2019-01-01

Computational Study Of Disease Related Proteins Form Electrostatic Point Of View

Chitra Bahadur Karki
University of Texas at El Paso

COMPUTATIONAL STUDY OF DISEASE RELATED PROTEINS FORM

ELECTROSTATIC POINT OF VIEW

CHITRA BAHADUR KARKI

Master's Program in Physics

APPROVED:

Lin Li, Ph.D., Chair

Jainjun Sun, Ph.D.

Jorge A. Munoz, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

COMPUTATIONAL STUDY OF DISEASE RELATED PROTEINS FORM ELECTROSTATIC POINT OF VIEW

by

CHITRA BAHADUR KARKI, M.S.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Physics

THE UNIVERSITY OF TEXAS AT EL PASO

August 2019

Abstract

In the last three decades, many giant DNA viruses have been discovered. The giant viruses present a unique and essential research frontier for the study of self-assembly and regulation of supramolecular assemblies. The question on how these giant DNA viruses assemble thousands of proteins so accurately to form their protein shell, the capsids, remains largely unanswered. Revealing the mechanisms of giant virus assembly will help to discover the mysteries of many self-assembly biology problems. Paramecium Bursaria Chlorella Virus 1 (PBCV-1) is one of the most intensively studied giant viruses. Based on the high-resolution structure of PBCV-1, we implemented a multi-scale approach to investigate the interactions among capsomers, which are giant virus capsid building units. An individual PBCV-1 capsomer consists of three major capsid proteins, each of which has two jelly-roll folds, giving the whole capsomer a pseudo-hexagonal shape. Electrostatic features calculated by DelPhi show that the six vertices of a single PBCV-1 capsomer are always positive while its six grooves are alternatively positive and negative. Based on the electrostatic potential analyses for capsomer-capsomer pairs around the relatively flat area at the icosahedral 2-fold axis, three binding modes with different strengths are found, two of which are attractive while the other one is repulsive. Furthermore, a capsomer structure manipulation tool package is developed to simulate the capsid assembly process and investigate the binding funnels. Results from this tool package shows that 1. long-range electrostatic forces indeed play important roles during and after the capsid assembly; 2. the binding funnels are obvious and can be quantitatively measured. In addition, the total binding free energy of each binding mode was calculated with the MM/PBSA method after molecular dynamic simulations. Interestingly, no matter if the electrostatic binding forces are attractive or repulsive, when counting the Van der Waals, non-polar solvation and other energies, the binding free energies of the three binding modes

are all negative indicating the capsid structure is very stable. Among all the binding modes, the weakest binding mode is located at the boundary of trisymmetron. This explains why the seam between two neighboring trisymmetrons becomes the breaking line when a giant virus capsid dissociates. Formulas generated for the total number of each binding mode within one capsid show the mode within trisymmetron are dominating the stabilization of the capsid which is consistent to previous observation. Results and tools generated in this work shed first light on the assembly of giant virus by providing quantitative analyses. Besides the viral capsid assembly, methods developed in this study pave the way for studying more complicated assembly process for other biomolecular structures.

One quarter of the world's population are infected by Mycobacterium tuberculosis (MTB), which is a bacterial pathogen and is one of the leading causes of death in humans. Recent evidences have demonstrated that two virulence factors, ESAT-6 and CFP-10, play crucial roles in cytosolic translocation. Many efforts have been made to study the ESAT-6/CFP-10 but the mechanism of how ESAT-6 and CFP-10 contribute to MTB cytosolic translocation and virulence is poorly understood. One recent interesting finding is that at low pH, the ESAT-6 with Post Translation Modification (PTM) dissociates from CFP-10 but the non-PTM ESAT-6 doesn't. This work focuses on the ESAT-6/CFP-10's dissociation mechanism. We found at low pH, the ionizable residues in both ESAT-6 and CFP-10 change their charge states significantly. The net charge of the ESAT-6/CFP-10 complex changes from -10e to +1e from pH 7 to 4. Due to the dramatic charge changes, the salt bridges at the interfaces in the complex are broken or weakened significantly, which may result in the increment of the contribution of electrostatic energy to the total binding energy of complex ESAT-6/CFP-20. The binding energy calculations using the MM/PBSA approach confirm that the electrostatic binding energy does increase at pH 4. However, the Van

der Waals (VDW) binding energy decreases at pH 4, which compensates the electrostatic binding energy and thus the total binding energy remains almost the same as in pH 7. Which is in good agreement with the experiments results where the binding affinity of the complex remain almost the same at both the pH conditions. Further investigation on the VDW energy shows that the N-terminal of ESAT-6 play a significant role in strengthening the VDW binding energy and stabilizing the complex structure at low pH. The structure used for this work was derived from E.coli where the PTM is absent but in experiments on native ESAT-6/CFP-10 from MTB bacteria from humans, where PTM is present, the dissociation of complex at low pH values was observed. This gives the hint that the PTM on N-terminal of ESAT-6 may play a role in the dissociation of the ESAT-6/CFP-10 complex.

After the dissociation of the complex ESAT-6/CFP-10 into the constituents, the ESAT-6 undergoes oligomerization process which then rupture the membrane to form a channel for bacterial cytosolic translocation. To simulate the oligomerization, process the M-ZDOCK and SymmDock algorithms were used to predict the structures with different number of symmetries. The predictions form docking algorithms were filtered based on some experimental facts and the final structures were embedded to membrane. The MM/PBSA analysis was implemented to obtain the binding free energies of these structures. No matter which docking algorithm, the oligomer with symmetry four had the least binding free energy per ESAT-6. So, oligomer with symmetry four could be the structure for membrane interaction.

Table of Contents

List of Tables

List of Figures

Chapter 1: The Role of Electrostatic Interactions in Capsid Assembly Mechanisms of Giant Viruses

1.1 INTRODUCTION

Since the early 1990s, many giant DNA viruses that infect eukaryotic cells have been discovered and studied [1]. Examples include Paramecium bursaria Chlorella virus 1 (PBCV-1) [2-7], Chilo iridescent virus (CIV) [3, 8], Phaeocystis pouchetii virus (PpV01) [9], Mimivirus [10-17], Megavirus [18], Cafeteria roenbergensis virus (CroV) [19, 20], Faustovirus [21, 22], Sambavirus [23, 24], as well as non-icosahedral-shaped giant viruses [25, 26]. Cellular genes found in all three kingdoms of life are present in giant virus genomes, reigniting the debate about the tree of life [27-34] as well as the definition of a virus [35, 36]. The size and complexity of the giant viruses have challenged current structural techniques [1] and present a unique research frontier for the study of self-assembly and regulation of supramolecular assemblies at an extremely large scale. Despite their size and complexity, the protein shell (capsid) of icosahedral giant DNA viruses still assemble symmetrically from simple and similar building blocks called capsomer [1]. Each of the donut-shaped capsomer consists of six "jelly-roll" folds [37-41]. The "jelly-roll" fold is a wedge-shaped β-barrel structure [41] that is used by all kinds of DNA or RNA viruses infecting species across all three kingdoms [42], varying in size from small parvoviruses [43], and plant viruses [37] to giant DNA viruses [1]. In giant icosahedral DNA viruses, two jelly-roll folds are linked together in their major capsid protein (MCP) and the capsomer is a trimer of the MCP (Figure 1.1a, b). These trimeric capsomers then pack closely to cover the icosahedral surface of the virus (Figure 1.1c). However, how giant icosahedral DNA viruses assemble their gigantic capsid from the capsomers so accurately is still largely unknown.

In 1969, Wrigley described large triangular and pentagonal arrays of capsomers resulting from broken Sericesthis iridescent virus (SIV) [44]. He defined these as trisymmetrons and pentasymmetrons and derived mathematical formulas for possible symmetron sizes [44]. Later, pentasymmetrons and trisymmetrons could be clearly identified on the capsid surface by the

discontinuous lines between them. This discontinuity occurs because three of the six jelly-rolls within one capsomer are positioned a little higher towards the exterior of the virus with longer surface loops (Figure 1.1a) in many giant viruses, conveying a truly trimeric appearance of the capsomer (Figure 1.1c, d). Within trisymmetrons, the capsomers are packed in the same orientation, but between neighboring trisymmetrons, the orientation of capsomers differs by 60° (Figure 1.1d), thus, generating visible discontinuous lines on the virus surface [1, 3, 5, 44]. Based on all these observations, it was assumed that the giant virus capsomers first assemble into these symmetrons, which then join together to form the capsid. However, electron tomography studies of thin sectioned cells infected by giant viruses have shown that the assembly starts from one 5-fold vertex and seems to be continuous [14, 45]. No large patches of capsomers have been observed inside host cells. A recent close analysis of the capsomer orientations around the icosahedral 5-fold axis among several giant icosahedral giant DNA viruses revealed a spiral pattern within the traditionally defined pentasymmetron [20]. A new spiral assembly hypothesis has been raised that is consistent with the continuous process observed inside the cells [20].

Computational simulations are widely used in viral capsid assembly research. Some of the simulations helped to resolve the capsid structures [46], others were aiming at revealing the mechanisms of the viral capsid assembly [47-51]. Even though the large size and the complexity of the viral capsids make it extremely challenging, many computational works have been done to understand the mechanisms of virus capsid assembly [47]. A significant portion of these simulation works were based on coarse-grained models [48-52], which selectively captured the main information of residues instead of focusing on some atomic details in simulations. Very few atomic level simulations were performed on a whole capsid [53]. Most of them focused on particular regions on the capsid assembly, such as the scaffold protein mediated capsomer-capsomer interactions [54].

Electrostatic interaction has been recognized as an important factor for protein-protein recognition and assembly [55-60]. In previous works, we have successfully calculated and illustrated the electrostatic potential distribution and electric field lines around a whole viral capsid

[61]. In this research, we investigated the possible interactions between capsomers by using PBCV-1 as the model system because the atomic structure of its MCP62 as well as a high resolution cryo-electron microscopically reconstructed map of the intact virus[2] are available. To study the binding mechanisms of capsomers, we used simulation methodology to calculate the interactions between each two capsomers among three chosen ones around the icosahedral 2-fold area of PBCV-1 capsid. Based on the electrostatic potential analyses, three binding modes were found among the three capsomers, two of which are attractive while the other one is repulsive. To mimic the assembly process, we manipulate one capsomer within the capsomer pair of all three binding modes in four different operations and then calculate the binding forces. It is found that the electrostatic interaction is critical to guide the direction and orientation of capsomers during the assembly process. Using the MM/PBSA (molecular mechanics with Poisson-Boltzmann and surface area solvation) method, we also obtained the total binding energies of these three binding modes including Van del Waals, non-polar solvation and other energies. Regardless attractive or repulsive electrostatic interactions, all three binding modes have negative total binding energies indicating that they all contribute to stabilize the capsid. Furthermore, formulas were derived to calculate total numbers of different binding modes within a single capsid. It leads to the conclusion that one of the binding modes that is within the trisymmetron provides the dominating stabilization interactions, which is consistent with experimental observations.

In summary, our results demonstrated that the electrostatic interactions play crucial roles in driving the assembly as well as stabilizing the final capsid after the assembly process of PBCV-1. These results shed the first light on the mechanisms of the giant virus assembly. Furthermore, various tools developed in this work can be widely utilized in studying protein-protein interactions beyond virus capsid assembly such as other self-assembly supramolecular complexes within cells.

1.2 METHODS

1.2.1 Capsid Structure Preparation

The capsomer structure used for this work was downloaded from the PDB bank (PDB ID 5TIP, resolution: 2.0 Å) [62]. In order to study the electrostatic interactions among PBCV-1 capsomers in the assembled virion, a PBCV-1 cryo-EM density map was used (EMDB Accession code 5378, resolution: 9.8 Å) [2]. The pseudo-atomic structure of PBCV-1 capsid was generated by fitting the capsomer structure62 into the high resolution cryo-EM map [2]. To study the electrostatic interactions, three capsomers around the icosahedral 2-fold area were selected (Figure 1.1d), which represent all the possible intra- and inter-trisymmetron interactions. Capsomer A and capsomer B are from the same trisymmetron with the same orientation. Capsomer C is from the neighboring trisymmetron and is 2-fold symmetric with capsomer A. However, because of the C3 symmetry of the capsomer, capsomer C's orientation can be described equivalently as 60°, 120° or 180° rotated to capsomer A. Due to the special orientation of capsomer C, it can be predicted that there are three types of interactions within the selected three capsomers. The binding mode between capsomers A and B is referred as binding mode 1, that between capsomers B and C is referred as binding mode 2, and that between capsomers A and C is referred as binding mode 3 (Figure 1.1e).

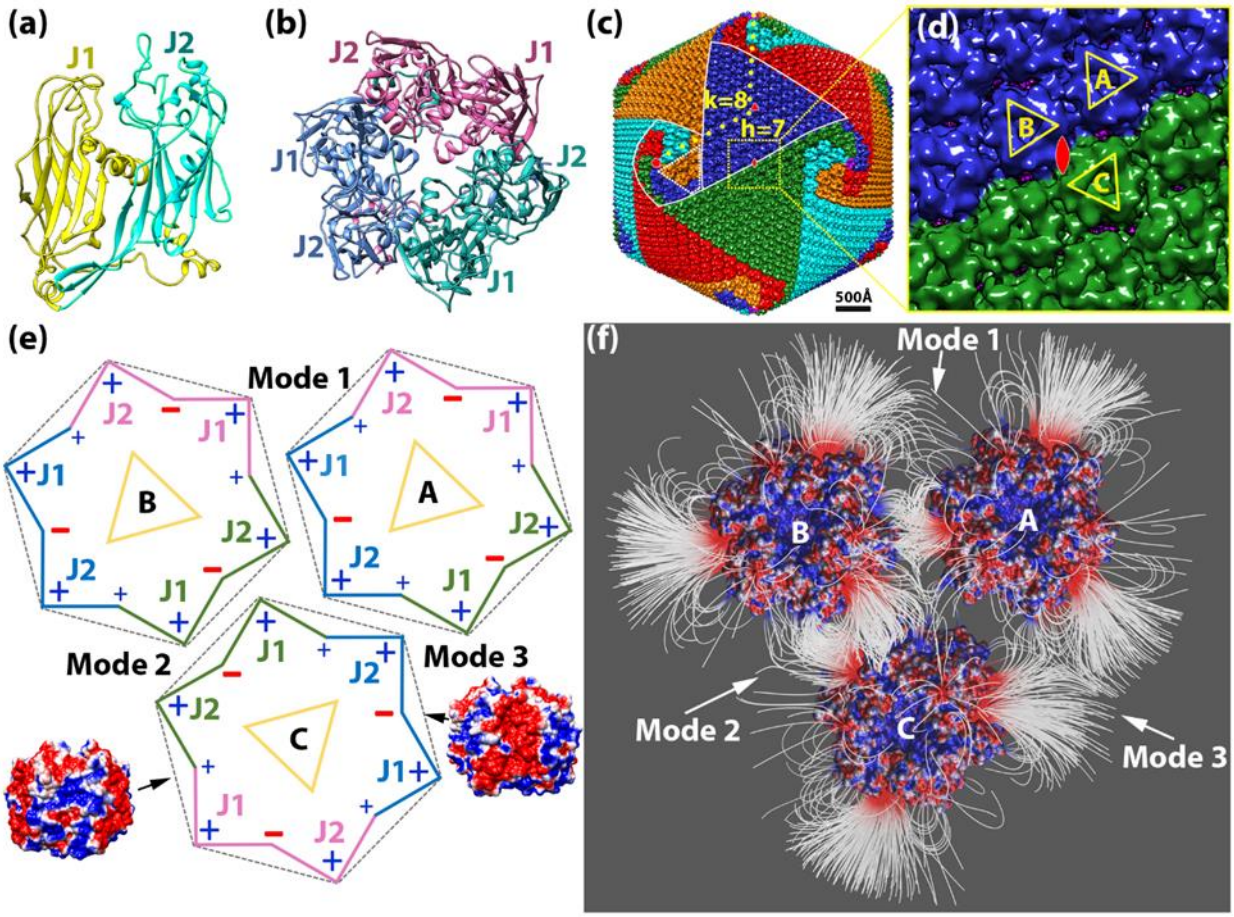

Figure 1.1: PBCV-1 Capsomers Structures and Arrangement. (a) A ribbon diagram of the PBCV-1 Major Capsid Protein (MCP) Vp54 (PDB ID 5TIQ). The two jelly-rolls are colored in yellow (J1) and cyan (J2), respectively. (b) a ribbon diagram of the PBCV-1 capsomer displaying a pseudo-hexagonal shape. Individual double jelly-roll MCPs are color-coded with pink, jade green, and azure; (c) Isosurface rendering of PBCV-1 cryo-EM map (EMD_5378). Capsomers are colored based on their orientation in red, blue, green, cyan, and orange; The boundaries of one trisymmetron and one pentasymmetron are outlined in white. One asymmetric unit within one pentasymmetron is outlined in dashed white lines. The yellow dots present the steps for calculating the Triangulation number with the h and k number labeled in yellow. One set of icosahedral symmetry symbols are labeled in red. (d) The magnified icosahedral 2-fold surface areas (outlined in yellow dashed line in (c)) and three selected capsomers (A, B, C). To show their orientations, these three capsomers are labeled by yellow triangles where the vertices of the triangle point to the three higher jelly-roll surface loops; (e) A schematic diagram of showing the three selected capsomers in (d) for highlighting their orientations, charge distributions and binding modes. The outlines of the capsomer are colored as the same color in (b) with J1 and J2 labeled. Three binding modes are labeled by 1, 2 and 3 closes to the corresponding interfaces. Two side views of PBCV-1 capsomer electrostatic surfaces are showing on the sides of capsomer C pointing to their corresponding surface. The electrostatic surfaces are rendered by Chimera[63] with color scale from -0.50 to -0.50. Negatively and positively charged areas are colored in red and blue respectively. (f) The electrostatic potential field lines between three selected PBCV-1 capsomers rendered by VMD[64]. Negatively and positively charged capsomer surface areas are colored as in (e) using the same parameters.

1.2.2 DelPhi Calculations of Electrostatic Potential

The PQR file of each capsomer was generated by PDB2PQR [65]. The protonation states of titratable functional groups were assumed to be standard, corresponding to pH = 7.0. To obtain the electrostatic potential and electric field at each desired position, the electrostatic potential map (phimap) of a single capsomer was generated by Delphi [66-68]. During DelPhi calculations, the resolution was set as 2 grids/Å. The dielectric constants were set as 2.0 for protein and 80 for water environment, respectively. The perfil was set to be 70. The probe radius for generating molecular surface was 1.4Å. Salt concentration was set as 0.15M. The boundary condition for the Poisson Boltzmann equation was set as dipolar boundary condition. The calculated electrostatic potential on surface was visualized with Chimera[69] (Figure 1.1e). In order to visualize interactions between capsomers, electric field lines were shown in VMD [64] (Figure 1.1f). Color scale range was set be from -1.00 to 1.00 kT/Å. The distance between selected capsomers was increased by 20Å so that the field lines could be illustrated more clearly.

1.2.3 Capsomer Manipulations to Simulate Capsid Assembly

A package of structure manipulation tools has been developed and used to manipulate the capsomer pairs and mimic the assembly process. Within each pair of capsomers, one capsomer was fixed, while the other one was manipulated. For the purpose of this study that focuses on capsomers at relatively flat areas around the icosahedral 2-fold area, four different types of geometric operations of each capsomer pair were carried out (Figure 1.2): (a) shifting away, the manipulated capsomer was shifted away from the fixed one in the direction along the vector of the mass centers of two capsomers (Figure 1.2a) from 5Å to 40Å in 1Å intervals; (c) shifting perpendicular: the manipulated capsomer was shifted along a vector that is perpendicular to the vector connecting the two mass centers from -60Å to 60Å in 5Å intervals after shifted 20Å (the approximate depths of the capsomer grooves) away from the fixed capsomer (Figure 1.2c); (b) spinning, the manipulated capsomer was spun from -30° to 30° in 2° intervals after being shifted

20Å away to avoid atom clashes from fixed capsomer (Figure 1.2b); and (d) rotating around, the manipulated capsomer was shifted 20Å away and then rotated around the fixed capsomer in the xy-plane (Figure 1.2d) from -30° to 30° in 2° intervals. For operations of binding modes 2 and 3, the capsomer C was manipulated while the capsomers A or B was fixed, because capsomers A and B were within the same trisymmetron.

1.2.4 Electrostatic Binding Forces between Capsomers

To investigate the roles of electrostatic interactions in guiding the assembly of the viral capsid, electrostatic binding forces in all three binding modes were calculated using DelphiForce [70, 71]. As mentioned above, the structures of each binding mode were obtained by displacing the manipulated capsomer in various distances and orientations from the fixed capsomer utilizing the four operations (Figure 1.2). For example, for binding mode 1, by shifting the capsomer away from fixed capsomer from 5Å to 40Å in 1Å intervals, 36 structures were obtained. The electrostatic force on the moving capsomer exerted from the fixed capsomer in each structure was calculated by DelPhiForce and compared so that the effects of capsomer distance could be analyzed (Figure 1.2a).Similarly, using the other operations, the effects of both translation and rotation were investigated (Figure 1.2). In DelPhiForce calculations, the resolution, dielectric constant, perfil, ionic strength of the solvent, and the boundary condition were set the same as DelPhi calculation mentioned above in section 2.2. Parallel operations and calculations were carried out with all three binding modes. The electrostatic binding forces calculated by DelphiForce were visualized with VMD and represented by arrows (Figure 1.2).

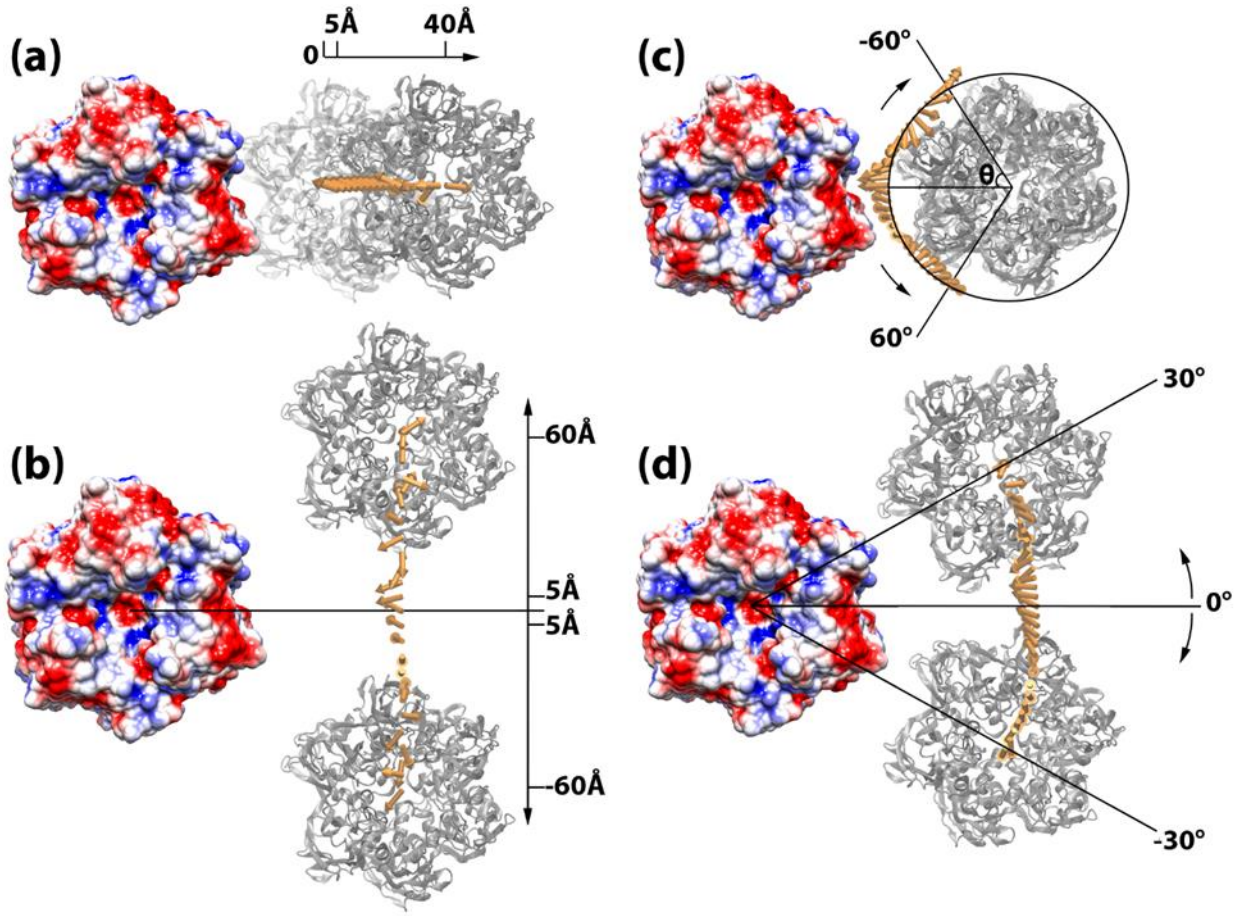

Figure 1.2: Operations of capsomers simulating assembly. In each of the four operations, the capsomer on the left (shown in electrostatic colored surface) is fixed, whereas the capsomer on the right (in grey) is manipulated by (a) shifting away 5Å to 40Å, (b) shifting perpendicular 5Å to 60Å up and down, (c) spinning -60° to 60°, and (d) rotating around -30° to 30°. The binding forces were presented by yellow arrows. The tail of each arrow is located at the mass center of the manipulated capsomer when it was displaced in the corresponding locations. In order to differentiate the binding force from different spinning degree in (c), the force arrows were translated by 40Å onto a circle where the spinning degrees were represented by the angles.

1.2.5 Molecular Dynamic (MD) Simulations for Capsomer Pairs

To predict the dynamic interactions between capsomers in all the three binding modes, MD simulations were carried out using the CHARMM [72] force flield. During each simulation, we set all interfacial residues to be flexible and other residues constrained. An interfacial residue is defined as a residue which has at least one contact with any atoms on the other capsomer. Such constraint saves CPU time significantly. Then we performed a 20ns MD simulation for each

binding mode. In all the simulations, the temperature of the system is set as 300 Kelvin and the pressure is set as standard using the Langevin dynamics. In order to identify the key residues that play important roles in electrostatic interactions, the salt bridges within the interface from each binding mode were analyzed with 4.0Å distance criterion from 4,000 frames. The strongest salt bridges in each binding mode were determined by calculating their formation frequency during MD simulation. For instance, in binding mode1, residue Asp324.B forms two stable salt bridges with Lys107.A and Arg104.A (Figure 1.3a).

1.2.6. Binding Energy Calculations using MM/PBSA

For energy calculations, 2,000 frames were abstracted from the last 10ns trajectory of each MD simulation. The MM/PBSA method was then implemented to calculate the average free binding energy of each binding mode. In this method, ΔGbind is calculated as:

$$\Delta G_{bind} = G_{Complex} - G_{cap1} - G_{cap2} \qquad (1.1)$$

Where $G_{Complex}$ is the total energy of complex of a capsomer pair; G_{cap1} and G_{cap2} are the total energies of the two individual capsomers. The total energy is calculated as:

$$G_{total} = G_{coul} + G_{polar} + G_{vdw} + G_{nonpolar} \qquad (1.2)$$

Where G_{coul} is the Coulombic energy, G_{polar} is the polar part of the electrostatic energy, G_{vdw} is the Van der Waals energy, and $G_{nonpolar}$ is the nonpolar part of the solvation energy. The Coulombic and polar electrostatic energies were calculated by DelPhi. The Van der Waals binding energy was calculated via NAMD[73]. The nonpolar term of solvation energy was calculated via the solvent accessible surface area method:

$$G_{nonpolar} = \gamma \cdot SA + b \qquad (1.3)$$

Where $\gamma = 0.0054\, kcal \cdot mol^{-1} \cdot A^{-2}$, $b = 0.92 kcal \cdot mol^{-1}$, and SA denotes the solvent accessible surface area, which is calculated using NACCESS 2.1.1 [74] program.

1.3 RESULTS AND DISCUSSIONS

1.3.1 Electrostatic Potential for the Capsomers

1.3.1.1 Single Capsomer

The charge distribution on a single capsomer showed an interesting pattern. Calculations from Delphi demonstrated that the net charge of the capsomer is -3. Note that negatively charged residues are predominant in between the two jelly-rolls within each MCP monomer; positively charged residues are mostly located at the interface between the neighboring MCP monomers; and the rest of positively charged residues are located on the outside edge of each jelly-roll (Figure 1.1f). The overall charge distribution is shown in a schematic diagram (Figure 1.1e). Negatively charged residues are dominant in every other groove of the pseudo-hexagon, whereas positively charged residues are most dominant on the vertices. This unique feature of charge distribution of PBCV-1 capsomer indicates that charge interactions on its side may provide strong and elegant forces to cause the biological assembly of PBCV-1 capsomers.

1.3.1.2 Interaction between Capsomers

Within the three capsomers, attractive electrostatic interaction was observed in binding mode 1 between capsomers A and B as well as in binding mode 2 between capsomers B and C. In binding mode 3, capsomer A and C were shown to have repulsive electrostatic interaction as expected since their interfaces are dominated with positively charged residues (Figure 1.1e). Furthermore, the electrostatic interaction in binding mode 2 between capsomers B and C was shown to be stronger with more field lines at the interface than binding mode 1 between capsomers A and B (Figure 1.1f). Almost no field lines are found at the interface of mode 3, which is consistent to the repulsive interaction (Figure 1.1f). Therefore, we referred these three binding modes as intermediate electrostatic attractive binding (mode 1), strong electrostatic attractive binding (mode 2), and electrostatic repulsive binding mode (mode 3).

1.3.1.3 Key Residues Contribute to the Electrostatic Interaction between Capsomers.

4, 2 and 2 salt bridges were found during molecular dynamics (MD) [75] simulations at the interfaces of binding modes 1, 2 and 3, respectively (Table 1.1). Salt bridge frequency analyses showed stable salt bridges are formed between residue Asp324.B and Lys107.A / Arg104.A in binding mode 1 (Figure 1.3a); between Glu97.B and Arg104.C (Figure 1.3c) in binding mode 2, and between Asp375.C and Arg62.A in binding mode 3 (Figure 1.3e). These residues play significant roles in stabilizing the capsomer-capsomer structures in different binding modes. Moreover, for each binding mode the total salt bridge number within each frame was also analyzed in order to determine the salt bridge population (Figure 1.3b, d, f). During MD simulation, binding mode 3 has the lowest average amount of salt bridges (Table 1.1). This result is consistent with the electrostatic potential calculations. However, binding mode 1 seems like maintaining a higher salt bridge population than binding mode 2, even though the electric field lines show that binding mode 2 has stronger binding interaction (Table 1.1 and Figure 1.3). It is noteworthy that we only count those close-range salt bridges in a distance of 4Å. Other charged interaction beyond this distance might still contribute for the electrostatic interactions significantly in binding mode 2, which can be verified by quantitative calculations in the section of binding force and binding energy calculation. On the other hand, this observation also shows that at close contact distance, mode 1 has more stable complimentary charged residue pairs compared to mode 2. It is also notable that a single residue of Asp324.B forms two salt bridges with very high frequency in binding mode 1, the dominant binding mode discussed below.

Table 1.2: Salt Bridges Formed with the 3 Binding Modes during MD Simulation

Binding modes	Salt bridges	Number	Percentage	Average*
Mode 1	Asp324.B-Lys107.A	1,934	48.35%	1.85
	Asp324.B-Arg104.A	3,604	90.10%	
	Asp375.A-Arg252.B	834	20.85%	
	Glu172.A-Arg62.B	865	21.63%	
Mode 2	Glu97.B-Arg104.C	3,970	99.25%	1.57
	Glu99.C-Arg104.B	2,226	55.65%	
Mode 3	Asp375.C-Arg62.D	1,670	41.75%	0.59
	Asp324.E-Lys317.C	684	17.10%	

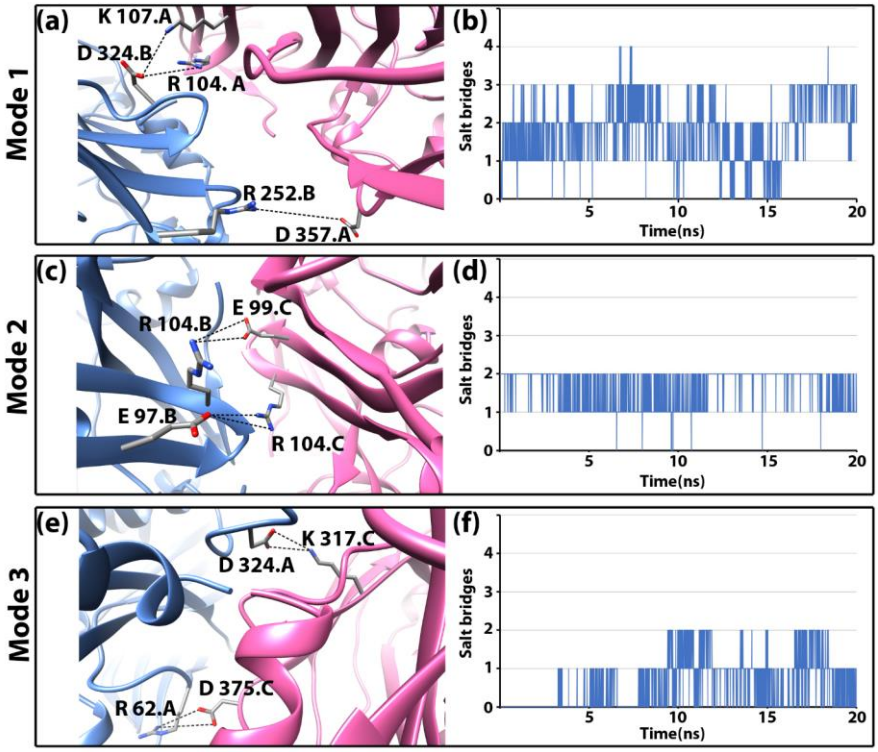

Figure 1.3: Key residues forming salt bridges at capsomer interfaces. The two capsomers are labeled in blue (left) and pink (right) to show the capsomer interface. (a), (c), and (e) show the strongest salt bridge found in modes 1, 2 and 3, respectively. (b), (d), and (f) show the total number of salt bridges counted at each frame during the 20nm simulation for the interface of modes 1, 2 and 3, respectively.

1.3.1.3 Electrostatic binding force between capsomers

To compare the strengths of the binding forces in all the three binding modes, the calculated total forces versus the distance from 5Å to 40Å for each binding mode were plotted (Figure 1.4a), which clearly demonstrated that the binding forces in binding modes 1 and 2 are attractive and in binding mode 3 is repulsive, consistent with the electric field line analyses above (Figure 1.1f). Furthermore, the attractive force in binding mode 2 is stronger, approximately 1.2 to 2.0 times at various distances, than that of binding mode 1. The maximum magnitude of repulsive mode 3 is 1/20 to 1/10 of that of mode 1 indicating its repelling contribution is minor. For binding modes 1 and 2, the forces are attractive up to 35Å, beyond which the directions of the forces become random while the magnitudes become insignificant (Figure 1.4a). This indicates the range of effective forces of modes 1 and 2 is around 35Å, which means that during the viral capsid assembling, capsomers need be brought together within this range when the electrostatic binding forces become effective.

For perpendicular shifting operations from -60Å to 60Å for all three binding modes, global maximum attractive forces for modes 1 and 2 and repulsive force for mode 3 are all located at the near native position (Figure 1.4b). The distribution of the forces shows attractive funnel between -10Å to +35Å, between -5Å to 25Å for mode 1 and 2, respectively (Figure 1.4b1, b2). Overall for mode 3, the repulsive forces form a funnel ranging from -35Å to +25Å and centering at near the native position (Figure 1.4b3).

For spinning operations of binding mode 1, one of the strongest forces exerted toward the fixed capsomer is observed at the native orientation. The range of attractive forces funnel is from -30° to 10° (Figure 1.4c1). For binding mode 2, the native orientation has the strongest binding force with the attractive funnel ranged -40° to 10° (Figure 1.4c2). For binding mode 3, consistent with all our previous observation, most of the binding forces are repulsive with the strongest one at around 30° (Figure 1.4c3).

The binding forces of rotating capsomer B around A in binding mode 1 shows the binding funnel width is 25°, from -10° to 15°. The center of the binding funnel is located at the near native position (Figure 14d1). For binding mode 2, one of the strongest attractive binding force is found at the native position. The range of the attractive funnel is from -5° to 25° (Figure 1.4d2). For binding mode 3, again consistent with all our previous observations, the binding forces remain repulsive mostly with two peaks located at -10° and 5°. The native position is near one of the peaks (Figure 1.4d3).

For each binding mode, binding funnels are investigated in various perspectives. These calculations of the effect of distance and orientations to the binding force show that when the manipulated capsomer is placed at a relatively close distance within 35Å to the fixed one, the electrostatic forces not only attract them together, but also adjust their orientations to achieve optimal binding for the assembly.

Based on the capsomer orientations (Figure 1.1d and 1.1e), within the trisymmetron, individual capsomer assembles using three mode 1 interactions which are all attractive. In contrast, at the boundaries of the trisymmetrons, individual capsomer has one intermediate attractive (mode 1), one strong attractive, and one repulsive interaction. In summary, at either situation, each capsomer should have similar overall binding affinity to its surrounding capsomers, which is further confirmed by the energy calculations discussed below.

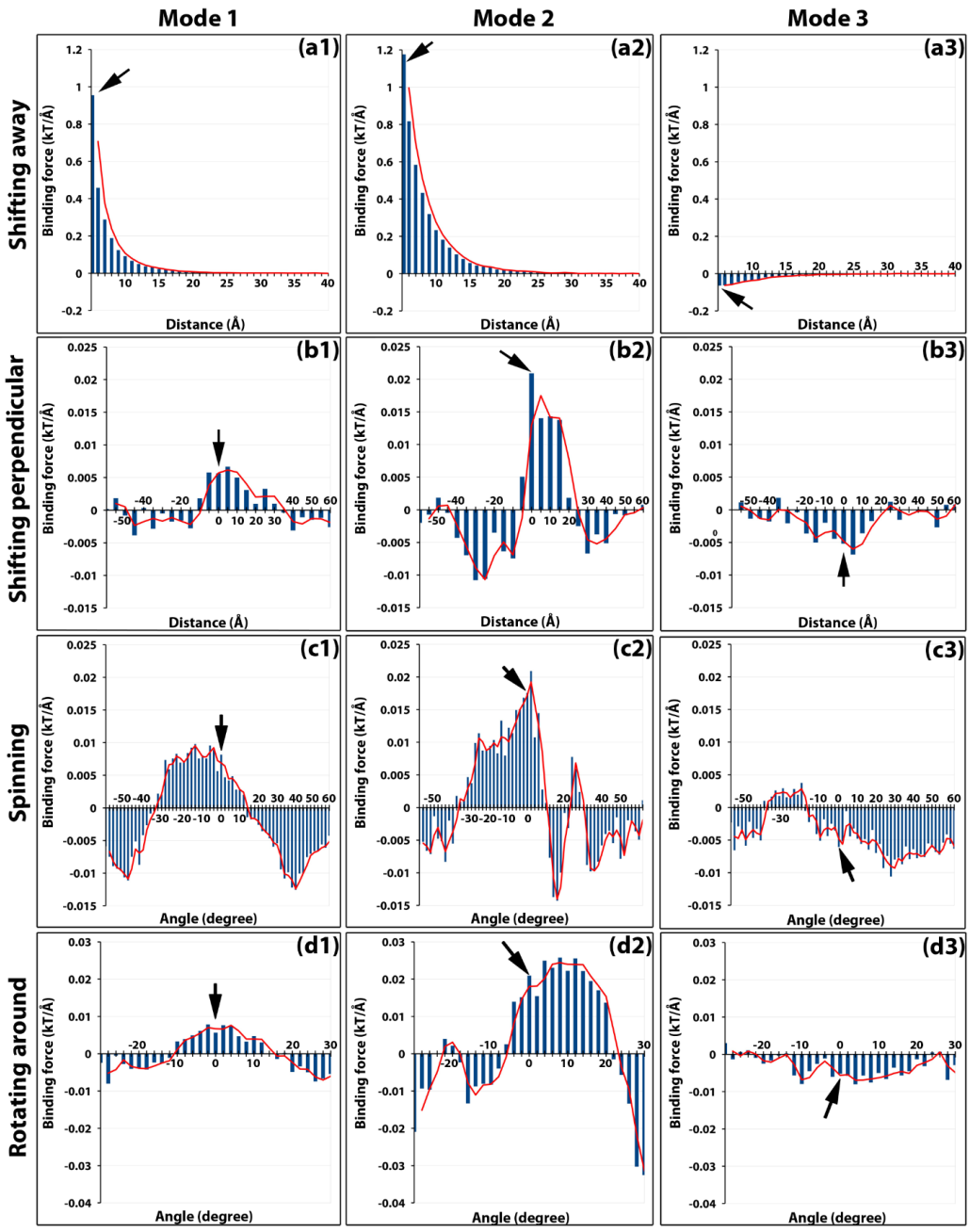

Figure 1.4: Binding forces of modes 1, 2 and 3 scanned by all four operations. The magnitudes of inding forces are presented as vertical histogram bars for each binding mode (panels in each column). Binding forces of each mode are scanned by four different operations (panels in each row). For example, the panels in the first column are binding forces of mode 1 scanned by (a1) shifting away 5Å to 40Å, (b1) shifting perpendicular 5Å to 60Å up and down, (c1) spinning -60° to 60°, and (d1) rotating around -30° to 30°. Positive forces indicate the attractive force while the negative forces are the repulsive forces. Black arrows on each panel point to the native position.

Table 1.2: Binding Energy for three Binding Modes

Bind-ing Mode	Polar Solvation Energy (kcal/mol)		Non-Polar Solvation Energy (kcal/mol)		Van der Waals Energy (kcal/mol)		Coulombic Energy (kcal/mol)		Total Binding Energy (kcal/mol)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	99.41	19.46	-14.58	0.73	-80.80	6.50	-64.89	25.22	-60.85	8.05
2	208.20	28.27	-19.17	1.05	-119.74	10.25	-167.72	32.72	-98.43	9.83
3	8.42	16.23	-9.67	0.71	-46.25	5.59	23.69	16.67	-23.81	5.65

1.3.1.4 Binding Energy Calculation Results from MM/PBSA Analysis

The binding free energy was calculated to be -60.85 kcal/mol for binding mode 1, -98.43 kcal/mol for binding mode 2, and -23.81 kcal/mol for binding mode 3 (Table 1.2). Overall, all three binding modes were shown to have attractive binding and would contribute to stabilize the viral capsid during and after its assembly. The Coulombic components are consistent to the electrostatic forces discussed above that binding mode 2 is strongly attractive (-167.72 kcal/mol), binding mode 1 is weakly attractive (-64.89 kcal/mol), and binding mode 3 is repulsive (23.69 kcal/mol). In the electrostatic repulsive binding mode 3, even though the charged residues in the interface are repelling each other, the total binding free energy is still negative mainly due to the contribution for the non-polar part of the solvation energy and Van der Waals energy. Binding mode 2 was shown to have the strongest binding. However, at the boundary of trisymmetrons, the combined energy of binding mode 2 with the weak binding mode 3 is similar to the summation of two binding energies from mode 1. Therefore, the assembly using binding mode 2 and 3 is possible at the boundary of trisymmetrons. After assembly, these two binding modes do contribute to the stabilization of the capsid. However, when disassembling, because mode 3 is the weakest mode with repulsive electrostatic interaction, those binding mode 3 interfaces along the boundary of trisymmetrons should be detached first. This further indicates that the boundaries of the

trisymmetrons should be the breaking lines, which explains the observations of previous experiments [44].

1.3.1.5 T-number and Total Contribution of each Mode in the Virion of Giant Virus

Caspar and Klug developed the mathematical formula, the Triangulation number (T number), to describe the assembly geometry of icosahedral viruses [76]. The T-number is a measure of how many monomeric triangle areas (e.g., jelly-roll domains) exist in one asymmetric unit of an icosahedron. Virus surfaces are tiled by triangular areas, most of which form hexameric shapes such as the double jelly-roll capsomer on the giant virus surface (Figure 1.1b, c). At the 5-fold axis of the icosahedral virus, there is a pentameric capsomer that only have five triangles (jelly-rolls). By "stepping" along capsomers from one 5-fold capsomer to the nearest one (Figure 1.1c) in the hexagonal array, the shortest route is to take h steps along one hexagonal axis (h), and k steps along the other hexagonal axis (k). The T-number can then be calculated using the equation: $T=h2 +h\cdot k +k2$. For example, the T-number of PBCV-1 is 169 (h=7, k=8; T=72 + 7 x 8 + 82) (Figure 1.1c). It is noteworthy that all giant viruses have their h number equal to 7 and have the same size pentasymmetrons with 31 capsomers: 1 pentameric capsomer located at the 5-fold axis and 30 (5 x 6) pseudohexameric capsomers [1]. Each asymmetric unit (triangular area encircled by white dotted lines in Figure 1.1c) inside the pentasymmetron has six capsomers in three layers radiating from the 5-fold axis, one in the first layer (most close to the 5-fold axis), two in the second layer, and three in the third layer. Therefore, the h number includes three steps in the pentasymmetron, and three steps in the trisymmetron and one more step to align with the pentameric capsomers (Figure 1.1c). The size difference among giant viruses only appears on their various sizes of trisymmetron1. The size of trisymmetron is linked to the k in the T-number. If the equilateral trisymmetron has n capsomers on its edge, due to the same three layers in a pentasymmetron asymmetric unit, n will be k plus 3 (Figure 1.1 and Figure 1.5).

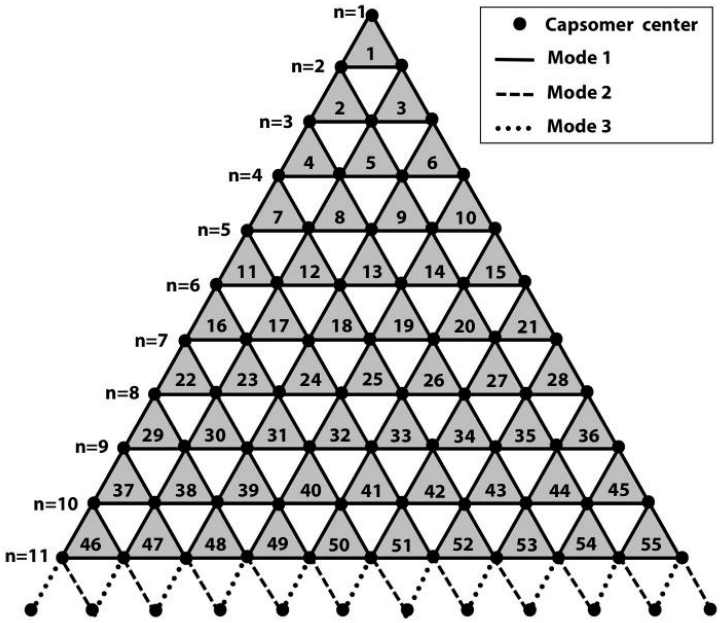

Figure 1.5: Schematic demonstration of the trisymmetron for calculating the population of all 3 binding modes. Each black dot represents one capsomer. Different types of lines represent different binding mode between two capsomers (dots). Within the trisymmetron, n (for PBCV-1, n=11) rows of capsomers were connected by binding mode 1 (solid line). At the bottom edge, extra row of dots (capsomers) from neighboring trisymmetron are shown connected by dashed line and dotted lines, representing binding modes 2 and 3, respectively. The capsomer numbers along one edge are labeled by their number of n.

Capsomers inside a trisymmetron with the same orientation interact with each other in binding mode 1, while capsomers at the boundary of trisymmetrons interact with each other in binding mode 2 or 3. Using PBCV-1 trisymmetron as an example (Figure 1.5), we can derive the formulas for the total number N of each mode of interactions within one virion.

For each trisymmetron, the number of shaded triangles in Figure 1.5 is n · (n-1)/2. For example, in PBCV-1, this number is 55 (11 x 10 / 2). Each of these shaded triangles are formed by three capsomers that are linked by three mode 1 bindings (solid lines in Figure 1.5), so the total number of mode 1 interactions for each trisymmetron Nmode1 will follow the formulae below.

$$N_{mode1} = n \cdot (n-1) \cdot \frac{3}{2} \qquad (1.4)$$

For binding modes 2 and 3 at each edge of trisymmetron, the number of their interaction per edge is n.

$$N_{mode2/3} = n \qquad (1.5)$$

Because each icosahedron giant virus has 20 trisymmetron, 30 trisymmetron edges, and as aforementioned, n = k + 3, formulas of total number of each mode for one intact virion can be derived as:

$$N_{mode1} = n \cdot (n-1) \cdot 3 \cdot 20/2 = (k+3) \cdot (k+2) \cdot 30 \quad (1.6)$$

$$N_{mode2/3} = n \cdot 30 = (k+3) \cdot 30 \quad (1.7)$$

Based on these formulas, the total numbers of binding modes 1, 2 and 3 in PBCV-1 are 12,600, 540, and 540, respectively. Using binding energy calculated in Table 1.2, we can estimate the total binding energies for three binding modes are -766,710 kcal/mol, 53,152 kcal/mol, and 12,857 kcal/mol, respectively. Clearly, binding mode 1 is the dominant interaction that stabilize the final assembled virion.

1.4 CONCLUSIONS

In this work, we investigated the capsomer-capsomer interactions at the icosahedral 2-fold region of PBCV-1 capsid. We demonstrated that due to the geometrical symmetry of a capsomer and charge distribution, there are three binding modes: binding mode 1, the dominant binding mode in the whole capsid, reside inside the trisymmetron where they are packed in the same orientation; binding modes 2 and 3 are formed at the boundary between two trisymmetrons where the capsomers have different orientations. The electrostatic interactions in all the binding modes are quite different. For the binding mode 1, the electrostatic interaction is attractive. However, the binding modes 2 and 3, which are at the boundary of two trisymmetrons, result in attractive and repulsive electrostatic interactions, respectively. The salt bridge analyses demonstrate that several key residues from the binding interfaces play significant roles for the capsomer interactions by forming stable inter-capsomer salt bridges. Simulations also show binding mode 1 forms highest average amount of stable salt bridges (Table 1.1 and Figure 1.3). For instance, Asp324.B forms two salt bridges with very high frequency. In order to demonstrate the importance of these residues for capsomer intermolecular interactions, these residues can be mutated and studied by simulations

and experiments in the future. Combining with DelPhiForce, the net forces between two capsomers in each binding mode are studied and it is shown that in binding modes 1 and 2, the electrostatic attractive binding forces are responsible for capsomer recognition. Binding funnels are found at the binding interfaces when manipulating capsomer pair in four different ways. Such binding funnels guide and rotate two capsomers towards the final assembled positions. The results demonstrate that the electrostatic forces play vital roles in viral capsid assembly by guiding the capsomers to interact with each other at the favored distance and orientation to maintain stable assembling.

Overall, when considering the total binding energies, all three binding modes were shown to have attractive binding interactions, among which the binding mode 2 is the strongest (Table 2). Even in electrostatic repulsive binding mode 3 (Figure 1.1e and 1.1f), the total binding energy is still negative mainly due to the contribution of the non-polar part of the solvation and Van der Waals energies (Table 1.2). Although binding mode 1 has intermediate binding energy, unlike binding modes 2 and 3 that only exist at the boundary between symmetrons, its total numbers, and more stable salt bridges (Table 1.1 and Figure 1.3), make it the dominating binding mode that stabilizes the capsid during and after the assembly. These results explain the phenomenon observed by Wrigley in 1969 where large icosahedral dsDNA viruses could be decomposed into trisymmetrons [44]. Our studies illuminated the elegant mechanisms of giant viral capsid assembly and disassembly from the aspects of energy and force.

In this study we focused on the dominant binding modes which are intra- and inter-trisymmetrons. However, the interactions among capsomers within the pentasymmetrons are also important because they are essential for initializing the assembly of the viral capsid. Structural manipulation tools with four operations have been developed to explore the electrostatic binding forces between two capsomers in each binding mode (Figure 1.2), which pave the way for future investigation of capsomer interaction among pentasymmetron of giant viruses. The spiral pattern of capsomer orientations and the protruding shape of the 5-fold vertices of the pentasymmetron

require more complicated calculations and more detailed examinations, which will be our future work.

Chapter 2: Stability of the key Protein Complex of Mycobacterium Tuberculosis at different pH: Van der Waals Energy Compensates the Electrostatic Energy

2.1 INTRODUCTION:

Tuberculosis (TB) is one of the oldest and the leading causes of death, globally. About one quarter of the world's population, ~1.7 billion people, are infected with latent TB and are at high risk of developing active TB. In 2017, worldwide, about 10 million people got sick with TB and 1.3 million deaths were reported [77]. The transition of latent to active TB is attributed to the maturation of the Mycobacterium tuberculosis (MTB) bacteria inside the phagosomal compartment followed by the secretion of the complex ESAT-6/CFP-10 (6-kDa early secreted antigenic target/10-kDa culture filtrate protein) from the ESX-1 locus of the MTB bacteria in a co-dependent manner [78] and forms a 1:1 heterodimeric complex, as in Figure (2.1). The interface of binding of the complex is essentially hydrophobic and the complex is stabilized by the favorable Van der Waals forces and the two salt bridges [79]. With the fact, pH changes are known to be key factor inside the phagosome [80], there are contradictory data reported in the literatures concerning the dissociation of the complex at acidic pH. The experimental results from de Jonge et.al, using the native protein form MTB bactera indicates the disunion of the ESAT-6/CFP-10 complex between pH 4 and pH 5 [80]. In contrast, at acidic pH the complex purified from the E. coli didn't exhibit the aggression and involve in membrane interaction, suggesting the constituents ESAT-6 and CFP-10 of the complex were not dissociated [81]. Further Lightbody et. al demonstrated through the circular dichroism study that the complex of ESAT-6/CFP-10 from the E. coli was too stable to be dissociated at low pH [82]. The native proteins derived for the MTB bacteria may comprise of distinctive features which are absent in the E. coli derived proteins, is the possible

explanation for the discrepancy [83]. These features might include post-translational modification, such as N-α-acetylation of THR-2 in ESAT-6 [83, 84].

To study the discrepancy mentioned above, the binding free energies at pH 4 and pH 7 were calculated for the complex of ESAT-6 and CFP-10. In order to prepare the complex in pH 4 the titration curve of the ionizable residues were obtained from DelPhipKa webserver [85]. The respective protonation state of ionizable residues were assigned with VMD (Visual Molecular Dynamics) [64]. Molecular dynamics (MD) [75] was performed for the complex at both pH values and Molecular Mechanics Possion-Boltzmann Surface Area (MM/PBSA) [86] method was adopted to obtain the binding free energies between ESAT-6 and CFP-10. The salt bridges analysis at the binding interface of the ESAT-6 and CFP-10 at both the pH values were performed. For pH 7 certain salt bridges were consistent with the previous results [79] and additional salt bridges were also obtained. Moreover, the salt bridges energies were also evaluated for the key salt bridges at both pHs. The electrostatic potential distribution over the surface and the electric field lines of the complex at both the pH were compared.

From the different aforementioned analysis, the direct evidence of dissociation of the complex at acidic pH was not obtained. The binding free energy in pH 4 become less favorable but the magnitude of difference is trivial. The key salt bridges at pH 7 were vanished at pH 4, also the frequencies and the salt bridges energy for salt bridges at pH 4 are insignificant. But the Van der Waals interaction became a bit stronger at pH 4, which is a key player in stabilizing the complex of ESAT-6/CFP-10.

2.2 METHODS

2.2.1 Prepare Structures at different pH Values

The structure of the complex ESAT-6/CFP-10 was obtained from Protein Data Bank with PDB ID 1WA8 [79], which is shown in Figure 2.1. Among all the 28 NMR models, the first model was used for this work. To study the effects of the pH values of the association of ESAT-6/CFP-10, we need to obtain the charge states for all the ionizable residues (GLU, ASP, HIS, LYS, ARG), which were calculated by the DelPhiPka web server [85] at pH values of 4 and 7. The titration curve obtained from the DelPhiPka web server gives the probability of ionizable residues to be in ionized +1e or -1e or neutral state at different pH. In this current work, to study the behavior of the complex at pH 7 and pH 4, the probability of ionizable residues at pH 4 are being compared relative to pH 7.

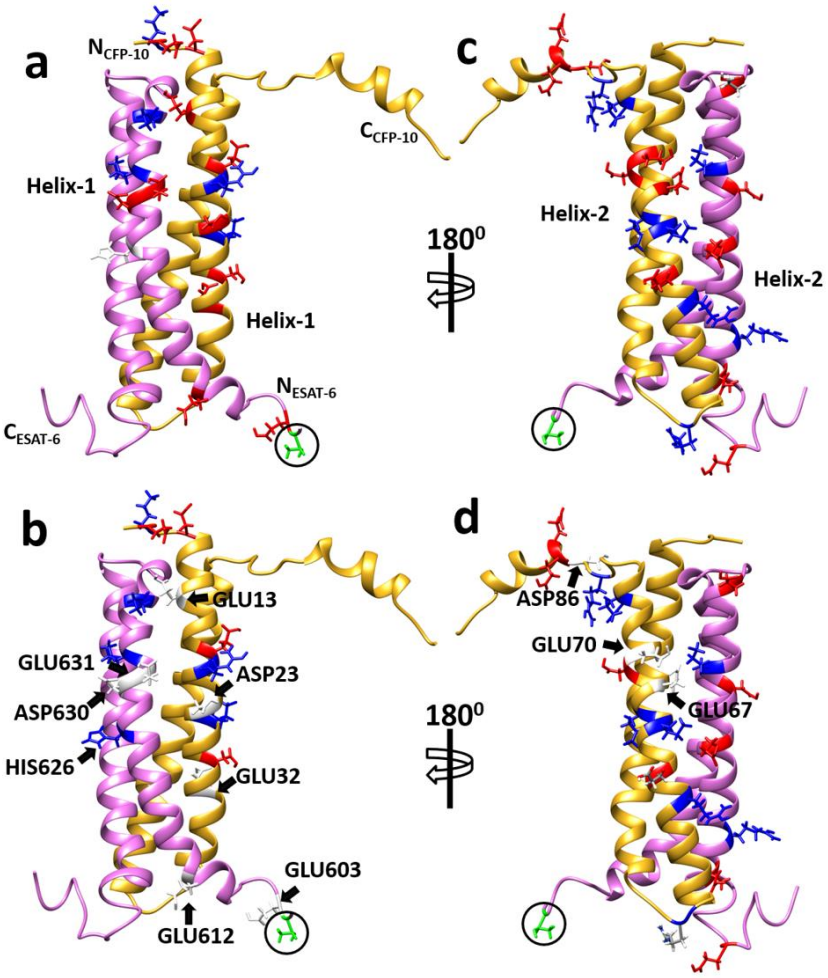

Figure 2.1: Structure of the complex ESAT-6/CFP-10, ionizable residues at pH 7 and pH 4 colored according to the ionization states of residues at different pHs. CFP-10 is colored gold and ESAT-6 as purple. Negative residues are colored in red, positive residues are blue and neutral residues are white. The residues whose ionization states changed at pH 4 are pointed with arrows. (a) Ionizable residues on the helix-1 of CFP-10 and ESAT-6 at pH 7. (b) The residues whose ionization state are changed are pointed with arrow and indicated by color change, at pH 4 for helix-1 of CFP-10 and ESAT-6. (c) Ionizable residues on the helix-2 of CFP-10 and ESAT-6 at pH 7, 180° anti-clock rotation of (a). (d) The residues whose ionization state are changed are pointed with arrows and indicated by color change at pH 4 for helix-2 of CFP-10 and ESAT-6. (a) and (c) are at pH 7 whereas (b) and (d) are at pH 4. The residue circled and highlighted in green is THR 602, which is predicted to have N-terminal acetylation for the complex derived from MTB [81, 83, 87].

2.2.2 MD Simulation

The residues whose ionization states were changed from pH 7 to pH 4 were patched during the conversion to protein structure file using Visual Molecular Dynamics (VMD). The information

about the respective patches are available in the Charmm topology file. After the proper patches and renaming of the residues to match the protonation states at pH 4, the complex was then solvated in water box with TIP3 [88] water model and ionized with NaCl with concentration 0.15 mol per liter with the help of solvate and autoionize plugins in VMD. The distance between the protein atoms and the wall of the solvation box was set to 20 Å. While preparing the system at pH 7, the protonation state of ionizable residues were kept standard. The simulations were performed for the complex at pH 7 and pH 4 with the molecular dynamics (MD) simulation program NAMD 2.12 [73]. Each simulation was performed for 20 ns with force field Charmm. 20,000 steps of energy minimization were carried out before the 20 ns simulation to obtain the local minimum. The simulations were performed using periodic boundary condition with the canonical (NPT) ensemble. The temperature and pressure were kept standard and constant, controlled by Langevin dynamics and Langevin Poston respectively. Van der Waals and electrostatic interactions were truncated at 12 Å. A switching function for Van der Waals at 10 Å was used whereas long-range electrostatic interaction was calculated using particle-mesh Ewald method with 1 grid point per Angstrom. Integration time step for both simulations was set to 1 fs. In every 5000 interval of steps a frame was saved producing 4000 frames for 20 million steps. Afterwards, these frames were employed for the different perspective of analysis.

2.2.3 MM/PBSA Analysis

The binding free energy was calculated by employing the MM/PBSA approach. For the binding free energy calculations, 200 frames were taken from the final 10 ns of each trajectory, at the interval of 0.05 ns. All the ions and water molecules were removed before the MM/PBSA analysis. All the energy terms were averaged over the 200 frames and respective standard deviations were calculated, as shown in Table 2. The total binding free energy is given as,

$$E = E_{coul} + E_{pol} + E_{vdw} + E_{np} \quad (2.1)$$

Where E_{coul} is Coulombic energy; E_{pol} is polar part of solvation energy; E_{vdw} is the Van der Waals energy; E_{np} is the non-polar part of solvation energy.

Both Coulombic Energy and polar solvation energy were obtained from Delphi [61, 66-68]. PDB2PQR [65] software package was used to assign the atomic charges and radii. For the complex structures at pH 7, the protonation states of titratable functional groups were assumed to be standard. For the complex structures at pH 4, the protonation states were modified corresponding to pH 4. During DelPhi calculations, the DelPhi calculation box was set to allow 70 percentage filled by protein (perfil). The dielectric constants were set as 2.0 for protein and 80 for water environment, respectively. The probe radius for generating molecular surface was 1.4 Å. Salt concentration was set as 0.15 M. The resolution was set as 2 grids/Å. The boundary condition for the Poisson Boltzmann equation was set as dipolar boundary condition.

Van der Waals energy (E_{vwd}) was calculated from NAMD 2.12 subjecting the structures to one step of equilibration with temperature 300 K implementing the implicit solvent model. E_{np} was obtained using solvent accessible surface area (SASA) method where SASA was obtained by NACCESS 2.1.1[74] and further E_{np} is given as

$$E_{np} = \alpha \cdot SASA + \beta \quad (2.2)$$

Where, $\alpha = 0.0054 \; and \; \beta = 0.92 \; kcal/mol$

The binding free energy for the complex of ESAT-6 and CFP-10 is calculated as

$$\Delta E_{binding} = E_{Complex} - E_{ESAT-6} - E_{CFP-10} \quad (2.3)$$

Where $E_{Complex}$ is the energy of the complex ESAT-6/CFP-10, E_{ESAT-6} and E_{CFP-10} are energies of ESAT-6 and CFP-10, respectively.

2.2.4 Salt Bridges and Salt Bridges Energy

The salt bridges formed by the interfacial residues of ESAT-6 and CFP-10 play crucial role in the stabilization of the complex. The salt bridge analysis was performed with VMD defining the cutoff of 3.2 Å between the Oxygen and Nitrogen atoms of the oppositely charged ionic residues. The frequency of occurrence of the salt bridges through the binding interface of the complex ESAT-6 and CFP-10 at both pHs were calculated based on the entire trajectory i.e. 20 ns, as shown in Table 3. Further the formation of salt bridges at each frame were also calculated to observe the more than one simultaneous salt bridges. The salt bridge energies were calculated from the 200 frames of last 10 ns trajectory. The residues pairs and the respective individual residues involved in the salt bridges interactions were extracted from these frames and salt bridges' energies were calculated using NAMD 2.12 subjecting to one step of equilibration at temperature 300 K. Only the non-bonded energy terms i.e. Van der Waals and Coulombic energies were considered for salt bridge energy calculation. The salt bridge energy, ΔE_{saltbr} for each pair of involved residues are calculated as:

$$\Delta E_{saltbr} = E_{res1.res2} - E_{res1} - E_{res2} \qquad (2.4)$$

Where $E_{res1.res2}$ is energy of the residue pair involved in salt bridges. E_{res1} and E_{res2} are the individual energies for the residues taking part in the salt bridge formation.

2.2.5 Potential Distribution over the Surface and Electric Field Lines

To characterize the effects of pH, on ESAT-6/CFP-10 complex, the surface map of electrostatic potential (phimap) of ESAT-6, CFP-10 and their complex at pH 4 and pH 7 were generated by Delphi, (Figure 2.5). The parameters used for the electrostatic potential calculations were same as in DelPhi calculation mentioned above. For the visualization of the electrostatic

interaction between the residues formed salt bridges, the distance between ESAT-6 and CFP-10 was increased by 20 Å during the electrostatic potential calculation of the complex. VMD was used to visualize the calculated electrostatic potential on surfaces for complexes at both conditions and VMD was used to illustrate electric field lines between salt bridges (Figure 2.5). Color scale range was set to be from -0.5 to 0.5 kT/Å.

2.3 RESULTS

2.3.1 Charge Distribution at different pH Values

The charge distributions at pH 4 and pH 7 are significantly different. At pH 7, the net charges for Esat-6 and CFP-10 are both -5e. However, when pH decreases to 4, the net charges of ESAT-6 and CFP-10 become neutral and +1e, respectively. Such significant net charge change comes from 11 ionizable residues, among which 5 residues are in ESAT-6 and the other 6 residues are in CFP-10. 6 out of 11 residues are at the interfaces of the complex. Such significant charge changes at pH 4, especially the charge changes at the interfaces, should result in quite different electrostatic binding forces between the ESAT-6 and CFP-10 proteins, which may break the stable structure of ESAT-6/CFP-10. Therefore, the salt-bridges at the interface of ESAT-6/CFP-10 were studied to analyze the electrostatic features in detail. The positions of the 11 residues and electrostatic charge surfaces are shown in Figure 2.6. The electrostatic potential on the surfaces show that the net charge changes do result in serious electrostatic potential changes, especially at the binding interfaces of the complex. The complementarity of the interfacial electrostatic potential is obviously decreased at pH 4, which may result in significant changes of interfacial salt-bridges.

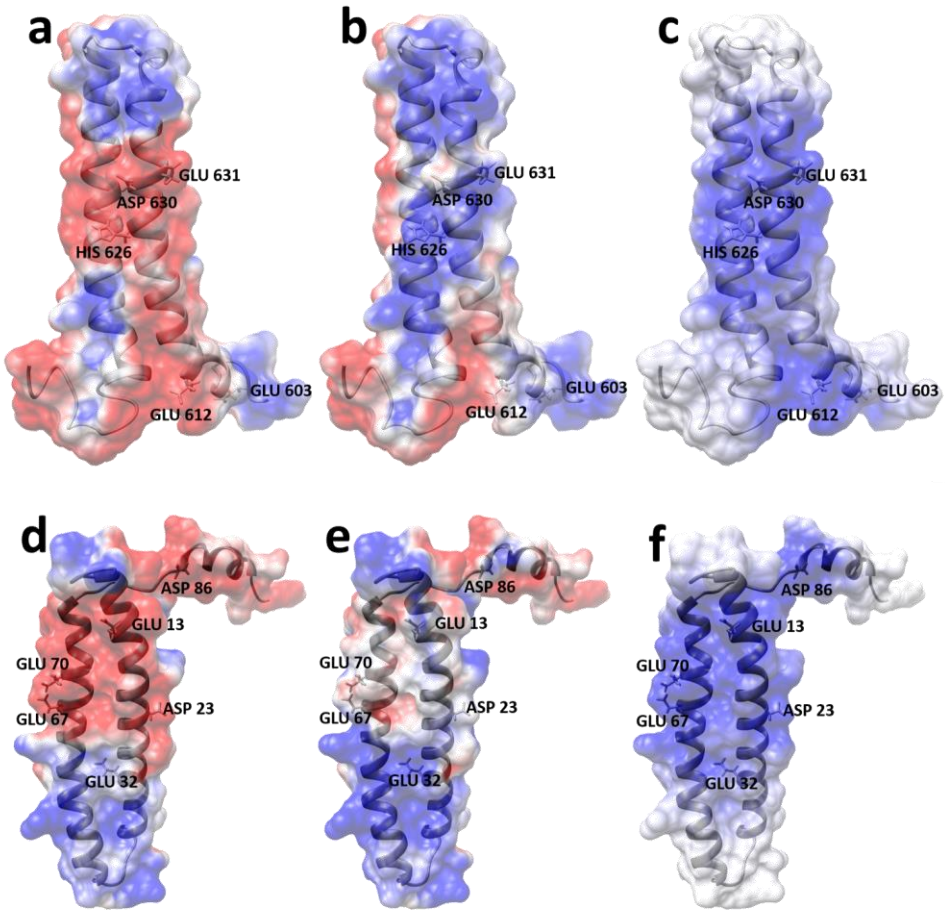

Figure 2.2: Electrostatic surface of ESAT-6 and CFP-10. The 11 ionizable residues which get charge changes are labeled. The panels show the electrostatic surfaces for: a. ESAT-6 at pH 7; b. ESAT-6 at pH 4; difference between pH 7 and 4 of ESAT-6; CFP-10 ant pH 7; CFP-10 ant pH 4; difference between pH 7 and 4 of CFP-10. Blue and red colors represent positive and negative colors. The thresholds for the colors are -3 and +3 kT/e in Chimera [63, 69].

2.3.2 Salt Bridges are Weakened at pH 4

To study the salt bridges at the interfaces of ESAT-6/CFP-10, 20 ns MD simulatinos are perfomed at both pHs. From each trajectory, 4,000 frames were abstracted and analyzed for number of interfacial salt-bridges. The simultaneous interfacial salt bridges formation in each frame for both pH values were analyzed (Figure 2.3). At pH 7, 357 frames contain no salt bridges, 1,270 frames contain 1, 1421 frames contain 2, 804 frames contain 3 and 148 frames contain 4 simultaneous salt bridges. One snapshot of the simulation with 4 simultaneous salt briges, which

has the shortest average salt bridge distance, is illustrated in Figure 2.4a and 2.4b. In case of pH 4, salt bridges were only observed in 691 frames, in which only one frame has 2 salt bridges while all the others contain only one salt brige. The frame at which 2 simultaneous salt birdges are formed is shown in Figure 2.4c, d. Compared with pH 7, it's obvious that the population and frequency of salt bridges at pH 4 are decreased significantly. Such a decrease of interfacial salt brige number may result in an increase of electrostatic energy, which may weaken the binding affinity.

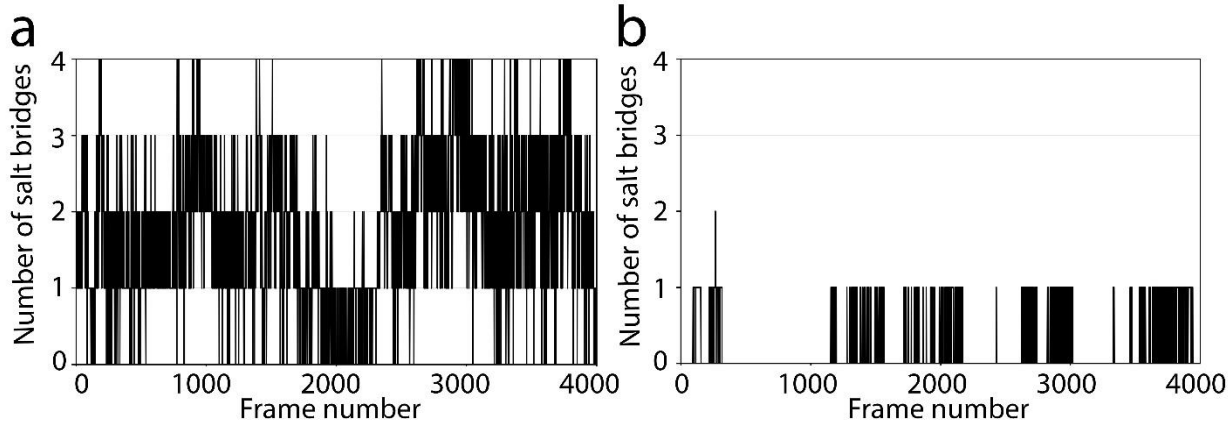

Figure 2.3: Populaion of key salt bridges. Number of simultaneous salt bridges formed in each frame at (a) pH 7 and (b) pH 4.

The formation of interfacial salt bridges of the ESAT-6/CFP-10 complex for 4000 frames were recorded at both pHs. Then the number of salt bridges were averaged over the total frames to obtain the salt bridge frequency (salt bridges per frame). The salt bridge energies of the respective salt bridges were calculated as descried in method section. The number of interfacial salt bridges among the 4000 frames, interfacial salt bridges frequency and average interfacial salt bridge energies from last 10 ns MD simulations are shown in Table 3.3.

Table 2.1: Residue pairs in salt bridges interactions, their occurrence in 4000 frames, frequency and averaged salt bridge energies over last 10 ns at both the pHs. Salt bridges with frequency less than 0.05 are not tabulated.

pH	Residue pairs	No. of frames	Salt bridges frequency	Average salt bridge energy (kcal/mol)
7	GLU13 – LYS638	1620	0.40	-1.52
	GLU67 – LYS657	1718	0.43	-2.46
	GLU70 – LYS657	1342	0.34	-2.18
	LYS63 – GLU664	2437	0.61	-2.17
4	GLU02 – LYS638	207	0.05	-0.08
	GLU70 – LYS657	486	0.12	-0.06

Figure 2.4 shows the positions of interfacial salt bridges at pH 7 and pH4. Four strong salt bridges were observed at pH 7. However, three out of the four salt bridges, GLU13–LYS638, GLU67–LYS657 and LYS63–GLU664, are vanished at pH 4. The salt bridge between GLU13 and LYS638 is vanished at pH 4 (due to the neutralization of GLU13, Figure 2.1), but a new salt bridge, GLU02–LYS638 is formed. As the GLU02 is located at the flexible N-terminal loop of CFP-10 (Figure 2.4d), the salt bridge GLU02–LYS638 is not very stable, which results in a salt bridge frequency of 0.05 (Table 2.3). The salt bridge GLU70–LYS657 is present at both the pHs but with significant reduction in salt bridge per frame and salt bridge energy at pH 4. Moreover, the residue LYS657 is involved in two salt bridges at pH 7 (Figure 2.4a) whereas only one in pH 4 (Figure 2.4c). It is clear that the salt bridges frequencies have significantly decreased while the salt bridge energies increased at pH 4, compared to pH 7. Figure 2.5 illustrates the electric field lines at the interface of ESAT-6 and CFP-10. At pH 7, there are clearly field lines connecting the interfacial salt bridges. At pH 4, the overall filed lines are less dense. The field lines between the three salt bridges, GLU13–LYS638, GLU67–LYS657 and LYS63–GLU664, are disappeared, which should result in dramatic electrostatic energy increase and significantly weaken the binding affinity. However, experimental evidences proved that the association rate at pH 4 and pH 7 are

almost the same without acetylation of ESAT-6. Thus, there must be some other factors which enhance the binding affinity at pH 4 to compensate the change of electrostatic binding energies. To investigate all the factors which may affect the binding affinities, MM/PBSA approach was implemented to perform the binding energy calculations at pH 7 and 4.

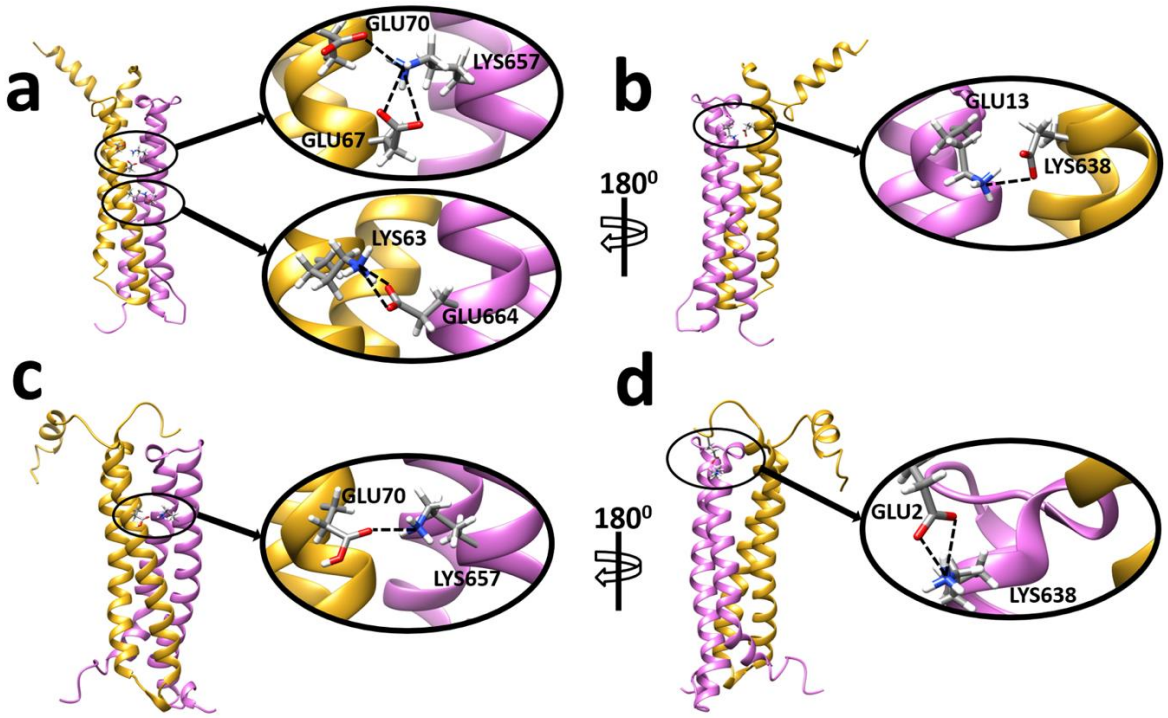

Figure 2.4: Sites of simultaneous and strongest salt bridges zoomed in bubbles. CFP-10 and ESAT-6 are indicated by gold and purple colors respectively. (a) and (b) 4 simultaneous salt bridges at pH 7 where (a) has 3 salt bridges at the Helix-2 of CFP-10 and ESAT-6 and (b) has 1 salt bridge at the Helix-1 of CFP-10 and ESAT-6. (c) and (d) 2 simultaneous salt bridges at pH 4 where (c) has one salt bridge at Helix-2 of CFP-10 and ESAT-6 and (d) has 1 salt bridge at Helix-2. The Helixes are same as in Figure 2.1. The blue and red color indicate residues with +1e and -1e charges respectively. The residue LYS657 is associated with residues GLU70 and GLU67 giving rise to 2 salt bridges at pH 7 and with GLU 70 at pH 4. 148 frames have four simultaneous salt bridges at pH 7 and 1 frame has 2 simultaneous salt bridges at pH 4. The frames for which the average salt bridge distances are least at both the pH values are displayed in the Figure 2.4.

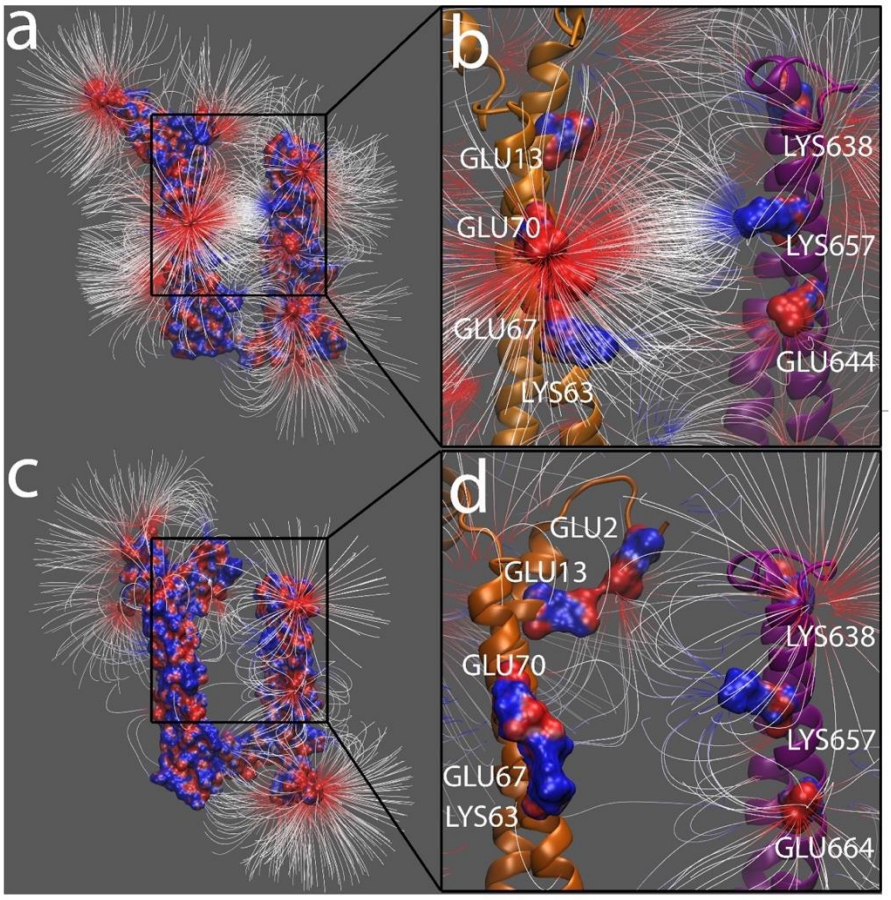

Figure 2.5: Electrostatic potential over the surface and field lines. The separation between the proteins was increased by 20 Å to visualize the potential surface and field lines. Blue and Red represents positive and negative potential surfaces respectively. (a) and (c) the proteins have surface representation and shows the density of electric field lines at pH 7 and pH 4 respectively. (b) and (d) the sites of the salt bridges focused at pH 4 and pH 7 respectively. The proteins have cartoon representation whereas the residues involved in the salt bridges having the least average salt bridges distances (same as in Figure 2.4) have surface representation.

2.3.3 MM/PBSA Binding Free Energy Calculations

Table 2.2 shows the details of the binding energy components of ESAT-6/CFP-10 at pH 7 and 4. The contribution of Non-Polar solvation energy remains almost the same for both the pHs as expected because it is independent from the charges on the residues and it only depends on solvent accessible surface area (details in method section). The average solvent accessible surface areas for the complex at pH 7 and pH 4 are 11632.11 Å² and 11822.13 Å², respectively.

As we expected, the electrostatic energy components change dramatically at pH 4, especially the Coulombic energy, which is increased from -177.25 to -72.42 kcal/mol from pH 7 to pH 4. The electrostatic energy, which includes Coulombic energy and the Polar Solvation Energy, was also increased at pH 4 as expected, because the net charges on ESAT-6 and CFP-10 have changed significantly and many salt bridges are broken at pH 4.

On the other hand, the Van der Waals energy decreased from -170.59 kcal/mol to -179.03 kcal/mol at pH 4. This is due to the net charge of the complex which let ESAT-6 and CFP-10 to be in more favorable position. Compared to the significant negative net charges (-10e) at pH 7, the near-neutral net charge (+1e) at pH 4 leads to the decrease of Van der Waals energy at pH 4. This is supported by the measurement of the distance between the center of masses of ESAT-6 and CFP-10 throughout the MD trajectory at both the pH values (Figure 2.6). Where the distance between the center of masses (COM) of ESAT-6 and CFP-10 has decreased at pH 4. Further study demonstrated that the conformational change at N-terminal of ESAT-6 makes the complex structure more compact and therefore results in a decrease of center of masses distance.

Our analysis demonstrated that the N-terminal plays a significant role in the binding affinity adjustment. Figure 2.7 shows the conformation change of N-terminal of ESAT-6 at pH 7 and 4. At pH 7, GLU603 and GLU613 are negatively charged, while the main structure of CFP-10 is also negatively charged. Therefore, the N-terminal of the ESAT-6 is repelled by the CFP-10 net charges therefore is not compactly binding with CFP-10 structure. However, at pH 4, the GLU603 and GLU613 become neutral so the repelling force between ESAT-6 N-terminal and CFP-10 is eliminated. Therefore, the Van der Waals interaction plays a role to attract the ESAT-6 N-terminal to CFP-10. Figure 2.7 clearly demonstrates that the N-terminal of ESAT-6 bends to have more contacts with CFP-10, which results in an enhancement of Van der Waals binding

energy. Note that this study doesn't take into account the post translation effect. A recent study indicates that with acetylation on N-terminal of ESAT-6, the binding affinity of ESAT-6/CFP-10 at pH 4 is lower than at pH 7, which results in dissociation between ESAT-6 and CFP-10. Our study also demonstrates that the N-terminal residues may have crucial roles in ESAT-6/CFP-10 binding.

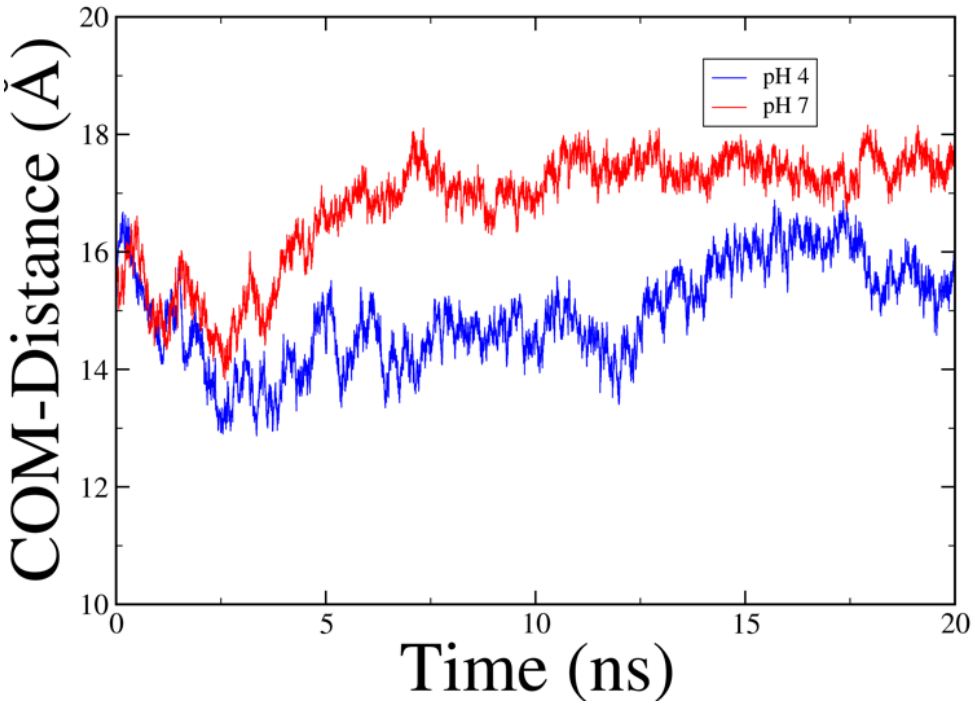

Figure 2.6: Distance between the center of masses (COM) of ESAT-6 and CFP-10 at pH 7 and pH 4. The COM distance at pH 4 has decreased meaning ESAT-6 and CFP-10 are in more favorable position and Van der Waals energy decreased.

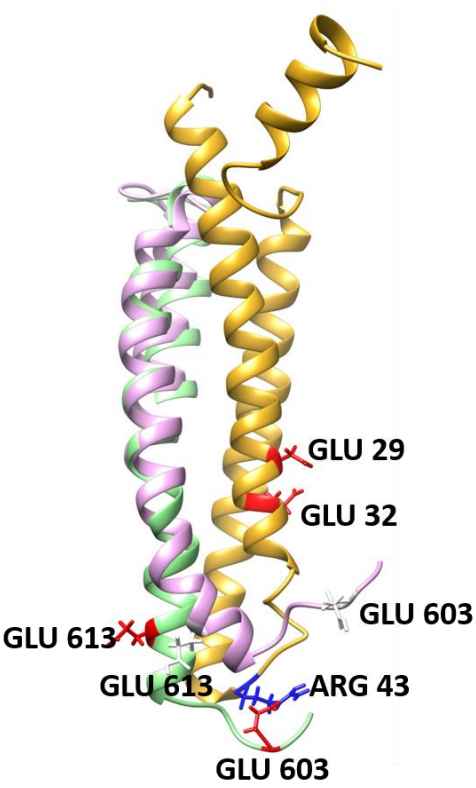

Figure 2.7: Conformation change of ESAT-6 from pH 7 to pH 4. The CFP-10 is shown in orange; ESAT-6 structures at pH 7 and 4 are shown in green and pink, respectively. Several ionizable residues close the N-terminal of ESAT-6 are shown in atomic representation. The red, blue and white color indicate negative, positive and neutral charge states, respectively. Residue Id for ESAT-6 have three digits and CFP-10 has two digits.

From MM/PBSA analysis the total binding energies of the complex of ESAT-6/CFP-10 for pH 7 and pH 4 were -179.98 kcal/mol and -175.61 kcal/mol, respectively (Table 2.2). The overall binding energy at pH 4 seems a bit weaker than at pH 7 but the difference in binding free energy is not significant. Even though the electrostatic binding energy has increased due to the changes of net charges and the interfacial salt bridges, the total energy doesn't change because the Van der Waals binding energy compensates the electrostatic binding energy, which can be observed in the Table 2.2.

Table 2.2: Binding energies and its constituent energy terms of the complex ESAT-6/CFP-10 at pH 7 and pH 4 and their respective standard deviations. The units of Energy are in kcal/mol.

pH	Polar Solvation Energy		Non-Polar Solvation Energy		Van der Waals Energy		Coulombic Energy		Total Binding Energy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
7	129.56	28.70	-21.70	0.66	-170.59	8.34	-117.25	34.07	-179.98	8.49
4	97.14	16.73	-21.31	0.73	-179.03	8.47	-72.42	17.60	-175.61	7.90

2.4 DISCUSSION:

2.4.1 Salt Bridges and Salt Bridges Energy linked to Stability of Complex.

It is widely believed fact that the salt bridges between the ionizable residues play a significant role in stabilizing the macromolecules including proteins. It can help stabilizing the structure of a protein or a protein-protein complex. In this study we have observed four key salt bridges at pH 7. Two of these salt bridges (GLU13–LYS638 and GLU70–LYS657) are also confirmed in, helping the upper regions of the helices to stabilize (Figure 2.4). Three salt bridges present in pH 7, GLU13–LYS638, GLU67–LYS657 and LYS63–GLU664, are broken in pH 4. Only one salt bridge, GLU70–LYS657, is survived but its occurrence in each frame and its energy has declined significantly. At pH 4, a new salt bridge GLU02–LYS638 has come in existence, again with very low occurrence frequency and energy, as shown in Table 2.3. Seemly, the salt bridges at pH 4 are broken evident form the population of the salt, occurrence in each frame and the energies, Table 2.3 and Figure 2.3 which may aid in dissociation. The salt bridges investigated in this study have cutoff of 3.2 Å. May be the salt bridges beyond this cutoff are acting to stabilize the interaction of ESAT-6 and CFP-10 at both pHs. The total salt bridge energy is -8.33 kcal/mol at pH 7 and -0.14 kcal/mol at pH 4 Table 2.3. And the difference of total salt bridges energies becomes 8.19 kcal/mol. Based on the discussion from the above section, the difference in electrostatic binding energy at both pHs is 12.41 kcal/mol. So, most of the gain in electrostatic

binding energy at pH 4 is fulfilled by the energy increase from the loss of salt bridges at pH 4. Moreover, the decrement of Van der Waals energy at pH 4 i.e. -8.44 kcal/mol, which compensates the electrostatic energy and keep the complex stable at pH 4.

2.4.2 MM/PBSA Analysis and Dissociation of the Complex

The MM/PBSA analysis in Table 2.2 shows that the total binding free energy has increased at pH 4. One can expect, if the complex is to dissociate at pH 7, the significant drop in the binding free energy which is not the case in our calculations. Renshaw *et al* mentioned that the interaction between the proteins ESAT-6 and CFP-10 is primarily based on the favorable Van der Waals contacts[79] which can be proved in our binding energy calculations. The major contribution to the binding free energy at both the pH values are from Van der Waals interactions. The Van der Waals energy has decreased and becomes more favorable for the interaction of the proteins at pH 4. As the net charges on the proteins has changed dramatically (ESAT-6 is neutral and CFP-10 is +1e) due to the protonation states of ionizable residues at pH 4, the Polar Solvation energy and the Coulombic energy has decreased. This may give the notion for the dissociation because for the stability of the complex of proteins, electrostatic interaction is a key factor. The electrostatic binding energy (sum of Coulombic and Polar solvation energy) at pH 7 and pH 4 are 12.31 kcal/mol and 24.72 kcal/mol respectively. The electrostatic binding energy has increased by 12.41 kcal/mol at pH 4. Similarly, the Van der Waals energy has decreased by -8.44 kcal/mol at pH 4 (Table 2.2). It becomes apparent that the increase in the electrostatic binding energy is being compensated by the decrement of Van der Waals energy and still holding the ESAT-6 and CFP-10 together. Further if dissociation taking place and proteins are moving away relative to each other, the significant increment in Non-polar solvation energy should be observed. The solvent accessible surface area would increase if the dissociation is taking place resulting in increment of

Non-polar solvation energy. Which is also not the case in our calculation. This study reinforced the hypothesis that the complex from E. coli lacks some distinctive features present in the complex from the MTB. Post Translational Modifications could be such features and the N-terminal Acetylation of THR602 of ESAT-6 is an example [80, 82, 86].

2.5 CONCLUSION

In this work we studied the behavior of the complex at pH 7 and pH 4. The binding free energies of the complex were obtained. The binding free energy of the complex at pH 4 increased by only 4.37 kcal/mol. Such similar binding energies of the complex at pH 7 and pH 4 are also proved with the experimental work from [81, 82] which indicates that the complex purified from the E coli. are stable and do not dissociate at acidic pH. If the complex is to be dissociated at pH 4 then the binding free energy should increase in a significant manner relative to pH 7. The distance between center masses between ESAT-6 and CFP-10 is reduced at pH 4 and the Van der Waals energy becomes more favorable. And at pH 4 the Van der Waals energy decreased by -8.44 kcal/mol. The electrostatic binding energy (sum of Polar solvation and Coulombic energy) increased by 12.42 kcal/mol at pH 4. The salt bridge energies increased by 8.19 kcal/mol at pH 4. The salt bridge energies are also the part of electrostatic energy as they are due to the interaction of ionic residues. So, here we observed that most of the increase in electrostatic energy at pH 4 is contributed by the increase in energy due to the breaking of the salt bridges at pH 4. Moreover, the decrease in Van der Waals energy and increase in salt bridges have almost the same magnitude. Therefore, the gain in Electrostatic energy at pH 4 is compensated by the decrement in Van der Waals energy and the ESAT-6 and CFP-10 are still stable at pH 4.

As discussed in [80], the ESAT-6 expressed from E. coli doesn't dissociate with CFP-10. While the ESAT-6 from MTB bacteria dissociation with CFP-10 at low pH environments. The reason of

this difference could be the Post Translational Modification, especially the N-terminal acetylation of the residue THR 602 (Figure 2.1) of protein ESAT-6 [89]. Besides, there can be other forms of Post Translational Modification which are yet to be discovered [80, 82, 86]. We will further study the Post Translation Modification in future works.

Chapter 3: A computational Model of ESAT-6 Complex in Membrane

3.1 INTRODUCTION

There are strong valid reasons to believe that after the dissociation, ESAT-6 undergoes the oligomerization to form a complex structure to interact with the membrane. In the experiments, neither the interaction of the complex i.e. ESAT-6/CFP-10 with the membrane was observed nor the CFP-10 but the interaction of ESAT-6 with membrane was witnessed [89]. Moreover, the length of helixes of ESAT-6 and the depth of the typical lipid bilayer were equivalent i.e. 50 Å [89]. So, its plausible to believe the ESAT-6 forms a membrane spanning pore. Which is further supported by [89], where certain residues of the ESAT-6 were labeled with NBD (Nitro-2-1,3-BenzoxaDiazol-4-yl) and trans-membrane domains were mapped. Among the labeled residues, residues Ser-635 and Ala-660 showed rapid and strong liposome dependent emission which advocates that these residues are lipid facing. In the complex form of ESAT-6/CFP-10, the residue Ser-35 is at the contact interface of ESAT-6 and CFP-10 and when it is inside the membrane it is facing the membrane. Which again supports the idea of dissociation of ESAT-6 form CFP-10 and ESAT-6 is involved in the membrane lysis activity. Further, the molecular size of a single ESAT-6 inside membrane is limited for the bacterial translocation or the exposure of the bacterial DNA to the DNA sensory mechanism in the cytosol of the host cell [89]. So, the aggregation of ESAT-6 for the formation of oligomer seems credible and is also observed in experiments [89]. But the number of the asymmetric monomers of ESAT-6 in the oligomeric assembly is largely unknown. On the basis of above discussions and the observations that most of the channels of proteins in members have cylindrical or barrel shape, we arrived at the conclusion that the oligomer of ESAT-6 has also some cylindrical symmetry. It is extremely challenging to determine the number of asymmetric units of ESAT-6 in the oligomer experimentally because the membrane proteins interact with membranes in a complicated manner and hence hard to purify and crystallize them [79-81, 83, 89]. In this regard computational methods can aid to great extent. So, in this work we

adopted comprehensive computational techniques to predict the model of oligomer of ESAT-6 interacting with membrane.

To imitate the process of oligomerization of ESAT-6, the docking algorithms, MZ-DOCK[90] and SymmDock [91] were implemented. The predicted structures were then filtered based on cylindrical symmetry and the residues Ser-635 and Ala-660 facing lipid as discussed above. The filtered structures were then subjected to clustering to obtain the best structures from both the docking algorithms and with different number of asymmetric units from 3 to 8. The molecular dynamic (MD) was performed for these structures inside the membrane to observe the interaction between them. The MM/PBSA analysis was then carried out to calculate the binding free energies based on the trajectories of MD simulations. Based on the MM/PBSA energy profile, the structure of oligomer with asymmetric unit 4 from M-ZDOCK stands out to be the best model for the oligomerization of ESAT-6 which spans trans-membrane pore, being binding free energy least among the other predictions from the docking algorithms and with different number of asymmetric units. The salt bridge analysis, H-bond analysis were further carried out to observe the stability of the best model while it interacts with the membrane.

Hence with the rigorous and comprehensive computational techniques, we successfully predicted the best model of the oligomer of ESAT-6 interacting with membrane to span a pore for the virulence of Tuberculosis based on the ground of various facts from experiments. Though the experimental verification of this prediction still needs to be achieved.

3.2 METHOD

3.2.1 Preparing the Structure of ESAT-6 for Docking

In order to obtain the structure of ESAT-6, the complex structure of ESAT-6/CFP-10 was downloaded from the Protein Data Bank (PDB ID 1WA8 [79]) and separated into individual proteins i.e. ESAT-6 and CFP-10. We employed docking algorithms including SymmDock and M-ZDOCK which predict the cyclically symmetric oligomers for given monomer. As these

algorithms treat proteins as rigid body and we have the flexible N- and C- terminal arms of monomer, ESAT-6 in our case, the algorithms may discard those predictions in which these arms are overlapped, and those predictions could be the predictions which fill our requirement. To avoid such situations the flexible arms were truncated beforehand of performing docking algorithms. The truncated flexible N- terminal arm includes residues 1 to 10 and C- terminal contains residues 79 to 95. Figure 3.1a shows the schematic diagram for the truncation.

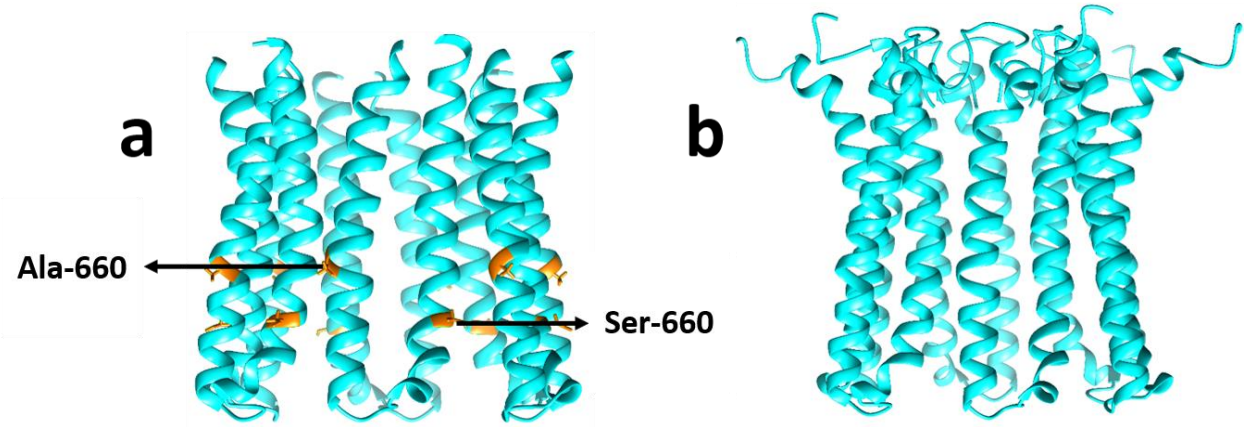

Figure 3.1: Truncation and reconstruction of flexible N- and C- terminals. (a) truncation of flexible terminals of oligomer with symmetry 6 for example. The lipid facing residues Ser-635 and Ala-660 are leveled. (b) reconstruction of the flexible terminals to the final structure of oligomer with symmetry 6 after filtration process

3.2.2 Filtration of the Predicted Structure

For the docking purpose the cyclic symmetries from 3 to 8 were considered. The SymmDock predicted 3624, 4271, 3020, 2502, 1808, 1504 for symmetries C3, C4, C5, C6, C7 and C8 respectively. On the other hand, MZ-DOCK predicted 1500 structures of oligomers for each cyclic symmetry i.e. C3 to C8. For filtration of the predicted structures of oligomers we adopted two facts. One is the cylindrical shape of the oligomer and the other is the experimentally determined fact that the residues Ser-635 and Ala-660 facing the membrane [89]. Chimera [63, 69] was used for the visualization and the filtration process. The clustering was then performed based on the root mean squared deviation (RMSD) of the predicted structures from docking algorithms. Structures will belong to a cluster if their RMSD is less than 10 Å. After the filtration

and clustering 6 final structures were obtained; one from each cyclic symmetry for SymmDock and similar for the MZ-DOCK. Meaning altogether 12 best predictions were obtained based on the creation we set for the filtration and clustering. The truncated flexible N- and C- arms were then reconstructed back to the final 12 oligomer structures using Chimera.

3.2.3 Preparation of the System for MD Simulation

The interactions of the oligomers of different cyclic symmetry with the membrane were then studied using Molecular Dynamic (MD) simulations [75]. The bi-lipid membrane of type POPC (Phosphatidylcholine) with hydrated headgroups was generated using Charmm [72] force field in Visual Molecular Dynamics (VMD) with dimensions 200 Å , 200 Å and 60 Å along X- , Y- and Z- axes respectively. The final structures of oligomers were then imbedded inside the membrane at the center with proper alignment such that the helixes were aligned along Z- axis of the membrane (3.2). The overlaps between the lipids and the oligomers of ESAT-6 and the lipids inside the cylindrical space of oligomers were removed as a result a channel is formed in the membrane. Due to the removal of lipids overlapped with oligomers certain gap was observed between the outer surface of oligomer and the membrane which will be removed during MD simulation. Then the system was solvated with TIP3 water model and ionized with Sodium Chloride (NaCl) with concentration 0.15 mol/liter to mimic the native environment. The dimension of the final system after the solvation and ionization was 200 Å , 200Å , 100 Å along X-, Y- and Z- axis respectively. The thickness of water layer was kept more on the side of the membrane were the flexible arms of ESAT-6 lies to avoid the situation were during the simulation the flexible arms stretch out of the water box (Figure 3.2).

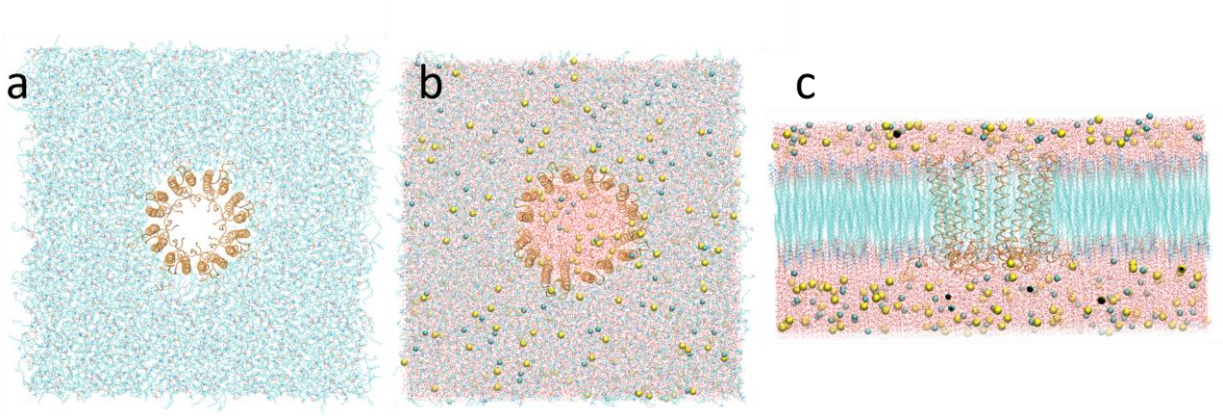

Figure 3.2: Preparation of system for the MD simulation. (a) insertion of oligomer of ESAT-6 in membrane with proper alignment (b) solvation of the system with water molecules and ionization with NaCl molecules, blue dots are Na ions and yellow dots are Cl (c) side view of b. X-axis is along horizontal, vertical is the Z-axis and the depth is the Y-axis.

3.2.4 MD Simulation

The simulations were performed with Molecular Dynamics Simulation package NAMD 2.12[73]. The simulations were performed in four stages and every step have different motives. The water molecules, ions the protein (oligomers) and the head groups of lipids were fixed in space while tails of the bi-lipid membrane were simulated in the first stage to obtain the fluid like bi-layer. The system was minimized for 20,000 steps followed by 2 ns simulation. In the second stage, the atoms which were fixed in stage one was released but the harmonic constraint was applied on the protein so that the proper packing of protein with membrane could be achieved. The membrane shrinked inside a little bit to fulfill the gap which was created by the removal of the overlapped lipids within and around the proteins in previous steps of preparation of the system for simulation. This stage has 20,000 steps of minimization and 1 ns of simulation. Also, the force on the water molecules were applied to the pump them out if they enter the hydrophobic regions of the membrane. The force is the user-defined forces in NAMD which makes use of the NAMD Tcl Forces modules [73]. In the third stage, the system was further equilibrated with releasing harmonic constraints on the protein but the restraint on the secondary structures was applied to

preserve the secondary structures. Also, the forces on the water molecules to pump them out from forbidden regions on membrane were removed. This stage contains 5 ns of simulation and no steps of minimization as all the components are already minimized in the previous stages. The fourth stage is the production run were the system was simulated for 20 ns. The area of the membrane along the x-y plane of the membrane was kept constant in this stage as the well equilibrated protein packed against the membrane is already achieved. The fluctuation of this area was needed in previous stages to have the packing of membrane with protein. The steps in the preparation of the system for simulations and the different stages of simulations were performed as guided by the Membrane Proteins Tutorial [92].

All the simulations for 12 final oligomers were performed with Molecular Dynamics Simulation package NAMD version 2.12. The first stage of simulation was performed with NVT ensemble and the other stages were performed with NPT ensemble. The temperature and pressure were kept standard. The periodic boundary conditions were applied to the solvation box. The long-range electrostatic interactions i.e. beyond the cutoffs were calculated with Particle Mesh Ewald method [93]. The cutoff was defined at 12 Å.

3.2.5 MM/PBSA Analysis

To quantify the binding free energy of the oligomers inside the membrane, MM/PBSA [86] analysis was carried out. 200 snapshots were taken form the final 10 ns of the production run. Water molecules and ions were removed before conducting the MM/PBSA analysis. According to this method, the total binding free energy is given as,

$$E = E_{coul} + E_{vwd} + E_{pol} + E_{np} \quad (3.1)$$

Where E_{coul} and E_{pol} are Coulombic and Polar Solvation Energies which were calculated employing DelPhi[61, 66, 67]. For DelPhi calculations, the dielectric for protein and lipids was set as 2 and 80 for water. The protein and the lipids could cover the 70 percentage of the DelPhi calculation box. The concentration of salt was 0.15 and the resolution for the grid was set to be 2

grids/Å. The Van der Waals energy (E_{vwd}) was calculated with NAMD 2.12 subjecting the structures to 1 step of equilibration. The Non-polar Solvation Energy (E_{np}) was calculated based on the solvent accessible surface area (SASA) where SASA was calculated via NACCESS[74] and energy associated is given as,

$$E_{np} = \alpha \cdot \text{SASA} + \beta \; ; \alpha = 0.0054 \; and \; \beta = 0.92 \text{ kcal/mol} \qquad (3.2)$$

The binding free energy of the oligomers inside the membranes were calculated with the formula,

$$\Delta E_{binding} = E_{PL} - E_L - E_{P1} - E_{P2} - E_{P3} - \ldots\ldots\ldots\ldots - E_{Pn} \qquad (3.3)$$

Where, P = Protein, L = Lipid and $P = P1 + P2 + P3 + \ldots\ldots\ldots$ Pn, Pn denotes number of ESAT-6 on oligomers.

3.2.6 Salt Bridge Analysis

The salt bridges among oppositely charged residues at the binding interface of the neighboring ESAT-6 play the significant role in stabilizing the structure of the oligomers inside the membrane. The salt bridge analysis was performed in VMD, setting the cutoff at 4 Å between the Oxygen and Nitrogen atoms of the oppositely charged residues, based on the final 10 ns simulation of production run. The salt bridges were also studied for cutoff 6 Å to account for the longer interactions.

3.2.7 Hydrogen Bond Analysis

The network of H-bond between the helixes in membrane proteins are important in their formation, structure, stability and functionality. So, to observe the network of H-bond among the helixes at the binding interfaces of the neighboring ESAT-6 in the oligomers, H-bond analysis was performed with HBond Plugin in VMD. The final 10 ns simulation of the production run was considered for the H-bonds analysis. The cutoff distance between the acceptor and donor atoms was set to 6 Å. Similarly, the cutoff angle between the receptor-hydrogen-donor was set to 20^0.

3.2.8 Residues Contacting the Lipids

From the experiments, certain residues like Ser-635 and Ala-660 are confirmed to face the membrane. To verify this the contact of the residues with lipid were counted with threshold 3 Å based on the last 10 ns simulation of production run. Here we define the contact as, if the atoms of the lipids come within the 3 Å of the residue of our interest. Total contacts were counted for the respective residues and divided with the number of frames i.e. 2000 in our case and again divided with number of symmetry because particular residue is present in every monomer of the oligomer to obtain the contact of a particular residue per frame.

3.3 Results and Discussions

The calculation of binding free energy of the oligomers in membrane with different number of symmetries and from different docking algorithms is shown in Table 3.1. Based on binding free energy calculation, no matter which docking algorithm, the contribution of the single ESAT-6 to the total binding energy is least for the cyclic symmetry C4. So, the oligomer with cyclic symmetry C4 could be the best model for membrane interaction. The C4 from M-ZDOCK has the least binding energy per monomer i.e. -266.87 kcal/mol amongst all. So, in this work, the investigations of the C4 from M-ZDOCK is further carried out to identify different properties responsible for making this model stable discussed below.

Table 3.1: Binding Energy Profile of the Oligomers from MM/PBSA Analysis for Oligomers with different Symmetries

Docking Symmetry		**Polar Solvation Energy**		**Non-Polar Solvation Energy**		**Van der Waals Energy**		**Coulombic Energy**		**Total Binding Energy**		**Binding Energy per monomer(ESAT-6)**
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
M-ZDOCK	3	117.25	67.90	-112.12	3.12	-790.87	22.95	191.66	69.39	-594.02	26.10	-198.01
	4	-95.51	39.89	-142.61	2.72	-1112.28	27.28	282.94	37.42	-1067.46	24.35	-266.87
	5	-636.30	141.22	-168.67	2.51	-1276.42	29.53	1032.95	155.33	-1048.44	29.44	-209.68
	6	-616.35	31.65	-172.62	2.21	-1328.07	26.05	822.92	36.80	-1294.11	25.43	-215.69
	7	-1592.07	103.29	-205.37	1.95	-1550.19	32.08	2071.43	106.96	-1276.22	34.61	-182.31
	8	-1147.99	42.12	-227.53	2.12	-1734.44	29.61	1468.71	43.36	-1641.25	29.50	-205.16
SymmDock	3	249.34	45.61	-106.82	1.70	-814.53	18.80	-38.29	48.70	-710.30	20.64	-236.77
	4	-21.22	33.32	-134.81	1.65	-1048.01	19.11	178.79	39.18	-1025.25	20.28	-256.31
	5	-609.85	85.73	-161.27	3.27	-1220.23	35.30	997.58	88.98	-993.76	22.21	-198.75
	6	-551.50	41.24	-190.49	2.22	-1448.27	36.33	810.56	39.06	-1379.71	41.71	-229.95
	7	-2090.36	86.41	-206.71	2.48	-1584.25	26.30	2580.17	89.81	-1301.13	28.36	-185.88
	8	-1158.56	47.56	-225.28	3.08	-1729.19	28.98	1487.39	49.46	-1625.64	28.52	-203.21

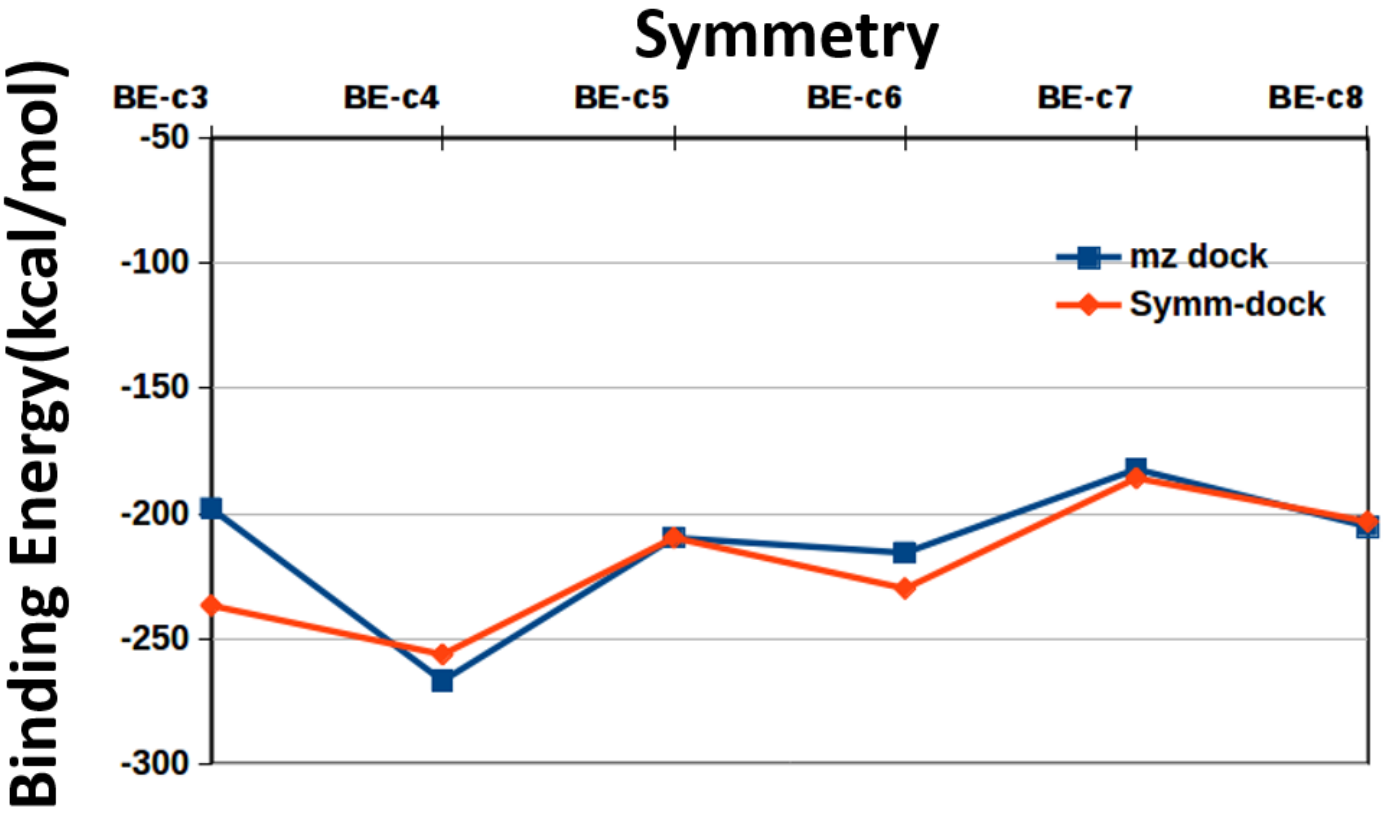

Figure 3.3: Binding Energy contribution of an individual ESAT-6 in oligomers for different number of symmetry.

3.3.1 Identification of Key Residues from salt Bridge Analysis

From the salt bridge analysis with cutoff 4 Å we identified two salt bridges ASP659 A – LYS638 D and GLU649 D – LYS638 C. The residue LYS638 is associated with both the salt bridges. Figure 3.4 shows these salt bridges are not formed coherently and are exclusive to each other.

Table 3.2: Salt Bridges and their cyclic occurrence

Cutoff	Salt bridge	Total	%
4 Å	ASP659 A – LYS638 D	631	31.55
	GLU649 D – LYS638 C	170	8.8
10 Å	GLU631 C – LYS657 D	497	24.85
	ASP659 A – LYS638 D *	1895	94.75
	ASP659 B – LYS638 A *	222	11.10
	GLU612 C – ARG674 D	1294	64.7
	GLU649 D – LYS638 C	1354	67.7
	GLU664 C – LYS638 B **	232	11.6
	GLU664 A – LYS638 D **	1017	50.85
	GLU664 B – LYS638 A **	207	10.35

The salt bridge ASP659 A – LYS638 D forms in the 2nd half of the last 10 ns simulation of production run whereas the salt bridge GLU649 D – LYS638 C forms in the 1st half. The distance between the residues involved the salt bridges were measured to understand the mechanism behind it (Figure 3.4c).The ESAT-6 (D) swinged towards the ESAT-6 (A) in the second half (Figure 3.4c,d) so that the relative distance between ASP659 A and LYS638D decreased and distance between GLU649 D and LYS638 C increased. So, the criterion for formation of salt bridge is favorable for ASP659 A and LYS638D not for GLU649 D and LYS638 C in the second half.

Though the C4 is a symmetric structure, the symmetric patterns was not observed in the formation of salt bridges. For an example the salt bridge ASP659 – LYS638 is observed only in the interface of ESAT-6 (A) and ESAT-6 (D) but remain absent in another interface of ESAT-6 i.e AB, BC, CD. One possible reason could be the cutoff. The cutoff, 4 Å, is small and encompasses only the strong salt bridges. The cutoff was increased to 10 Å to account for weak salt bridges and if certain symmetric patterns could be observed for the salt bridges. Certainly, some cyclic patterns

are observed but still salt bridges for certain interfaces are missing (Table 3.2). But then, at least all the binding interfaces between the ESAT-6 in the C4 symmetry have salt bridges which advocates the stability of the oligomer with the symmetry C4 in membrane. Moreover, the residue LYS638 is involved in most of the salt bridges (Table 3.2) so this residue should be the key for the stability of the oligomer with symmetry C4 in the membrane.

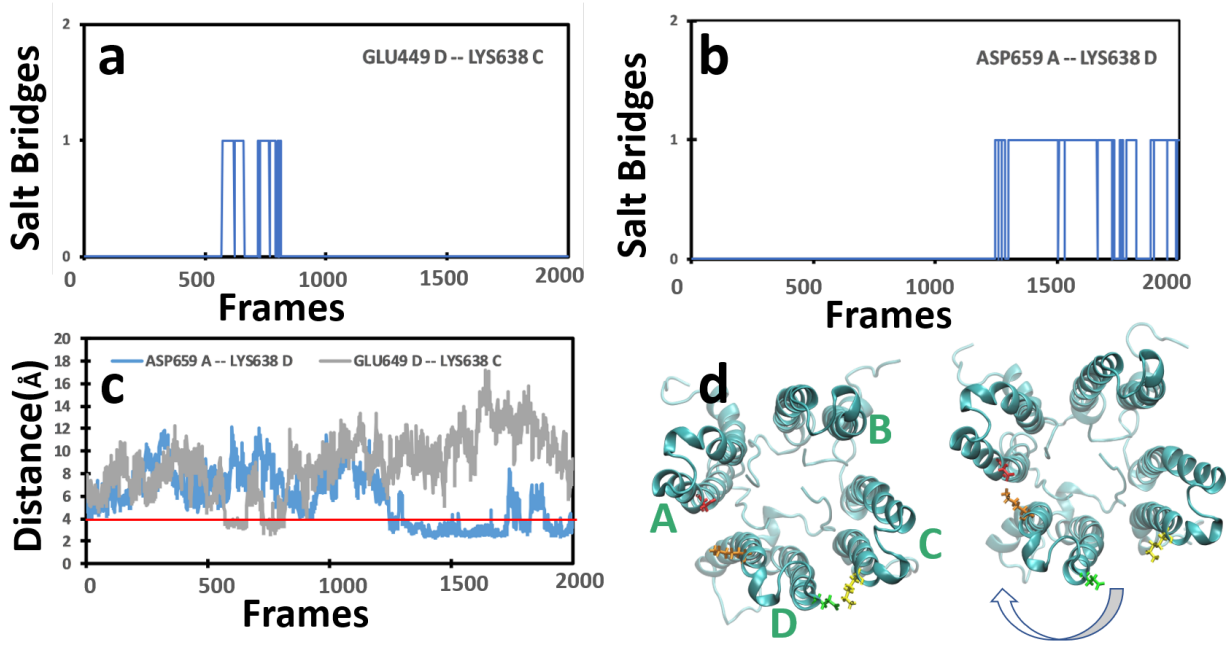

Figure 3.4: Example of relative motion between the ESAT-6 in oligomer (a) formation of salt bridge when ESAT-6 , D is close to C. (b) formation of salt bridge when ESAT-6 , D is close to A (c) measure of distance between the residues involved in salt bridges while D swings between C and A (d) depiction of swinging of D when it is at closest distance to C and D.

Initially, the symmetry C4 from the docking had perfect symmetry but once subjected to molecular dynamic simulation with membrane the dynamics is really random due the random forces, which lead to the distortion in the symmetry. The swing of ESAT-6 (D) is an example and there could be other different possible motions like sliding up or down, shifting, rotation, etc.

3.3.2 H-bonds:

The H-bonds per frame for every binding interface of ESAT-6 in the oligomer of symmetry C4 is depicted in Figure 3.5a. The presence of H-bonds in every interface indicates that the H-

bonds have significant role in stabilizing the oligomer inside the membrane. But the interface BC has the least H-bonds per frame. The Table 3.2 shows the salt bridges at the interface AB and BC have least occurrences compared to others. In the interface AB there are two salt bridges i.e. ASP659 B-LYS638 A and GLU664 B - LYS638 A and has greater H-bond per frame compared to the interface BC. Furthermore, the interface BC has only one salt bridge i.e. GLU664 C-LYS638 B with less occurrence and has least H-bonds per frame. So, the interface BC can be considered as the weakest interface and if the oligomer breaks the interface BC could be the fault zone.

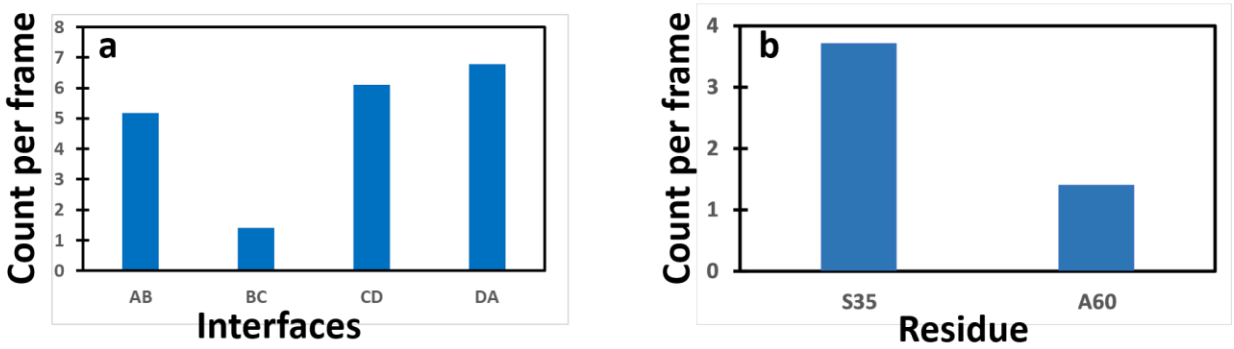

Figure 3.5: H-bond and contact of residues to lipid of the oligomer with symmetry 4. (a) population of H-bonds on every binding interface of oligomer with symmetry 4 from MZ-Dock which is the best model for oligomer to interact with membrane according our prediction (b) frequency of contact of residues S35 and A60 to lipid in every fame for each ESAT-6 .

3.3.3 Conformation of A60 and S35 facing membrane:

From the experiments performed by Yue Ma et al. [89] the residues Ala-660 and Ser-635 are membrane facing. If they do so, the lipids of the membrane and the residues should be in close vicinity. To verify this, we defined a contact as mention in method part, and the contact of these residues per frame is as shown in Figure 3.5b. The residue Ser-635 has greater contact per frame means it is well facing the membrane and has stronger interaction with membrane in comparison to residue Ala-660, which is consistent with the experimental observations.

3.4 CONCLUSION

The interaction of the ESAT-6 with membrane is considered as key for the virulance of Tuberculosis. If the molecular mechanism of the interaction is well understood, it would be helpful to understand the disease well so that we can fight back well. In this work we predicted the best model of oligomer of the ESAT-6 for interaction with the membrane which is the C4 symmetry. The salt bridges and H-bonds were also observed for this symmetry which contribute a lot for the stability of the oligomer. Further the experimental result, residues Ser-635 and Ala-660 facing the lipid was verified by counting their contacts with lipids. Hence, the key of this work is to predict the model of oligomer for membrane interaction.

References

1. Xiao, C. and M.G. Rossmann, *Structures of giant icosahedral eukaryotic dsDNA viruses.* Current opinion in virology, 2011. **1**(2): p. 101-109.
2. Zhang, X., et al., *Three-dimensional structure and function of the Paramecium bursaria chlorella virus capsid.* Proceedings of the National Academy of Sciences, 2011. **108**(36): p. 14837-14842.
3. Yan, X., et al., *Structure and assembly of large lipid-containing dsDNA viruses.* Nature Structural & Molecular Biology, 2000. **7**(2): p. 101.
4. Van Etten, J.L., L.C. Lane, and R.H. Meints, *Viruses and viruslike particles of eukaryotic algae.* Microbiology and Molecular Biology Reviews, 1991. **55**(4): p. 586-620.
5. Simpson, A.A., et al., *Structural analyses of Phycodnaviridae and Iridoviridae.* Acta Crystallographica Section D: Biological Crystallography, 2003. **59**(12): p. 2053-2059.
6. Nandhagopal, N., et al., *The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus.* Proceedings of the National Academy of Sciences, 2002. **99**(23): p. 14758-14763.
7. Cherrier, M.V., et al., *An icosahedral algal virus has a complex unique vertex decorated by a spike.* Proceedings of the National Academy of Sciences, 2009. **106**(27): p. 11085-11089.
8. Yan, X., et al., *The capsid proteins of a large, icosahedral dsDNA virus.* Journal of molecular biology, 2009. **385**(4): p. 1287-1299.
9. Yan, X., et al., *The marine algal virus PpV01 has an icosahedral capsid with T= 219 quasisymmetry.* Journal of virology, 2005. **79**(14): p. 9236-9243.
10. La Scola, B., et al., *A giant virus in amoebae.* Science, 2003. **299**(5615): p. 2033-2033.
11. Raoult, D., et al., *The 1.2-megabase genome sequence of Mimivirus.* Science, 2004. **306**(5700): p. 1344-1350.
12. La Scola, B., et al., *Mimivirus in pneumonia patients.* Emerging infectious diseases, 2005. **11**(3): p. 449.
13. Xiao, C., et al., *Cryo-electron microscopy of the giant Mimivirus.* Journal of molecular biology, 2005. **353**(3): p. 493-496.
14. Zauberman, N., et al., *Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus.* PLoS biology, 2008. **6**(5): p. e114.
15. Xiao, C., et al., *Structural studies of the giant mimivirus.* PLoS biology, 2009. **7**(4): p. e1000092.
16. Kuznetsov, Y.G., et al., *Atomic force microscopy investigation of the giant mimivirus.* Virology, 2010. **404**(1): p. 127-137.
17. Seibert, M.M., et al., *Single mimivirus particles intercepted and imaged with an X-ray laser.* Nature, 2011. **470**(7332): p. 78.
18. Arslan, D., et al., *Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae.* Proceedings of the National Academy of Sciences, 2011. **108**(42): p. 17486-17491.
19. Fischer, M.G., et al., *Giant virus with a remarkable complement of genes infects marine zooplankton.* Proceedings of the National Academy of Sciences, 2010. **107**(45): p. 19508-19513.
20. Xiao, C., et al., *Cryo-EM reconstruction of the Cafeteria roenbergensis virus capsid suggests novel assembly pathway for giant viruses.* Scientific reports, 2017. **7**(1): p. 5484.

21. Klose, T., et al., *Structure of faustovirus, a large dsDNA virus.* Proceedings of the National Academy of Sciences, 2016. **113**(22): p. 6206-6211.
22. Reteno, D.G., et al., *Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae.* Journal of virology, 2015. **89**(13): p. 6585-6594.
23. Campos, R.K., et al., *Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon.* Virology journal, 2014. **11**(1): p. 95.
24. Schrad, J., et al., *Microscopic characterization of the Brazilian giant samba virus.* Viruses, 2017. **9**(2): p. 30.
25. Philippe, N., et al., *Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes.* Science, 2013. **341**(6143): p. 281-286.
26. Legendre, M., et al., *Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology.* Proceedings of the National Academy of Sciences, 2014. **111**(11): p. 4274-4279.
27. Claverie, J.-M. and H. Ogata, *Ten good reasons not to exclude giruses from the evolutionary picture.* Nature Reviews Microbiology, 2009. **7**(8): p. 615.
28. Hegde, N.R., et al., *Reasons to include viruses in the tree of life.* Nature Reviews Microbiology, 2009. **7**(8): p. 615.
29. Koonin, E.V., T.G. Senkevich, and V.V. Dolja, *Compelling reasons why viruses are relevant for the origin of cells.* Nature Reviews Microbiology, 2009. **7**(8): p. 615.
30. Ludmir, E.B. and L.W. Enquist, *Viral genomes are part of the phylogenetic tree of life.* Nature Reviews Microbiology, 2009. **7**(8): p. 615.
31. Moreira, D. and P. López-García, *Ten reasons to exclude viruses from the tree of life.* Nature Reviews Microbiology, 2009. **7**(4): p. 306.
32. Navas-Castillo, J., *Six comments on the ten reasons for the demotion of viruses.* Nature Reviews Microbiology, 2009. **7**(8): p. 615.
33. Raoult, D., *There is no such thing as a tree of life (and of course viruses are out!).* Nature Reviews Microbiology, 2009. **7**(8): p. 615.
34. Dickey Zakaib, G., *The challenge of microbial diversity: out on a limb.* Nature News, 2011. **476**(7358): p. 20-21.
35. Raoult, D. and P. Forterre, *Redefining viruses: lessons from Mimivirus.* Nature Reviews Microbiology, 2008. **6**(4): p. 315.
36. Claverie, J.-M. and C. Abergel, *Mimivirus: the emerging paradox of quasi-autonomous viruses.* Trends in Genetics, 2010. **26**(10): p. 431-437.
37. Harrison, S., et al., *Tomato bushy stunt virus at 2.9 Å resolution.* Nature, 1978. **276**(5686): p. 368.
38. Abad-Zapatero, C., et al., *Structure of southern bean mosaic virus at 2.8 Å resolution.* Nature, 1980. **286**(5768): p. 33.
39. Rossmann, M.G., et al., *Structure of a human common cold virus and functional relationship to other picornaviruses.* Nature, 1985. **317**(6033): p. 145.
40. Hogle, J., M. Chow, and D. Filman, *Three-dimensional structure of poliovirus at 2.9 A resolution.* Science, 1985. **229**(4720): p. 1358-1365.
41. Rossmann, M.G. and J.E. Johnson, *Icosahedral RNA virus structure.* Annual review of biochemistry, 1989. **58**(1): p. 533-569.
42. Benson, S.D., et al., *Does common architecture reveal a viral lineage spanning all three domains of life?* Molecular cell, 2004. **16**(5): p. 673-685.

43. Kaufmann, B., A.A. Simpson, and M.G. Rossmann, *The structure of human parvovirus B19.* Proceedings of the National Academy of Sciences, 2004. **101**(32): p. 11628-11633.
44. Wrigley, N., *An electron microscope study of the structure of Sericesthis iridescent virus.* Journal of General Virology, 1969. **5**(1): p. 123-134.
45. Suárez, C., et al., *Open membranes are the precursors for assembly of large DNA viruses.* Cellular microbiology, 2013. **15**(11): p. 1883-1895.
46. Zhao, G., et al., *Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics.* Nature, 2013. **497**(7451): p. 643.
47. Hagan, M.F., *Modeling viral capsid assembly.* Advances in chemical physics, 2014. **155**: p. 1.
48. Hagan, M.F. and R. Zandi, *Recent advances in coarse-grained modeling of virus assembly.* Current opinion in virology, 2016. **18**: p. 36.
49. Hagan, M.F., *Controlling viral capsid assembly with templating.* Physical Review E, 2008. **77**(5): p. 051904.
50. Roos, W., et al., *Squeezing protein shells: how continuum elastic models, molecular dynamics simulations, and experiments coalesce at the nanoscale.* Biophysical journal, 2010. **99**(4): p. 1175-1181.
51. Arkhipov, A., P.L. Freddolino, and K. Schulten, *Stability and dynamics of virus capsids described by coarse-grained modeling.* Structure, 2006. **14**(12): p. 1767-1777.
52. Koehl, P., *Large Eigenvalue Problems in Coarse-Grained Dynamic Analyses of Supramolecular Systems.* Journal of chemical theory and computation, 2018. **14**(7): p. 3903-3919.
53. Freddolino, P.L., et al., *Molecular dynamics simulations of the complete satellite tobacco mosaic virus.* Structure, 2006. **14**(3): p. 437-449.
54. Jiang, J., et al., *Early stage P22 viral capsid self-assembly mediated by scaffolding protein: Atom-resolved model and molecular dynamics simulation.* The Journal of Physical Chemistry B, 2015. **119**(16): p. 5156-5162.
55. van der Schoot, P. and R. Bruinsma, *Electrostatics and the assembly of an RNA virus.* Physical Review E, 2005. **71**(6): p. 061928.
56. Šiber, A. and R. Podgornik, *Role of electrostatic interactions in the assembly of empty spherical viral capsids.* Physical Review E, 2007. **76**(6): p. 061906.
57. Li, L., L. Wang, and E. Alexov, *On the energy components governing molecular recognition in the framework of continuum approaches.* Frontiers in molecular biosciences, 2015. **2**: p. 5.
58. Li, L., J. Alper, and E. Alexov, *Cytoplasmic dynein binding, run length, and velocity are guided by long-range electrostatic interactions.* Scientific reports, 2016. **6**: p. 31523.
59. Li, L., J. Alper, and E. Alexov, *Multiscale method for modeling binding phenomena involving large objects: application to kinesin motor domains motion along microtubules.* Scientific reports, 2016. **6**: p. 23249.
60. Li, L., et al., *Forces and Disease: Electrostatic force differences caused by mutations in kinesin motor domains can distinguish between disease-causing and non-disease-causing mutations.* Scientific reports, 2017. **7**(1): p. 8237.
61. Li, C., et al., *Highly efficient and exact method for parallelization of grid-based algorithms and its implementation in DelPhi.* Journal of computational chemistry, 2012. **33**(24): p. 1960-1966.

62. De Castro, C., et al., *Structure of the chlorovirus PBCV-1 major capsid glycoprotein determined by combining crystallographic and carbohydrate molecular modeling approaches.* Proceedings of the National Academy of Sciences, 2018. **115**(1): p. E44-E52.
63. Goddard, T.D., C.C. Huang, and T.E. Ferrin, *Visualizing density maps with UCSF Chimera.* Journal of structural biology, 2007. **157**(1): p. 281-287.
64. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**(1): p. 33-38.
65. Dolinsky, T.J., et al., *PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations.* Nucleic acids research, 2004. **32**(suppl_2): p. W665-W667.
66. Li, L., et al., *DelPhi: a comprehensive suite for DelPhi software and associated resources.* BMC biophysics, 2012. **5**(1): p. 9.
67. Li, L., et al., *On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi.* Journal of chemical theory and computation, 2013. **9**(4): p. 2126-2136.
68. Chakravorty, A., et al., *A new DelPhi feature for modeling electrostatic potential around proteins: Role of bound ions and implications for zeta-potential.* Langmuir, 2017. **33**(9): p. 2283-2295.
69. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.
70. Li, L., A. Chakravorty, and E. Alexov, *DelPhiForce, a tool for electrostatic force calculations: Applications to macromolecular binding.* Journal of computational chemistry, 2017. **38**(9): p. 584-593.
71. Li, L., et al., *DelPhiForce web server: electrostatic forces and energy calculations and visualization.* Bioinformatics, 2017. **33**(22): p. 3661-3663.
72. Vanommeslaeghe, K., 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): p. 671-690.
73. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of computational chemistry, 2005. **26**(16): p. 1781-1802.
74. Hubbard, S., *NACCESS: program for calculating accessibilities.* Department of Biochemistry and Molecular Biology, University College of London, 1992.
75. Lyubartsev, A.P. and A. Laaksonen, *M. DynaMix–a scalable portable parallel MD simulation package for arbitrary molecular mixtures.* Computer physics communications, 2000. **128**(3): p. 565-589.
76. Caspar, D.L. and A. Klug. *Physical principles in the construction of regular viruses*. in *Cold Spring Harbor symposia on quantitative biology*. 1962. Cold Spring Harbor Laboratory Press.
77. Organization, W.H., *The World health report: 2004: changing history.* 2004.
78. Berthet, F.-X., et al., *A Mycobacterium tuberculosis operon encoding ESAT= 6 and a novel low-molecular-mass culture filtrate protein (CFP-10).* Microbiology, 1998. **144**(11): p. 3195-3203.
79. Renshaw, P.S., et al., *Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6.* The EMBO journal, 2005. **24**(14): p. 2491-2498.

80. De Jonge, M.I., et al., *ESAT-6 from Mycobacterium tuberculosis dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity.* Journal of bacteriology, 2007. **189**(16): p. 6028-6034.
81. De Leon, J., et al., *Mycobacterium tuberculosis ESAT-6 exhibits a unique membrane-interacting activity that is not found in its ortholog from non-pathogenic Mycobacterium smegmatis.* Journal of Biological Chemistry, 2012. **287**(53): p. 44184-44191.
82. Lightbody, K.L., et al., *Molecular features governing the stability and specificity of functional complex formation by Mycobacterium tuberculosis CFP-10/ESAT-6 family proteins.* Journal of Biological Chemistry, 2008. **283**(25): p. 17681-17690.
83. Peng, X. and J. Sun, *Mechanism of ESAT-6 membrane interaction and its roles in pathogenesis of Mycobacterium tuberculosis.* Toxicon, 2016. **116**: p. 29-34.
84. Okkels, L.M., et al., *CFP10 discriminates between nonacetylated and acetylated ESAT-6 of Mycobacterium tuberculosis by differential interaction.* Proteomics, 2004. **4**(10): p. 2954-2960.
85. Wang, L., M. Zhang, and E. Alexov, *DelPhiPKa web server: predicting p K a of proteins, RNAs and DNAs.* Bioinformatics, 2015. **32**(4): p. 614-615.
86. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models.* Accounts of chemical research, 2000. **33**(12): p. 889-897.
87. Medie, F.M., et al., *Homeostasis of N-α-terminal acetylation of EsxA correlates with virulence in Mycobacterium marinum.* Infection and immunity, 2014. **82**(11): p. 4572-4586.
88. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of chemical physics, 1983. **79**(2): p. 926-935.
89. Ma, Y., V. Keil, and J. Sun, *Characterization of mycobacterium tuberculosis EsxA membrane insertion roles of N-and C-terminal flexible arms and central helix-turn-helix motif.* Journal of Biological Chemistry, 2015. **290**(11): p. 7314-7322.
90. Pierce, B., W. Tong, and Z. Weng, *M-ZDOCK: a grid-based approach for C n symmetric multimer docking.* Bioinformatics, 2004. **21**(8): p. 1472-1478.
91. Schneidman-Duhovny, D., et al., *PatchDock and SymmDock: servers for rigid and symmetric docking.* Nucleic acids research, 2005. **33**(suppl_2): p. W363-W367.
92. Aksimentiev, A., M. Sotomayor, and D. Wells, *Membrane proteins tutorial.* University of Illinois at Urbana-Champaign, 2012.
93. Essmann, U., et al., *A smooth particle mesh Ewald method.* The Journal of chemical physics, 1995. **103**(19): p. 8577-8593.

Vita

My name is Chitra Bahadur Karki. I completed my undergraduate and graduate study form Tribhuvan University (TU), Nepal. I joined Liverpool International College as Physics lecture after my graduate studies and worked there for a year. In the fall of 2017, I joined The University of Texas at El Paso (UTEP) as a graduate student in Physics department. Here in UTEP, I joined Biophysics and Bioinformatics lab, Dr. Lin Li being research advisor. Under his supervision I was able to publish a paper and had couple of oral and poster presentations at the workshops in New Mexico State University both the years . I want to express my gratitude to Dr. Lin Li and all the members of Physics department for their valuable support and guidelines.