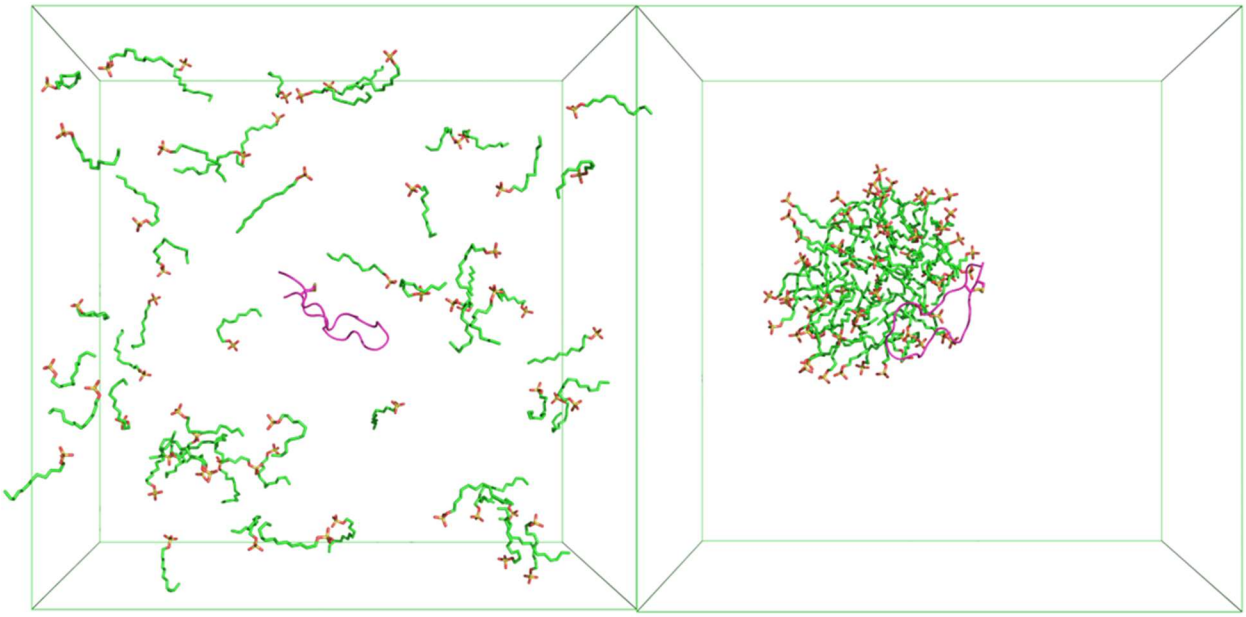


Figure S 4.5 Unfolded 2JNI in the presence of random SDS simulation. Initial conformation (0ns, left box) and final conformation (2μs, right box).

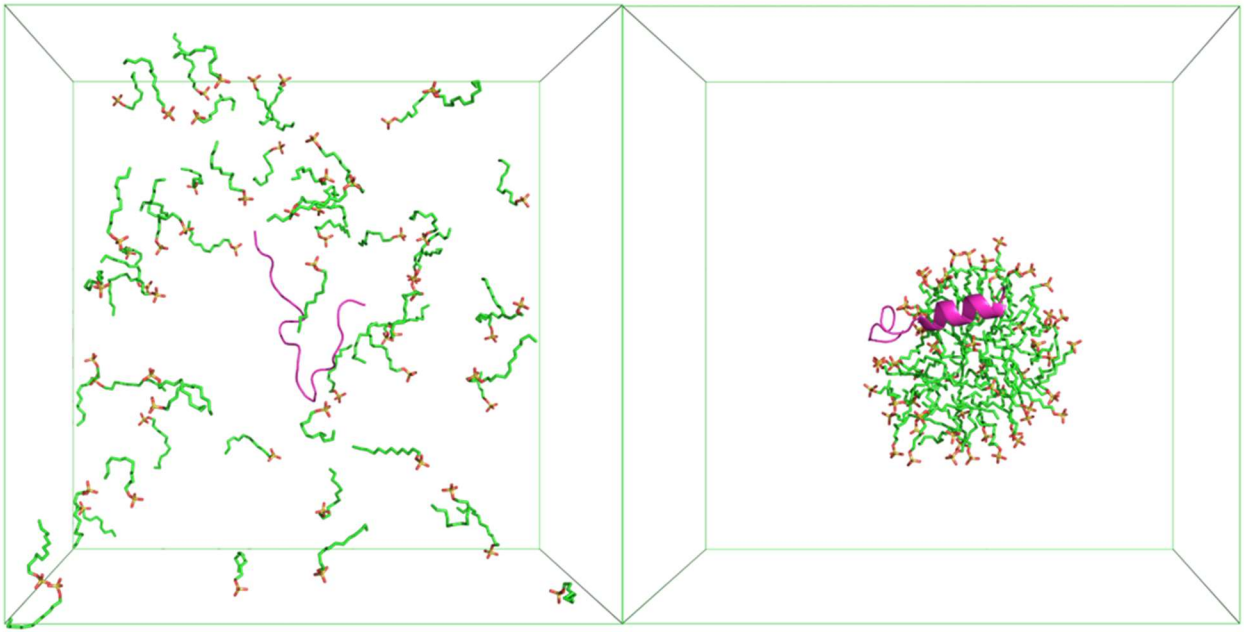


Figure S 4.6 Unfolded 2OJM in the presence of random SDS simulation. Initial conformation (0ns, left box) and final conformation (2μs, right box).

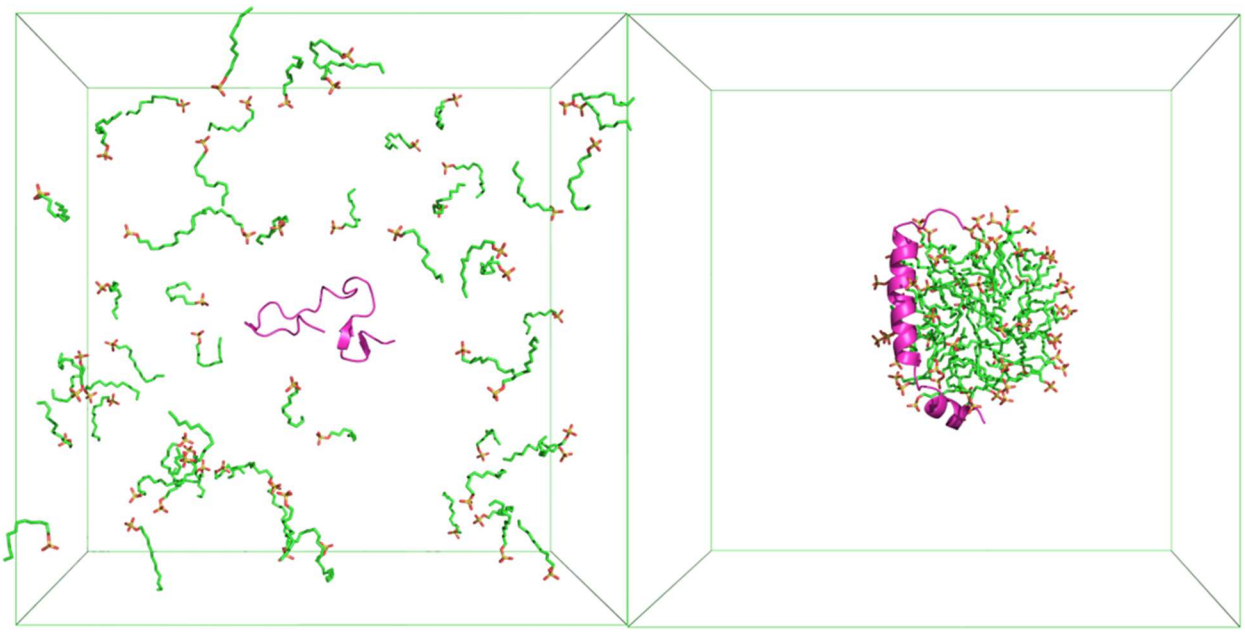


Figure S 4.7 Unfolded 2RMY in the presence of random SDS simulation. Initial conformation (0ns, left box) and final conformation (2μs, right box).

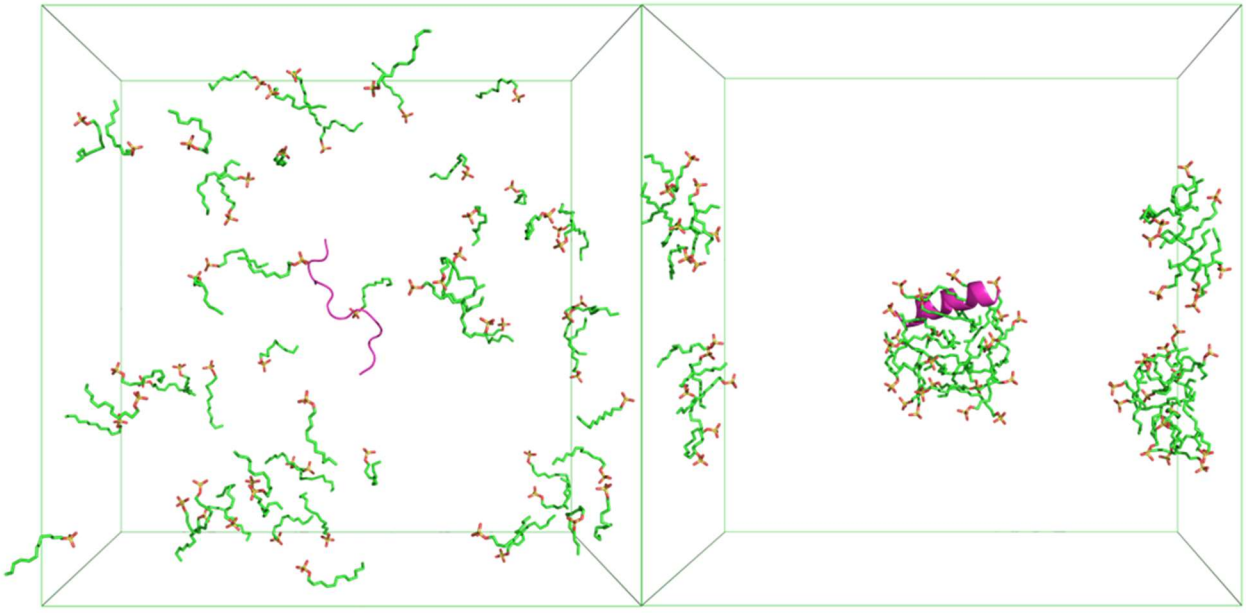


Figure S 4.8 Unfolded 6WPB in the presence of random SDS simulation. Initial conformation (0ns, left box) and final conformation (1μs, right box).

4.6 Reference

(1) Walker, J. M. *The Protein Protocols Handbook*; Humana Press, 2002.

(2) Gallagher, S.; Wiley, E. A. *Current Protocols Essential Laboratory Techniques*; Johh Wiley & Sons, Inc, 2008.

(3) Mahmood, T.; Yang, P.-C. Western Blot: Technique, Theory, and Trouble Shooting. *North American Journal of Medical Science* **2012**, *4* (9), 429-434.

(4) Majeed, S.; Ahmad, A. B.; Sehar, U.; Georgieva, E. R. Lipid Membrane Mimetics in Functional and Structural Studies of Integral Membrane Proteins. *Membranes* **2021**, *11* (9). DOI: 10.3390/membranes11090685.

(5) Carlson, M. L.; Young, J. W.; Zhao, Z. Y.; Fabre, L.; Jun, D.; Li, J. N.; Li, J.; Dhupar, H. S.; Wason, I.; Mills, A. T.; et al. The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. *Elife* **2018**, *7*. DOI: 10.7554/eLife.34085.

(6) Franzin, C. M.; Gong, X. M.; Thai, K.; Yu, J. H.; Marassi, F. M. NMR of membrane proteins in micelles and bilayers: The FXYD family proteins. *Methods* **2007**, *41* (4), 398-408. DOI: 10.1016/j.ymeth.2006.08.011.

(7) Kermani, A. A. A guide to membrane protein X-ray crystallography. *Febs Journal* **2021**, *288* (20), 5788-5804. DOI: 10.1111/febs.15676.

(8) Nielsen, M. M.; Andersen, K. K.; Westh, P.; Otzen, D. E. Unfolding of beta-sheet proteins in SDS. *Biophysical Journal* **2007**, *92* (10), 3674-3685. DOI: 10.1529/biophysj.106.101238.

(9) Sun, Y.; Oseliero, P. L.; Oliveira, C. L. P. alpha-Lactalbumin and sodium dodecyl sulfate aggregates: Denaturation, complex formation and time stability. *Food Hydrocolloids* **2017**, *62*, 10-20. DOI: 10.1016/j.foodhyd.2016.07.031.

(10) Pedersen, J. N.; Lyngso, J.; Zinn, T.; Otzen, D. E.; Pedersen, J. S. A complete picture of protein unfolding and refolding in surfactants. *Chemical Science* **2020**, *11* (3), 699-712. DOI: 10.1039/c9sc04831f.

(11) Naidu, K. T.; Prabhu, N. P. Protein-Surfactant Interaction: Sodium Dodecyl Sulfate-Induced Unfolding of Ribonuclease A. *Journal of Physical Chemistry B* **2011**, *115* (49), 14760-14767. DOI: 10.1021/jp2062496.

(12) Sun, Y.; Filho, P. L. O.; Bozelli, J. C.; Carvalho, J.; Schreier, S.; Oliveira, C. L. P. Unfolding and folding pathway of lysozyme induced by sodium dodecyl sulfate. *Soft Matter* **2015**, *11* (39), 7769-7777. DOI: 10.1039/c5sm01231g.

(13) Valstar, A.; Brown, W.; Almgren, M. The lysozyme-sodium dodecyl sulfate system studied by dynamic and static light scattering. *Langmuir* **1999**, *15* (7), 2366-2374. DOI: 10.1021/la981234n.

(14) Sokalingam, S.; Raghunathan, G.; Soundrarajan, N.; Lee, S. G. A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. *Plos One* **2012**, *7* (7). DOI: 10.1371/journal.pone.0040410.

(15) Xu, Q.; Keiderling, T. A. Effect of sodium dodecyl sulfate on folding and thermal stability of acid-denatured cytochrome c: A spectroscopic approach. *Protein Science* **2004**, *13* (11), 2949-2959. DOI: 10.1110/ps.04827604.

(16) Sau, A. K.; Currell, D.; Mazumdar, S.; Mitra, S. Interaction of sodium dodecyl sulfate with human native and cross-linked hemoglobins: a transient kinetic study. *Biophysical Chemistry* **2002**, *98* (3), 267-273. DOI: 10.1016/s0301-4622(02)00076-5.

(17) Tulumello, D. V.; Deber, C. M. SDS Micelles as a Membrane-Mimetic Environment for Transmembrane Segments. *Biochemistry* **2009**, *48* (51), 12096-12103. DOI: 10.1021/bi9013819.

(18) Giehm, L.; Oliveira, C. L. P.; Christiansen, G.; Pedersen, J. S.; Otzen, D. E. SDS-Induced Fibrillation of alpha-Synuclein: An Alternative Fibrillation Pathway. *Journal of Molecular Biology* **2010**, *401* (1), 115-133. DOI: 10.1016/j.jmb.2010.05.060.

(19) Rangachari, V.; Reed, D. K.; Moore, B. D.; Rosenberry, T. L. Secondary structure and interfacial aggregation of amyloid-beta(1-40) on sodium dodecyl sulfate micelles. *Biochemistry* **2006**, *45* (28), 8639-8648. DOI: 10.1021/bi060323t.

(20) Winogradoff, D.; John, S.; Aksimentiev, A. Protein unfolding by SDS: the microscopic mechanisms and the properties of the SDS-protein assembly. *Nanoscale* **2020**, *12* (9), 5422-5434. DOI: 10.1039/c9nr09135a.

(21) Jalili, S.; Akhavan, M. Study of the Alzheimer's A beta 40 peptide in SDS micelles using molecular dynamics simulations. *Biophysical Chemistry* **2011**, *153* (2-3), 179-186. DOI: 10.1016/j.bpc.2010.11.007.

(22) Ploetz, E. A.; Karunaweera, S.; Bentenitis, N.; Chen, F.; Dai, S.; Gee, M. B.; Jiao, Y.; Kang, M.; Kariyawasam, N. L.; Naleem, N.; et al. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 2964-2990. DOI: 10.1021/acs.jctc.1c00075.

(23) Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B.; team, G. d. GROMACS User Manual version 2016. 2016.

(24) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463-1472. DOI: 10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.

(25) Miyamoto, S.; Kollman, P. A. SETTLE - An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992**, *13* (8), 952-962. DOI: 10.1002/jcc.540130805.

(26) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large System. *Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397.

(27) Montserret, R.; McLeish, M. J.; Bockmann, A.; Geourjon, C.; Penin, F. Involvement of electrostatic interactions in the mechanism of peptide folding induced by sodium dodecyl sulfate binding. *Biochemistry* **2000**, *39* (29), 8362-8373. DOI: 10.1021/bi000208x.

(28) Lee, S.-A.; Kim, Y. K.; Lim, S. S.; Zhu, W. L.; Ko, H.; Shin, S. Y.; Hahm, K.-S.; Kim, Y. Solution structure and cell selectivity of piscidin 1 and its analogues. *Biochemistry* **2007**, *46* (12), 3653-3663, Article. DOI: 10.1021/bi062233u.

(29) Low, C.; Weininger, U.; Lee, H.; Schweimer, K.; Neundorf, I.; Beck-Sickinger, A. G.; Pastor, R. W.; Balbach, J. Structure and Dynamics of Helix-0 of the N-BAR Domain in Lipid Micelles and Bilayers. *Biophysical Journal* **2008**, *95* (9), 4315-4323, Article. DOI: 10.1529/biophysj.108.134155.

(30) Gomes, I. P.; Santos, T. L.; de Souza, A. N.; Nunes, L. O.; Cardoso, G. A.; Matos, C. O.; Costa, L. M. F.; Liao, L. M.; Resende, J. M.; Verly, R. M. Membrane interactions of the anuran antimicrobial peptide HSP1-NH2: Different aspects of the association to anionic and zwitterionic biomimetic systems. *Biochimica Et Biophysica Acta-Biomembranes* **2021**, *1863* (1), Article. DOI: 10.1016/j.bbamem.2020.183449.

(31) Ovchinnikova, T. V.; Shenkarev, Z. V.; Nadezhdin, K. D.; Balandin, S. V.; Zhmak, M. N.; Kudelina, I. A.; Finkina, E. I.; Kokryakov, V. N.; Arseniev, A. S. Recombinant expression,

synthesis, purification, and solution structure of arenicin. *Biochemical and Biophysical Research Communications* **2007**, *360* (1), 156-162. DOI: 10.1016/j.bbrc.2007.06.029.

Chapter 5 – The Behavior of Amylin at Different Interfaces

5.1 Introduction

Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) are proteins, or sections of proteins, that lack a stable three-dimensional structure under physiological conditions.[1] Unlike most proteins that fold into specific structures to carry out their functions, IDPs remain unfolded or adopt a highly dynamic ensemble of conformations. IDPs are common in nature, and they play crucial roles in a wide range of cellular processes, including signaling, regulation, and transcription.[2] Due to their unique properties, IDPs are involved in many diseases, including cancer, Alzheimer's, and Parkinson's disease. IDPs are challenging to study experimentally due to their dynamic nature and the absence of a well-defined structure.[3] Two methodologies have generally been employed to link biological process with proteins that may be structured, that may be IDPs, or that may use IDRs. One method involves performing targeted experiments on specific proteins to investigate their functions in cellular processes and their relationships between structure and function. By examining the relationship between structure and function, it may be possible to discover certain functions that are carried out by disordered regions of proteins (IDPs or IDRs) or by interactions between disordered regions and structured protein partners.[4] An alternative method involves performing extensive computational analyses to uncover new connections between the function and structure of IDRs. Subsequently, additional investigations are carried out to confirm the significance of these connections.[5]

Amylin, also known as islet amyloid polypeptide (IAPP), is a hormone that is produced and secreted by the beta cells of the pancreas along with insulin. Amylin is considered an IDP.[6-8] Amylin is a relatively small peptide consisting of 37 amino acids, and it is characterized by a high degree of structural flexibility and a propensity to aggregate into insoluble amyloid fibrils. In fact,

the intrinsically disordered nature of amylin is thought to play a key role in its aggregation and may contribute to the pathogenesis of type 2 diabetes.[9] Amylin aggregation occurs both within the cellular matrix and at interfaces between different materials, such as air-water or lipid-water interfaces.[10-12] In the cellular matrix, amylin aggregates can form within pancreatic islets and contribute to the destruction of β cells, leading to the development of type 2 diabetes. Amylin aggregates have also been found in the brain tissue of Alzheimer's disease patients and are associated with the formation of amyloid plaques.[13] Amylin molecules can adhere to and aggregate at surfaces such as air-water or lipid-water interfaces.[14] These interfaces provide a unique environment that can promote the formation of amyloid fibrils, which are associated with the development of diseases such as type 2 diabetes and Alzheimer's disease.

There have been numerous studies focused on exploring the structure of amylin at various interfaces. Nanga *et al*. used detergent micelles to stabilize an intermediate structure of human IAPP (hIAPP) at neutral pH and solve its 3D structure in a membrane environment, because the direct application of NMR experiments for lipid bilayers is difficult due to the instant aggregation of the peptide.[15] Their results for an amidated form of hIAPP in SDS micelles demonstrated an overall kinked helix motif with specific residues in a helical conformation, and the angle between the N- and C-terminal helices is constrained to 85°. The differences between the structures of hIAPP with and without amidation and their impact on amyloid formation has been discussed. They also suggested that the structured C-terminus can significantly modulate the kinetics of amyloid formation in the presence or absence of a membrane. Thus, it is important to understand its role in both types of aggregation. The presence of partially structured intermediates can reduce the entropic cost of the initial step of amyloid formation, and helical intermediates have been directly observed for some amyloidogenic proteins.[16] Inhibitors have been constructed that use this

fact to stop amyloid formation.[17] The presence of 3_{10} helices are particularly suited for undergoing a helix to β-sheet transition due to their conformational similarities with β-sheet structures and ease of partial unfolding.[18]

Molecular dynamics (MD) simulations can be a useful tool for studying the behavior of biomolecules at interfaces. MD simulations allow for the investigation of the dynamic behavior of proteins and lipids at a molecular level, providing insights into their structural, thermodynamic, and kinetic properties. In contrast to experiments, MD simulations offer a detailed view of the molecular interactions that occur at the interface, such as the interactions between amylin and lipids. Additionally, MD simulations allow for the study of mutations and their effect on the behavior of amylin at the interface, which can be difficult or impossible to perform experimentally.[19] Molecular dynamics (MD) simulations have greatly aided our understanding of amylin. Notably, a study by Wang *et al*.[20] employed atomistic discrete MD simulations to study amyloidogenic and non-amyloidogenic amylin oligomerization dynamics and conformations. Another study used MD simulations to compare different amylin protofibril models proposed by NMR and X-ray techniques, revealing the steric zipper interactions' crucial role in amyloid aggregate stability.[21] Despite these advances, there's still a gap in knowledge about amylin's behavior at the lipid interface.

In this chapter, we will investigate the conformational changes of Amylin at various interfaces, including air/water, SDS/water, and POPC/water, using our recently developed force fields. The conformational behavior of Amylin will provide detailed information concerning the initial steps required for the formation of alternative structures that may lead to fibril formation.

5.2 Simulation methods

5.2.1 System preparation: Amylin in aqueous solution

The system was prepared using the GROMACS 2016 software package.[22] Three independent simulations were performed using initial structures taken from 3 folded models of the NMR structure of Amylin in the presence of SDS (PDB ID 2L86).[23] It has a helix-kink-helix structure, which matches the color of the configuration and one-letter sequence, as shown in Figure 5.1. The peptides were simulated using the KBFF20 force field.[24] Each system was solvated with SPC/E water molecules in a 10 x 10 x 10 nm cubic box. The system was neutralized by adding an appropriate amount of counter ions (Na^+ or Cl^-) and an additional physiological concentration (0.15M) of NaCl was added.

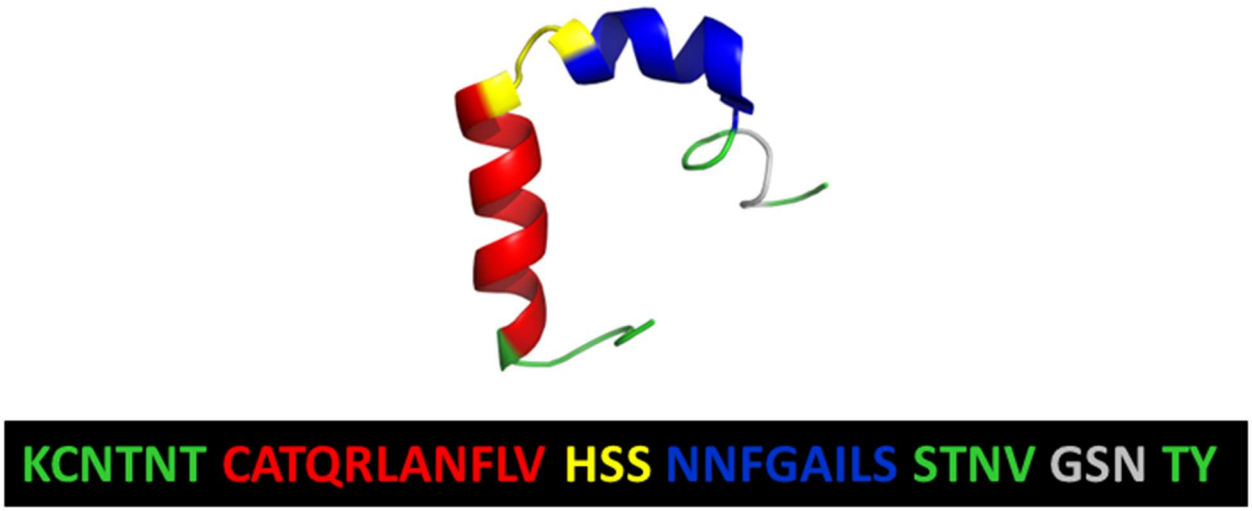


Figure 5.1. 2L86 configuration in the presence of SDS micelle obtained from PDB databank. Top configuration shows helix-kink-helix structure and colored letter at the bottom matches with its structure.

5.2.2 System preparation: Amylin with SDS

Model 1 of 2L86 was placed in a cubic box of size 10 x 10 x 10 nm^3 in the presence of 60 SDS molecules. Three different initial SDS structures - micelle SDS, split half micelle, and random SDS in solution – were used, as shown in the earlier schematic diagram in Figure 4.1. Each box

was then solvated with SPC/E water and an appropriate number of counter ion were added (Na^+ or Cl^-) to neutralize the system. Additional Na^+ and Cl^- ions were added to mimic physiological conditions (0.15M).

5.2.3 System preparation: Unfolded amylin at different interface

Several simulations were also performed starting with unfolded Amylin structures. Three random unfolded structures were selected from the results of the aqueous simulations. Each unfolded structure was placed in a simulation box with dimensions of 10 x 10 x 30 nm^3 with the middle 10 x 10 x 10 nm^3 filled with water to create two interfaces of air/water. Salt was added to neutralize the system and achieve physiological conditions. The simulations were run for 2 μs. One unfolded structure of amylin was placed with 60 random SDS in a simulation box with dimensions of 10 x 10 x 10 nm^3 filled with water. Salt was added to neutralize the system and achieve physiological conditions. The simulations were run for 2 μs.

5.2.4 System preparation: Amylin with POPC

A POPC bilayer containing 512 lipids was constructed and centrally positioned within a 12 x 12 x 12 nm^3 simulation box. The random coil structure of two Amylin peptides selected from a previous aqueous simulation was then placed near the top and bottom of the POPC bilayer. The simulation box was then solvated with SPC/E water, and counter ions (Na^+ or Cl^-) are added to neutralize the system. Additional Na^+ and Cl^- ions are added to simulate physiological conditions (0.15M).

5.2.5 General simulation methods

All simulations in this study maintained an isothermal isobaric ensemble (NpT) at 298K temperature and 1 atm pressure. The temperature and pressure were weakly coupled to a bath with relaxation times of 0.1 ps and 0.5 ps, respectively. A time step of 2 fs was used to integrate the

equations of motion. All bonds in the simulations, including those of the lipids and the amylin peptide, were constrained using the LINCS algorithm, with the highest order in the expansion of the constraint coupling matrix set to 4, and with 4 iterations in the final step of LINCS. [25] To calculate electrostatic interactions, the particle mesh Ewald technique was used with a 1.0 nm cut off distance for real space calculations, as well as 1.0 nm cutoff was used for the van der Waals interactions.[26] An energy minimization using the steepest descent method followed by 100ps of equilibration was performed before the production run. Production runs of 500ns were used to calculate ensemble averages for all the peptides and interfaces without SDS. For the production simulations for the peptide in the presence of SDS, Cα restraints were applied during the first 100ns of the simulation. Additional 500ns simulations were then performed without Cα constraints. All the configurations were saved for analysis every 1 ns.

5.3 Results and Discussion

5.3.1 Amylin in aqueous solution

A series of three independent aqueous simulations were performed for Amylin starting from the initial 2L86 structure determined by NMR in presence of SDS. These initial structures are characterized by a helix-kink-helix conformation as displayed in Figure 5.1. The simulations performed in aqueous solution displayed a progressive unfolding of the Amylin structure, as evidenced by the helicity analysis presented in Figure 5.2. While the simulations provide slightly different helical percentages for the same residues, due to incomplete sampling, the helix-kink-helix structure is clearly unstable in aqueous solution using the KBFF20 model. Furthermore, there was no residual (helical) structure for any residues that was consistent between the three simulations. This agrees with experiment that clearly indicates that Amylin is an IDP.[23] The analysis also indicates that over-stabilization of folded structures of Amylin is not a problem for

KBFF20.

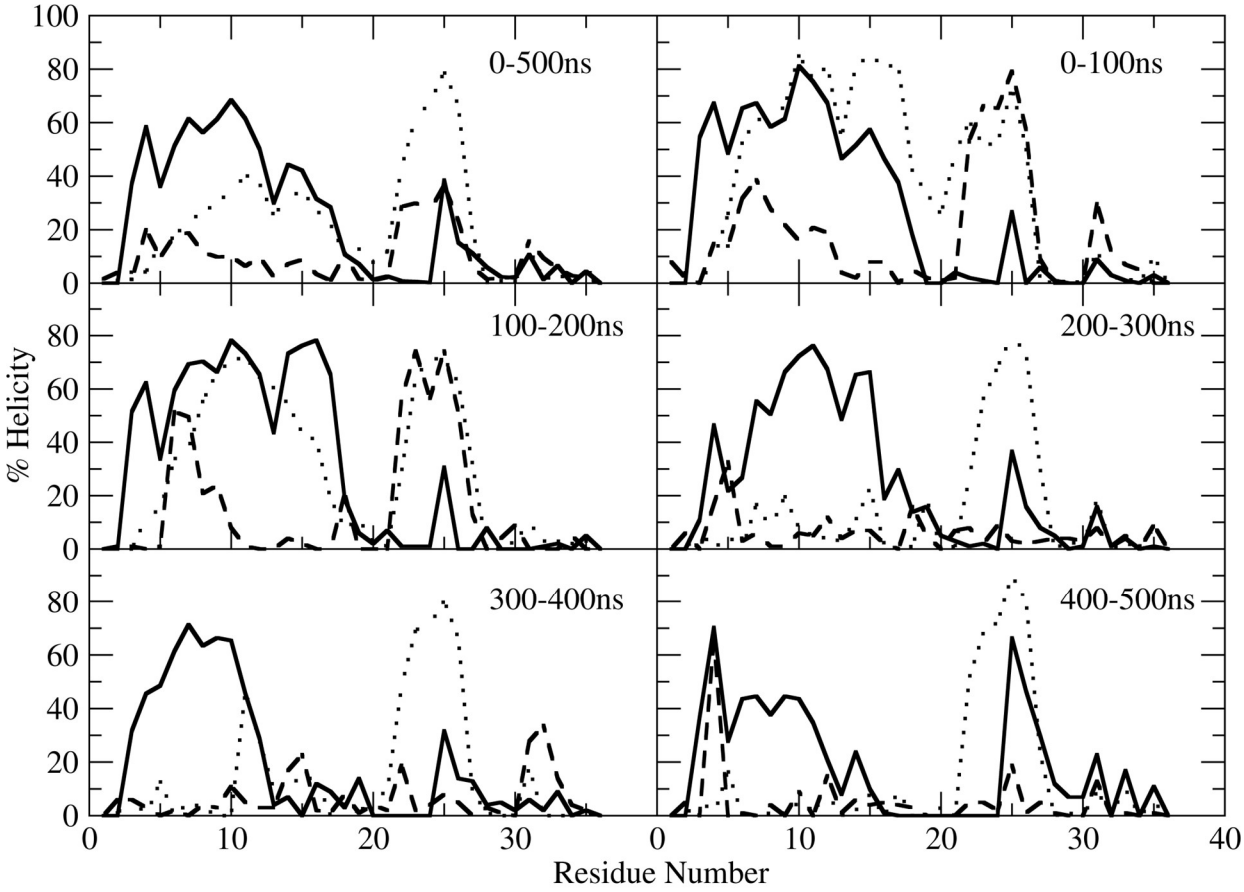


Figure 5.2 Percent Helicity of 2L86, in aqueous solution. Black solid, dotted and dashed lines are represented three different initial folded configurations.

5.3.2 Amylin in the presence of SDS

Three more simulations of Amylin in the presence of SDS were performed starting from the folded helix-kink-helix structure. The resulting helix percentages for the 2L86 models in the presence of SDS are provided in Figure 5.3. These simulations maintained the initial overall helix-kink-helix structure, but with a loss in helix percentage for the C-terminal (27-37) residues. Similar results were obtained for all three simulations: including those associated with an intact micelle (black solid line), the peptide initially situated between two half-micelles (black dotted line), and the peptide in the presence of randomly distributed SDS molecules (black dash line). These

simulations indicate that Amylin is stable in the presence of SDS when using the KBFF20 peptide and SDS force fields.

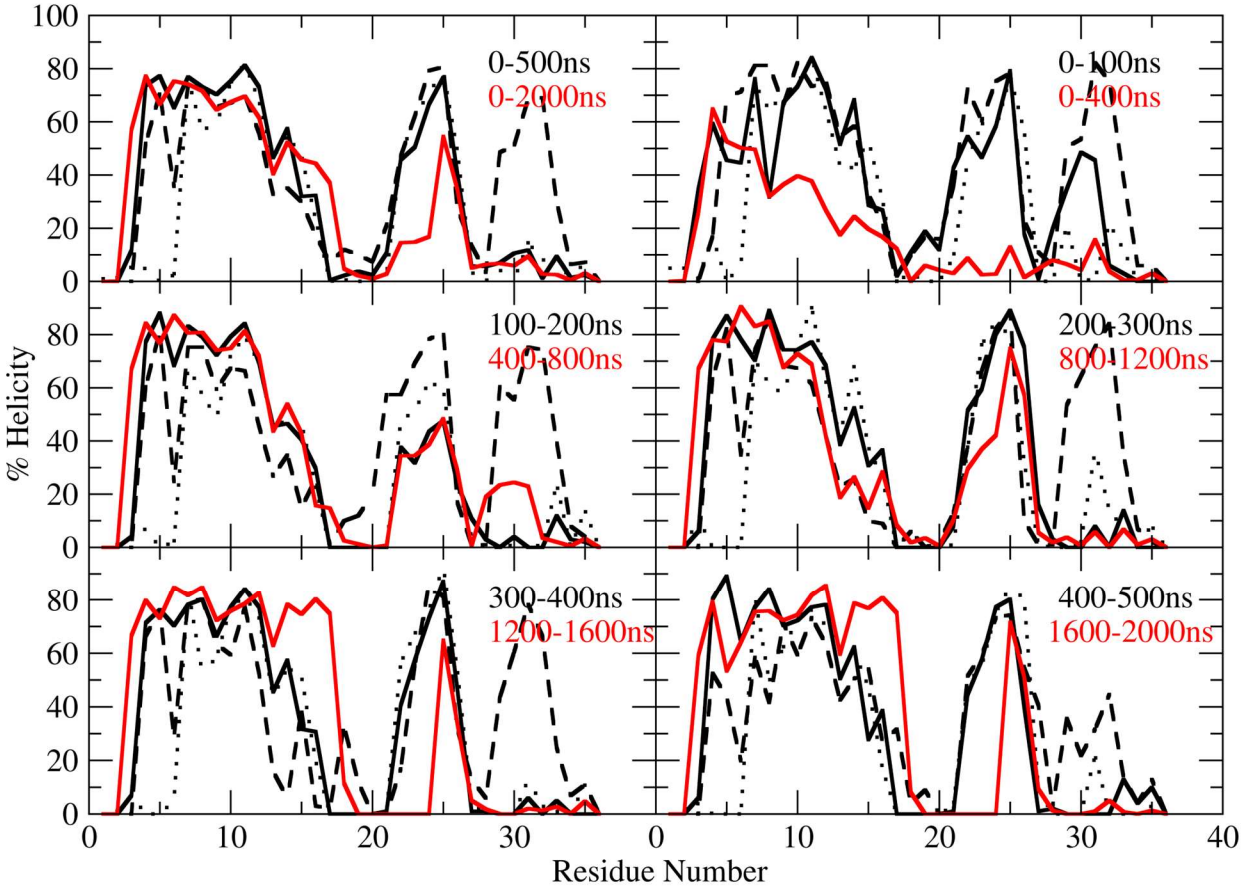


Figure 5.3 Percent helicity of 2L86 in the presence of SDS. All black lines are helix-kink-helix structure with micelle (solid), split micelle (dotted), and random SDS. Each configuration have been simulated for 500ns. Red solid line represents percent helicity results for 2μs simulation of unfolded 2L86 with random SDS as initial configuration.

As an additional test of the reliability of the force fields we have also simulated Amylin in the presence of a set of random SDS molecules starting from an unfolded peptide structure. The results are also shown in Figure 5.3. Here, we observed folding of the Amylin peptide in the presence of SDS. The N-terminal helix formed within the first 500 ns, while towards the end of the 2 μs simulation we also observed moderate C-terminal helix formation. The initial and final configurations for this simulation are shown in Figure 5.4.

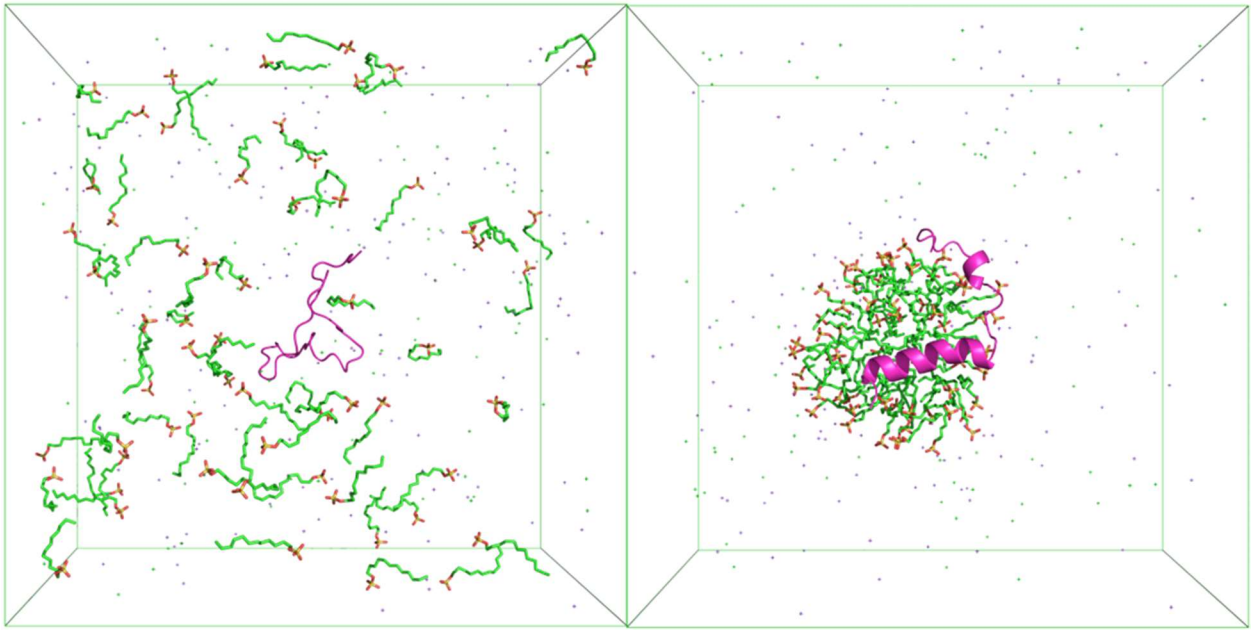


Figure 5.4 Unfolded 2L86 in the presence of random SDS simulation without Cα Constraints. The left box is at 0ns and right box is at 2μs.

An analysis of the radius of gyration and end to end distributions in aqueous solution, and in the presence of SDS, are provided in Figure 5.5. The presence of SDS changes the radius of gyration distribution so it populated a more limited range of possibilities, due to the presence of increased secondary structure. However, the presence of SDS has little effect on the corresponding end to end distribution, indicating that Amylin still maintains a high degree of overall flexibility.

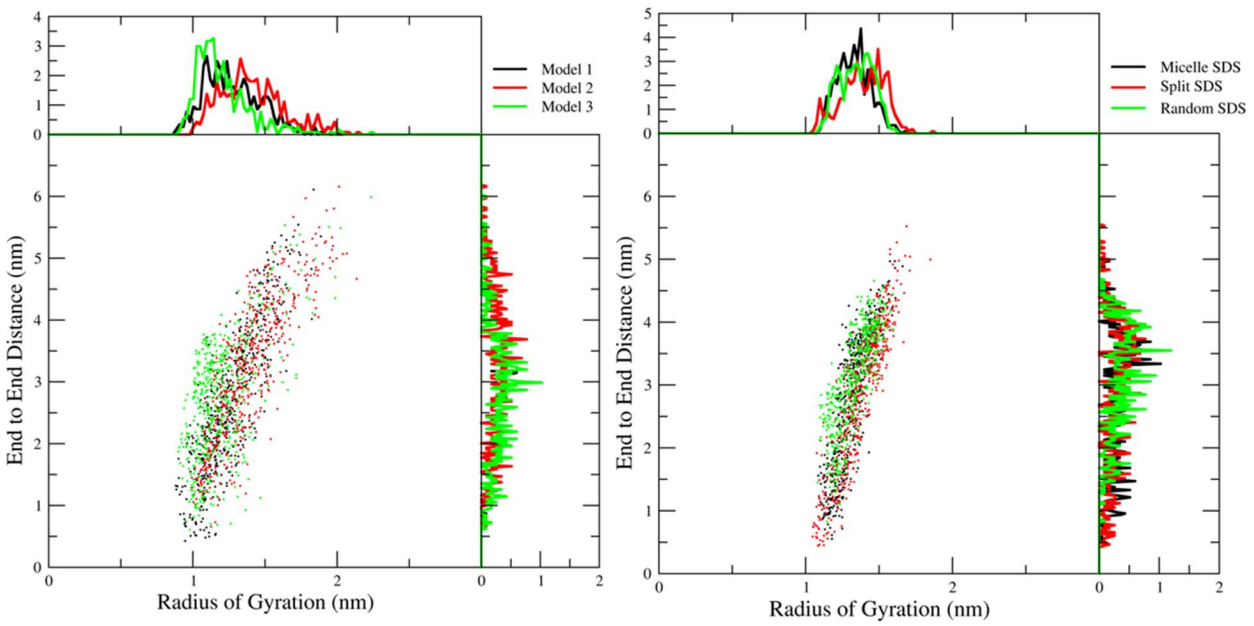


Figure 5.5 Analyses results for radius of gyration and end to end distance. Left graph represents 3 different model of 2L86 in absence of SDS. Right graph represents 2L86 in the presence of SDS. Both simulations started with α-helix structure as initial configuration.

5.3.3 Amylin at air/water interface

The previous comparison of Amylin in aqueous and SDS environments suggest that the structure of Amylin is sensitive to its environment, as observed experimentally, and that the KBFF20 and SDS models provide a good description of the properties of Amylin in these two different environments. We have performed additional simulations of Amylin in the presence of a vacuum/water interface. Three independent simulations of Amylin were performed starting from different unfolded peptide structures located in the center of the aqueous environment. The simulations were run for 2 μs. After 500 ns, all three amylin peptides had migrated towards the air/water interface, as depicted in Figure 5.6. Notably, once the Amylin peptides had reached the interface they remained there, suggesting that the peptide has an affinity for the interface compared to the bulk aqueous solution.

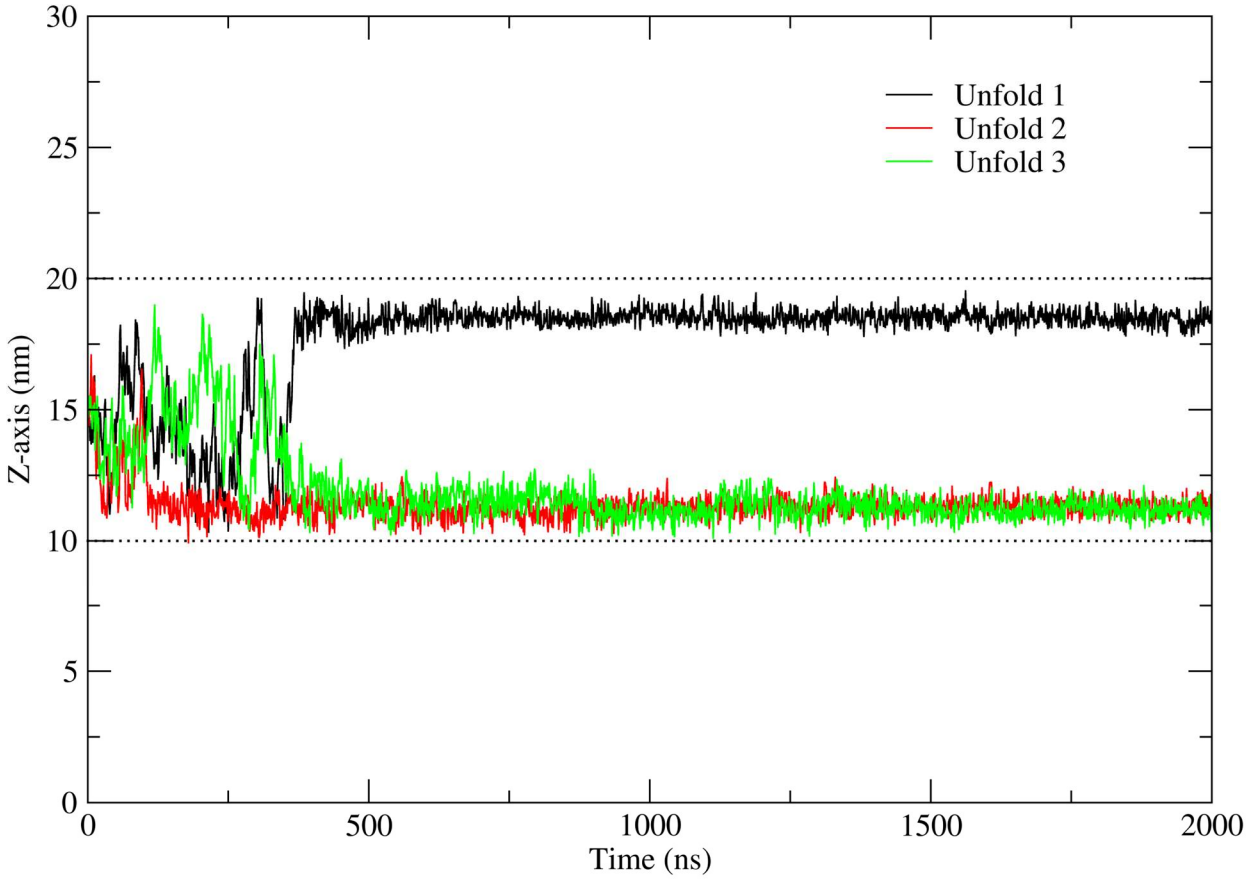


Figure 5.6 Z-axis coordinates of N atom of K1 residue. For 3 unfolded initial structure of amylin during the simulation. Black dotted line is approximation of Air/Water interface location at Z-axis.

Analysis of the simulation data revealed that the helix percentages increased after 500 ns, as depicted in Figure 5.7. This observation suggests that when amylin interacts with the air/water interface, it has a propensity to adopt an α-helical conformation. However, it was noted that the helical conformation adopted by amylin did not exhibit the exact same helix-kink-helix structure observed in the original amylin structure in the presence of SDS micelles. Interestingly, the amyloidogenic region involving the NFGAIL sequence, from residue N22 to L27, exhibits a high percentage of helicity for all three configurations. However, it is clear that additional sampling would be required to obtain converged results between all three simulations. Experimental data from Fu *et al.*[27] demonstrate that hIAPP, when situated at the air/water interface and in the absence

of DPPG, displays a vibrational band around 1650 cm^{-1}. This frequency is indicative of α-helical structures. Such observations suggest that hIAPP adopts a stable, surface-active α-helical conformation in these conditions, a finding that aligns well with the outcomes of our simulation studies.

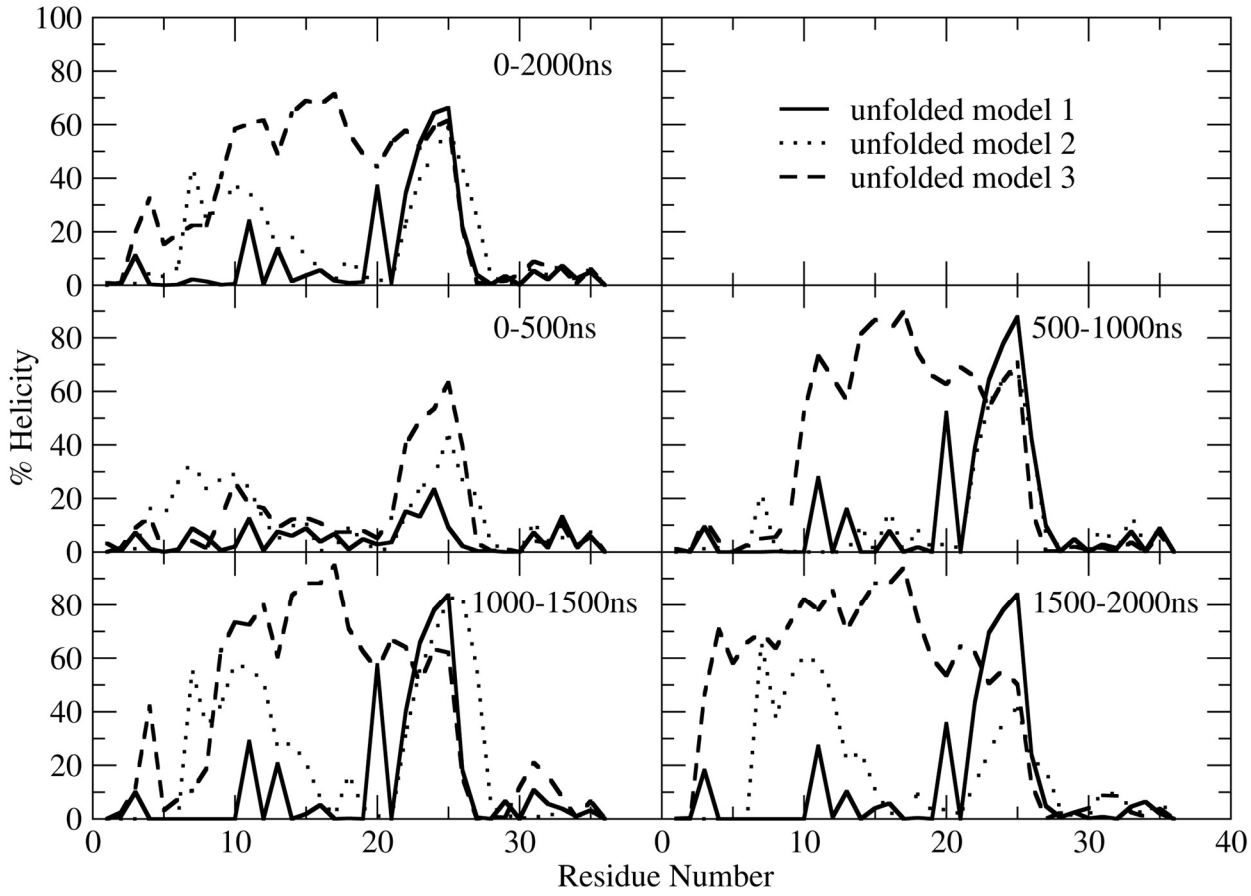


Figure 5.7 Percent helicity for 3 unfolded configuration of 2L86 at air/water interface.

5.3.4 Amylin at lipid bilayer

Two identical unfolded configurations of amylin were placed on the top and bottom of a POPC lipid bilayer and simulated for a duration of 2 μs. After the simulation started, amylin approached the lipid bilayer quickly and then remained on the surface of lipid bilayer. The helicity percentage remained low throughout the simulation period for peptides on both sides of the bilayer. As shown in Figure 5.8, one of the peptides started increasing in helicity after 500ns for residues

between 20-27, which correspond to the region important for the formation of fibril structure. In a study by Khemtemourian *et al*.,[28] the behavior and structural evolution of human Islet Amyloid Polypeptide (hIAPP) at membrane interfaces was explored, particularly focusing on the initial stages of hIAPP aggregation. The research revealed that hIAPP undergoes significant structural changes during its interaction with membrane surfaces, transitioning from an initially disordered structure to parallel β-sheets. Interestingly, prior to the emergence of these β-sheet structures, an intriguing transient chiral N-H stretching signal was detected. This signal was attributed to the formation of an α-helical intermediate structure. The sequential observation of these signals suggests a complex misfolding process, wherein hIAPP first assumes an α-helical structure before transitioning into β-sheets. The final conformations after 2 μs are shown in Figure 5.9. As simulation started, amylin approached lipid bilayer quickly and remained on the surface of lipid bilayer. Thus, it appears we can conclude that amylin prefers to stay on the surface of lipid bilayer, with a slight propensity for increased helical content, but the simulations are not converged.

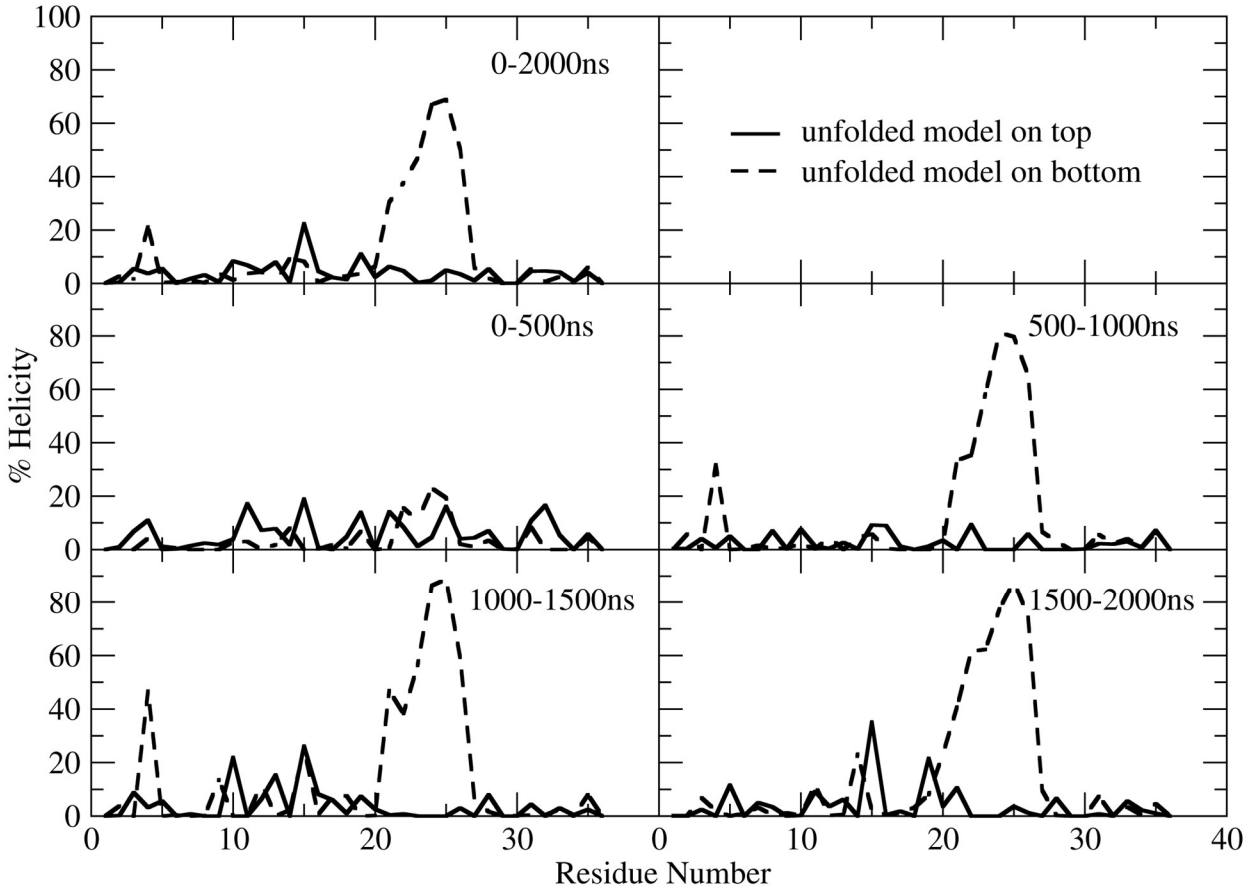


Figure 5.8 Percent helicity of unfolded 2L86 on lipid bilayer. Two are identical configuration located on top or bottom of lipid bilayer.

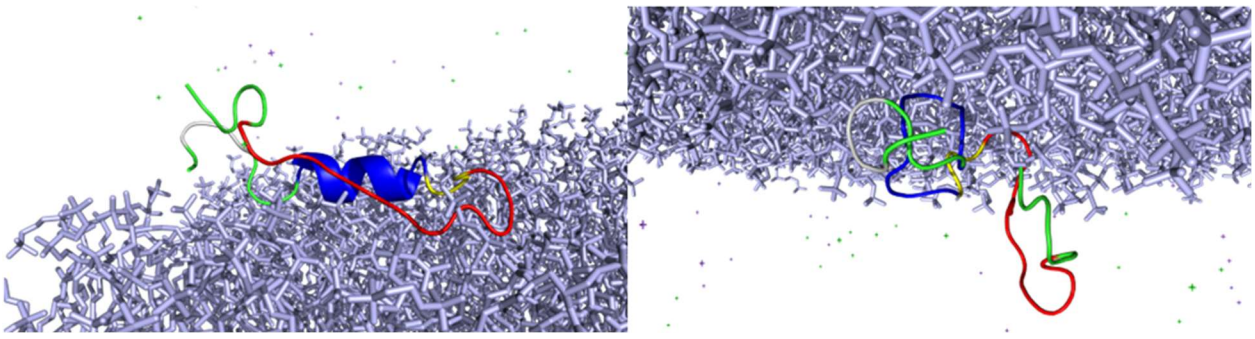


Figure 5.9 Final conformation of unfolded 2L86 after 2s simulation. Left is 2L86 on top of lipid bilayer and right is 2L86 on the bottom of lipid bilayer. Grey color represents lipid bilayer.

5.4 Conclusions

Here, we have investigated the conformational properties of the amylin peptide in the presence of different environments. In this study, it was found that the helix-kink-helix form of

amylin was maintained in the presence of SDS, but gradually unfolded in aqueous solution over the course of the simulations. Amylin prefers to be located at an SDS, air/water or a lipid/water interface, presumably so it can better accommodate a mixture of polar and nonpolar amino acid sidechains. Although some helix formation was observed at the interfaces, they were not always identical to the original kink-helix-kink structure of 2L86 from the PDB database. These exploratory simulations clearly suggest that the conformation of Amylin is dependent on its environment. The absence of structure in aqueous solution, together with the helical structure in the presence of SDS micelles, seem well reproduced by the length of simulations used here. However, the simulations at the vacuum/water interface displayed slower dynamics, and this problem appeared to be even worse for the lipid/water interface. Clearly, additional sampling would therefore be required for a deeper quantitative analysis of the different structures available to the Amylin peptide.

.

5.5 Supporting Information

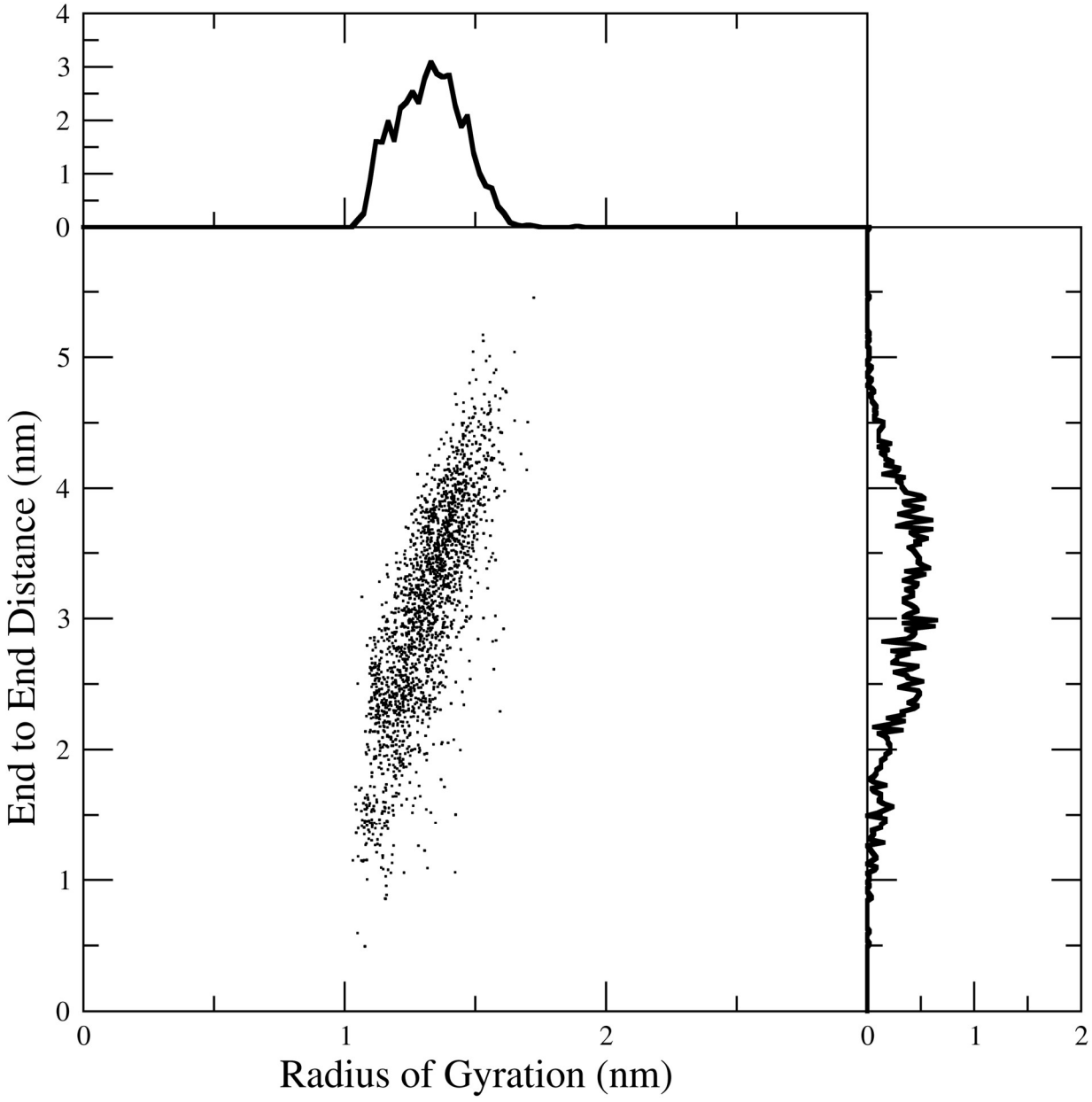


Figure S 5.1 Analysis results for the radius of gyration and end-to-end distance were obtained for the 2μs simulation of unfolded 2L86 with randomly distributed SDS.

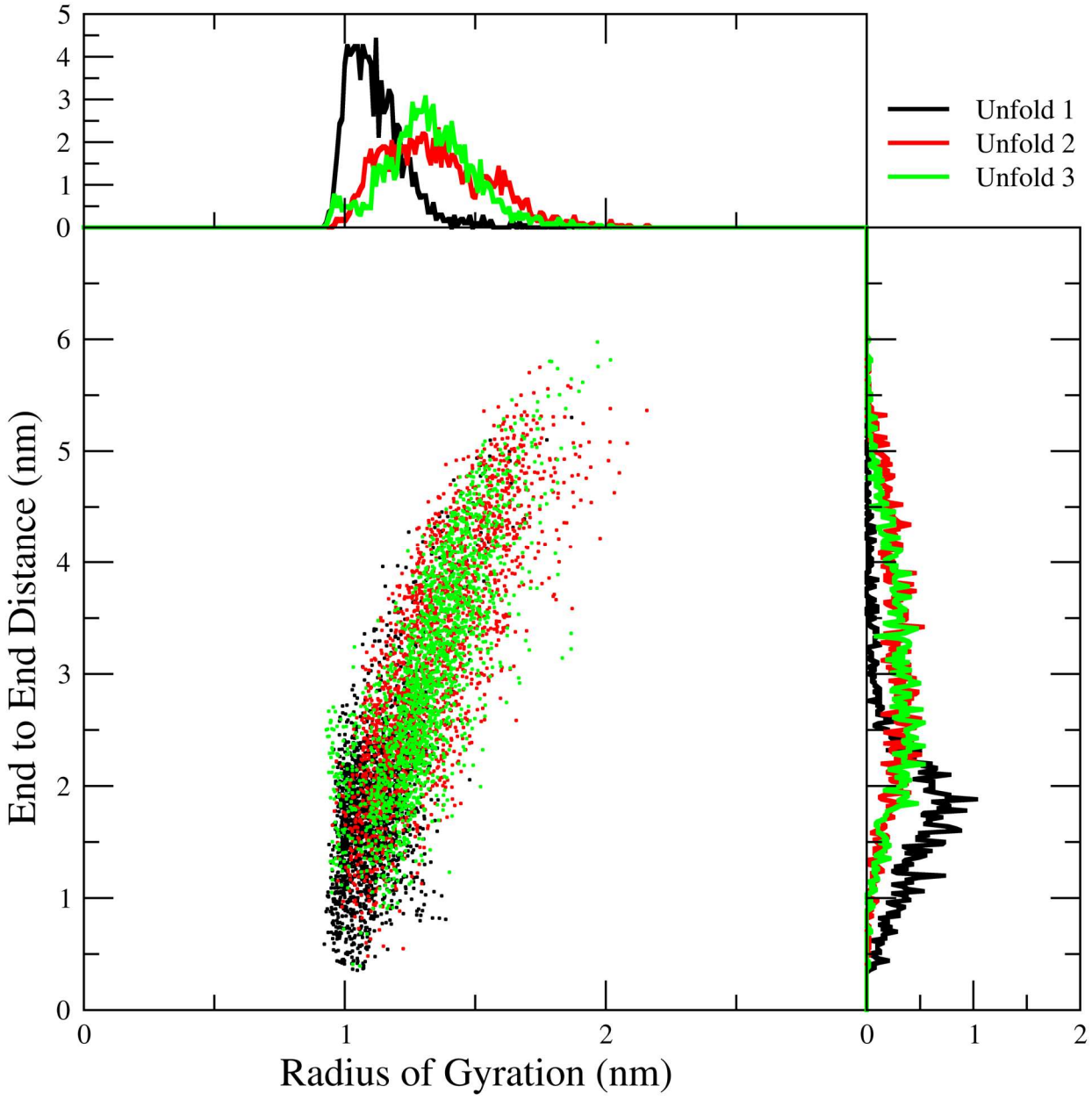


Figure S 5.2 Analysis results for the radius of gyration and end-to-end distance were obtained for the 2μs simulation of 3 unfolded 2L86 at air/water interface.

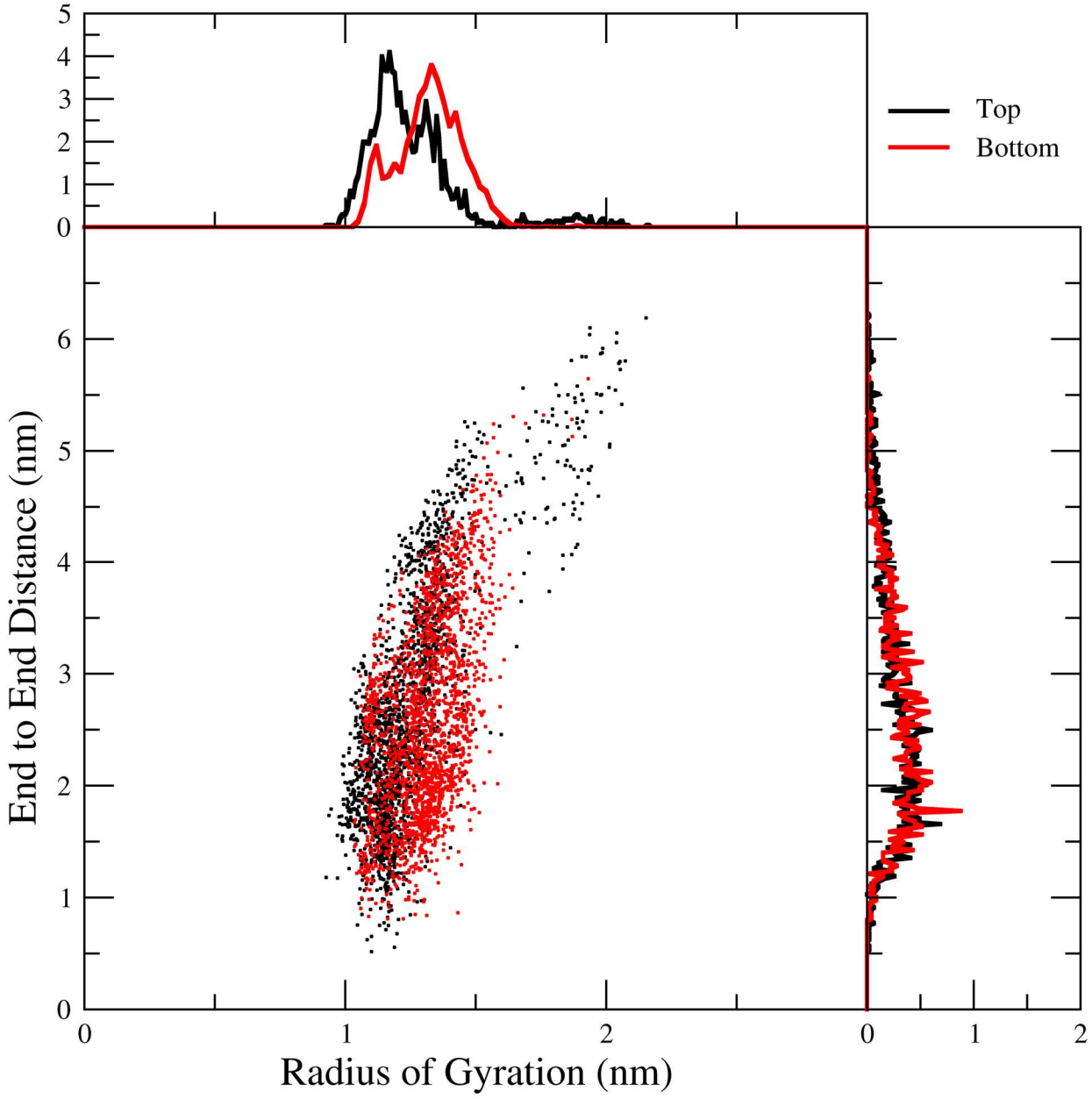


Figure S 5.3 Analysis results for the radius of gyration and end-to-end distance were obtained for the 2μs simulation of 2 unfolded 2L86 at lipid/water interface.

5.6 Reference

(1) Trivedi, R.; Nagarajaram, H. A. Intrinsically Disordered Proteins: An Overview. *International Journal of Molecular Sciences* **2022**, *23* (22). DOI: 10.3390/ijms232214050.

(2) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **2015**, *16* (1), 18-29. DOI: 10.1038/nrm3920.

(3) Uversky, V. N. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D-2 concept. *Expert Review of Proteomics* **2010**, *7* (4), 543-564. DOI: 10.1586/epr.10.36.

(4) Liu, Z. R.; Huang, Y. Q. Advantages of proteins being disordered. *Protein Science* **2014**, *23* (5), 539-550. DOI: 10.1002/pro.2443.

(5) Han, B. Q.; Ren, C. J.; Wang, W. D.; Li, J. S.; Gong, X. Q. Computational Prediction of Protein Intrinsically Disordered Region Related Interactions and Functions. *Genes* **2023**, *14* (2). DOI: 10.3390/genes14020432.

(6) Cao, P.; Marek, P.; Noor, H.; Patsalo, V.; Tu, L. H.; Wang, H.; Abedini, A.; Raleigh, D. P. Islet amyloid: From fundamental biophysics to mechanisms of cytotoxicity. *Febs Letters* **2013**, *587* (8), 1106-1118. DOI: 10.1016/j.febslet.2013.01.046.

(7) Fernandez, M. S. Human IAPP amyloidogenic properties and pancreatic beta-cell death. *Cell Calcium* **2014**, *56* (5), 416-427. DOI: 10.1016/j.ceca.2014.08.011.

(8) Milardi, D.; Gazit, E.; Radford, S. E.; Xu, Y.; Gallardo, R. U.; Caflisch, A.; Westermark, G. T.; Westermark, P.; La Rosa, C.; Ramamoorthy, A. Proteostasis of Islet Amyloid Polypeptide: A Molecular Perspective of Risk Factors and Protective Strategies for Type II Diabetes. *Chemical Reviews* **2021**, *121* (3), 1845-1893. DOI: 10.1021/acs.chemrev.0c00981.

(9) Kanatsuka, A.; Kou, S.; Makino, H. IAPP/amylin and beta-cell failure: implication of the risk factors of type 2 diabetes (vol 9, pg 143, 2018). *Diabetology International* **2018**, *9* (4), 266-266. DOI: 10.1007/s13340-018-0363-1.

(10) Lopes, D. H. J.; Meister, A.; Gohlke, A.; Hauser, A.; Blume, A.; Winter, R. Mechanism of islet amyloid polypeptide fibrillation at lipid interfaces studied by infrared reflection absorption spectroscopy. *Biophysical Journal* **2007**, *93* (9), 3132-3141. DOI: 10.1529/biophysj.107.110635.

(11) Tan, J. J.; Zhang, J. H.; Luo, Y.; Ye, S. J. Misfolding of a Human Islet Amyloid Polypeptide at the Lipid Membrane Populates through beta-Sheet Conformers without Involving alpha-Helical Intermediates. *Journal of the American Chemical Society* **2019**, *141* (5), 1941-1948. DOI: 10.1021/jacs.8b08537.

(12) Bleem, A.; Daggett, V. Structural and Functional Diversity Among Amyloid Proteins: Agents of Disease, Building Blocks of Biology, and Implications for Molecular Engineering. *Biotechnology and Bioengineering* **2017**, *114* (1), 7-20. DOI: 10.1002/bit.26059.

(13) Jackson, K.; Barisone, G. A.; Diaz, E.; Jin, L. W.; DeCarli, C.; Despa, F. Amylin Deposition in the Brain: A Second Amyloid in Alzheimer Disease? *Annals of Neurology* **2013**, *74* (4), 517-526. DOI: 10.1002/ana.23956.

(14) Relini, A.; Marano, N.; Gliozzi, A. Probing the interplay between amyloidogenic proteins and membranes using lipid monolayers and bilayers. *Advances in Colloid and Interface Science* **2014**, *207*, 81-92. DOI: 10.1016/j.cis.2013.10.015.

(15) Nanga, R. P. R.; Brender, J. R.; Xu, J. D.; Veglia, G.; Ramamoorthy, A. Structures of Rat and Human Islet Amyloid Polypeptide IAPP(1-19) in Micelles by NMR Spectroscopy. *Biochemistry* **2008**, *47* (48), 12689-12697. DOI: 10.1021/bi8014357.

(16) Hall, D.; Hirota, N.; Dobson, C. M. A toy model for predicting the rate of amyloid formation from unfolded protein. *Journal of Molecular Biology* **2005**, *351* (1), 195-205. DOI: 10.1016/j.jmb.2005.05.013.

(17) Kim, J.; Harada, R.; Kobayashi, M.; Kobayashi, N.; Sode, K. The inhibitory effect of pyrroloquinoline quinone on the amyloid formation and cytotoxicity of truncated alpha-synuclein. *Molecular Neurodegeneration* **2010**, *5*. DOI: 10.1186/1750-1326-5-20.

(18) Harrison, R. S.; Sharpe, P. C.; Singh, Y.; Fairlie, D. P. Amyloid peptides and proteins in review. *Reviews of Physiology, Biochemistry and Pharmacology, Vol 159* **2007**, *159*, 1-77. DOI: 10.1007/112_2007_0701.

(19) Tofoleanu, F.; Yuan, Y.; Pickard, F. C.; Tywoniuk, B.; Brooks, B. R.; Buchete, N. V. Structural Modulation of Human Amylin Protofilaments by Naturally Occurring Mutations. *Journal of Physical Chemistry B* **2018**, *122* (21), 5657-5665. DOI: 10.1021/acs.jpcb.7b12083.

(20) Wang, Y.; Liu, Y. Y.; Zhang, Y.; Wei, G. H.; Ding, F.; Sun, Y. X. Molecular insights into the oligomerization dynamics and conformations of amyloidogenic and non-amyloidogenic amylin from discrete molecular dynamics simulations. *Physical Chemistry Chemical Physics* **2022**, *24* (36), 21773-21785. DOI: 10.1039/d2cp02851d.

(21) Xu, W. X.; Su, H. B.; Zhang, J. Z. H.; Mu, Y. G. Molecular Dynamics Simulation Study on the Molecular Structures of the Amylin Fibril Models. *Journal of Physical Chemistry B* **2012**, *116* (48), 13991-13999. DOI: 10.1021/jp308708h.

(22) Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B.; team, G. d. GROMACS User Manual version 2016. 2016.

(23) Nanga, R. P. R.; Brender, J. R.; Vivekanandan, S.; Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane

environment. *Biochimica Et Biophysica Acta-Biomembranes* **2011**, *1808* (10), 2337-2342. DOI: 10.1016/j.bbamem.2011.06.012.

(24) Ploetz, E. A.; Karunaweera, S.; Bentenitis, N.; Chen, F.; Dai, S.; Gee, M. B.; Jiao, Y. F.; Kang, Y. S.; Kariyawasam, N. L.; Naleem, N.; et al. Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 2964-2990. DOI: 10.1021/acs.jctc.1c00075.

(25) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463-1472. DOI: 10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.

(26) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large System. *Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397.

(27) Fu, L.; Ma, G.; Yan, E. C. Y. In Situ Misfolding of Human Islet Amyloid Polypeptide at Interfaces Probed by Vibrational Sum Frequency Generation. *Journal of the American Chemical Society* **2010**, *132* (15), 5405-5412. DOI: 10.1021/ja909546b.

(28) Khemtemourian, L.; Gazit, E.; Miranker, A. Recent Insight in Islet Amyloid Polypeptide Morphology, Structure, Membrane Interaction, and Toxicity in Type 2 Diabetes. *Journal of Diabetes Research* **2016**, *2016*, Editorial Material. DOI: 10.1155/2016/2535878.

Chapter 6 – The Effects of Amylin Mutations on the Behavior at Interfaces

6.1 Introduction

The process of amylin aggregation, which leads to fibril formation, is challenging to study due to its large size and lack of permanent structure in solution. However, similar fibril formation has been observed for small peptide fragments derived from Amylin, specifically involving residues 22-27 or NFGAIL, which is known as the amyloidogenic region.[1, 2] The importance of each amino acid for fibril formation has been quantified through experiments, and NFGAIL sequences have been crystallized in beta sheet arrangements, and their aggregation has been studied through simulation.[3, 4]

Amylin aggregation occurs within the cellular matrix and at interfaces between different materials such as air-water or lipid-water interfaces.[5] Our simulation results in Chapter 5 show that amylin prefers to stay at interfaces, highlighting the importance of investigating the factors that contribute to amylin aggregation at these interfaces.

Comparisons involving amylin from different species and even with other peptides from the calcitonin superfamily have enabled a better understanding of some regions in this molecule. The amyloidogenic sequence N20-S29 has been suggested as the primary region responsible for the aggregation function.[6-9] However, given the complexity of amylin aggregation, exploring the role of other regions and investigating the effect of mutations on amylin aggregation may be necessary.

Surface-selective sum frequency generation (SFG) spectroscopy has been applied to study the secondary structures of hIAPP and rIAPP at the air/water interface.[10] In the absence of dipalmitoylphosphoglycerol (DPPG), the SFG spectra of both hIAPP and rIAPP are fitted into

single peaks centered at ~1650 cm^{-1}, which contain contributions from both random-coil and α-helical structures. Fu et al.[10] suggested that both hIAPP and rIAPP adopt α-helical and random-coil structures at the air/water interface in the absence of DPPG. However, upon interactions with lipid molecules at the air/water interface, hIAPP gradually folds into parallel multistranded β-sheets, together with the turn structures between the strands of β-sheets, with the long axis of the β-sheets aligned parallel to the air/water interface. In contrast, rIAPP does not undergo such structural changes at the air/water interface.

Various structures of IAPP in different membrane environments have been reported,[11-14] including SDS micelles at neutral pH for hIAPP, rat IAPP, and pramlintide (a modified Amylin). These structures generally agree but differ in the bend angle between stable and dynamic α-helix segments. Factors like low pH and the rat IAPP sequence result in lower bend angles,[13, 15] while hIAPP at neutral pH favors a formation with a more pronounced bend. The kink in hIAPP structure in SDS aligns with the β-turn region in amyloid fibrils, suggesting a correlation between membrane-induced bends and β-turn formation.[11] Molecular dynamics simulations support this, showing higher aggregation propensity with more hairpin structures. Indeed, the S20G mutation exhibits a more pronounced bend, facilitating amyloid formation.[16]

This study builds upon previous work and investigates the conformational changes of hIAPP mutations at different interfaces, including the vacuum/water and SDS micelle/water interfaces, by studying the unfolded full-length hIAPP sequence, the rat/mice amylin sequences that do not form fibrils, as well as pramlintide and a deaminated form of Amylin. The goal of this study is to gain insights into the structural properties of amylin and its variants at different interfaces. The use of molecular dynamics simulations provides a detailed view of the molecular interactions that occur. Our findings could provide valuable insights into the initial mechanism of

amylin aggregation and inform the development of therapeutic strategies to prevent or treat diseases associated with amylin aggregation.

6.2 Simulation Methods

6.2.1 System preparation: Amylin mutation

Unfolded amylin, as described in Chapter 5, was selected from the previous study. We then modified the sequence to generate rat/mice amylin sequences, as well as pramlintide and deaminated forms, as shown in Figure 6.1 using Pymol molecular modeling software.[17]

Human	:	KCNTNTCATQ RLANFLVHSS NNFGAILSST NVGSNTY-NH_2
Pramlintide	:	---------- ---------- ----P--PP- --------NH_2
Rat/Mouse	:	---------- -------R-- --L-PV-PP- --------NH_2
Deamidated	:	---------- ---------- ---------- -------

Figure 6.1 This Figure depicts the modified amylin sequence for mutation analysis. In Chapter 5, the original human amylin sequence was studied, and in this chapter, the sequence has been modified as illustrated in this figure.

6.2.2 System preparation: Amylin mutation in aqueous solution

Unfolded structures of three mutants (deamidated C-terminal, amylin with rat sequence, and pramlintide) were placed in each simulation box with dimensions of 10 x 10 x 10 nm^3 filled with water. Salt was added to neutralize the system and achieve physiological conditions and simulated for 500ns.

6.2.3 System preparation: Amylin mutation in the presence of SDS micelle

Each mutant was also placed in the presence of SDS micelle in a simulation box with dimensions of 10 x 10 x 10 nm^3. Salt was added to neutralize the system and achieve physiological conditions. The simulations were run for 1 μs.

6.2.4 System preparation: Amylin mutation at Air/Water interface

Each mutation structure was placed in a simulation box with dimensions of 10 x 10 x 30 nm^3 with the middle 10 x 10 x 10 nm^3 filled with water to create two interfaces of air/water. Salt was added to neutralize the system and achieve physiological conditions. The simulations were run for 2 μs.

6.2.5 Simulation methods

In this study, all simulations were conducted under an isothermal isobaric ensemble (NpT) at a temperature of 298 K and a pressure of 1 atm. The temperature and pressure were weakly coupled to a bath with relaxation times of 0.1 ps and 0.5 ps, respectively.[18] A time step of 2 fs was used to integrate the equations of motion. The LINCS algorithm[19] was employed to constrain all bonds in the simulations.

For electrostatic interactions, the particle mesh Ewald technique[20] was utilized with a real space cutoff distance of 1 nm. Van der Waals interactions were calculated with a cutoff distance of 1.0 nm. Prior to the production run, an energy minimization using the steepest descent method was performed, followed by a 100 ps equilibration phase.

The production run for the peptide without SDS lasted 500 ns, during which ensemble averages were calculated. On the other hand, the simulation for the peptide in the presence of SDS micelle was conducted for 1 μs. Configurations were saved for analysis every 1 ns in both cases. Additionally, the production simulation for the peptide at the air/water interface was simulated for 1 μs, with configurations saved every 1 ns as well.

6.3 Results and Discussion

6.3.1 Amylin mutation in aqueous solution

In the present study, we conducted additional 500 ns simulations of a series of unfolded mutant structures of the Amylin peptide in an aqueous solution. The resulting observed simulated percent helicity is illustrated in Figure 6.2. The Pramlintide and rat/mouse mutants included several substitutions for proline, a known helix breaking residue, while the deamidated structure involves a single amino acid change. Hence, it is not surprising that all the simulations displayed little to no statistically significant helix propensity. These results align with the nature of human amylin as wild type (WT), which is widely recognized as an IDP in aqueous solution.

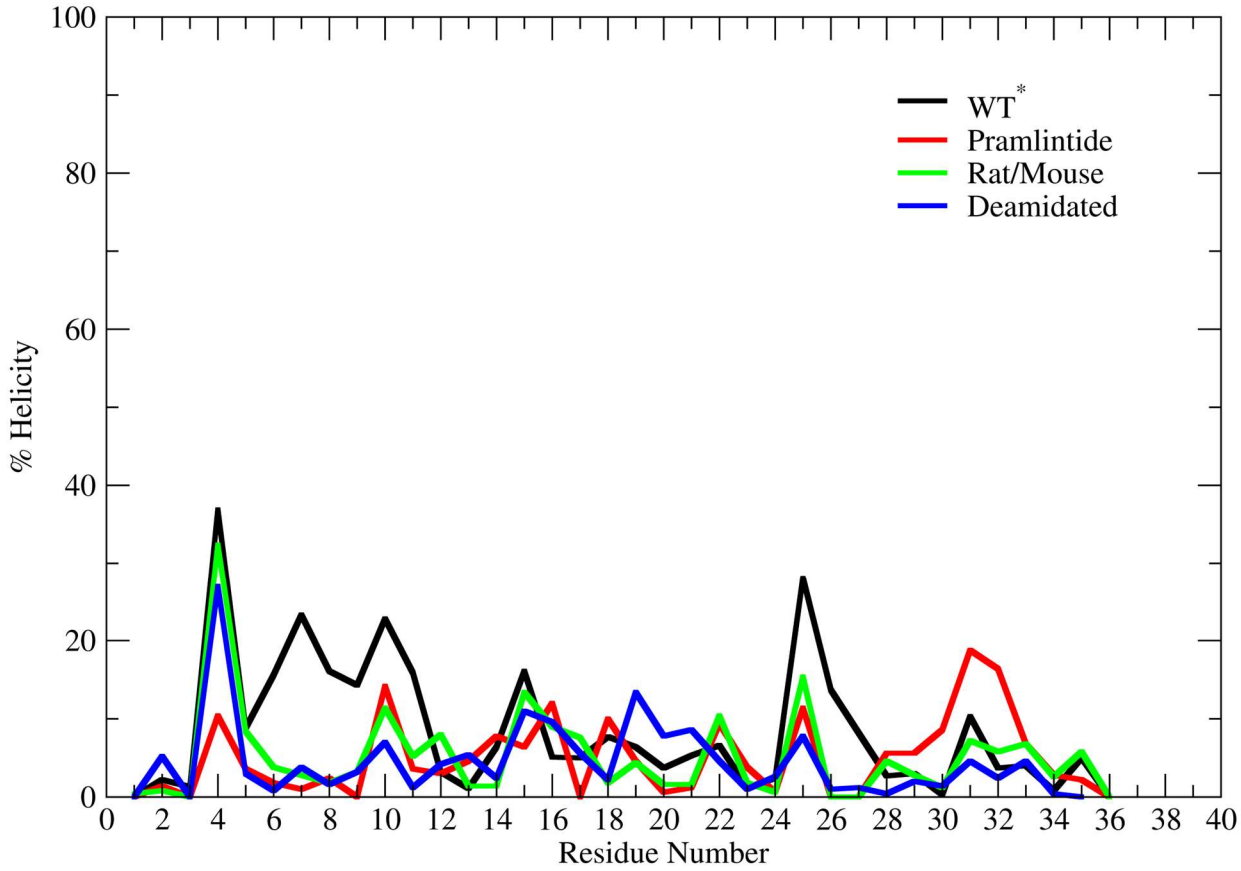


Figure 6.2 % Helicity of unfolded 2L86 mutation structure in aqueous solution. Simulation duration is 500ns. * WT simulated for 1μs.

The radius of gyration and end-to-end distributions for each mutant in aqueous solution were analyzed and are provided in Figure 6.3. It was observed that all mutants exhibited a wide range of radius of gyration values, ranging from 1.0 to 2.5 nm, as well as end-to-end distance distributions spanning from 0.50 to 6 nm, both characteristic of IDPs. Notably, however, the deamidated form displayed distinct peaks at smaller radius of gyration values, centered around 1.2 nm (blue line in Figure 6.3). It is possible that the generation of a charged C-terminal residue in this mutant increases the probability of N-terminal contacts with the K1 residue that bears two positive charges. However, despite its compact nature, the deamidated mutant does not exhibit a pronounced helical structure, as indicated by the low percent helicity observed throughout the simulation (see Figure 6.2). Conversely, the Pramlintide and rat/mouse mutations also display wide ranges of radius of gyration and end-to-end distances, indicating their unfolded state without well-defined structures. These findings support the notion that all mutations lack stable secondary structure elements and remain in a disordered conformation.

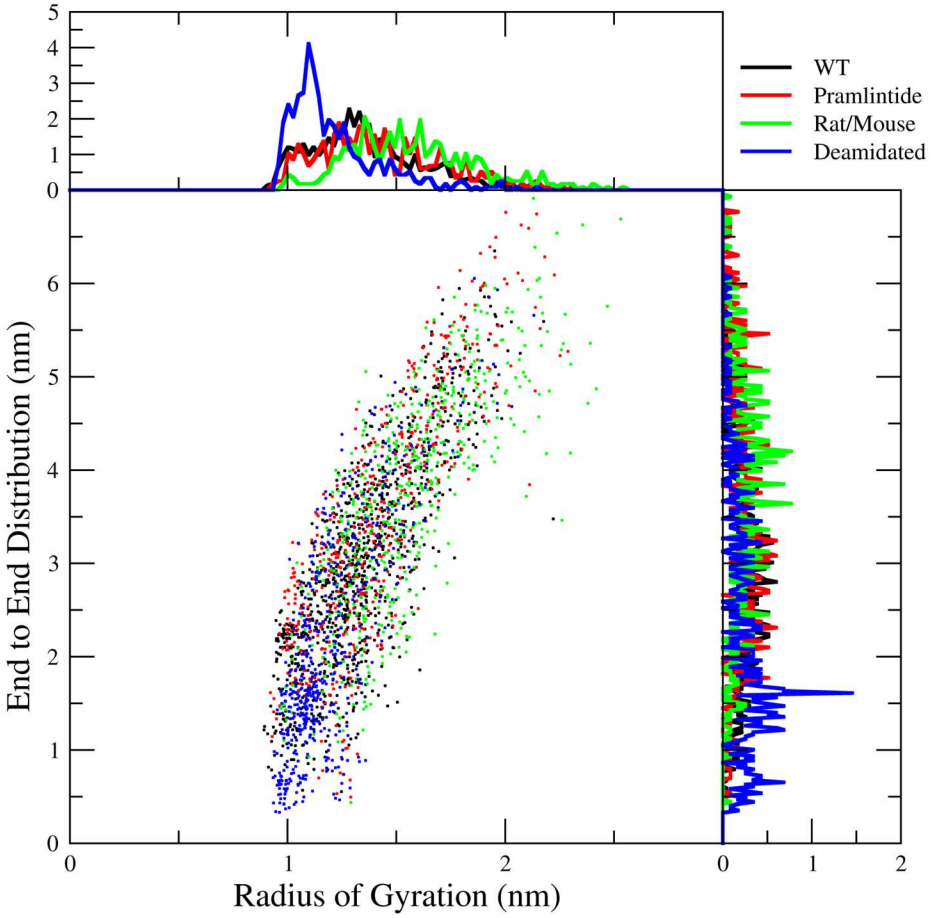


Figure 6.3 Analysis results for the radius of gyration and end-to-end distance were obtained for the 500ns simulation of 3 unfolded mutation in aqueous solution. WT simulated for 1μs.

6.3.2 Amylin mutation at air/water interface

All the mutants were simulated in the presence of a vacuum/water interface. Similar to human amylin, as discussed in Chapter 5, all the mutants exhibited a pronounced preference for localizing at the interface. This behavior is clearly depicted in Figure 6.4, where we observe that after 200 ns of simulation, all mutants diffused to, and remained at, the air/water interface. The corresponding helix propensities are provided in Figure 6.5. At the air/water interface, the Pramlintide sequence (red line in Figure 6.5) exhibited a generally low percentage of helicity, indicating it remained in an unfolded conformation. In contrast, the Rat/Mouse sequence (green

line in Figure 6.5) shows a higher percent helicity for the residues between N5 and F15. Also, the helicity gradually increased after 500 ns simulation (data not shown).

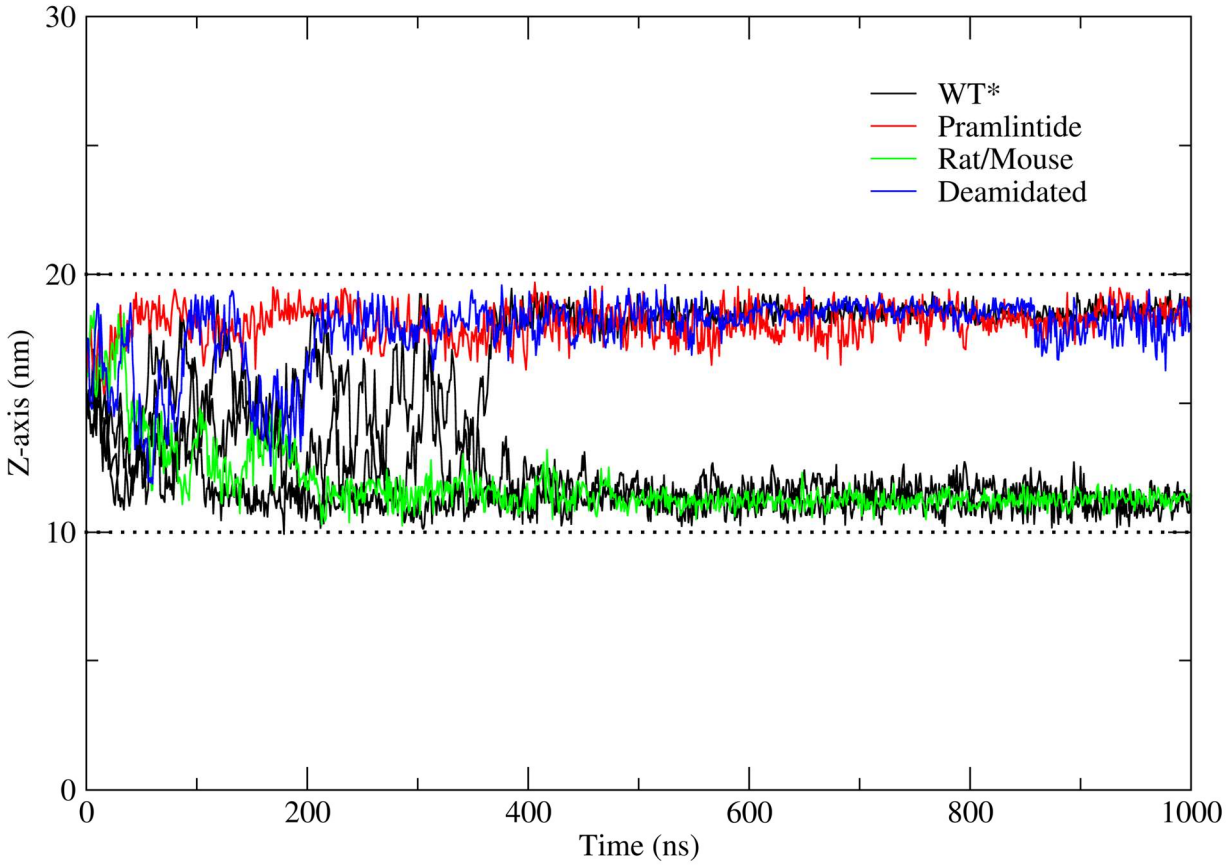


Figure 6.4 Z-axis coordinates of N_{ξ} atom of K1 residue, for each mutation. Black dotted line is approximation of Air/Water interface location at z-axis. 3 different initial configuration of WT was simulated for 2μs. After 400ns all WT stays at the interface though out the simulation as shown in Chapter 5 Figure 5.6.

Notably, the Deamidated Amylin sequence demonstrated helical conformations specifically within the amyloidogenic region as depicted by the blue line in Figure 6.5. In Figure 6.5, the data represented in black indicate the average percent helicity of human amylin (WT), derived from three separate simulations, each initiated from a distinct unfolded state. WT simulation exhibited higher percentages of helicity compared to the mutations, particularly within the amyloidogenic region. This observation suggests the relevance of the amyloidogenic region in determining the structure of Amylin at the interface. However, it should be noted that

differentiating between the effects of WT and deamidation solely based on MD simulations can be challenging due to the different simulation periods between mutant and WT simulations.

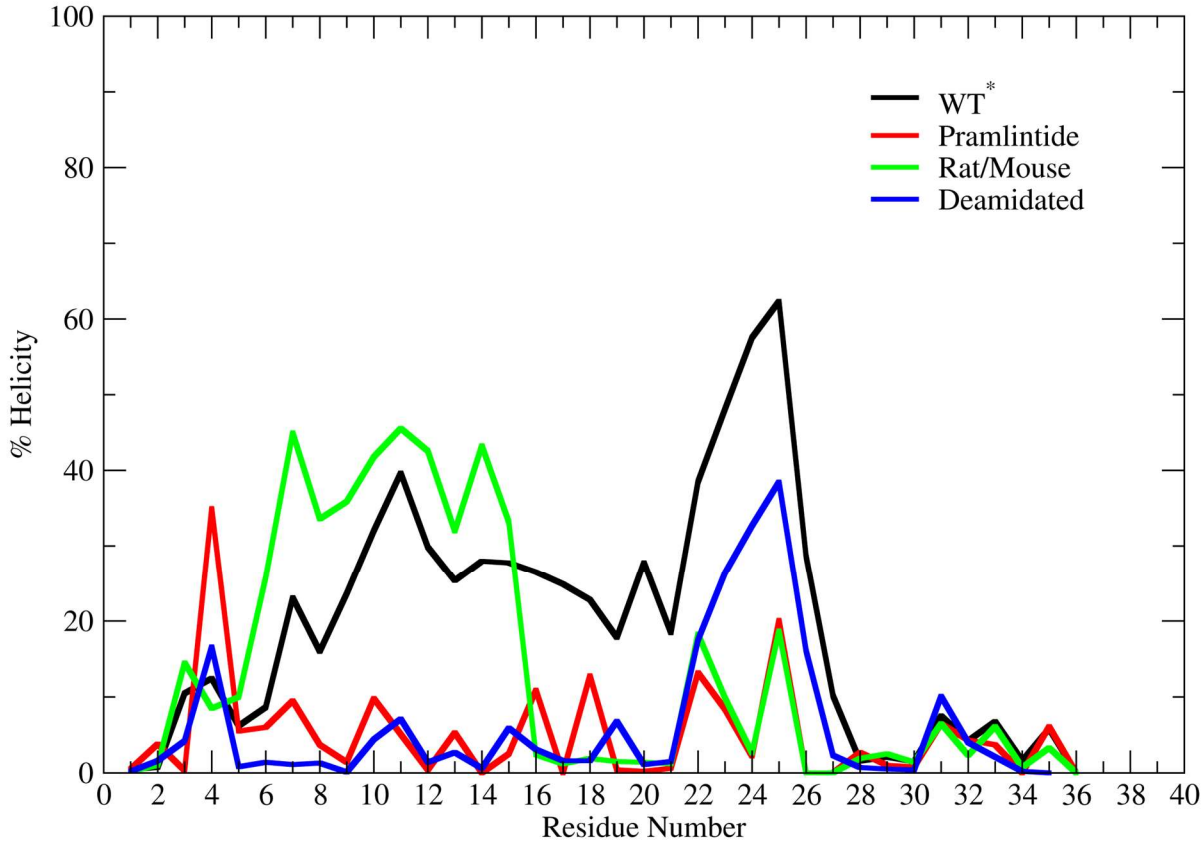


Figure 6.5 % Helicity of unfolded 2L86 mutation structure at Air/Water interface. Simulation duration is 1μs. * WTs simulated for 2μs. 3 WT simulation with different initial configuration are averaged out. Individual percent helicity for WT are shown in Chapter 5 Figure 5.3.

The resulting radius of gyration and end to end distributions are provided in Figure 6.6. Notably, the WT peptide at the Air/Water interface exhibited its highest peak at a shorter radius of gyration distance. Specifically, one of the WT unfolded structures showed a peak at 1 nm, indicating the formation of a long continuous helical structure at the Air/Water interface. The other two WT structures also displayed peaks at shorter distances compared to Pramlintide, with consistent helical structures in the amyloidogenic region. The analysis of end-to-end distance showed similar trends to the radius of gyration results. Although the WT exhibited higher percent

helicity compared to the mutations, it is important to note that the WT simulations were significantly longer than the mutation simulations. Therefore, it is challenging to draw definitive conclusions, but the WT structures appeared to demonstrate a higher propensity for helicity at the Air/Water interface compared to the mutants.

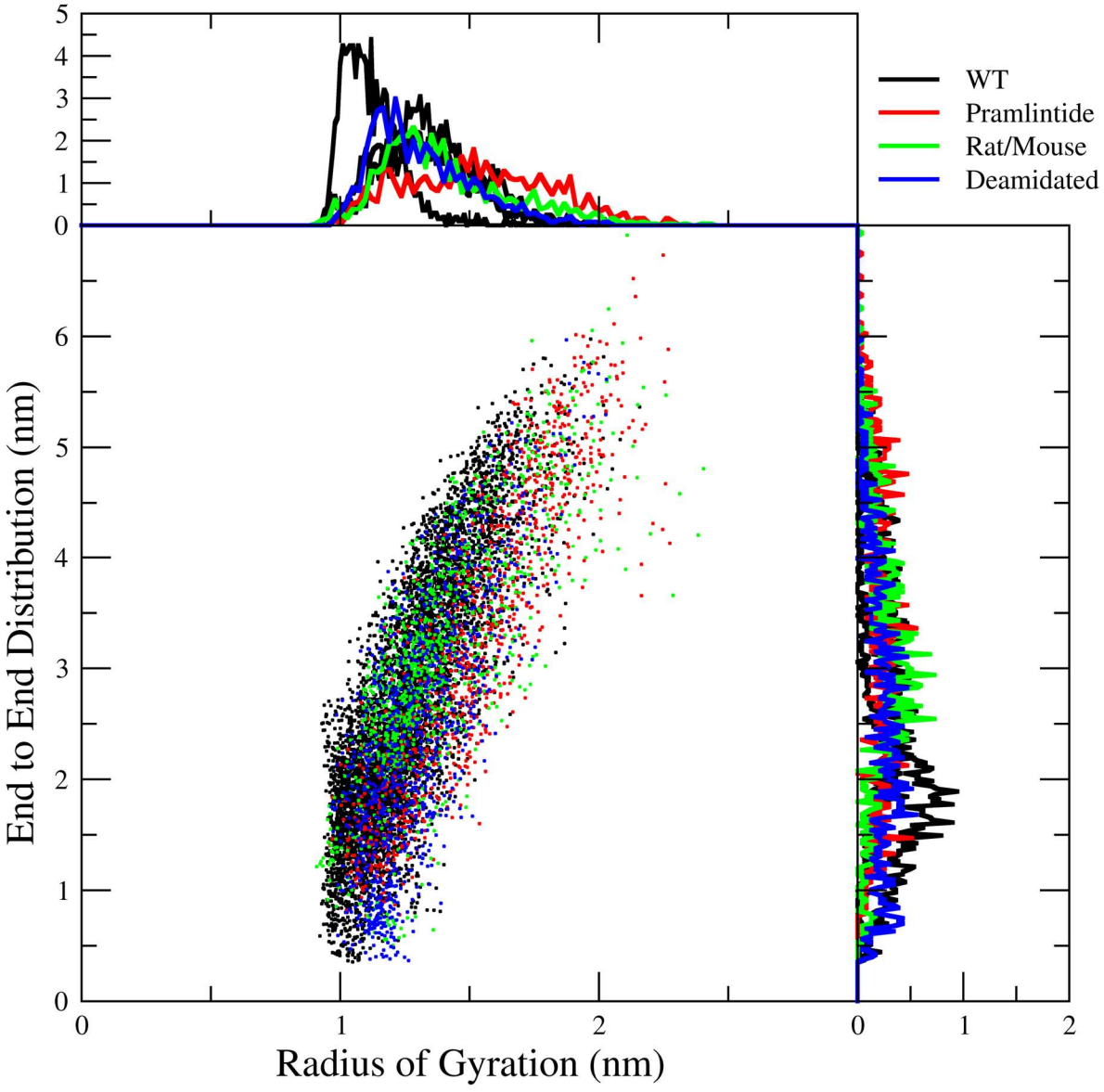


Figure 6.6 Analysis results for the radius of gyration and end-to-end distance were obtained for the 1μs simulation of 3 unfolded mutation at Air/Water interface.

6.3.3 Amylin mutation in the presence of SDS

In the presence of SDS micelles, Pramlintide displayed a significantly higher percentage of helicity, particularly for the unmutated region, as illustrated by the red solid line in Figure 6.7. Conversely, the Rat/Mouse sequence maintained a predominantly unfolded conformation throughout the simulation, with a slight increase in helicity observed around residue N22 after 500 ns. However, overall, the Rat/Mouse sequence remained largely unfolded on the SDS micelle, as depicted by the green line in Figure 6.7. This is probably due to the presence of several non-helix forming proline residues in this mutant. Notably, the Deamidated sequence exhibited a substantial percentage of helicity at the amyloidogenic region in the presence of SDS micelles, as shown by the blue line in Figure 6.7. Additionally, some helicity was observed near residue Q10. These observations provide valuable insights into the helical tendencies of Pramlintide, the Rat/Mouse sequence, and Deamidated Amylin at the air/water interface and in the presence of SDS micelles, highlighting the influence of environmental factors on their structural behavior.

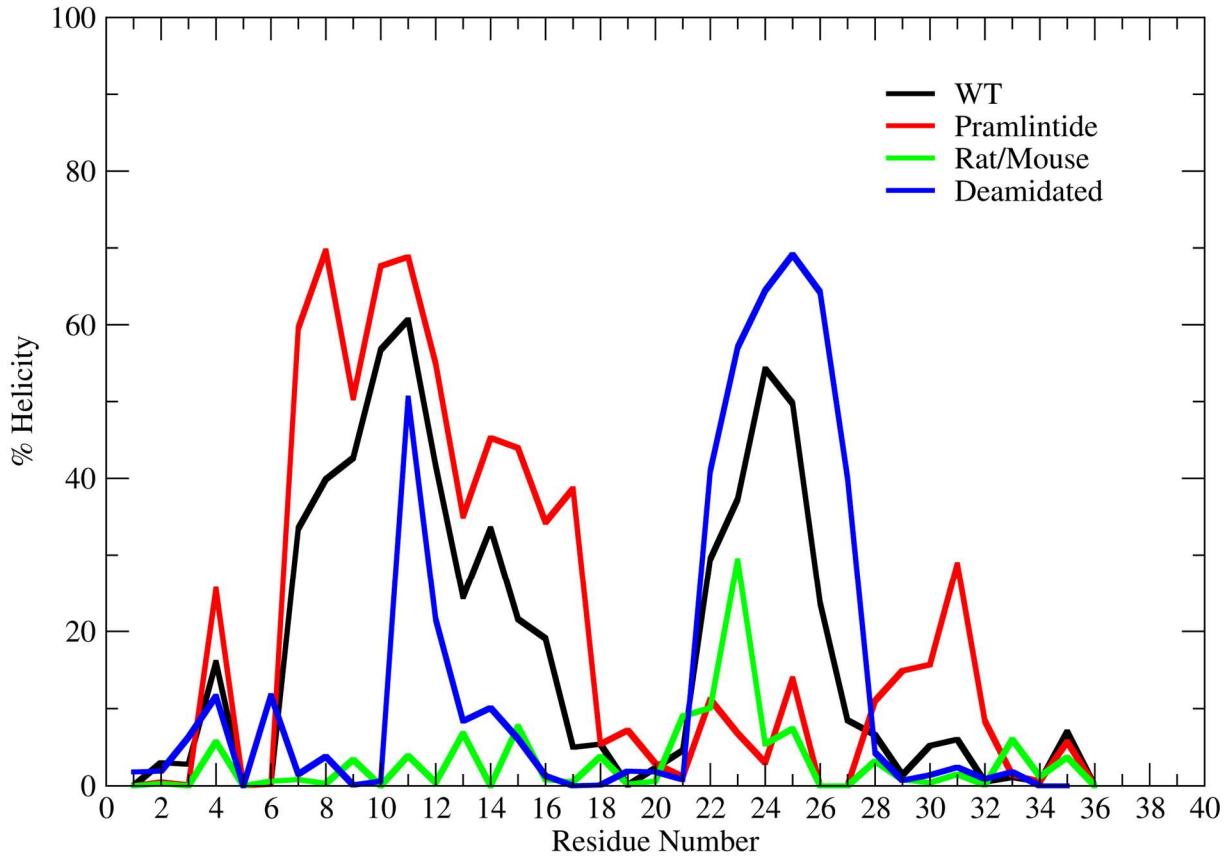


Figure 6.7 Percent helicity of amylin mutations in the presence of SDS micelle. All mutation and WT were simulated for 1μs.

An analysis of the radius of gyration and end-to-end distance for the WT and three mutants in the presence of a SDS micelle is provided in Figure 6.8. The radius of gyration provided a relatively narrow distribution, primarily centered around 1-2 nm. On the other hand, the histogram for the end-to-end distance presents a broader distribution, spanning from 1 to 7 nm. These observations suggest a peptide structure that features a compact or folded region within its central region, while providing greater flexibility at the terminal residues. Pramlintide stands out with its two distinctive peaks at a shorter radius of gyration values, aligning with the higher helicity observed for the unmutated region as displayed in Figure 6.6, where the helicity surpasses even that of the wild type. The condensed radius of gyration histogram indicates a well-structured core

within the peptide, with the majority of conformations falling within the 1-2 nm range. This suggests a compact assembly for the peptide backbone and side chains, alluding

to the presence of structured secondary elements and interactions that contribute to the peptide's overall stability and compactness.

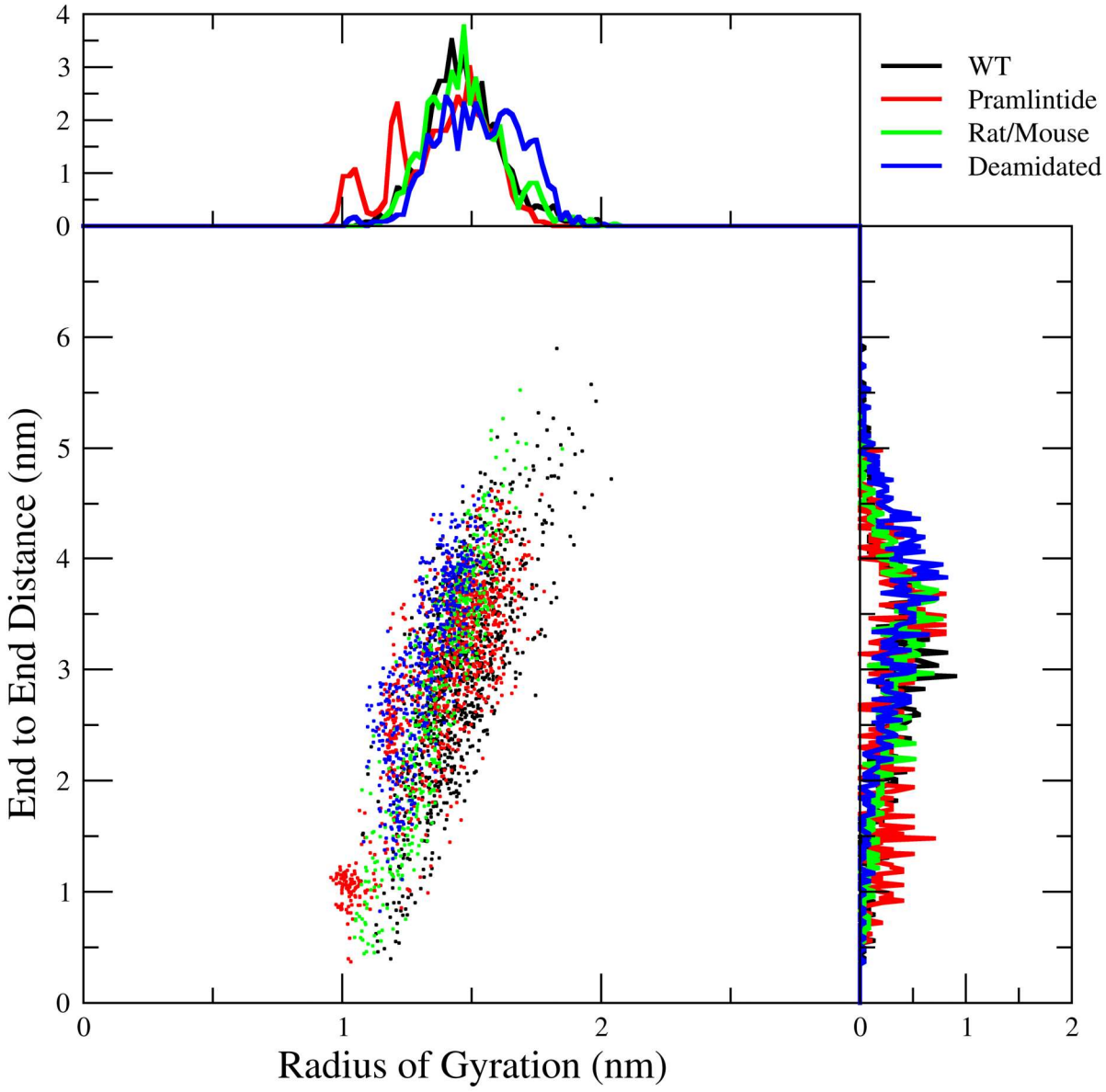


Figure 6.8 Analysis results for the radius of gyration and end-to-end distance were obtained for the 1μs simulation unfolded structures in the presence of SDS micelle.

6.4 Conclusions

In aqueous solution, all mutations were observed to remain in an unfolded state. However, in the presence of a SDS micelle, each mutation exhibited helical conformations at different residues. It is worth noting that none of the mutations formed a helix-kink-helix structure similar to that of human amylin. At the air/water interface, certain regions of the peptides exhibited partial helical formation, although the overall percentage of helicity was lower compared to the helical content observed in the presence of SDS. Nevertheless, it is evident that all peptides displayed a clear preference for localizing at the interface, suggesting an affinity for the interfacial region.

6.5 Reference

(1) Tracz, S. M.; Abedini, A.; Driscoll, M.; Raleigh, D. P. Role of aromatic interactions in amyloid formation by peptides derived from human amylin. *Biochemistry* **2004**, *43* (50), 15901-15908. DOI: 10.1021/bi048812l.

(2) Zanuy, D.; Ma, B. Y.; Nussinov, R. Short peptide amyloid organization: Stabilities and conformations of the islet amyloid peptide NFGAIL. *Biophysical Journal* **2003**, *84* (3), 1884-1894. DOI: 10.1016/s0006-3495(03)74996-0.

(3) Miller, Y.; Ma, B.; Nussinov, R. Polymorphism in Alzheimer A beta Amyloid Organization Reflects Conformational Selection in a Rugged Energy Landscape. *Chemical Reviews* **2010**, *110* (8), 4820-4838. DOI: 10.1021/cr900377t.

(4) Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (2), 315-320. DOI: 10.1073/pnas.0406847102.

(5) Sofinska, K.; Lupa, D.; Chachaj-Brekiesz, A.; Czaja, M.; Kobierski, J.; Seweryn, S.; Skirlinska-Nosek, K.; Szymonski, M.; Wilkosz, N.; Wnetrzak, A.; et al. Revealing local molecular distribution, orientation, phase separation, and formation of domains in artificial lipid layers: Towards comprehensive characterization of biological membranes. *Advances in Colloid and Interface Science* **2022**, *301*. DOI: 10.1016/j.cis.2022.102614.

(6) Hamodrakas, S. J. Protein aggregation and amyloid fibril formation prediction software from primary sequence: towards controlling the formation of bacterial inclusion bodies. *Febs Journal* **2011**, *278* (14), 2428-2435. DOI: 10.1111/j.1742-4658.2011.08164.x.

(7) Bemporad, F.; Taddei, N.; Stefani, M.; Chiti, F. Assessing the role of aromatic residues in the amyloid aggregation of human muscle acylphosphatase. *Protein Science* **2006**, *15* (4), 862-870. DOI: 10.1110/ps.051915806.

(8) Westermark, P.; Engstrom, U.; Johnson, K. H.; Westermark, G. T.; Betsholtz, C. Islet Amyloid Polypeptide - Pinpointing Amino Acid Residues Linked to Amyloid Fibril Formation. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87* (13), 5036-5040. DOI: 10.1073/pnas.87.13.5036.

(9) Kapurniotu, A.; Schmauder, A.; Tenidis, K. Structure-based design and study of non-amyloidogenic, double N-methylated IAPP amyloid core sequences as inhibitors of IAPP amyloid formation and cytotoxicity. *Journal of Molecular Biology* **2002**, *315* (3), 339-350. DOI: 10.1006/jmbi.2001.5244.

(10) Fu, L.; Wang, Z. G.; Batista, V. S.; Yan, E. C. Y. New Insights from Sum Frequency Generation Vibrational Spectroscopy into the Interactions of Islet Amyloid Polypeptides with Lipid Membranes. *Journal of Diabetes Research* **2016**, *2016*. DOI: 10.1155/2016/7293063.

(11) Asthana, S.; Mallick, B.; Alexandrescu, A. T.; Jha, S. IAPP in type II diabetes: Basic research on structure, molecular interactions, and disease mechanisms suggests potential intervention strategies. *Biochimica Et Biophysica Acta-Biomembranes* **2018**, *1860* (9), 1765-1782. DOI: 10.1016/j.bbamem.2018.02.020.

(12) Pappalardo, G.; Milardi, D.; Magri, A.; Attanasio, F.; Impellizzeri, G.; La Rosa, C.; Grasso, D.; Rizzarelli, E. Environmental factors differently affect human and rat IAPP: Conformational preferences and membrane interactions of IAPP17-29 peptide derivatives. *Chemistry-a European Journal* **2007**, *13* (36), 10204-10215. DOI: 10.1002/chem.200700576.

(13) Nanga, R. P. R.; Brender, J. R.; Xu, J. D.; Hartman, K.; Subramanian, V.; Ramamoorthy, A. Three-Dimensional Structure and Orientation of Rat Islet Amyloid Polypeptide Protein in a Membrane Environment by Solution NMR Spectroscopy. *Journal of the American Chemical Society* **2009**, *131* (23), 8252-8261. DOI: 10.1021/ja9010095.

(14) Nanga, R. P. R.; Brender, J. R.; Xu, J. D.; Veglia, G.; Ramamoorthy, A. Structures of Rat and Human Islet Amyloid Polypeptide IAPP(1-19) in Micelles by NMR Spectroscopy. *Biochemistry* **2008**, *47* (48), 12689-12697. DOI: 10.1021/bi8014357.

(15) Nanga, R. P. R.; Brender, J. R.; Vivekanandan, S.; Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochimica Et Biophysica Acta-Biomembranes* **2011**, *1808* (10), 2337-2342. DOI: 10.1016/j.bbamem.2011.06.012.

(16) Duan, M. J.; Fan, J.; Huo, S. H. Conformations of Islet Amyloid Polypeptide Monomers in a Membrane Environment: Implications for Fibril Formation. *Plos One* **2012**, *7* (11). DOI: 10.1371/journal.pone.0047150.

(17) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

(18) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - A New Molecular-Dynamics Method. **1981**, *52* (12), 7182-7190. DOI: 10.1063/1.328693.

(19) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463-1472. DOI: 10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h.

(20) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large System. *Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397.

Chapter 7 – Summary and Future Directions

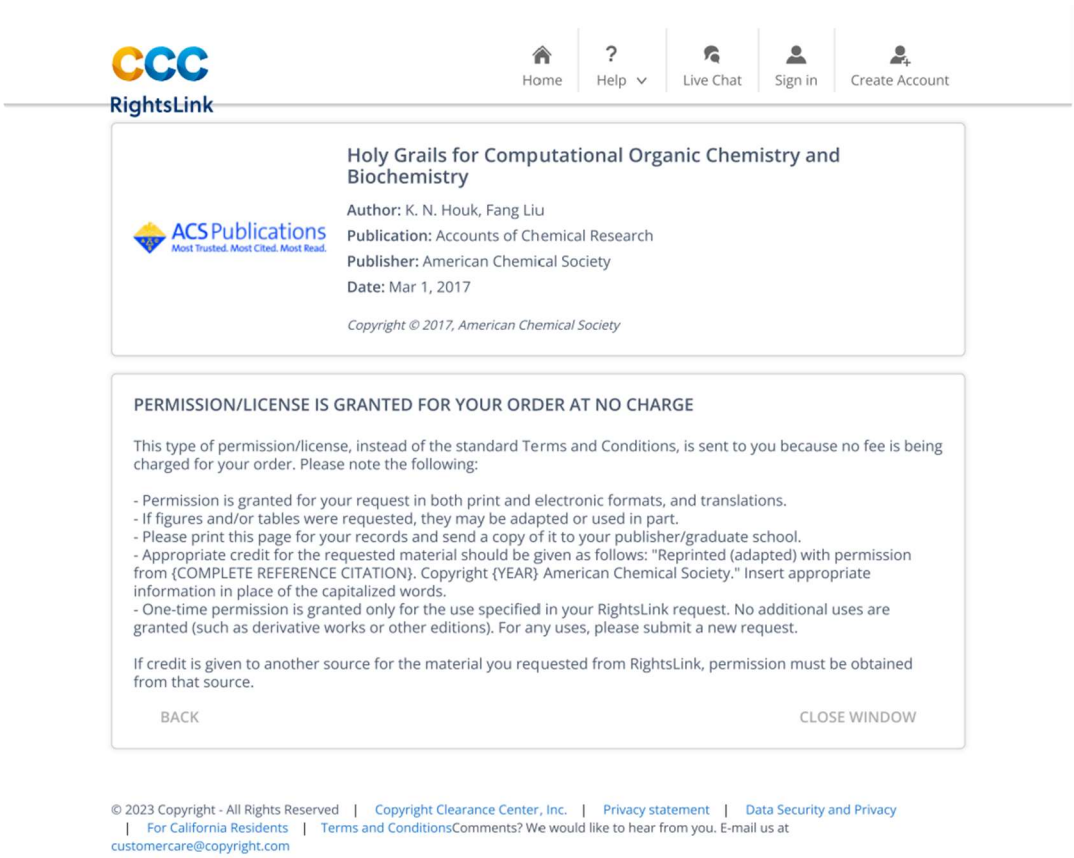
In Chapter 2 & 3 we presented new force fields for performing classical molecular dynamics simulations of zwitterionic amino acids, sulfates, alkylsulfates, and SDS. These force fields were developed using the KBFF approach, where simulated Kirkwood-Buff integrals (KBIs) are compared to those extracted from experimental thermodynamic data. The KBIs provide a way to quantify solute-solute, solute-solvent and solvent-solvent affinities in the condensed phase. Good agreement between simulated and experimental KBIs ensures a good balance between these different affinities. Hence, the problem of under solvation that appears in many traditional biomolecular force fields can be mitigated. The force fields developed here can be used by a variety of researchers to investigate systems containing similar types of molecules and ions.

In Chapter 4 we further investigated the behavior of the new force field for SDS as there is little experimental data for SDS itself available for comparison. Several small peptides that display secondary structure in the presence of SDS, but are structureless in pure water, were chosen for study. Simulations of these peptides, treated with our established KBFF20 force field for peptides and proteins, in the presence of SDS, treated with the new force field developed here, were performed. The results confirmed the experimentally observed disordered behavior in aqueous solution, with secondary structure formation in the presence of SDS. These observations, and the reproducibility of the results, strongly suggest that the current KBFF20 and new SDS force fields provide accurate descriptions of these types of systems. This, in turn, provides confidence in future results for similar systems.

In Chapters 5 & 6 we performed simulations of wild type and mutant Amylin in the presence of several different environments. These simulations shed light on the behavior of this important peptide. Simulations of Amylin in aqueous solution provided a disordered conformation

in agreement with experiment. In the presence of SDS Amylin forms a folded helix-kink-helix motif very similar to that provided by NMR studies in the presence of SDS. Simulations of Amylin in the presence of an air/water or lipid/water interface indicated that Amylin always prefers to be at the interface. At both interfaces there is significant helical structure, especially for the NNFGAIL region of the peptide. This behavior could be important for understanding the subsequent aggregation behavior of Amylin in biological environments, which is an active area of research.

Appendix – Copyright Clearance



CCC RightsLink

Home | Help | Live Chat | Sign in | Create Account

ACS Publications
Most Trusted. Most Cited. Most Read.

Holy Grails for Computational Organic Chemistry and Biochemistry

Author: K. N. Houk, Fang Liu
Publication: Accounts of Chemical Research
Publisher: American Chemical Society
Date: Mar 1, 2017

Copyright © 2017, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

Prospects and challenges for computer simulations of monolayer-protected metal clusters

Author: Sami Malola et al
Publication: Nature Communications
Publisher: Springer Nature
Date: Apr 13, 2021

Kirkwood–Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20

Author: Elizabeth A. Ploetz, Sadish Karunaweera, Nikolaos Bentenitis, et al
Publication: Journal of Chemical Theory and Computation
Publisher: American Chemical Society
Date: May 1, 2021

Study of surface properties of aqueous solutions of sodium dodecyl sulfate in the presence of hydrochloric acid and heavy metal ions

Author: Anna Wołowicz,Katarzyna Staszak

Publication: Journal of Molecular Liquids

Publisher: Elsevier

Date: 1 February 2020

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Jul 12, 2023

This Agreement between Kansas State University -- Kyung-Shin Suh ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5586610926152
License date	Jul 12, 2023
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Journal of Raman Spectroscopy
Licensed Content Title	Raman study of cation effect on sulfate vibration modes in solid state and in aqueous solutions
Licensed Content Author	Marc D. Fontana, Hassen Aroui, Thomas H. Kauffmann, et al
Licensed Content Date	Sep 11, 2013
Licensed Content Volume	44
Licensed Content Issue	11
Licensed Content Pages	6
Type of use	Dissertation/Thesis
Requestor type	University/Academic

Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	Graduate Student
Institution name	Kansas State University
Expected presentation date	Jul 2023
Portions	Figure 5
Requestor Location	Kansas State University 1600 Hillcrest Dr. Apt06 MANHATTAN, KS 66502 United States Attn: Kansas State University
Publisher Tax ID	EU826007151
Total	0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR

A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives,

and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jul 12, 2023

This Agreement between Kansas State University -- Kyung-Shin Suh ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5586611376263
License date	Jul 12, 2023
Licensed Content Publisher	Elsevier
Licensed Content Publication	Biochemical and Biophysical Research Communications
Licensed Content Title	Recombinant expression, synthesis, purification, and solution structure of arenicin
Licensed Content Author	Tatiana V. Ovchinnikova,Zakhar O. Shenkarev,Kirill D. Nadezhdin,Sergey V. Balandin,Maxim N. Zhmak,Irina A. Kudelina,Ekaterina I. Finkina,Vladimir N. Kokryakov,Alexander S. Arseniev
Licensed Content Date	Aug 17, 2007
Licensed Content Volume	360
Licensed Content Issue	1
Licensed Content Pages	7
Start Page	156

Print This Page