12-2021

FAST & EFFICIENT SIMULATION FOR LONG POLYMERIC CHAINS

Naeyma N. Islam
nnislam@uno.edu

Fast & Efficient Simulation For Long Polymeric Chains

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Chemistry

by

Naeyma Nahida Islam

M.S. Middle Tennessee State University, 2016

December, 2021

Abstract

The assembly of large polymeric systems in water involves a range of length scales, from the angstrom scale of a water molecule to the tens of nanometer scale of the solute, a range of interactions from strong waterwater interactions to the weaker interactions between different groups, and a range of time scales. These effects combine to cause a challenge for molecular simulations. Polypeptoids are a type of peptidomimetic polymers that are highly tunable, and hence an ideal model system to analysis self-assembly as a function of chemical groups in aqueous soft matter systems. The focus of our work to develop a computer model that can simulate the large scale of macromolecules, like polypeptoids, faster and efficiently. For this purpose, we have developed the model for the different side chain groups, hydrophobic, hydrophilic, and charges, and the backbone of polypeptoid separately. These are combined to allow the simulation a varietys of polypeptoids.

We also saw how successfully the model we developed captures the water behavior near the hydrophobic group in terms of thermodynamic properties and order parameters.

Since the coronavirus pandemic hit, several works have been done targeting SARS-CoV-2, which is responsible for coronavirus, to discover potential inhibitors. The helicase protein and main protease are two of the most promising target of SARS-CoV-2. Through a combination of molecular docking, molecular dynamics simulations, we proposed a list of potential inhibitors of both the helicase and main protease.

Keywords: Coarse-Grained; Stillinger-Weber potential; Hydrophobic effect; SARS-CoV-2

Contents

List of Figures

List of Figures

List of Tables

List of Tables

1

Chapter 1

COARSE-GRAINED MODELS FOR CONSTANT pH SIMULATIONS OF CARBOXYLIC ACIDS

1.1 Introduction

Polyacidic pH-responsive polymers display large reversible structural changes provoked by pH and have a wide range of applications [2]. To determine this structure by successfully reproducing right balance of the different interactions and how they shift as pH is changed, the model need to be computationally efficient for large scale simulations of multiple macromolecules. The coarse grained model based on the Stillinger-Weber(SW) potential [3] gains interest due to an elimination of interaction sites on hydrogen atoms and a significant reduction in the range of interaction. Despite the lack of hydrogen sites and Coulombic interaction the SW has been shown successfully model water [4], hydrophobic interaction [5], aqueous salt solution [6] and water soluble polymer [7]. The model uses a combination of two- and three- body interactions to produce nearest-neighbor coordination numbers, geometries, and energy scales.

In constant pH simulation [8], the protonation states are permitted to change in response to pH and structural changes. The simplest and most commonly used CpH method uses a continuum model for water and Monte Carlo moves between different protonation state are used [9]. The SW model are well suited for the CpH simulation

because the lack of hydrogen sites makes the water response quicker to the change of protonation state since it involves just translational motion, no orientational motion.

In this work we have developed SW models for protonated and deprotonated carboxylic acid, which we apply to poly(methacrylic acid) (PMAA) as an example. In the PMAA studies, we demonstrate how the protonation states, tacticities, and counterions influence the structure of the polymer. We also developed a CpH model in which transitions between protonated and deprotonated states are made through discrete jumps, with one intermediate state using the Metropolis algorithm. In CpH studies, we further explore the effects of local polymer structure and counterions in the pH response.

1.2 Method

1.2.1 Potential Model

The StillingerWeber potential [3] uses two- and three-body potentials, as given by

$$E = \sum_i \sum_{j>i} \phi_2(r_{ij}) + \sum_i \sum_{j\neq i} \sum_{k>j} \phi_2(r_{ij} r_{jk} \theta_{ijk}) \tag{1.1}$$

where r_{ij} is the distance between atoms i and j and θ is the angle between atom i and its neighbors j and k.

The two-body term has this form:

$$\phi_2(r_{ij}) = A \in_{ij} [B(\sigma_{ij}/r_{ij})^4 - 1] \exp(\sigma_{ij}/(r_{ij} - a_{ij}\sigma_{ij})) \tag{1.2}$$

and the three-body term has this form:

$$\phi_3(r_{ij}, r_{jk}, \theta_{ijk}) = \lambda_{ijk} \in_{ijk} [\cos\theta_{ijk} - \cos\theta_0]^2 \exp(\frac{\gamma\sigma_{ij}}{r_{ij} - a_{ij}\sigma_{ij}}) \exp(\gamma\sigma_{ik}/(r_{ik} - a_{ik}\sigma_{ik})) \tag{1.3}$$

The three-body term is necessary for particles with highly directional interactions, such as hydrogen bonds. There are no Coulombic interactions for the model, but for atoms with formal charges (sodium ions and the oxygens on the carboxylate group) a Yukawa potential is used between pairs with the same charge, to treat charge repulsion. The Yukawa potential has the form [10]

$$\phi_Y(r_{ij}) = A_Y \frac{e^{-k r_{ij}}}{r_{ij}} r_{ij} < r_c \tag{1.4}$$

The parameters A, B, a_{ij}, γ, k, and r_c are taken from previous studies(A=7.049556277, B=0.6022245583, a_{ij} =1.80, γ= 1.20, k = 1.80 $^{-1}$, $and r_c$ = 7.0 [11] The well depth (ϵ) and length scale (σ) parameters for ϕ_2, the angle (θ_0) and energy scale (ϵ_{ijk} and

λ_{ijk}) parameters for ϕ_3, and the energy scale (A_Y) parameter for ϕ_Y are taken to be adjustable. The bond stretch ($E = k_b(r - r_0)^2$) and bond angle ($E = k_\theta(\theta - \theta_0)^2$) terms were taken from OPLS-UA values [12]

1.2.2 Parameter Optimization

Parameters for water [4], the carbon atom types CHn and CT, [5] and the sodium ion [6] are taken from previous studies. Nonbonded parameters for CT were taken to be equivalent to CH2. Here, the atom types in the protonated and deprotonated carboxylic groups of the PMAA are CO, COA, OH, O, and OA, where OH refers to the protonated oxygen, O to the oxygen with double bond, OA to the carboxylate oxygen, CO to the carbon of the protonated acid, and COA to the carbon atom of the deprotonated acid, as shown in Figure 1. To reproduce the radial distribution function between water and solute by SW potential TIP4P/2005 [12] and OPLS-AA12 are used for all atom water and solute respectively. The free energy of hydration for acetic acid was also used to parametrize the model, against the experimental value [14, 15]. The interaction between OH-O, OH-OH, and OH-OA are taken to be slightly stronger than OH-water interaction because of stronger hydrogen bonds between carboxylic group stabilize the self-assembly of PMAA at low pH suggested by experimental [16, 17]and other computational studies [18, 19]. Interactions between OA atoms are taken to be purely repulsive, using the Yukawa potential.

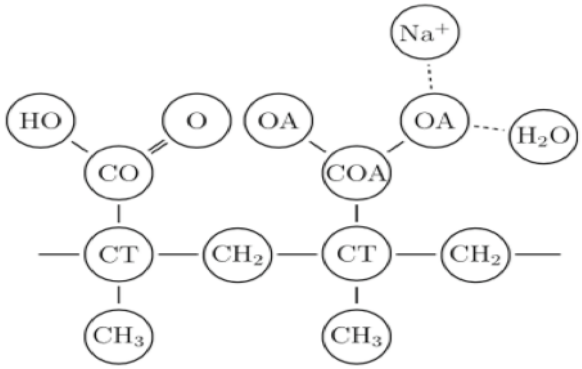

Figure 1.1: Two units of the poly(methacrylic acid) structure, showing the protonated and deprotonated forms and the coarse-grained groups (circles)

1.2.3 Constant pH simulations

For each carboxylic acid group, there are two states: a protonated state with potential energy E_0, and a deprotonated state with energy E_1. With the λ- dynamics method, [20] for continuous values of λ, the potential energy function is

$$E(\lambda) = (1 - \lambda)E_0 \; + \lambda E_1 \; - \lambda \Delta G_m \; + \lambda k_B T ln\,(10)\,(pH - pK_{a,m}) \qquad (1.5)$$

where ΔG_m is the precalculated free energy difference between the protonated and deprotonated states for the isolated monomer, k_B is Boltzmann's constant, and pKa, m is the experimental pKa of the isolated monomer. λ represent the intermediate which is allowed to have discrete values instead of being a continuous variable. In this work, it is found that the method works with only one intermediate state. Intermediate states are defined with interaction parameters interpolated between the protonated and deprotonated parameters:

$$\alpha_\lambda = (1-\lambda)\alpha_0 + \lambda\alpha_1 \tag{1.6}$$

where α represents a potential parameter, which includes ϵ, σ, and the three-body parameters.

The simulation jumps to one state to another state if there is lower energy deference given by

$$E = \{E_0 E_\lambda - \Delta G_\lambda + \lambda k_B T \ln(10)(pH - pK_{am})E_1 - \Delta G_1 + k_B T \ln(10)(pH - pK_{am})\} \tag{1.7}$$

where ΔG_λ is the free energy of state λ relative to state 0 and ΔG_1 is the free energy of state 1 relative to state 0. (E_0, E_λ, and E_1 refer to potential energy functions, with different force field parameters.) The free energies, ΔG_λ and ΔG_1, for the monomer in water are found from standard free energy calculations. These were done using free energy perturbation theory with 11 intermediate states, each simulated for 1 ns, using the CG model. The values are $\Delta G_\lambda = 35.15$ kcal/mol and $\Delta G_1 = 41.95$ kcal/mol. The monomer is taken to be trimethylacetic acid, for the PMAA calculations, using the experimental value of 5.03 for $pK_{a,m}$.

1.2.4 Simulation Details

- The CG simulations used LAMMPS, [21] with a 5 fs time step in the isothermalisobaric (TPN) ensemble, with temperature and pressure controlled using a Nose Hoover thermostat, with a 100 fs damping constant, and a barostat, with a damping parameter of 1000 fs, at a temperature of 300 K and a pressure of 1 bar, respectively.

- The all-atom simulations used the Gromacs simulation package [22] using TIP4P/2005 [12] water and OPLS-AA solute models.

- The free energy calculations for the hydration free energy of acetic acid were done using finite difference thermodynamic integration (FDTI), [23] using a soft core potential to avoid singularities.

- The constant pH simulations were carried out using bash shell scripts, which call LAMMPS with the appropriate input files for the specific protonation state

1.3 Result and Discussion

Two body parameters are obtained from the Lorentz Berthelot combining rules:

$$\varepsilon_{AB} = \left(\varepsilon_{AA}\varepsilon_{BB}\right)^{1/2} \tag{1.8}$$

and

$$\sigma_{AB} = {}^{(\sigma_{AA}+\sigma_{BB})}\!/_{2} \tag{1.9}$$

For aqueous acetate and acetic acid (Figure 1.2), the coarse-grained models were parameterized to reproduce pair correlations of all atom model. And the CG model does fairly well at reproducing the water structure near the three different oxygen types, in terms of the positions of the first solvation shell and the peak heights. The pair correlation function between the methyl carbon on (protonated) trimethylacetic acid and water can be used to check how the model captures the hydrophobic interaction (Figure 2D). The free energy of hydration for acetic acid was also for parameterized the model. Model gives this value of 6.6 ± 0.1 kcal/mol whereas the experimental value is 6.69 kcal/mol. [14]. On the length scale of the entire polymer, the structural changes are captured by radius of gyration. The fraction of proximal (fproximal) neighbors was used to see the changes in structure at the local and the nearest neighbor scale. If the carbonyls are on the same side then conformation is known as the "proximal", or known as the "distal" if they are on the opposite sides (at a dihedral angle equal to 120^o). If that distance between the carbonyl carbon atom is less than 3.7 A, then the conformation is proximal; otherwise it is distal. The properties of PMAA with 20 repeat units show that the model undergoes a structural transition induced by the change in the protonation states (Table 1). The radius of gyration, R_G, changes from about 6 for the protonated state to 10 for the deprotonated state. Similar changes have been reported using all-atom simulations. [24] Table 2 compares our results with simulations of PMAA of the same length, but different force fields, tacticities, and counterion concentrations. [24] The R_G for the deprotonated form shows a small dependence on sodium concentration, decreasing as the sodium concentration increases. The local structure changes from mostly proximal for the isotactic form and mostly distal for the syndiotactic form to about mostly distal for both tacticities. The fraction of proximal values for both protonation states and tacticities are similar to those from all-atom studies. [24] Sodium ions are taken to be associated with the polymer if they are less than 4.0 from a carboxylate carbon. The amount of sodium ion in contact with the polymer increases from about 3 at 0.25M to about 5 at 0.5M.

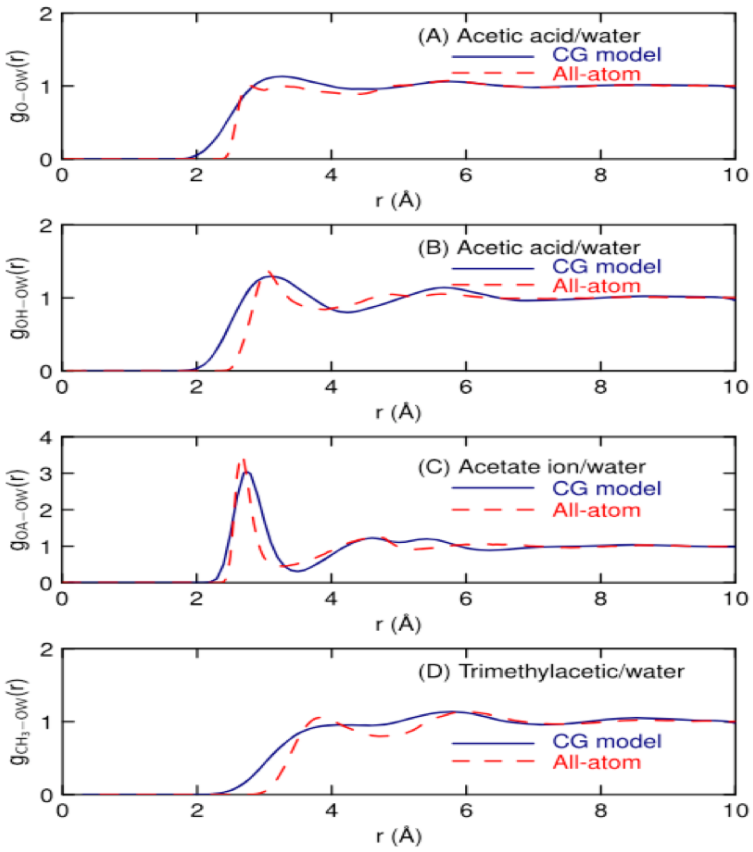

Figure 1.2: Coarse-grained and all-atom pair correlation functions between water oxygen and oxygen atoms on the acetic acid (A and B) and the acetate ion (C), and between water oxygen and the methyl carbon on trimethylacetic acid (D)

Table 1.1: Properties of a 20 Unit Long PMAA for the Protonated and Deprotonated Form at Various Sodium Concentration

	[Na+]	Radius of gyration ()		Fraction proximal		Number of contact Na+	
		Syn	Iso	Syn	Iso	Syn	Iso
Protonated	0 m	5.78± 0.04	5.67± 0.04	0.28± 0.01	0.77± 0.02	N/A	N/A
De-protonated	0 m	10.8± 0.6	9.7± 0.4	0.20± 0.02	0.30± 0.02	N/A	N/A
De-protonated	0.25m	10.7± 0.5	9.7± 0.3	0.19± 0.03	0.35± 0.02	3.0± 0.5	3.2± 0.3
De-protonated	0.50m	10.2± 0.4	8.8± 0.4	0.22± 0.03	0.39± 0.02	5.6± 0.5	4.9± 0.3

Table 1.2: Radius of Gyration for n = 20 PMAA from Different Simulation Studies

tacticity	[Na+] (M)	radius of gyration ()	
		protonated	deprotonated
syndiotactic, this work	0.25	5.80 ± 0.03	10.7 ± 0.5
isotactic, this work	0.25	5.73 ± 0.05	9.7 ± 0.3
syndiotactic [3]	0	7.3 ± 0.1	12.0 ± 0.1
isotactic [4]	0	7.0 ± 0.5	12.4 ± 0.1
atactic [5]	0.2	6.4, 6.7	9.9, 10.0

1.3.1 Constant pH Simulations

Using one intermediate state the method can jump between protonation states

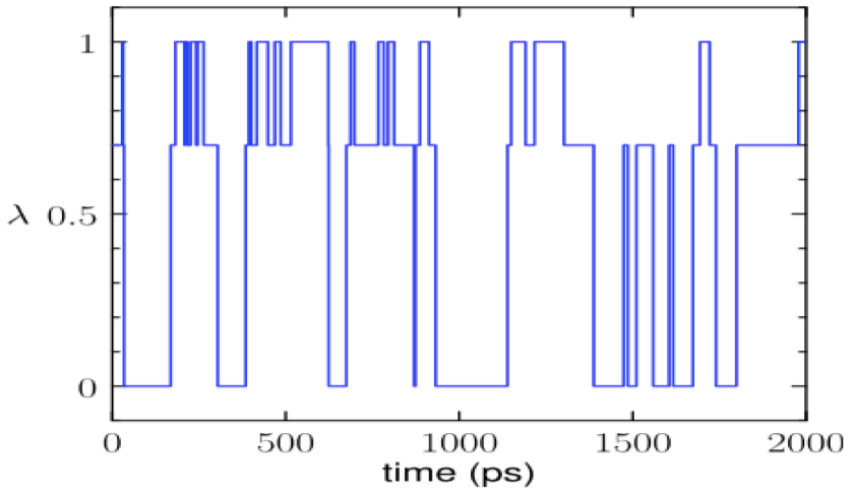

Figure 1.3: Protonation state variable, λ, as a function of time for aqueous trimethylacetic acid.

We found that a value of equal to 0.70 gives even transitions between the two states using trimethylacetic acid as a test case. The titration curves are shown in Figure 1.4. The results are fit to the Hill equation:

$$fractiondeprotonated = 1/\left(10^{n(pK_a - pH)} + 1\right) \tag{1.10}$$

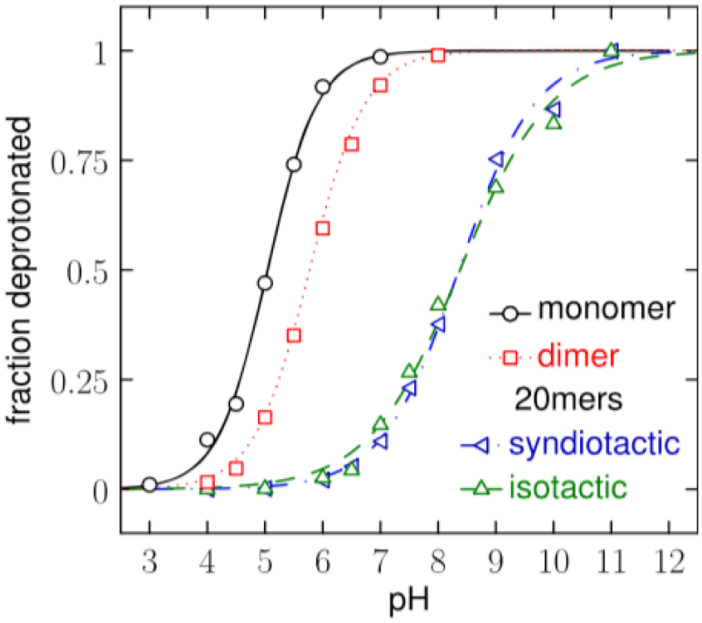

Figure 1.4: Titration curves for oligomers of various lengths.

where pK_a is minus the log of the acid dissociation constant and n gives the degree of cooperativity. The pK_a of the monomer, trimethylacetic acid, is, by construction, close to the experimental value of 5.03.

Sodium ions also affect the pKa. By comparing structures which had a sodium ion near a carboxylate carbon with those that do not, the pKa shows a large decrease when a sodium ion is near. For the 20-mers, this is over 2.5 pH units.

Structural changes for the 20-mer as a function of pH are shown in Figure 6. The radius of gyration shows an abrupt change near the pKa. The fraction of proximal neighbors show a more gradual change. The local change, as given by the fraction proximal, begins at a lower pH than the global change, as given by the radius of gyration.

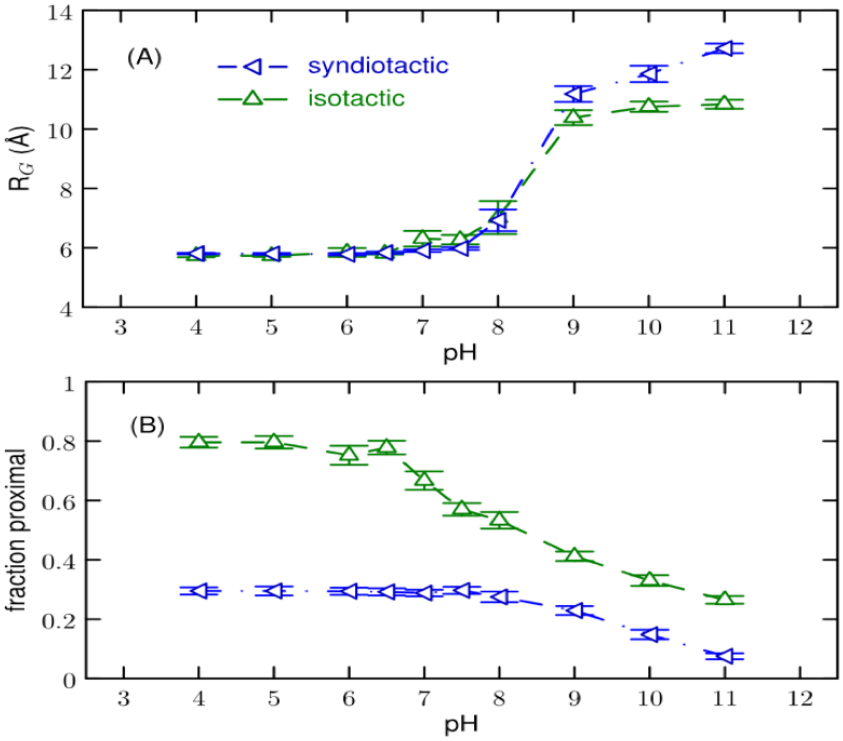

Figure 1.5: Radius of gyration (A) and fraction proximal nearest neighbors (B) as a function of pH for the isotactic and syndiotactic 20-mer.

1.4 Conclusion

The CG model is able to reproduce properties of both the protonated and deprotonated states of carboxylic acid despite the lack of hydrogen interaction sites and the lack of any long-ranged interactions The CpH method enables Monte Carlo moves between the protonated and deprotonated states using one state in between using this model.The models presented here can be combined with similar models for water [4], ions, [6] alkanes, [5] ethers, [7] and peptoids, [25] to allow the simulation a larger range of polymeric materials.

2

Chapter 2

WATER HYDROGEN DEGREES OF FREEDOM AND THE HYDROPHOBIC EFFECTS

2.1 Introduction

Insights into the hydrophobic effect, as exemplified by the addition of small non-polar solutes to water, can be gained by investigating the associated thermodynamic or structural changes. [27, 28, 29, 30] Structural interpretation of hydrophobic hydration relate the large entropy change to ordered water structure surrounding the solute.

The ordering of atoms in water involves the oxygen atom (or translational) degrees-of-freedom or hydrogen atom (or rotational) degrees-of-freedom. The tetrahedral parameter [31] is used to characterize the oxygen order and the hydrogen order is characterized by a hydrogen bond analysis. Different model finds different water behavior in terms of the ordering of water near the hydrophobic solute. However, most of them finds more tetrahedral order structure of water near hydrophobic solute. [32, 33] The translational and orientational order of water structure mainly contribute to the entropy of solvation. Some works say that they have equal contribution to entropy, other says orientational contribution to entropy is small. [34]

The mW model of Molinero and Moore has no hydrogen interaction sites and long-ranged interactions. [35] For all the molecular properties that are considered special for water (hydrogen bonds and large dipole and quadrupole moments), the only one

the mW contains is tetrahedality, and the model has no rotational response. Despite its simplicity, the model can capture most of the properties of liquid water about as accurately as more complex models, as can be captured by some complex models.

In this paper, we examine the predictions of the mW model for the solvation properties of methane and ethane. These properties include the temperature dependence of solvation free energy, from which entropy, enthalpy, and heat capacity can be found, and the structural properties of water near the solute, including the tetrahedral order parameter and the Voronoi volume.

2.2 Method

For simulations, the mW model was used for water [35] and united-atom, short-ranged models were used for methane [36] and ethane. [37] The Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) program was used for simulations, [38] with a 5 fs time step in the isothermal-isobaric (TPN) ensemble, with temperature and pressure controlled using a Nose-Hoover thermostat, with a 100 fs damping con- stant, and a barostat, with a damping parameter of 1000 fs. All simulations were run at a pressure of 1 bar, except certain constant volume simulations.

The entropy at 300 K was found during simulations at two additional temperatures and using

$$S\left(r\right) \; = \; -\left(w_{T_2}\left(r\right) \; - w_{T_1}\left(r\right)\right) / \left(T_2 - T_1\right), with T2 = 320K and T1 = 280K \qquad (2.1)$$

The pure water calculations used 512 molecules, the free energy calculations used 256 water molecules plus the solute.

The value of the thermal pressure coefficient, γ_V , for the mW model is calculated through the fluctuation formula

$$\gamma_V = \left(1/k_B T\right)^2 (\langle PE\rangle V - \langle P\rangle V \langle E\rangle V) \qquad (2.2)$$

where E is the potential energy of the system and $< ... >_V$ indicates a canonical, constant TVN, ensemble average.

The geometries of a water molecule's first solvation shell can be described using the tetrahedral order parameter, Q, of Errington and Debenedetti [39]

$$Q_i = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos\theta_{ijk} + \frac{1}{3}\right)^2 \qquad (2.3)$$

where the sums are over the four nearest neighbors of molecule i and θ_{jjk} is the angle between the three molecules, with molecule i at the center.

A Voronoi tessellation was used to estimate the volume occupied by the water molecules in the simulation. The voronoi method assumes that all atoms are of equal

size. The different molecular sizes can be taken into account to generate alternative Voronoi volumes by assigning each molecule a radius. [40, 41] In this surface-based (S-cell) method, the distance from a point in space to the surface of an atom is used to assign volumes. Based on the radial distribution functions atomic radii were assigned. The Monte Carlo integration was used to calculate both the normal and the S-cell Voronoi volumes by generating 106 random points per configuration and assigning the point to a specific atom.

2.3 Result and Discussion

The density and thermal pressure coefficient of the pure liquid as a function of temperature are shown in Figure 2.1.

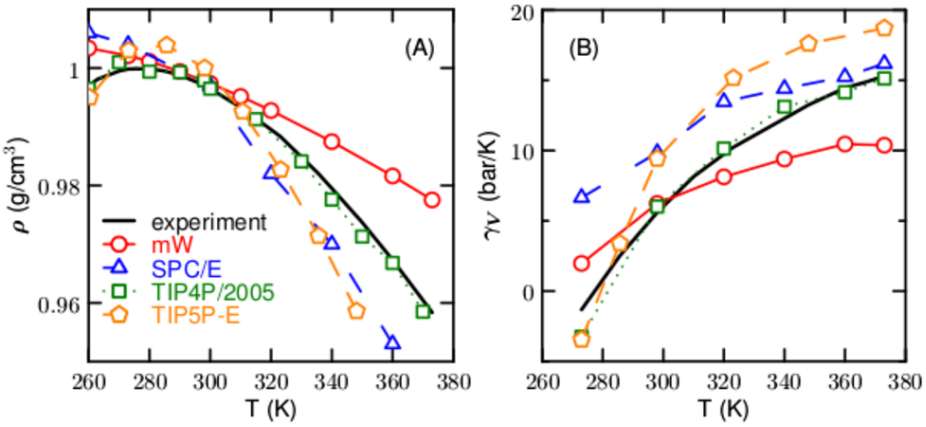

Figure 2.1: Density (A) and thermal pres- sure coefficient (B) from experiment (black solid line) and various models: mW (red solid line and circles), SPC/E (blue dotted-dashed line and triangles), TIP4P/2005 (green dotted line and squares), and TIP5P-E (orange dashed line and pentagons)

SPC/E, [42] TIP4P/2005, [43] and TIP5P-E, [44] which represent optimal 3- site, 4-site, and 5-site non-polarizable models, respectively are chosen for comparison to the mW model. The mW model does fairly well in reproducing the experimental data, [45] with an accuracy that is close to SPC/E, particularly at lower temperature. For temperature dependence of γ_V , the increase from the mW model is similar to SPC/E. At 300 K, the slope of the γ_V versus the T curve is about 0.13 bar/K_2 for the mW model and 0.23 bar/K_2 from the experiment.

The free energies of solvation are shown in Figure 2 . The other models from Ashbaugh et al. [32] and the experimental data [46, 47] are also shown for comparing free energy of solvation to mW model. The values of ΔH, ΔS, and ΔC_P from

mW model, along the values calculated for the other models [7] and the experiment, [46, 47] are given in Table 2.1.

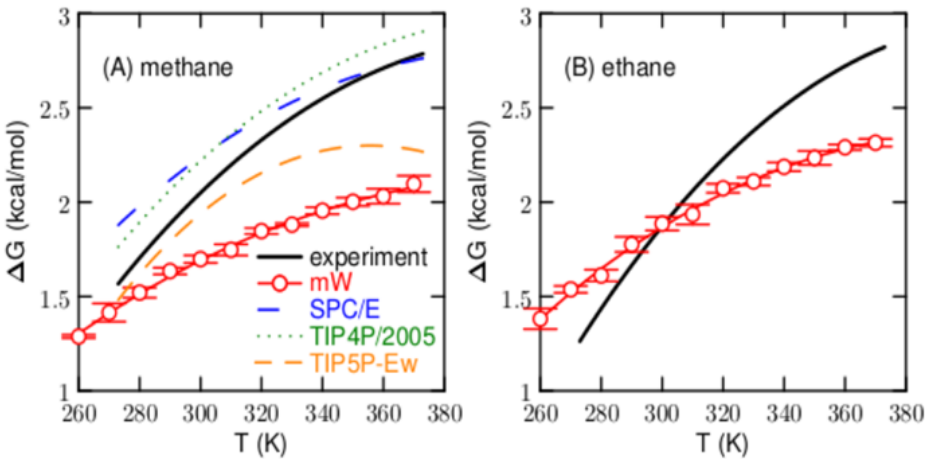

Figure 2.2: Free energies of solvation for methane (A) and ethane (B)

Table 2.1: Solvation thermodynamic properties of methane and ethane at a temperature of 298 K and a pressure of 1 atm

	ΔG	ΔH	ΔS	ΔC_P
	(kcal/mol)	(kcal/mol)	(cal/mol K)	(cal/mol K)
Methane				
Experiment	2.02	-2.72	-15.9	48.8
mW	1.70±0.02	-0.73±0.04	-8.1±0.1	23±3
SPC/E	2.23	-1.3	-11.8	39
TIP4P/2005	2.21	-2.28	-15	47.2
TIP5P-E	1.93	-2.2	-13.8	78.2
Ethane				
Experiment	1.83	-4.17	-20.15	60
mW	1.89 ± 0.03	-1.12 ± 0.09	-10.0 ± 0.2	32 ± 2

For both methane and ethane, the mW model underestimates the magnitude of S and CP by about a factor of two. The enthalpy is underestimated as well, which compensates for the entropy to give a free energy that is close to the experiment.

The structure of the solvating water is characterized by the tetrahedral order parameter, Q, and Voronoi volumes (Figures 2.3 and 2.4)

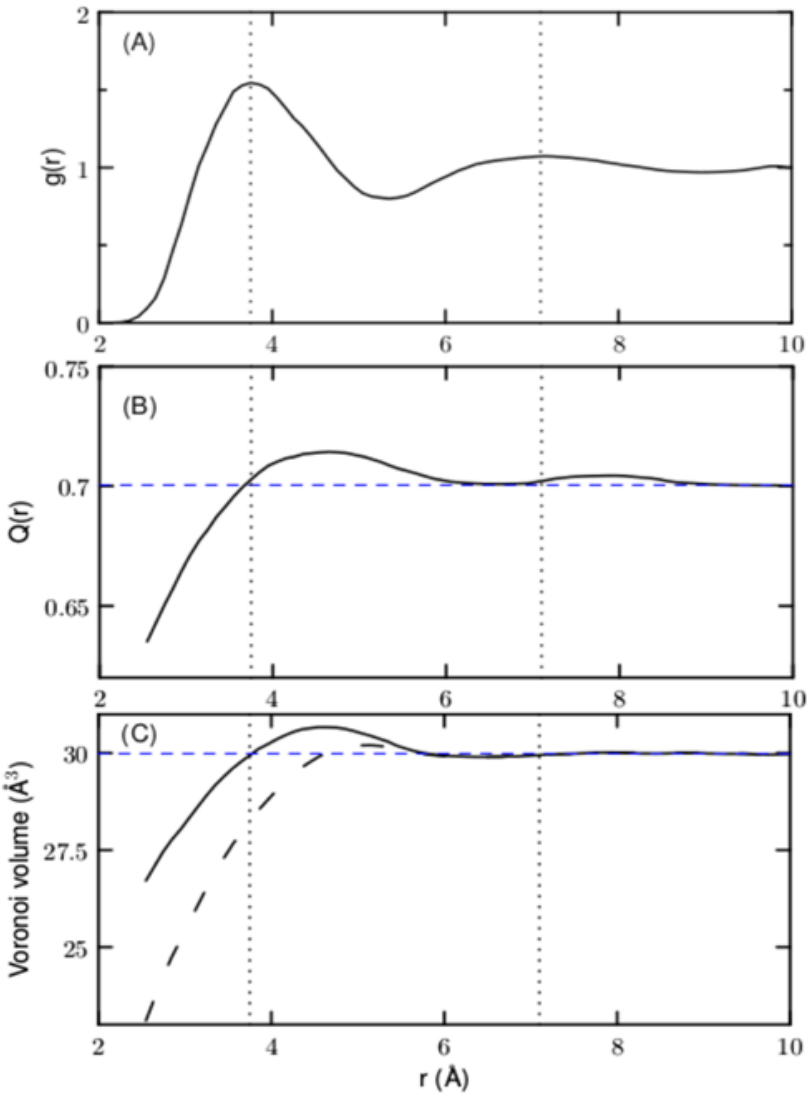

Figure 2.3: Methane-water structure. (A) Radial distribution function between water oxygen and methane carbon. (B) Tetrahedral order parameter for water as a function of distance from methane. (C) Voronoi volume of water as a function of distance from methane, with the solid line showing the ordinary Voronoi cell volume and the dotted-dashed line the S-cell volume. The dotted lines show the positions of the first and second nearest neighbor peaks.

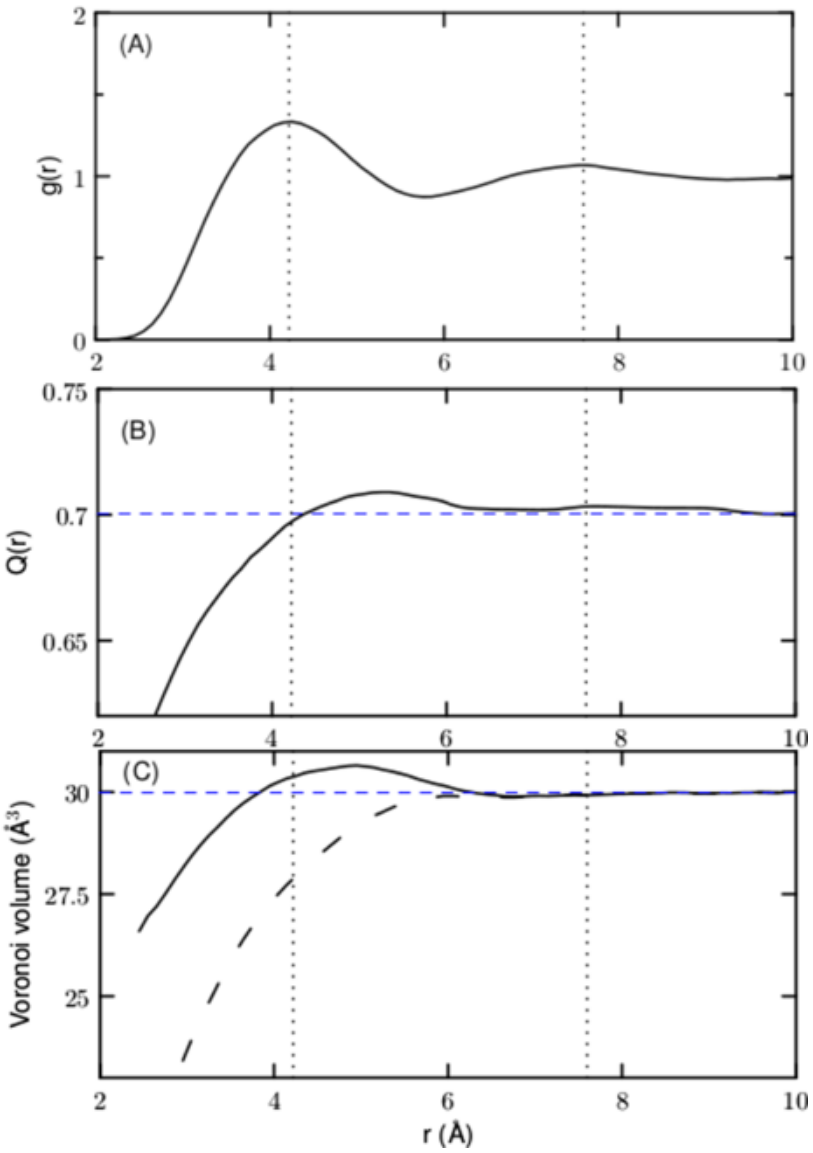

Figure 2.4: Ethane-water structure. (A) Radial distribution function between water oxy- gon and methyl carbon. (B) Tetrahedral order parameter for water as a function of distance from the methyl carbon. (C) Voronoi volume of water as a function of distance from the methyl carbon, with the solid line showing the ordinary Voronoi cell volume and the dot-dashed line the S-cell volume. The dotted lines show the positions of the first and second nearest neighbor peaks.

For both ethane and methane, Q shows a Small increases at the outer region of the first solvation shell, which is very similar to the all-atom simulations of Ashbaugh et al. [32] and the ab initio simulations of Grdadolnik et al. [48] for methane . Overall, the increase of Q is very small. The average Q for a water molecule in the first two solvation shells is 0.702, and in the bulk, Q is 0.701. The Voronoi tessellation results [Figures 3(c) and 4(c)] indicate that the solvating water molecules may occupy a different volume than the bulk. There is an increase in cell volume for Ordinary Voronoi tessellation, where Q increases at the outer part of the first solvation shell, and a decrease at close distances. The S-cell tessellation, where it is assumed that the sizes of atoms are different, results are different from the ordinary tessellation. Overall the addition of the solute decreases the S-cell volumes for both methane and ethane.

Two approaches we took into account to improve the thermodynamic results from mW model capture the half of the entropy solvation. First, from Figure 1(A), the density of water as predicted by the mW model is slightly different from the experiment. Except near 300 K, the density is too large. We ran a set of calculations at a constant volume, rather than at a constant pressure, where the volume was adjusted so that the density corresponds to the experimental density at that temperature. The TVN results show that correcting the density does not improve the free energy results (Figure 2.5)

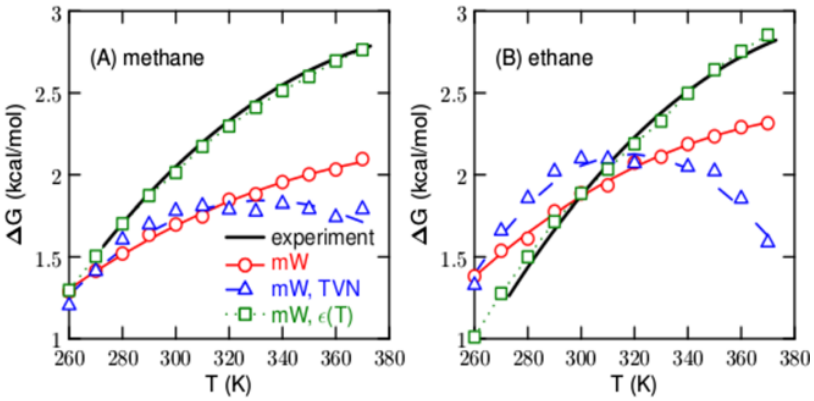

Figure 2.5: Free energies of solvation for methane (A) and ethane (B), compar- ing the mW results (circles) with the TVN simulations using the experimental density (triangles) and the temperature dependent potential (squares)

A second approach is designed to build in the entropy left out of the coarse-grained model by making the potential temperature dependent. A simple way to do this is to make the well-depth parameter, ϵ_{CO} , between the carbon and oxygen atoms dependent on temperature, as $\epsilon_{CO}(T) = \epsilon^0_{CO} + \delta\epsilon_{CO}T$. This was done only for

the solute-solvent interactions. The parameters ϵ^0 and $\delta\epsilon_{CO}$ were found by finding the values of ϵ_{CO} that gave the experimental ΔG of solvation at 300 K and at 340 K. (For ethane, ϵ_{CO} at 300 K, does not need to adjusted.) Fitting to ΔG at 300 and 340 K results in good agreement for the free energy over the whole range of temperatures, indicating that the linear temperature dependence for can capture the missing entropy.

2.4 Conclusion

The mW coarse-grained water models fairly accurate overall as compared to the commonly used water models like SPC/E but the temperature dependences of entropy of solvation of methane and ethane differ from the experiment by roughly a factor of two. There are slightly increased tetrahedral order and slightly decreased Voronoi volume of water near the non-polar solutes .The neglected entropy can be added to the coarse- grained model by treating the methane-water interaction as a potential of mean force with a temperature independent or enthalpic part.

3

Chapter 3

COARSE GRAINED MODEL ON THE AGGREGATION BEHAVIOR OF POLYPEPTOIDS AT VARIOUS AGGREGATION NUMBERS

3.1 Introduction

Amphiphiles consist of at least two differently interacting groups, for example, a hydrophobic and a hydrophilic region. The self-assembly of polypeptoid amphiphiles (PAs) plays a important role to design a very interesting structure, micelle, whereby their hydrophilic part is directed toward the aqueous phase, while their hydrophobic tails point to the center of the micelle, in aqueous solutions at concentrations higher than their critical micelle concentration (CMC). It has drawn lots of attention for scientific research as it can be programmed for biological activity that leads to applications in drug delivery vehicles, [49, 50] immunomodulators, [51] and molecular reporters. [52]

The CMC depends on the relative length of the hydrophobic and hydrophilic blocks. For a given soluble polypeptoid length, an increase of the hydrophobic length leads to a decrease of CMC. [53, 54, 55] Several techniques have been used to determine the aggregation number and the CMC, such as fluorescence quenching, [56, 57] static light scattering, [58, 59, 60, 61] neutron scattering, [62] and a combination of viscometry and dynamic light scattering. [63] However, these different techniques usually do not generate supportive results. Static light scattering calculated a lower

aggregation number than did the dynamic light scattering method, while static fluorescence gave the lowest aggregation number.16 Molecular dynamics (MD) simulation is a powerful tool in evaluating micelle formation. However, the computational cost restricts simulation times and system size to a few hundred nanoseconds and nanometers, respectively. Coarse-grained (CG) models, on the other hand, are computationally cheaper choices wherein atomistic level details are removed while retaining the relevant important physics, thus making it easier to study more complicated systems, including self-assembly in amphiphilic polypeptoid at different aggregation number. [64] By reducing some degrees of freedom and only representing the most important ones, CG modeling can span scales cost-effectively.

Using the SW based coarse grained model for different molecule like water, [65] ions [66] alkanes, [67] ethers, [68] and peptoids, [69] developed by the different members in our group we successfully capture the aggregation behavior of polypeptoid at different aggregation number(N) of 2, 5, 10, 15, 20, 27, leads to forms the micelle at CMC and also provide an estimate for the CMC by conducting simulations of free energy profile or the potential of mean force(PMF) required to extract individual molecules from a micelle.

3.2 Method

The simulations used LAMMPS, [70] with a 5 fs time step in the isothermalisobaric (TPN) ensemble, with temperature and pressure controlled using a Nose Hoover [71] thermostat, with a 100 fs damping constant, and a barostat, with a damping parameter of 1000 fs, at a temperature of 300 K and a pressure of 1 bar, respectively.

The equilibrated configuration at the end of an NPT trajectory was used as the initial structure for a "pulling" [74] simulation, to generate the initial configurations for umbrella sampling within each window.The center of mass (COM) distances between each polypeptoid and the rest of the polypeptoid were calculated, and the polypeptoid with the shortest COM distance from the rest of the polypeptoid was selected and then pulled away from the micelle so that the COM distance is 40 using a spring constant of 50 kcalmol^{-1-2}. To create a set of starting configurations, one for each "window" to be used in the weighted histogram free energy calculations, targeted molecular dynamics was done(TMD) [75] to get the intermediate COM distance till 40A as a reaction coordinate for umbrella sampling. A total of 112 windows were used with a spacing of 0.25 A COM separation between them. In each window, a 20 ns MD simulation was performed for umbrella sampling. A harmonic force with a force constant of 50 kcalmol^{-1-2} was applied for each umbrella sampling window. The Weighted Histogram Analysis Method (WHAM) [76] is used to get the PMF.

3.3 Result and Discussion

The results of the simulations, including the structure and self-assembly of Polypeptoid (Figure 3.1) at different aggregation number(N) are shown in Figure 2 where the hydrophobic and hydrophilic groups are represented by red and blue color respectively

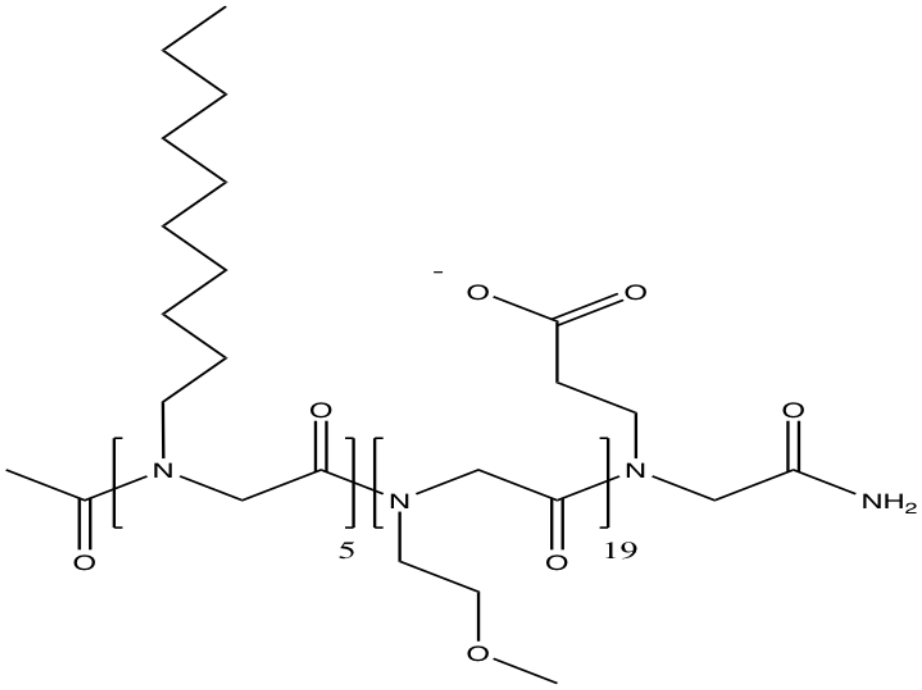

Figure 3.1: Polypeptoid

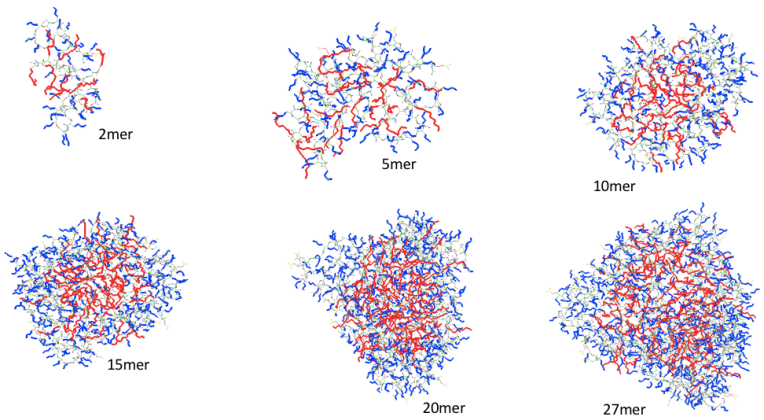

Figure 3.2: Self-assembly of Polypeptoid at different aggregation number(N)

3.3 Result and Discussion

The stability of the micelle can be further characterized by examining the eccentricity defined as. [75]

$$e = 1 - \frac{I_{\min}}{I_{avg}} \tag{3.1}$$

where I_{min} is the moment of inertia along the x, y, or z axis with the smallest magnitude and I_{avg} is the average of all three moments of inertia. From the Figure 3.3, it is shown that the shape of the micelle gets more spherical (smaller value of e) as the aggregation number gets higher

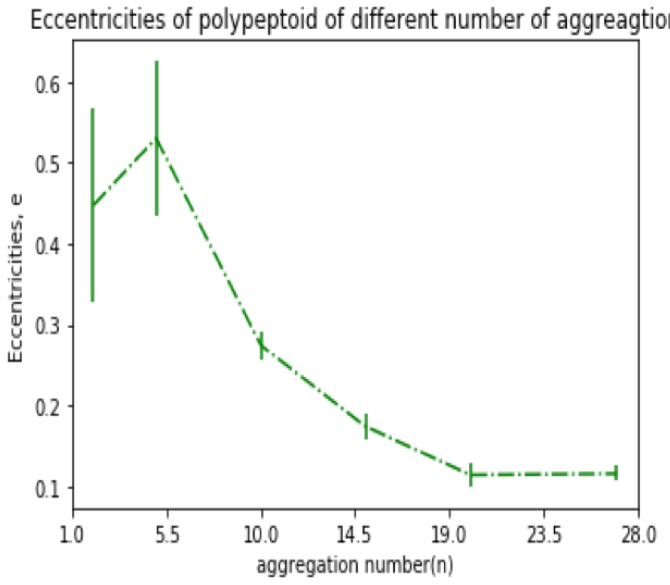

Figure 3.3: Eccentricity of polypeptoids as a function of different aggregation number

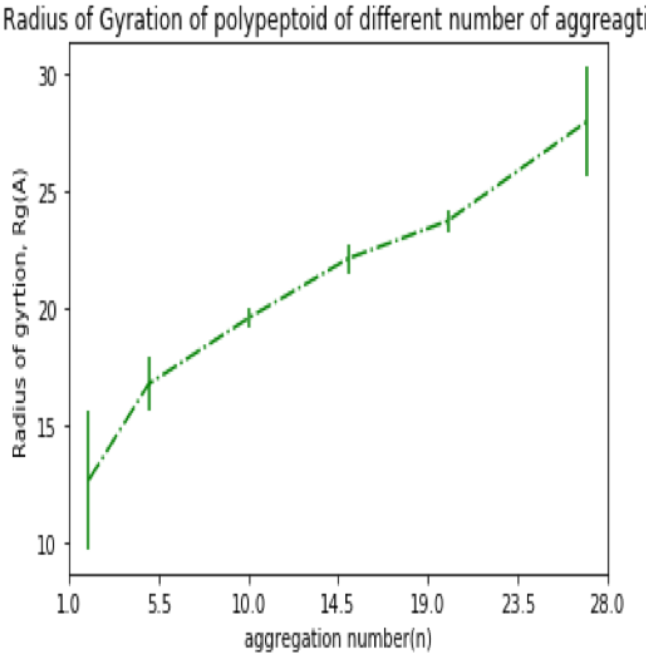

Figure 3.4: Radius of Gyration of Polypeptoids as a function of different aggregation number.

On the other hand, it is noticeable that the radius of gyration of the aggregated structure increases with the aggregation number (Figure 3.4). This is what we expect. The totally opposite scenario is seen in shape as aggregation number increases. We also calculated the probability density distribution of the distances of hydrophilic(Figure 3.5) and hydrophobic group (Figure 3.6) from center of mass of polypeptoid. When N is 2 hydrophilic groups are closer to center of mass than hydrophobic groups, the same feature is noticeable for N equals 5. However, this is changes when N is 10 where hydrophobic groups are closer to the center of mass than hydrophilic groups and the same thing is visible for N larger than 5. This indicates that the polypeptoids start self-assembling to micelle when the N is 10.

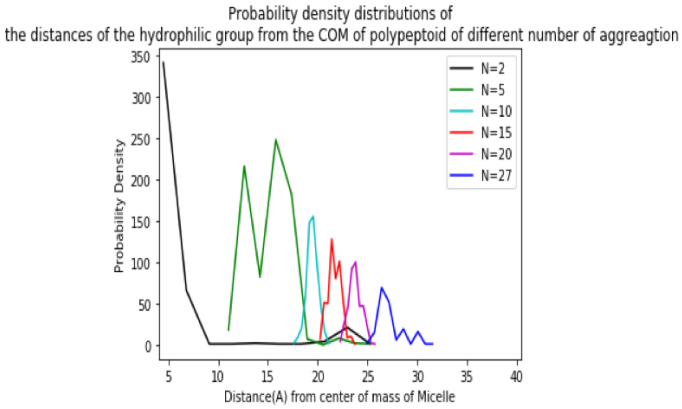

Figure 3.5: Probability density distribution of the center of mass distance of hydrophilic group of polypeptoid at different number of aggregations.

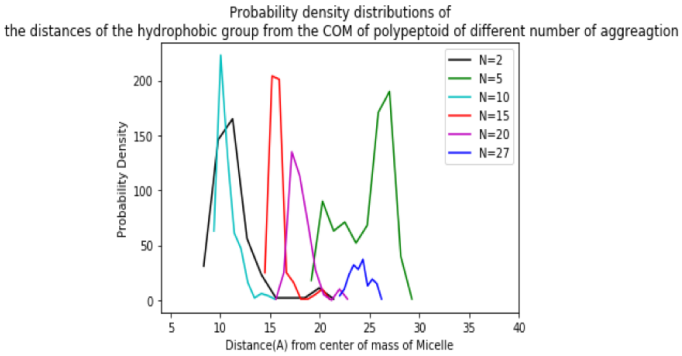

Figure 3.6: Probability density distribution of the center of mass distance of hydrophobic group of polypeptoid at different number of aggregations.

From the radial distribution function (rdf) between hydrophobic groups and water (Figure 3.7) it is noticeable that there is a large peak around the hydrophobic group for both the aggregation number 2 and 5, indicating more water contacts at close distances. But rdf peak is higher between hydrophilic and water for the case of 10, 15 ,20, 27 (Figure 3.8) This figure again implies that the most probable aggregation number of polypeptoid to form micelle is 10.

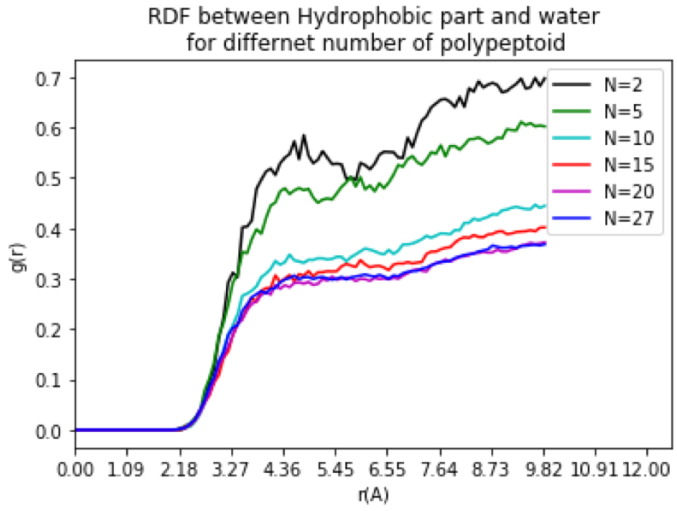

Figure 3.7: Radial distribution function (RDF) between hydrophobic groups and water for different number of Polypeptoid

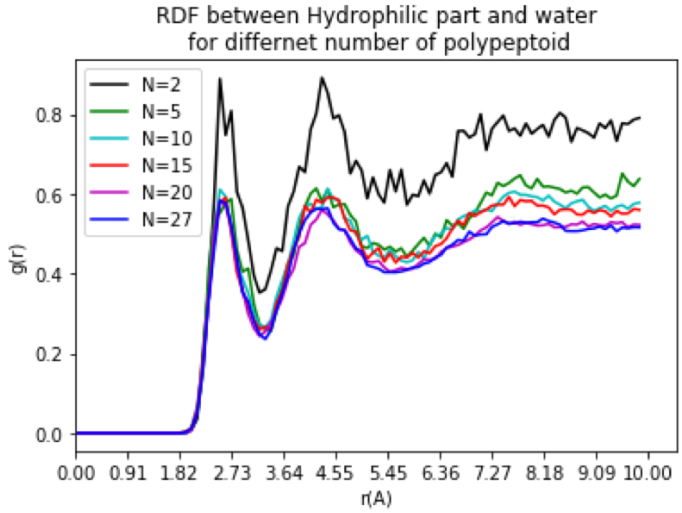

Figure 3.8: Radial distribution function (RDF) between hydrophilic group and water for different number of Polypeptoid.

We calculated PMF by pulling out a single PA from the preformed structure or micelle by umbrella sampling for different aggregation number to estimate the critical micelle concentration. In Figure 3.9. PMF of separating the monomers from the dimer and fivemer is shown.

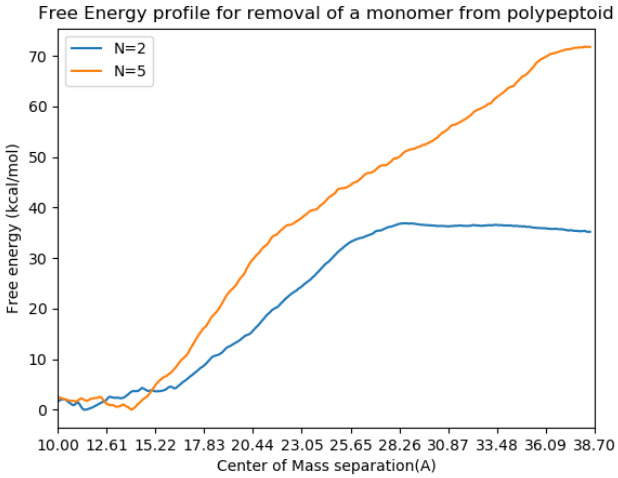

Figure 3.9: Potentials of mean force as a function of distance of pulled polypeptoid center of mass R from micelle center of mass for rest of the polypeptoid.

3.4 Conclusion

SW based Coarse Grained model are successful for large-scale simulation of a large range polymeric materials like polypeptoids in a very fast and efficient way. For our case, self assembly of 27mer polypepoid to a micelle took 2 to 3 months of computer time, whereas all atom model may take years to do this job.

4

Chapter 4

ALLOSTERIC INHIBITOR OF SARS-CoV-2 HELICASE BY MOLECULAR SIMULATION AND VIRTUAL SCREENING

4.1 Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has significantly affected the health and socioeconomic worldwide. SARS-CoV-2 helicase (Nsp13) plays an essential role in viral replication and is the most conserved nonstructural protein within the coronavirus family, making it a very promising target. SARS-CoV-2 helicase comprises 596 residues. And It is 99 percent homologous to SARS-CoV helicase. The X- ray structure from SARS-CoV (PDB ID 6JYT.2.A) shows five domains. [76] The three domains, IA, 1B, and 2A make up a triangular base with the N-terminal zinc binding domain (ZBD) at the top, and a stalk region connecting the base to the ZBD (Figure 4.1)

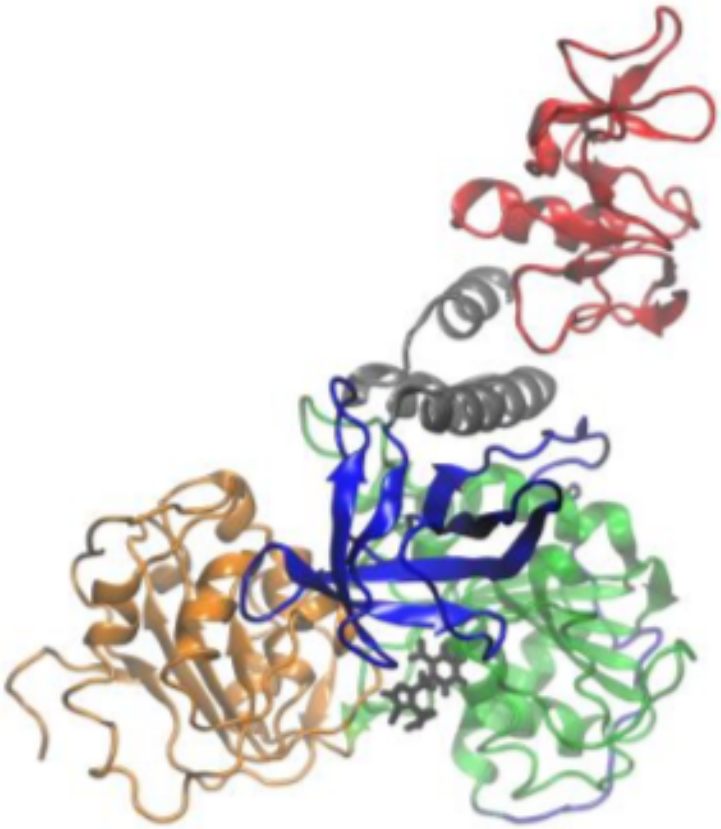

Figure 4.1: SWISS-MODEL structure of SARS-CoV-2 helicase, showing domain 1A in green, domain 1B in blue, domain 2A in orange, the stalk domain in grey, the ZBD in red, and a potential inhibitor in grey near the bottom.

SARS-CoV-2 helicase change its conformation, gaining the energy hydrolyzing ATP to perform its catalytical function.The flexibility of the protein, key to its function, might make helicases a challenge for structural-based drug design. [77]. However, many works have been done to identify potential SARS-CoV-2 inhibitors using computational docking and molecular dynamics (MD) simulations. [78]. The competitive inhibitors are designed at the ATP binding site of helicase located between domains 1A, 1B, and 2A. A better understanding of the flexibility of helicase can benefit the Inhibitor design. [78, 79]. MD simulation with enhanced sampling methods can be done to explore all the conformational areas.

In this work, the allosteric site away from ATP binding site and the nucleic acid-binding channel was determined by principal component analysis from enhanced MD simulation. The non-competitive inhibitors at the allosteric site are also proposed by molecular docking to stop helicase's flexibility, hence its function.

4.2 Method

4.2.1 Molecular simulation

For simulations, the recently developed a99SB-disp model [80] was used for protein since it has been shown that it can accurately reproduce the structure of a range of folded and disordered proteins, [80] as well as a protein that contains both folded and disordered domains. [81].A new water model, a99SB-disp water with slightly stronger Lennard-Jones dispersion was used with the a99SB-disp model. [80]. For the three zinc ions in the ZBD and the complexing cysteine and histidine residues,the Zinc Amber Force Field (ZAFF) was used [87]. The Generalized Amber Force Field (GAFF)51 parameters was used for the potential inhibitor, using AM1-BCC charges [89].

The simulations were setup using the tleap and antechamber programs of AMBER 16 [90, 91]. Those coordinate and topology files were converted to GROMACS input files using the ACPYPE program [92]. The simulations used the Replica Exchange with Dynamic Scaling (REDS) method as implemented in the GROMACS simulation package, version 5 [93].

The structure was minimized using steepest decent for 1000 steps, followed by constant molecule number, volume, and temperature (NVT) simulations for 100 ps and then 100 ps at constant N, pressure, and T (NPT) for each of the temperatures.

Gmx covar utility in Gromacs was used for Principal component analysis [94].

4.2.2 Molecular docking and scoring

To find potential allosteric inhibitors, docking into these sites was done using the program Autodock Vina, [79] using 3868 compounds from the SWEETLEAD [80] library. The SWEETLEAD library consists of approved pharmaceuticals, illicit drugs, and traditional medicinal herbs. Spatial data file (SDF) input files were converted to PDBQT (protein data bank format plus partial charges and atom types) input for Autodock using OpenBabel [81].

Additional binding pockets were explored using PrankWeb, [77] a web interface for the P2Rank ligand-binding site prediction method [78] for the structure resulting from the enhanced MD simulation.

4.3 Result and Discussion

4.3.1 Finding alosteric site

To quantify the motions of the protein, a principle component (PC) analysis was done. The first principal component of motion involves a bending motion of domain 2A, towards domains 1A and 1B. The second principal component involves a twisting motion of domain 2A. These motions are shown in Figure 4.2. The PC eigenvalues are 35.2 nm2 and 15.2 nm2 for the first and second components, respectively. The eigenvalues sum to 117.0 nm2, so the first two PCs represent 43 percent of all protein fluctuations. The projection of all the configuration (after the 50 ns of equilibration) for the apo and complex simulations onto the first two PC eigenvectors (from the apo simulations) are given in Figure 4.2. The complex and the apo protein cover similar regions of PC2, but the complex does not explore very negative regions of PC1, corresponding to an open structure, with a large separation between domains 2A and 1B.

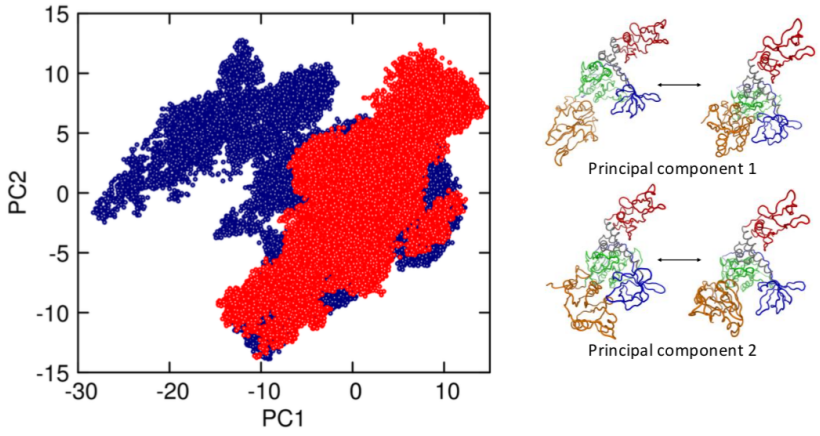

Figure 4.2: Projection of the trajectory for the apo protien (blue dots) and the complex (red dots) onto the first and second principle components. The structures on the left correspond to negative values of the PC, so that negative values of PC1 correspond to a larger distance between domains 1B and 2A and positive values represents a rotation of the C-terminus part of domain 1B towards domain 2A

The PC map analysis shows that the large scale motion of the protein involves movement of domain 2A, relative to the other domains. To restrain the flexibility of the helicase an inhibitor can be designed to bind at the interface between domains 1A and 2A, in a region away from both the ATP binding site and the nucleic acid binding channel.Using P2Rank [78] potential drug binding sites were identified. Four,

out of 17, binding sites were found at the interface between domains 1A and 2A, as shown in Figure 4.3(A). In order to inhibit motion of domain 2A, the inhibitor would ideally be at a site with residues that are involved in domain motion but moving in opposite directions. The four binding sites can be assessed based on the motions of the residues in those binding sites using the results of the PCA. P2Rank returns the points of the solvent accessible surface (SAS) for each binding site. For a point j on the surface, the nearest residue is found, which is labeled i_j. The motion of that residue can be characterized from $q_\alpha(i_j)$ the eigenvector of principal component α for residue i_j. The total motion of the binding site can be found from

$$q_\alpha(total) = \sum_{j=1}^{N} q_\alpha(i_j)/N \tag{4.1}$$

where N is the number of SAS points for a specific pocket. The angle between the direction of motion of one point on the surface relative the average direction is a way to determine if a principal component of motion is moving the whole surface together, or if parts of the surface move in different directions. The cosine of this angle, given by

$$d_\alpha(i) = q_\alpha(i_j).q_\alpha(total)/(|q_\alpha(i_j)||q_\alpha(total)|) \tag{4.2}$$

can be mapped onto the SAS.

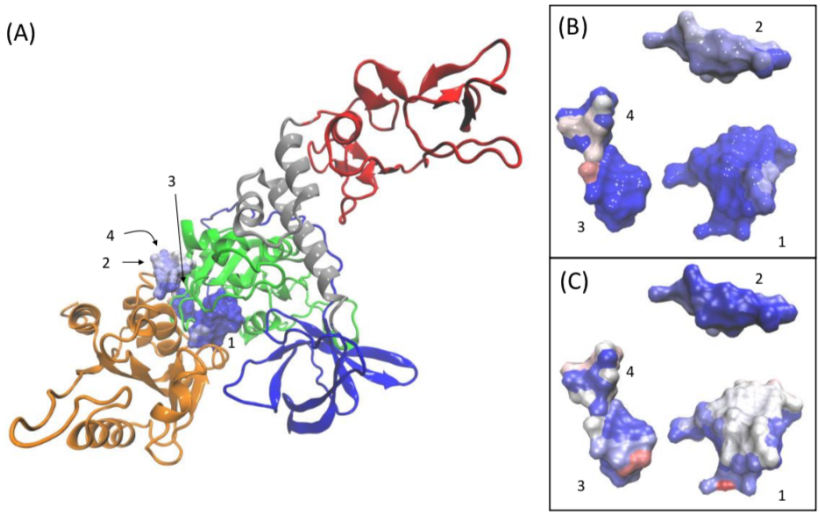

Figure 4.3: (A) Four potential binding sites at the interface between domains 1A and 2A. Map for the movement of principal component 1 (B) and 2 (C) onto the surface of the binding sites, with blue denotes a positive values and red negative d(i) values

In the Figure 4.3 (B) and (C) areas of the surface that move in the same direction as the average (positive $d_\alpha(i)$) of the pocket are shown in blue and regions that move

in the opposite direction as shown in red. The PC map analysis suggests that pockets 3 and 4 might be good targets for a non-competitive inhibitor, which could limit the flexibility of domain 2A since pocket 4 shows anisotropy for both PC and pocket 3 shows anisotropy for second PC. Whereas pocket 2 does not show any anisotropy neither for 1st PC nor 2nd PC and pocket 1 is ATP binding site.

4.3.2 Find inhibitor at alostering site by molecular docking

Pocket 4, while promising in terms of inhibiting domain motion, is not ranked highly by P2Rank (13th out of 17 potential pockets). Docking into this pocket gave weakly binding compounds, with Autodock Vina scores at or above -7 kcal/mol. Pocket 3, ranked fourth out of the 17 pockets, is more promising from a drug binding perspective. Nine compounds from the SWEETLEAD database80 had Vina scores below -7 kcal/mol. The scores and several pharmacokinetic factors are given in Table 4.1. The chemical structures of these compounds and the docked position of the top scoring molecule are shown in Figure 4.4. The top scoring compound is raltegravir, an approved antiviral medication used to treat HIV/AIDS, as an HIV integrase inhibitor.85 The molecule makes hydrogen bonds with ASN 265, THR 440, and ARG 442, as well as hydrophobic contacts between benzyl group and ALA 446, VAL 449, and the methyl group on THR 440. It, therefore, makes significant contacts with residues on both domains 1A and 2A.

Table 4.1: Docking score, molecular weight (MW), octanol water partition coefficient (logP), polar surface area (PSA), the number of hydrogen bond donors, and the number of hydrogen bond acceptors for the top nine scoring compounds for binding pocket 3

Compounds	Score (kcal/mol)	MW (g/mol)	logP	PSA ($Å2$)	H-bond donors	H-bond aacceptors
SW03354	-8.6	443.4	-3.9	155.1	2	11
SW03379	-8.3	533.5	-3.3	215.5	6	13
SW04090	-8.2	721.6	0.6	220.9	3	14
SW03819	-8.1	552.5	1.9	172.6	3	10
SW03493	-7.9	463.6	1.2	86.5	6	6
SW03107	-7.8	414.4	7.1	68.3	0	4
SW03768	-7.7	529.5	5.0	97.6	2	8
SW03357	-7.6	444.4	-2.4	192.8	3	10
SW03122	-7.5	415.5	0.5	91.2	1	6

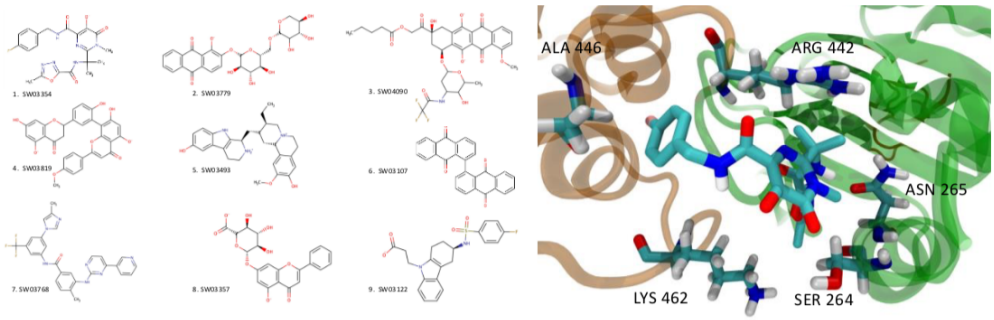

Figure 4.4: Chemical structures of inhibitors for allosteric pocket 3 and docked structure of SW03354

4.4 Conclusion

The flexibility of the helicase might be altered with an inhibitor to bind at the interface between domains 1A and 2A, in a region away from the ATP and the nucleic acid active sites. Four potential drug binding pockets between domains 1A and 2A were identified. We sought a binding site that could change the flexibility by mapping the atoms' PC motion on each pocket's surface onto the pocket's surface, as shown in Figure 4.3. If a pocket is at the boundary between regions that tend to move in different directions, then the PC map would show negative and positive areas on the surface. One pocket was identified as a promising allosteric site from the PC map involved domain motion. The virtual screening revealed that this site could bind drug-like molecules. Further studies would be needed to determine if molecules designed to bind to the allosteric site are effective helicase inhibitors.

5

Chapter 5

REPURPOSING APPROVED DRUGS AS INHIBITORS OF MAIN PROTEASE OF SARS-CoV-2 FROM MOLECULAR DYNAMIC SIMULATION AND VIRTUAL SCREENING

5.1 Introduction

SARS-CoV-2, responsible for COVID-19 pandemic [94], is known as a coronavirus (CoV) from the Coronaviridae family. Its RNA genome is 82 percent identical to SARS-CoV [95], causing severe acute respiratory syndrome (SARS) pandemic in 2003 [96]. Two cysteine proteases: the chymotrypsin-like cysteine or main protease, known as 3CLpro or Mpro, and the papain-like cysteine protease, PLpro are encoded by SARS-CoV-2. They catalyze the proteolysis of polyproteins translated from the viral genome into nonstructural proteins which is essential for packaging the nascent virion and viral replication. [97] Therefore, the replication of the virus can be stopped by inhibiting the activity of this protease. The peptide bond in Leu-Gln-Ser-Ala-Gly recognition sequence of the substrate hydrolyzed by Mpro is distinct from the peptide sequence recognized by other human cysteine proteases so far. [98] Thus, Mpro is considered a promising target for anti SARS-CoV-2 drug design; it has been the center of attention of several studies since the pandemic has started. [99] An X-ray crystal structure of Mpro (PDB ID: 5R82)shows that it comprises 306 residues. [95]

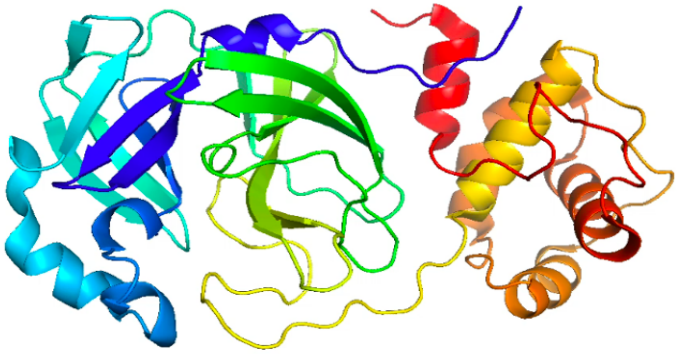

Figure 5.1: Main Protease of SARS-CoV-2.

The amino acid sequence of the active site of Mpro is highly conserved among coronaviruses. [100] From the literature survey, it is found that residues involved in the binding of substrates are Phe140, His163, Met165, Glu166, and Gln189 (Figure 5.2) [101]

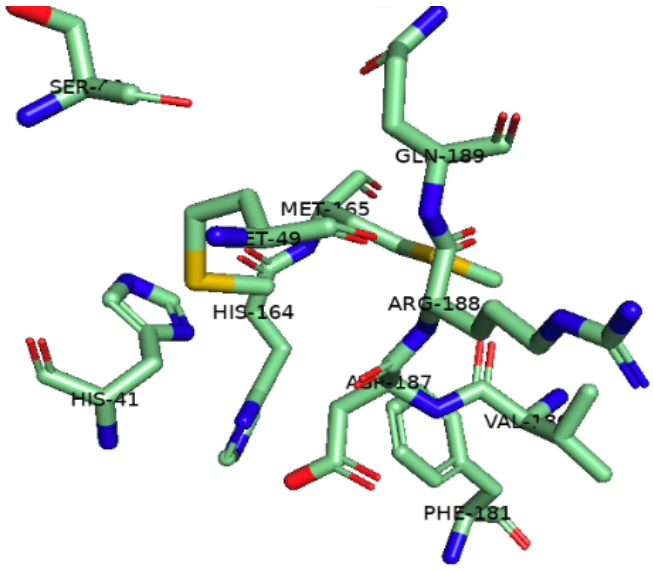

Figure 5.2: Active site of Main Protease of SARS-CoV-2.

When we started this work, the vaccine was not invented; in that case, drug repurposing can be an important strategy for immediate response to the COVID-19 pandemic. So, the primary goal of this computational work is to find existing drugs that might be effective against SARS-CoV-2 main protease. We basically choose three

significant drug molecules from existing drugs: Ribavirin, an anti-hepatitis C drug; Azelastine, an anti-allergic drug; and luteolin, an anti-inflammatory drug (Figure 5.3). In addition, several reports have indicated that HIV-1 protease inhibitors can design SARS- CoV proteinase inhibitors (Zhang and Yap, 2004; Yamamoto et al., 2004; Jenwitheesuk and Samudrala, 2003). In particular, the proteinase inhibitor Kaletra, a mixture of the proteinase inhibitors ribavirin and ritonavir (Fig. 1), exhibits encouraging signs of being partially effective against the SARS virus (Vastag, 2003). Therefore, a deeper understanding of known proteinase inhibitors as anti-SARS drugs can be used as a starting point to design and discover better or more specific inhibitors for treating SARS and any mutations by predicting resistance mechanisms and new drugs for such mechanisms in advance.

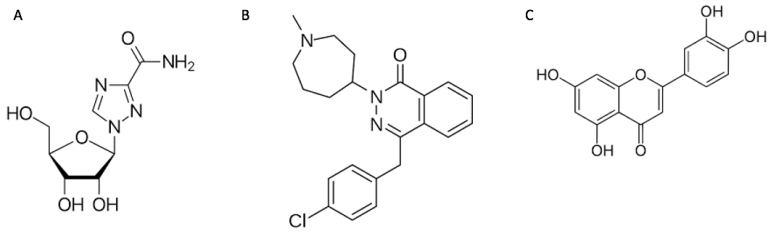

Figure 5.3: A)Ribavirin, B) Azelastine, C)Luteolin.

In this work, two docking algorithms Autodockvina [79] and SwissDock [102], are used for a consensus virtual screening using the three FDA-approved drugs (Ribavirin, Azelastine, Luteolin). In addition, further studies molecular dynamic simulation are done to see whether the complex structure of the main protease of SARS-CoV-2 and drugs from virtual screening is stable or not.

5.2 Method

5.2.1 Starting structure

Pubchem database is used to get the SDF files of 3D stucture of drug molecules. [103] Pubchem CID for ribavirin, azelastine, luteolin are 37542, 2267, 5280445 respectively.

5.2.2 Molecular docking

To mitigate the biases, we performed two independent runs of proteinligand docking with three approved drugs using AutoDock Vina[79] and Swissdock. [102]

SwissDock is an online docking web server. Before uploading into SwissDock, the input structure file of protein and ligand was prepared using UCSF Chimera [104].

Swissdock returns the results by e-mail.The docked pose with the minimum fullfitness score is considered for futher analysis. LigPlot (Pettersen et al., 2004), a molecular visualization tool, was used to visualize the results obtained from the server.

AutoDock Vina [79] is a widely used open-source program for molecular docking. The PDBQT files from SDF and PDB files of compounds and the protein generated by The AutoDockTools (ADT) [79] software6. Non-polar hydrogen atoms were removed. An affinity grid box with a size of 18 × 21 × 18 Å was generated and centered on the active site. The default docking parameters were used, except for the number of modes that was set to 9.

5.2.3 Molecular Dynamic Simulation

The GROMACS software, version 2018a, was used for performing all molecular dynamics (MD) simulations. [105] OPLS-AA/M force field. [106] was used for the protonated Mpro dimer, with a net charge of -8 e. TIP4P was used for water solvent. [107] Sodium counterions were added to neutralize the system. The OPLS/CM1A force field was used for the selected ligand candidates, [108] using d the LigParGen Python code.13 The parameters were converted to GROMACS format using LigParGen. [109] . A triclinic simulation box with 10-Å was used padding. Energy minimization was then performed using the steepest descent algorithm. A cutoff radius of 12 Å was used to calculate non-bonded interactions explicitly. Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) algorithm. [110] All covalent bonds along with hydrogen atoms were constrained at their equilibrium lengths using the LINCS algorithm. [111] Each system was initially simulated for 1 ns in the canonical ensemble (NVT) for the solvent to relax and the temperature to equilibrate. Initial velocities were sampled from a Maxwell- Boltzmann distribution at 310 K. The coupling constant of the thermostat was set to 2.0 ps. The system was then equilibrated for 1.5 ns in the isothermal-isobaric ensemble (NPT). The pressure was kept at 1 bar by the Berendsen barostat [112].

5.3 Result and Discussion

5.3.1 Molecular Docking

In comparing the ΔG value from both autodockvina and swissdock, which gives the estimated free energy of binding, we can infer that the azelastine has the highest binding affinity compared to Ribavirin and Luteolin drug compounds towards SARS-CoV-2 main protease. However, the binding affinity of azelastine is very close to that of Luteolin from autodock vina (Figure 5.4).

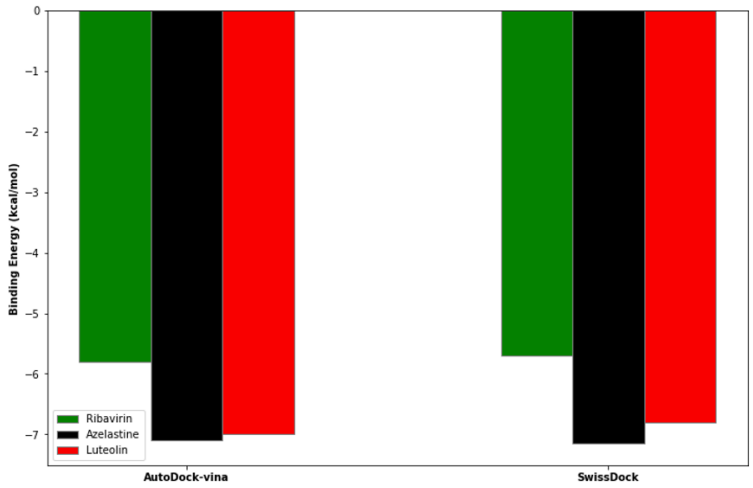

Figure 5.4: Binding Affinity of Drug Molecule for Main protease from Docking

The docked pose of the minimum energy conformers of the three drugs, ribavirin, azelastine, and luteolin, along with their corresponding 2D interaction plots, are shown in Figure 5.5, 5.6, and 5.7, respectively. The docked poses clearly indicate that the drugs molecules bind within the active site of the SARS-CoV-2 Mpro macromolecular structure.

Figure 5.5 shows that ribavirin binds firmly through two conventional hydrogen bonds with residues Gln189 and Ser144 also form a significant number of hydrophobic interactions.

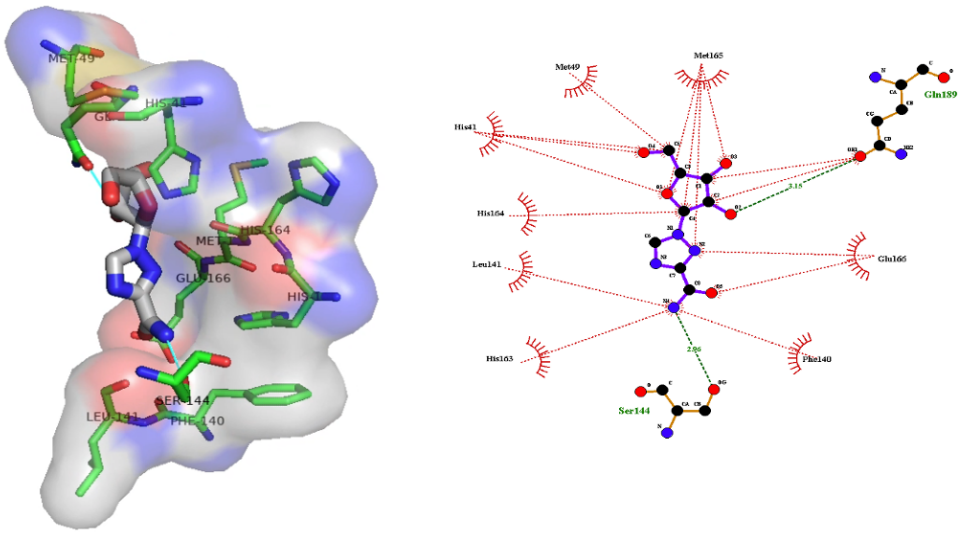

Figure 5.5: Ribavirin in the binding site of main protease of SARS-CoV-2 and Molecular Interaction Between Them from docking

Azelastine stabilizes in the active site through all hydrophobic interactions with residues. It does not create any H-interaction even though it shows the highest binding affinity, as depicted in the 2D plot of Figure 5.6

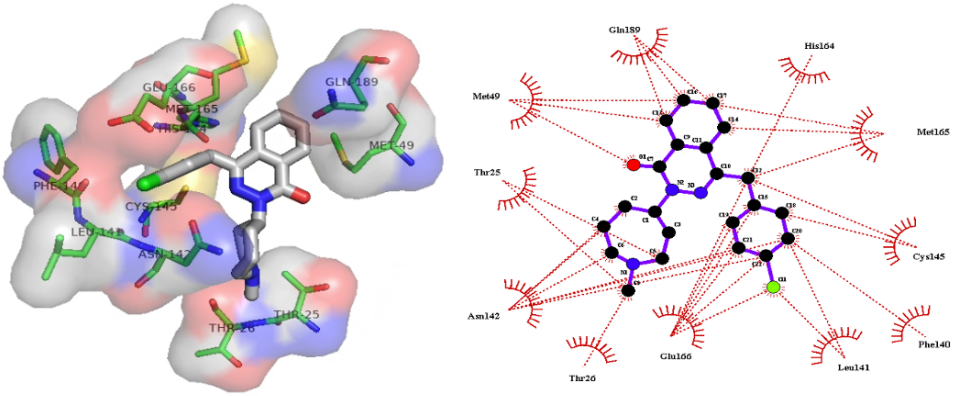

Figure 5.6: Azilastine in the binding site of main protease of SARS-CoV-2 and Molecular Interaction Between Them from docking

Luteolin, a promising candidate for the treatment of the current pandemic due to SARS-CoV-2, showing 2nd highest binding affinity among the three-drug candidate (Gautret et al., 2020), has been found to interact with the active site of the protease through H-bond interaction with Gly 143, His 41, Thr 25, Cys 44 and Glu 66 as shown in Figure 5.7. Besides H-bond, it also forms some hydrophobic interaction.

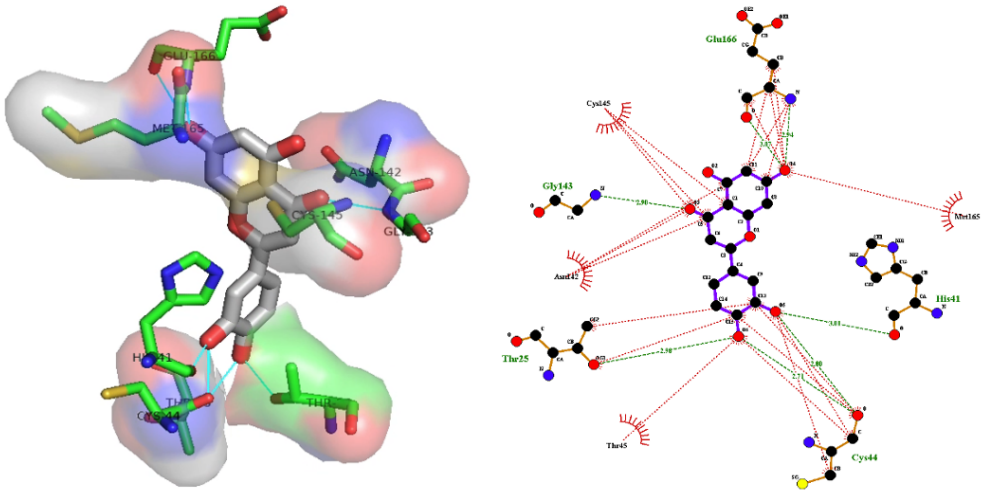

Figure 5.7: Luteolin in the binding site of main protease of SARS-CoV-2 and Molecular Interaction Between Them from docking

5.4 Molecular Dynamic Simulation

Usually, the analysis of RMSD and RMSF can infer important information about the stability and flexibility of the receptor-ligand complex. Weak stability may be seen in the high deviation and fluctuation of proteins during a simulation(Ghosh Chakraborty, 2020). RMSDs during the whole simulation of SARS-CoV-2 Mpro in complexed with three drugs were not found in the significant difference. The RMSD in Figure 5.8 (A) shows SARS-CoV-2 Mpro is equilibrated at the last 5ns of MD simulation for all systems in which the values were 0.28 nm with a standard deviation of 0.02 nm for the three complexes. This shows a stable binding of drug molecules with Mpro. Likewise, the RMSF per residue using their C-alpha atoms (Figure 5.8(B)) was calculated to determine and identify the Main protease's flexibility. Comparing the RMSF of the ligands and main protease complexes, it appears that the complex for Ribavirin involves the highest fluctuation than the other two ligand-complex. The radius of gyration (Figure 5.8(C)) was used to characterize the compactness of the main protease (Figure 5.8(C)). The Main protease complexed with Azelastine and Luteolin ligands presents similar compactness along with the entire MD simulation. For Ribavirin, the Rg of the protein remains low fluctuation at the last 7ns of the trajectory. Overall, the main protease in all systems is compact with an Rg value of 4.35nm. In general, the RMSD, RMSF, and Rg show that the main protease has low conformation change at the last 6ns of the MD trajectories. Thus, as expected, the main protease present a similar solvent accessible surface area (Figure 5.8(D)),

retaining its accessibility in the last 6 ns of the MD simulations.

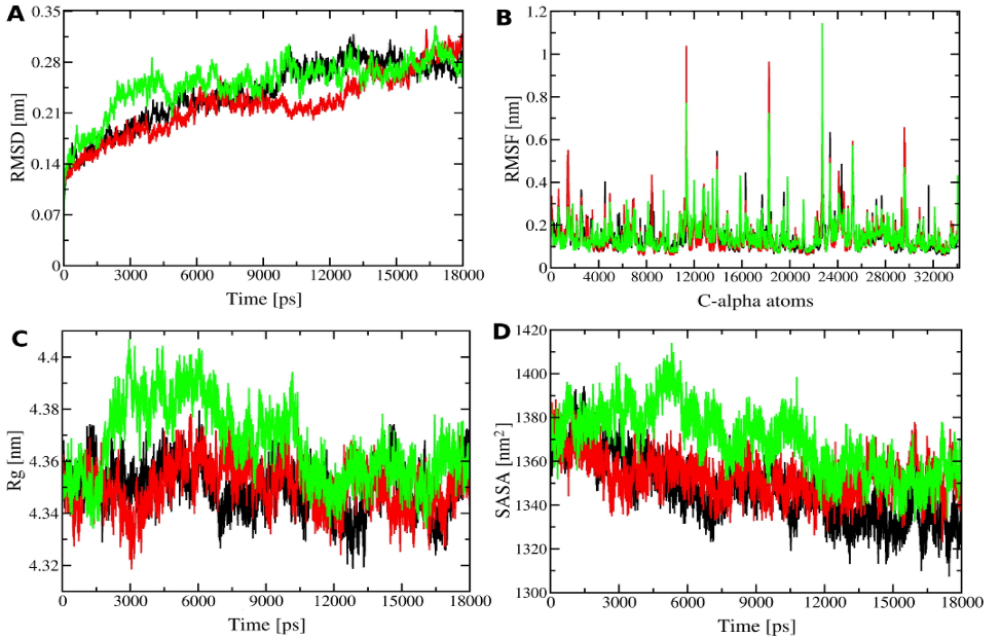

Figure 5.8: A)RMSD , B)RMSF and C)radius of gyration (Rg) of the C-alpha atoms, and D)solvent accessible surface area (SASA,) of the Mpro complexed with Ribavirin(green line), Azilastine (black line) and Luteolin (red line) ligands

5.5 Conclusion

Molecular dynamics (MD) simulation and docking calculations are used to study the SARS-CoV- 2 Main protease to obtain potential drug candidates for repurposing against COVID-19. Overall, the Main protease does not undergo significant structural change along the 5 ns MD simulation equilibration in a water solvent.

Docking shows Azilastine has a high binding affinity to the Main protease.MD simulations of the three ligands@complexes show that the ligands remain bounded in the binding region along the simulation.

6
References

Bibliography

[1] Islam, N. N., Sharma, A., Gyawali, G., Kumar, R., Rick, S. W. Coarse-grained models for constant pH simulations of carboxylic acids. *Journal of chemical theory and computation*, (2019), 15(8), 4623-4631.

[2] Dai, S.; Ravi, P.; Tam, K. C. pH-responsive polymers: synthesis, properties, andapplications. *Soft Matter* 2008, 4, 435449.

[3] Stillinger, F. H.; Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B: Condens. Matter Mater. Phys.* 1985, 31, 52625271.

[4] Molinero, V.; Moore, E. B. Water modeled as an intermediate element between carbonand silicon. *J. Phys. Chem.* B 2009, 113, 40084016.

[5] Jacobson, L. C.; Molinero, V. A methane-water model for coarse-grained simulations ofsolutions and clathrate hydrates. *J. Phys. Chem. B* 2010, 114, 73027311.

[6] DeMille, R. C.; Molinero, V. Coarse-grained ions without charges: Reproducing the solvation structure of NaCl in water using short-ranged potentials. *J. Chem. Phys.* 2009,131, 034107.

[7] Raubenolt, B.; Gyawali, G.; Tang, W.; Wong, K. S.; Rick, S. W. Coarsegrainedsimulations of aqueous thermorepsonsive polyethers. *Polymers* 2018, 10, 475.

[8] Chen, W.; Morrow, B. H.; Shi, C.; Shen, J. K. Recent development and application of constant pH molecular dynamics. *Mol. Simul.* 2014, 40, 830838.

[9] Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics ingeneralized Born implicit solvent. *J. Comput. Chem.* 2004, 25, 20382048.

[10] Yukawa, H. On the interaction of elementary particles. *I. Proc. Phys.-Math. Soc. Jpn.*,3rd Ser. 1935, 17, 4857.

[11] Jacobson, L. C.; Kirby, R. M.; Molinero, V. How short is too short for the interactions of a water potential? Exploring the parameter space of a coarse-grained water model usinguncertainty quantification. *J. Phys. Chem. B* 2014, 118, 81908202.

[12] Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water:TIP4P/2005. *J. Chem. Phys.* 2005, 123, 234505.

[13] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetis and properties of organic liquids. *J. Am.Chem. Soc.* 1996, 118, 1122511236.

[14] Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. Thermodynamics of solute transfer from water tohexadecane. *J. Chem. Soc., Perkin Trans.* 2 1990, 2, 291300.

[15] Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of absolute free energies of hydration using continuum methods: Accuracy of partial charge models and optimization of nonpolar contributions. *J. Chem. Theory Comput.* 2006, 2, 128139.

[16] Eliassaf, J.; Silberberg, A. The gelation of aqueous solutions of polymethacrylic acid.*Polymer* 1962, 3, 555564.

[17] Higashi, N.; Matsumoto, T.; Niwa, M. Chain length discrimination of water-soluble polymers by a poly(methacrylic acid) or poly(oxyethylene) segment-carrying assembliesat the air- water interface. *Langmuir* 1994, 10, 46514656.

[18] Peljhan, S.; Zagar, E.; Cerkovnik, J.; Kogej, K. Strong intermolecular association between short poly(ethacrylic acid) chains in aqueous solutions. *J. Phys. Chem. B* 2009,113, 23002309.

[19] Chockalingam, R.; Natarajan, U. Self-association behaviour of atactic poly-methacrylic acid in aqueous solution investigated by atomistic molecular dynamics simulations. *Mol.Simul.* 2015, 41, 11101121.

[20] Wallace, J. A.; Shen, J. K. Continuous constant pH molecular dynamics in explicitsolvent with pH-based replica exchange. *J. Chem. Theory Comput.* 2011, 7, 26172629.

[21] limpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.*1995, 117, 119.

[22] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C.GROMACS: Fast, flexible and free. *J. Comput. Chem.* 2005, 26, 17011718.

[23] Mezei, M. The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water. *J. Chem.Phys.* 1987, 86, 7084 7088.

[24] Sharma, A.; Smith, J. D.; Walters, K. B.; Rick, S. W. Constant pH simulations of pHresponsive polymers. *J. Chem. Phys.* 2016, 145, 234906.

[25] Du, P.; Rick, S. W.; Kumar, R. Towards a coarse-grained model of the peptoid backbone: the case of N,N-diemethyl acetamide. *Phys. Chem.Chem.Phys* .2018, 20, 2338623396.

[26] Islam, N., Flint, M., Rick, S. W. Water hydrogen degrees of freedom and the hydrophobic effect. *The Journal of Chemical Physics*, 150(1), 014502 (2019)..

[27] R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.* 83, 8069 (1986).

[28] N. T. Southall, K. A. Dill, and A. D. J. Haymet, *J. Phys. Chem. B* 106, 521(2002).

[29] L. R. Pratt, *Annu. Rev. Phys. Chem.* 53, 409 (2002).

[30] R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.* 110, 1670 (2013).

[31] J. R. Errington and P. G. Debenedetti, *Nature* 409, 318 (2001).

[32] H. S. Ashbaugh, J. W. Barnett, A. Saltzman, M. E. *Langrehr*, and H. Houser, J.Chem. Phys. 145, 201102 (2016).

[33] Y. L. A. Rezus and H. J. Bakker, *Phys. Rev. Lett.* 99, 148301 (2007).

[34] M. Kinoshita, *J. Chem. Phys.* 128, 024507 (2008).

[35] V. Molinero and E. B. Moore, *J. Phys. Chem. B* 113, 4008 (2009)

[36] B. Song and V. Molinero, *J. Chem. Phys.* 139, 054511 (2013).

[37] G. Gyawali, S. Sternfield, R. Kumar, and S. W. Rick, *J. Chem. Theory Comput.* 13, 3846 (2017).

[38] S.Plimpton,*J. Comput. Phys.* 117, 1 (1995).

[39] J. R. Errington and P. G. Debenedetti, *Nature* 409, 318 (2001).

[40] S. V. Anishchik and N. N. Medvedev, *Phys. Rev. Lett.* 75, 4314 (1995).

[41] V. P. Voloshin, N. N. Medvedev, M. N. Andrews, R. R. Burri, R. Winter, and A. Geiger, *J. Phys. Chem. B* 115, 14217 (2011).

[42] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Chem. Phys.* 91, 6269 (1987).

[43] J. L. F. Abascal and C. Vega, *J. Chem. Phys.* 123, 234505 (2005).

[44] S. W. Rick, *J. Chem. Phys.* 120, 6085 (2004).

[45] G. S. Kell, *J. Chem. Eng. Data* 20, 97 (1975).

[46] E. Wilhelm, R. Battino, and R. J. Wilcock, *Chem. Rev.* 77, 219 (1977).

[47] A. Ben-Naim and Y. Marcus, *J. Chem. Phys.* 81, 2016 (1984).

[48] J. Grdadolnik, F. Merzel, and F. Avbelj, *Proc. Natl. Acad. Sci. U. S. A.* 114, 322 (2017).

[49] Rosler, A.; Vandermeulen, G. W.; Klok, H.-A. Advanced drug delivery devices via self-assembly of amphiphilic block copolymers. *Adv. Drug Delivery Rev.* 2001, 53, 95108

[50] Noble, G. T.; Stefanick, J. F.; Ashley, J. D.; Kiziltepe, T.; Bilgicer, B. Ligand-targeted liposome design: challenges and fundamental considerations. *Trends Biotechnol.* 2014, 32, 3245.

[51] Wen, Y.; Collier, J. H. Supramolecular peptide vaccines: tuning adaptive immunity. *Curr. Opin. Immunol.* 2015, 35,7379.

[52] Mulder, W. J.; Strijkers, G. J.; Van Tilborg, G. A.; Cormode, D. P.; Fayad, Z. A.; Nicolay, K. Nanoparticulate assemblies of amphiphiles and diagnostically active materials for multimodality imaging. *Acc. Chem. Res.* 2009, 42, 904914.

[53] Alexandridis, P.; Lindman, B. Amphiphilic Block Copolymers: Self- Assembly and Applications; *Elsevier*: Amsterdam, The Netherlands, 2000.

[54] Kelarakis, A.; Yang, Z.; Pousia, E.; Nixon, S. K.; Price, C.; Booth, C.; Hamley, I. W.; Castelletto, V.; Fundin, J. Association properties of diblock copolymers of propylene oxide and ethylene oxide in aqueous solution. the effect of p and e block lengths. *Langmuir* 2001, 17, 8085 8091

[55] Renou, F.; Nicolai, T.; Nicol, E.; Benyahia, L. Structure and viscoelasticity of mixed micelles formed by poly (ethylene oxide) end capped with alkyl groups of different length. *Langmuir* 2009, 25, 515 521

[56] Vorobyova, O.; Yekta, A.; Winnik, M. A.; Lau, W. Fluorescent probe studies of the association in an aqueoussolution of a hydrophobically modified poly(ethylene oxide). *Macromolecules* 1998, 31, 89989007.

[57] Elliott, P. T.; Xing, L.-l.; Wetzel, W. H.; Glass, J. E. Influence of terminal hydrophobe branching on the aqueous solution behavior of model hydrophobically modified ethoxylated urethane associative thickeners. *Macromolecules* 2003, 36, 84498460.

[58] Daoud, M.; Cotton, J. Star shaped polymers: a model for the conformation and its concentration dependence. *J. Phys.* 1982, 43, 531538.

[59] Chassenieux, C.; Nicolai, T.; Durand, D. Association of hydrophobically end-capped poly (ethylene oxide). *Macromolecules* 1997, 30, 49524958.

[60] Gourier, C.; Beaudoin, E.; Duval, M.; Sarazin, D.; Maitre, S.; Francois,J.Alight scattering study of the association of hydrophobi- cally -and , -end-capped poly (ethylene oxide) in water. *J. Colloid Interface Sci.* 2000, 230, 4152.

[61] Beaudoin, E.; Borisov, O.; Lapp, A.; Billon, L.; Hiorns, R. C.; Francois, J. Neutronscattering of hydrophobically modified poly (ethylene oxide) in aqueous solutions. *Macromolecules* 2002, 35, 74367447

[62] Serero, Y.; Aznar, R.; Porte, G.; Berret, J.-F.; Calvet, D.; Collet, A.; Viguier, M. Associating polymers: from "flowers" to transient networks. *Phys. Rev. Lett.* 1998, 81, 5584.

[63] Pham, Q.; Russel, W.; Thibeault, J.; Lau, W. Micellar solutions of associative triblock copolymers: Entropic attraction and gas-liquid transition. *Macromolecules* 1999, 32, 29963005.

[64] Vorobyova, O.; Lau, W.; Winnik, M. A. Aggregation number determination in aqueous solutions of a hydrophobically modified poly (ethylene oxide) by fluorescence probe techniques. *Langmuir* 2001, 17, 13571366.

[65] Molinero, V.; Moore, E. B. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem.* B 2009, 113, 40084016.

[66] DeMille, R. C.; Molinero, V. Coarse-grained ions without charges: Reproducing the solvation structure of NaCl in water using short-ranged potentials. *J. Chem. Phys.* 2009, 131, 034107.

[67] Gyawali, G.; Sternfield, S.; Kumar, R.; Rick, S. W. Coarse- grained models of aqueous and pure liquid alkanes. *J. Chem. Theory Comput.* 2017, 13, 38463853.

[68] Raubenolt, B.; Gyawali, G.; Tang, W.; Wong, K. S.; Rick, S. W. Coarse-grained simulations of aqueous thermorepsonsive polyethers. *Polymers* 2018, 10, 475.

[69] Du, P.; Rick, S. W.; Kumar, R. Towards a coarse-grained model of the peptoid backbone: the case of N,N-diemethylacetamide. *Phys. Chem.Chem.Phys.* 2018, 20, 23386-23396

[70] Silani, M., Talebi, H., Arnold, D., Ziaei-Rad, S., Rabczuk, T. (2012). On the coupling of a commercial finite element package with lammps for multiscale modeling of materials. *Steel Research International, Special issue*, 1, 1371-1374.

[71] Evans, D. J., Holian, B. L. (1985). The nose–hoover thermostat. *The Journal of chemical physics*, 83(8), 4069-4074.

[72] Vijayaraj, R., Van Damme, S., Bultinck, P., Subramanian, V. (2012). Molecular dynamics and umbrella sampling study of stabilizing factors in cyclic peptide-based nanotubes. *The Journal of Physical Chemistry B*, 116(33), 9922-9933.

[73] Pedretti, A., Labozzetta, A., Monte, M. L., Beccari, A. R., Moriconi, A., Vistoli, G. (2011). Exploring the activation mechanism of TRPM8 channel by targeted MD simulations. *Biochemical and biophysical research communications*, 414(1), 14-19.

[74] Grossfield, A. (2003). WHAM.

[75] Yuan, F., Larson, R. G. (2015). Multiscale molecular dynamics simulations of model hydrophobically modified ethylene oxide urethane micelles. *The Journal of Physical Chemistry B*, 119(38), 12540-12551.

[76] Jia, Z.; Yan, L.; Ren, Z.; Wu, L.; Wang, J.; Guo, J.; Zheng, L.; Ming, Z.; Zhang, L.; Lou, Z.; Rao, Z., Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res.* 2019, 47, 6538-6550.

[77] Datta, A.; Brosh, R. M., Jr., New insights Into DNA helicases as druggable targets for cancer therapy. *Front. Mol. Biosci.* 2018, 5, 59.

[78] Mirza, M. U.; Froeyen, M., Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. *J Pharm Anal* 2020.

[79] Ligat, G.; Da Re, S.; Alain, S.; Hantz, S., Identification of Amino Acids Essential for Viral Replication in the HCMV Helicase-Primase Complex. *Front Microbiol* 2018, 9, 2483.

[80] Robustelli, P.; Piana, S.; Shaw, D. E., Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci U S A* 2018, 115, E4758-E4766.

[81] Kuzmanic, A.; Pritchard, R. B.; Hansen, D. F.; Gervasio, F. L., Importance of the Force Field Choice in Capturing Functionally Relevant Dynamics in the von Willebrand Factor. *J Phys Chem Lett* 2019, 10, 1928-1934.

[82] Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M., Jr., Structural survey of zinc containing proteins and the development of the zinc AMBER force field (ZAFF). *J Chem Theory Comput* 2010, 6, 2935-2947.

[83] Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 2002, 23, 1623-41.

[84] Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A., AMBER 2016, University of California, San Fransisco. 2016.

[85] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 2006, 25, 247-60.

[86] Sousa da Silva, A. W.; Vranken, W. F., ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* 2012, 5, 367.

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

[88] Daura, X.; van Gunsteren, W. F.; Mark, A. E., Folding-unfolding thermodynamics of a beta- heptapeptide from equilibrium simulations. *Proteins* 1999, 34, 269-80.

[89] Jendele, L.; Krivak, R.; Skoda, P.; Novotny, M.; Hoksza, D., PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Res* 2019, 47, W345-W349.

[90] Krivak, R.; Hoksza, D., P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J Cheminform* 2018, 10, 39.

[91] Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* 2010, 31, 455-61.

[92] Novick, P. A.; Ortiz, O. F.; Poelman, J.; Abdulhay, A. Y.; Pande, V. S., SWEETLEAD: an in silico database of approved drugs, regulated chemicals, and herbal isolates for computer-aided drug discovery. *PLoS One* 2013, 8, e79568.

[93] O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *J Cheminform* 2011, 3, 33.

[94] Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.- H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.- Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* 2020, 579, 265269.

[95] Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* 2020, 582, 289293.

[96] Drosten, C.; Gunther, S.; Preiser, W.; van der Werf, S.; Brodt, H.-R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A.M.;Berger,A.;Burguiere,A.-M.;Cinatl,J.;Eickmann,M.;Escriou, N.; Grywna, K.; Kramme, S.; Manuguerra, J.-C.; Muller, S.; Rickerts, V.; Sturmer, M.; Vieth, S.; Klenk, H.-D.; Osterhaus, A. D. M. E.; Schmitz, H.; Doerr, H. W. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* 2003, 348, 1967 1976.

[97] Jin, Z.; Zhao, Y.; Sun, Y.; Zhang, B.; Wang, H.; Wu, Y.; Zhu, Y.; Zhu, C.; Hu, T.; Du, X.; Duan, Y.; Yu, J.; Yang, X.; Yang, X.; Yang, K.; Liu, X.; Guddat, L. W.; Xiao, G.; Zhang, L.; Yang, H.; Rao, Z. Structural Basis for the Inhibition

of SARS-CoV-2 Main Protease by Antineo- plastic Drug Carmofur. *Nat. Struct. Mol. Biol.* 2020, 27, 529532

[98] Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved -Ketoamide Inhibitors. *Science* 2020, 368, 409412.

[99] Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J. Boceprevir, GC- 376, and Calpain Inhibitors II, XII Inhibit SARS-CoV-2 Viral Replication by Targeting the Viral Main Protease. *Cell Res.* 2020, 30, 678692.

[100] Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. R. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemBioChem* 2020, 21, 730738.

[101] Mohammad M. Ghahremanpour, Julian Tirado-Rives, Maya Deshmukh, Joseph A. Ippolito, Chun-Hui Zhang, Israel Cabeza de Vaca, Maria-Elena Liosi, Karen S. Anderson, and William L. Jorgensen *ACS Medicinal Chemistry Letters* 2020 11 (12), 2526-2533

[102] Grosdidier, Aurélien et al. "SwissDock, a protein-small molecule docking web service based on EADock DSS." *Nucleic acids research vol.* 39,Web Server issue (2011): W270-7. doi:10.1093/nar/gkr366

[103] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021 Jan 8;49(D1):D1388-D1395. doi: 10.1093/nar/gkaa971. PMID: 33151290; PMCID: PMC7778930.

[104] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." *J. Comput. Chem.* 25(13):1605-1612 (2004).

[105] Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High- Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* 2013, 29, 845–854.

[106] Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* 2015, 11, 3499– 3509.

[107] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983, 79, 926–935.

[108] Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. Proc. *Natl. Acad. Sci.* 2005, 102, 6665–6670.

[109] Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucleic Acids Res.* 2017, 45, 331–336.

[110] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 1995, 103, 8577–8593.

[111] Hess, B.; Bekker, H.; Berendsen, H. J. C. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* 1997, 18, 1463-1472.

[112] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* 1984, 81, 3684–3690.

7
Vita

VITA

The author was born in Dhaka, Bangladesh. She obtained her Bachelor's degree in Applied Chemistry & Chemical Engineering from University of Dhaka, Bangladesh in 2013. She then moved to USA in 2014 to pursue her Master's degree in Chemistry at Middle Tennessee State University and received her degree in 2016. She joined the University of New Orleans chemistry graduate program to pursue a PhD in computational chemistry,and became a member of Professor Steven Rick's research groups in 2017.