Structural Dynamics of Membrane Interacting Viral Proteins

Nisha Bhattarai
nbhat006@fiu.edu

Follow this and additional works at: https://digitalcommons.fiu.edu/etd

Part of the Biological and Chemical Physics Commons

Recommended Citation
https://digitalcommons.fiu.edu/etd/4795

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.
FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

STRUCTURAL DYNAMICS OF MEMBRANE INTERACTING VIRAL PROTEINS

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Nisha Bhattarai

2021
To:  Dean Michael R. Heithaus  
    College of Arts, Sciences and Education  

This dissertation, written by Nisha Bhattarai, and entitled Structural Dynamics of Membrane Interacting Viral Proteins, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_______________________________________
Yuan Liu

_______________________________________
Jin He

_______________________________________
Prem P Chapagain, Co-Major Professor

_______________________________________
Bernard S Gerstman, Co-Major Professor

Date of Defense: May 28, 2021

The dissertation of Nisha Bhattarai is approved.

_______________________________________
Dean Michael R. Heithaus  
    College of Arts, Sciences and Education  

_______________________________________
Andrés G. Gil  
    Vice President for Research and Economic Development and Dean of the University Graduate School  

Florida International University, 2021
ACKNOWLEDGMENT

Firstly, I would like to wholeheartedly thank my two amazing mentors, Professor Bernard S Gerstman and Professor Prem P Chapagain for their invaluable and continuous support, motivation and encouragement during my graduate career. I feel myself very lucky to have such wonderful mentors who not only gave me insights into research work but helped me in so many ways professionally and personally. Under their mentorship, I learned various skills in problem solving, critical thinking, writing manuscripts and professional presentation. They always motivated me and gave a wonderful lab environment and resources to accomplish my projects. My Ph.D. experience was enjoyable and productive because of their ideas and their valuable time, and I am very grateful to them.

I would like to express my appreciation to my committee members, Prof. in He and Prof. Yuan Liu for their suggestions and comments which helped to improve my research and to develop new ideas which led to successful completion of this dissertation. I would like to sincerely thank my collaborator Prof. Robert Stahelin from Purdue University for his collaboration on Ebola and Marburg virus project. His experimental finding on virus helped to understand and interpret the results. I am also thankful to Dr. Jin He and Dr. Miksovska for providing me opportunity to collaborate in their project. Furthermore, I am thankful to FIU Castic team, Barry Branch and FIU HPC team for providing computer resources and helping in solving various technical issues. I would also like to acknowledge FIU Physics Department Faculty and staffs. I am also grateful to have friendly past and current lab members Jeevan, Tim, Rudra, Prabin, Lokman, Elumalai, Micheal, Hugo, Tej for their ideas and suggestions during my research. I greatly appreciate University graduate
school at FIU for providing me dissertation year fellowship which helped me to wrap up my thesis on time.

I do not have any words to express my emotions for my parents. Their love, sacrifice and support in every step of my life encouraged me to accomplish my dream. I am very thankful to my siblings, relatives, and friends for their continued help. Last but not the least, I am very thankful and feel proud to have such a loving, understanding and inspiring husband who was always there for me in my ups and downs, encouraged me, gave me positive mindset every time.
ABSTRACT OF THE DISSERTATION

STRUCTURAL DYNAMICS OF MEMBRANE INTERACTING VIRAL PROTEINS

by

Nisha Bhattarai

Florida International University, 2021

Miami, Florida

Professor Bernard S Gerstman, Co-Major Professor

Professor Prem P Chapagain, Co-Major Professor

Viruses do not possess complete cellular machinery but have the ability to reproduce by utilizing cellular machinery inside host cells. They are nanoscale machines that rapidly modify (evolve) their molecular components to cause disease and death. Therefore, emergence of deadly infectious viruses is a monumental health concern and understanding how viruses can enter, replicate, assemble and egress from the host cell is important to mitigate the threat.

A fully active, infectious viral structure is known as a virion. A virion contains genetic material and is enclosed by a capsid. The capsid is a protein shell, and some viruses also are coated by a lipid membrane. My research focuses on viral proteins that interact with lipid membranes in host cells. The lipid molecules can be part of the cellular membrane or part of lipid structures within a cell, such as the endoplasmic reticulum. In my research, I used molecular dynamics computational techniques to study the interactions of the filovirus matrix protein, also known as VP40, of Ebola and Marburg viruses with the lipid molecules in the human plasma membrane. VP40 proteins associate in the inner layer of the plasma membrane and oligomerize to form the matrix that gives the shape of
the virion particle. My research focuses on the membrane association, membrane transportation of VP40, which represents the early stage of the virus’ assembly inside the cell. I have identified the amino acids playing important roles in both membrane association and conformational flexibility. I also investigated the Zika virus NS1 protein association on the outer layer of endoplasmic reticulum, which is where the Zika virus forms the virion. The aim of my research is to use molecular level understanding of the virus life cycle to develop improved molecular interventions such as designing drug molecules that disrupt the functions of the VP40 and NS1 proteins, for prevention and cure of viral diseases.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CHAPTER 1: INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Viruses</td>
<td>1</td>
</tr>
<tr>
<td>1.2 VP40 Proteins in the <em>Filovirus</em> Genus (Ebola and Marburg Viruses)</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Flaviviruses (Zika virus)</td>
<td>5</td>
</tr>
<tr>
<td><strong>CHAPTER 2: METHODS</strong></td>
<td>7</td>
</tr>
<tr>
<td>2.1 Molecular dynamics (MD) simulations:</td>
<td>7</td>
</tr>
<tr>
<td>2.1.1 Bonded and non-bonded interactions</td>
<td>7</td>
</tr>
<tr>
<td>2.1.2 Initial atomic coordinates</td>
<td>8</td>
</tr>
<tr>
<td>2.1.3 Simulation of atomic motion</td>
<td>9</td>
</tr>
<tr>
<td>2.1.4 All atom molecular dynamics (AAMD) simulation</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Steered Molecular Dynamics (SMD)</td>
<td>11</td>
</tr>
<tr>
<td>2.3 Dynamic Network Analysis</td>
<td>12</td>
</tr>
<tr>
<td>2.4 Principal Component Analysis (PCA)</td>
<td>13</td>
</tr>
<tr>
<td>2.5 Normal Mode Analysis and Dynamic Cross Co-relation:</td>
<td>13</td>
</tr>
<tr>
<td><strong>CHAPTER 3: Plasma Membrane Association Facilitates Conformational Changes in mVP40 Dimer</strong></td>
<td>15</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>15</td>
</tr>
<tr>
<td>3.2 Materials and methods</td>
<td>18</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>19</td>
</tr>
<tr>
<td>3.3.1 Plasma membrane association of the mVP40 dimer</td>
<td>22</td>
</tr>
<tr>
<td>3.3.2 Lipid-protein interactions and lipid specificity</td>
<td>24</td>
</tr>
<tr>
<td>3.3.3 Conformational changes in mVP40 dimer</td>
<td>28</td>
</tr>
<tr>
<td>3.4 Conclusion</td>
<td>34</td>
</tr>
<tr>
<td><strong>CHAPTER 4: Conformational Flexibility of the Protein-Protein Interfaces of the Ebola Virus VP40 Structural Matrix Filament</strong></td>
<td>36</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>36</td>
</tr>
<tr>
<td>4.2 Materials and methods</td>
<td>39</td>
</tr>
<tr>
<td>4.3 Results and discussion</td>
<td>41</td>
</tr>
<tr>
<td>4.3.1 Flexibility of the interface structures</td>
<td>41</td>
</tr>
<tr>
<td>4.3.2 Bending Flexibility of Interfaces</td>
<td>47</td>
</tr>
<tr>
<td>4.3.3 Interaction of the interface structures</td>
<td>48</td>
</tr>
<tr>
<td>4.3.4 Dynamical network analysis</td>
<td>49</td>
</tr>
<tr>
<td>4.3.5 Normal mode analysis</td>
<td>54</td>
</tr>
<tr>
<td>4.3.6 Principal component analysis</td>
<td>56</td>
</tr>
<tr>
<td>4.4 Conclusions</td>
<td>58</td>
</tr>
</tbody>
</table>
CHAPTER 5: Ebola Protein VP40 Binding to Sec24c for Transport to the Plasma Membrane ................................................................. 59
  5.1 Introduction ........................................................................................................................................................................ 59
  5.2 Protein-protein docking .......................................................................................................................................................... 62
  5.3 Results and Discussion ........................................................................................................................................................... 63
    5.3.1 Sec24c and VP40 docking ............................................................................................................................................. 64
    5.3.2 Molecular interactions determined from MD computational simulations ........................................................................ 66
    5.3.3 H-bond analysis for the Sec24c-VP40 wild-type complex .............................................................................................. 69
    5.4.4 Biding region Dynamic Network Analysis: ......................................................................................................................... 71
    5.5.5 Comparison of Sec24c-VP40-WT complex vs VP40 point mutations .......................................................... 72
    5.5.6 Additional Docking complex ........................................................................................................................................... 74
  5.4 Conclusion .................................................................................................................................................................................. 77

CHAPTER 6: Dynamics and Membrane Association of Zika Virus Protein NS1: A Molecular Dynamics Study ................................................................. 79
  6.1 Introduction .................................................................................................................................................................................. 79
  6.2 System Set-up ........................................................................................................................................................................... 81
  6.3 Structure of NS1: ...................................................................................................................................................................... 82
  6.4 Result .......................................................................................................................................................................................... 83
    6.4.1 NS1 - Membrane localization: ........................................................................................................................................... 83
    6.4.2 Insertion of Hydrophobic residues ...................................................................................................................................... 84
    6.4.3 Root mean square Deviation .............................................................................................................................................. 86
    6.4.4 Lipid-count and Relative Percentage of H-bonds ............................................................................................................. 88
    6.4.5 Root Mean square fluctuation .......................................................................................................................................... 90
    6.4.6 Interaction Energy between protein and membrane: ...................................................................................................... 90
    6.4.7 Change in electrostatic potential distribution ............................................................................................................... 91
    6.4.8 Conformational changes in ZIKV Dimer ............................................................................................................................ 93
  6.5 Conclusions .................................................................................................................................................................................. 95

CHAPTER 7: Overall Summary ............................................................................................................................................................ 97

VITA ............................................................................................................................................................................................................. 116
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Figure 1.1</strong> Crystal Structure of: a) Ebola VP40 where NTD and CTD are colored in silver and purple respectively, b) Zika NS1 protein where different domains are colored differently</td>
</tr>
<tr>
<td><strong>Figure 2.1</strong> Bonded interactions between atoms of a molecule</td>
</tr>
<tr>
<td><strong>Figure 3.1</strong> Structure of the a) mVP40 dimer and b) eVP40 dimer. The loop 1 residues are highlighted in blue, and loop 2 residues are highlighted in red. C) Electrostatic (Eel) and van der Waals (Evdw) energy between the monomers in both eVP40 and mVP40 dimers, calculated using NamdEnergy plugin in VMD</td>
</tr>
<tr>
<td><strong>Figure 3.2</strong>. (a-c) Snapshots of the mVP40 dimer association with the plasma membrane at different times. d) Various lipid types interacting with the basic loop 1 and basic loop 2 residues at 300 ns. The lipids are colored as: POPS-cyan, POPI-green, POPC-gray, POPE-purple</td>
</tr>
<tr>
<td><strong>Figure 3.3</strong> Time evolution of structural parameters during mVP40 dimer-membrane association. a) distance (along the z-axis) between the center of masses of the protein and the lipid bilayer (calculated using the center of mass of the phosphorous atoms). b) The number of hydrogen bonds between the lipid head groups and the protein. The red and blue curves represent the hydrogen bonds formed with loops 1 and 2, respectively</td>
</tr>
<tr>
<td><strong>Figure 3.4</strong>. E260 interactions with the PM. a) Lipid molecules clustered around E260 (purple: POPE, cyan: POPS). Extensive network of hydrogen bonds formed between E260, POPE, and POPS. C) Lipid-protein interactions between the basic patch residues in the mVP40-CTD and the plasma membrane lipids</td>
</tr>
<tr>
<td><strong>Figure 3.5</strong>. Lipid-protein interactions for the mVP40 dimer at the lower leaflet of the PM. a) The number of contacts between the lipids and protein heavy atoms for various lipid types. b) Relative percentage of the hydrogen bonds between the lipids and each interacting protein residues. Inset: Relative percentage of the hydrogen bonds made with protein by each lipid type</td>
</tr>
</tbody>
</table>
Figure 3.6. Time evolution of the angle (in degrees) between the monomers in the presence (purple) and absence (green) of lipid interactions for the a) mVP40 dimer and b) eVP40 dimer.

Figure 3.7a. Root-mean-squared fluctuations (rmsf) for all the residues in the mVP40 dimer at various time windows. As the protein associates with the membrane, the flexibility of all the residues decreases in general. The lipid interactions affect the CTD rmsf more than the NTD.

b) Time course of the flexibility of the basic patch residues (K210, K211, R215, K218, K259, K264, K265, and R266). For each point, the rmsf values of the basic loop residues were calculated for a time window of 5 ns and averaged over the residues of the basic patch.

Figure 4.1. (A) A section of the EBOV VP40 composing of two hexamers connected through end-to-end ‘unsprung’ CTDs. Three distinct interfaces are highlighted by dashed circles: (B) NTD-NTD oligomerization to connect dimers, (C) CTD-CTD interface to connect hexamers end-to-end, and (D) NTD-NTD dimer interface between monomers to form a dimer. For clarity, the ‘sprung CTDs’ are not shown in ‘B’ for the dimer-dimer oligomer interface, the NTDs are not shown in ‘C’ for the CTD-CTD hexamer interface, and the ‘unsprung’ CTDs are not shown in ‘D’ for the monomer-monomer dimer interface.

Figure 4. 2. The RMSD of the different interfaces as a function of MD simulation time.

Figure 4.3. Fluctuations in the angle between the best-fit line of the Cα atom coordinates of the two domains for each interface as a function of time.

Figure 4.4 Fluctuations in the distance between the structural elements of each group as a function of simulation time.

Figure 4.5. (A, B, C) Visualization of the structural elements located at the three interfaces; (D, E, F) fluctuations in the angle between the best-fit line of the Cα atom coordinates of the structural elements of each group as a function of simulation time. The group elements are highlighted in A, B and C, and the corresponding colors are used in the plot.

Figure 4.6 Root-mean-square fluctuations (RMSF) computed from the last 100 ns trajectory for the Cα atoms of the interface structural elements.
Figure 4.7. (A) Schematic showing the transect line set-up for using SMD to measure the resistance to flexing at the domain interface. (B) An amino on the transect line at the outer edge of each domain is immobilized and an amino acid of each domain at the interface are both pulled perpendicular to the transect line. (C) The pulling force necessary to flex each interface at the same constant speed ..........................................................48

Figure 4.8. Inter-domain interaction energy computed between domains for the interface structures as a function of simulation time. (A) electrostatic, (B) van der Waals ..........................................................49

Figure 4.9. Dynamic network community analysis from the last 100 ns of the MD trajectories. Different communities are shown in different colors. The line thickness between the amino acids represents how strongly the motions of residues within the community are correlated. (A) CTD-CTD, (B) dimer, (C) oligomer. Amino acid communities formed by residues that span both domains are shown in (D) CTD-CTD, (E) Dimer, (F) Oligomer ........................................................................50

Figure 4.10. One of the low frequency internal modes for each interface. Each figure is the superposition of the Cα NMA trajectory and the gray shading shows regions with large motion. ................................55

Figure 4.11. Dynamical cross-correlation matrix demonstrating the correlation of motion between Cα atoms of residues in one domain (horizontal axis) with respect to residues on the other domain (vertical axis). The circles highlight the strong interdomain correlations between amino acids that are listed in Table 4.1 and Figure 4.5 .................................................................56

Figure 4.12. From Principal Component Analysis, projections of the trajectories onto planes by the first three eigenvectors. (A) Projection of trajectories into PC1, PC2, and PC3 for the CTD-CTD interface, (D) Projection of trajectories into PC1, PC2, and PC3 for the Dimer interface and (G) Projection of trajectories into PC1, PC2, and PC3 for the Oligomeric interface. The color scale from blue to white to red shows that there are periodic jumps between these conformers through the trajectory. (B) Visualization of the molecular motions along (B) PC1 and (C) PC2 for CTD-CTD, (E) PC1 and (F) PC2 for the Dimer, and (H) PC1 and (I) PC2 for the Oligomeric interfaces. The gray shading shows regions with large motion. ..................................................57
Figure 5.1. (A) Crystal structure of the human Sec24c trimer (PDB: 3EH2); the three chains are colored differently. (B) Crystal structure of the Ebola virus VP40 dimer: NTDs (green) and CTDs (pink). (C) Monomer of Sec24c: zinc finger (yellow), trunk domain (gray), beta-sandwich (purple), helical-domain (blue), gelsolin-like domain (red). (D) Monomeric structure of the VP40 protein.

Figure 5.2. A) Sequence alignment of Sec24 (blue) and Sec24c (red) highlighting different binding sites. Sec23 (yellow) binding to Sec24 in the trunk domain region is also shown. (B) Docking complex of Sec24c (red) with VP40-CTD (brown).

Figure 5.3. For (A) and (B), residues involved in the inter-protein interactions of the complex are highlighted: Sec24c (yellow) and VP40 (green). (A) Docked Sec24c-VP40 complex at 0 ns. (B) Sec24c-VP40 complex after 500 ns of MD simulation. (C) Surface representation of (B).

Figure 5.4. (A) Root mean square deviation of Ca atoms of Sec24c and VP40 with respect to time. (B) Interaction energy between Sec24c and VP40 as a function of time: Total Energy (black), Electrostatic (cyan), van der Waals (VDW) (grey). (C) The distance between the N and O atoms in the residues involved in four important salt-bridges. (D) Number of hydrogen bonds between Sec24c and VP40 continues to increase after the system has stabilized.

Figure 5.5. (A) Percentage of hydrogen-bond occupancy for different inter-protein amino acid pairs where charged interactions are colored in yellow. Inset: Hydrogen-bond occupancy of specific residues on VP40 (left column) and Sec24c (right column). (B) Residues involved in important hydrogen-bonding and salt-bridge interactions for the Sec24c-VP40-Wt complex. VP40 is on the left and Sec24c on the right.

Figure 5.6. (A): VP40-CTD (brown) in the binding pocket of Sec24c(white). The trunk domain of Sec24c is displayed in red. (B) Residues in VP40 interacting in the binding pocket of Sec24c. (c) DNA communities between Sec24c-VP40-WT, where each community is colored differently (green, blue, yellow).

Figure 5.7. Sec24c and VP40 complex A) WT B) M305A C) I307A D) L303A where Sec4c is represented in red color and VP40 in
brown. The mutated residue on VP40 (Figs. B-D) is colored differently. E) Number of inter-protein hydrogen bonds as a function of time for all four complexes. F) Total interaction energy between Sec24c and VP40 with respect to simulation time.

Figure 5.8. Sec24c-VP40-AD complex with highlighting of the VP40 binding region (green) and the Sec24c binding region (yellow). MD structure of the complex at: A) 0 ns, B) 500 ns. C) Important residues at the protein-protein interface. Hydrophobic residues (gray) form a strong hydrophobic core that contributes to the stability of the complex. The polar, negative, and positive residues at the interface are shown in green, red, and blue surface representation, respectively. D) View showing some interprotein residue-residue interactions between VP40 and Sec24c.

Figure 6.1: Crystal structure of NS1 ZIKV dimer, beta-ladder(gray), wing domain(yellow), intertwined loop (green) and beta-roll (magenta). The residues N130 and N207 are glycosylated with sugar group (BGL group). Important hydrophobic residues for membrane localization are highlighted.

Figure 6.2: Snapshots of protein association with membrane at different trajectory, where different regions of NS1 are shown; wing domain (yellow and green), beta ladder domain (white), beta roll of two monomers (purple and blue) for: (a) 5 ns (b) 1000ns. c) Few residues inserted in lipid bilayer after protein-membrane localization d) Lipids interacting with different residues of protein within 3.5 A of protein.

Figure 6.3: Amino acids important for membrane stabilization and association in (a) Wing domain of NS1 with intertwined loop (green), (b) NS1 beta roll.

Figure 6.4: (a) Residues penetrating in the lipid bilayer. Orientation of NS1 where cyan represents phosphate atoms (b) before lipid interaction begins and (c) after lipid interaction where, dotted circle indicates the region of void formation where residues begin penetrating are also shown.

Figure 6.5: (a) Time evolution of number of hydrogen bonds for Chain A (cyan), Chain B (gray) and Protein (red). (b) Root mean square deviation (rmsd) of ZIKV dimer with respect to time.
Figure 6.6: (a) Lipid counts between ZIKV dimer and membrane with respect of simulation time. (b) Histogram representing relative percentage of hydrogen bonds formed by different residues of ZIKV dimer with lipid molecules. c) Different residues in beta roll region interacting with lipids d) Amino-acids interacting with lipids in wing domain e) Hydrogen-bond formed by Y122 with lipid. .................................................................89

Figure 6.7. Root mean square fluctuation (rmsf) for a) Chain A b) Chain B and c) Wing domain of NS1 protein dimer. .................................................................90

Figure 6.8. Interaction energy between protein and membrane with a function of time. .................................................................91

Figure 6.9. Electrostatic potential distribution of NS1 protein at 0ns and at around 800 ns with top and bottom view shown........................................92

Figure 6.10. Salt bridge between D157 and K191 and K227 and D234 at a) 0 ns and 1000 ns. b) Center of mass distance between D157 and K191(green), D234-K227(red) with respect to simulation time. ........................................................................................94

Figure 6.11. Mass density profile of NS1 for first 10ns and last 10ns at X, Y and Z-direction. .................................................................95
CHAPTER 1: INTRODUCTION

1.1 Viruses

Viruses are contagious agents which are only capable of replicating their genes inside a living host cell\(^1\). Since a virus does not possess complete cellular structure but does have the ability to reproduce by utilizing cellular machinery inside living cells, they are sometimes described as “organisms at the edge of life”\(^2\). Viruses are nanoscale machines that rapidly modify (evolve) their molecular components to create disease and death. Viruses can also be engineered to perform beneficial functions. However, the harm created by deadly infectious viruses is a monumental health concern. In my research, I used molecular dynamics computational techniques to study the interaction with lipid molecules of the Ebola and Marburg virus’ VP40 proteins and the Zika NS1 protein. The aim is to use molecular level understanding of the virus life cycle to develop improved molecular interventions for prevention and cure of viral diseases. These investigations are on the forefront of bio-nanophysics and molecular medicine. The elegant simplicity of the laws of physics when combined with the random events of evolutionary natural selection allows viruses to use a range of different molecular interactions to create proteins that facilitate viral life cycles. This is a specific example of the underlying tension of life: organismal strength versus adaptability, molecular stability versus flexibility, thermodynamic enthalpy versus entropy.

Different viruses have different shapes: filamentous, helical, and icosahedral. Spheroidal viruses are generally 20 to 300 nm in size whereas filamentary viruses can be up to 14,000 nm in length (e.g. filoviruses)\(^3\). A fully active, infectious viral structure is
known as a virion. A virion contains nucleic acids (RNA/DNA) and is enclosed by a capsid. The capsid is a protein shell, and some viruses also are coated by a lipid membrane. My research focuses on viral proteins that interact with lipid membranes in host cells. The lipid molecules can be part of the cellular membrane or part of lipid structures within a cell, such as the endoplasmic reticulum.

In addition to the proteins in the capsid shell, viruses contains other proteins that assist in various stages of the viral life cycle such as: viral entry inside the host cell, viral RNA/DNA replication within the cell, and viral budding\(^4,\,5\) when new viruses leave the host cell. Many viruses have viral proteins which interact with host cell lipid membranes. When these viral proteins bind to the host lipid molecules, the lipid binding can induce viral protein conformational changes that modulate the protein’s functions\(^6-\,11\). These lipids binding viral proteins are also commonly called membrane-associated proteins. For example, Zika and Dengue NS1 proteins associate with the cell’s endoplasmic reticulum membrane, and the Ebola and Marburg VP40 proteins interact with the cell’s bounding plasma membrane. The binding to lipid molecules causes structural rearrangements of the proteins that leads to higher order oligomeric structures\(^12,\,13\). Some of these membrane-associated proteins, are also major “matrix proteins” and are the major structural building block for the virus. One such example is the VP40 proteins found in filoviruses. The protein is the building block for the structural matrix of the virus\(^12\). It is also responsible for wrapping the virus with a lipid envelope using the membrane of the host cell. Other examples include the influenza membrane proteins hemagglutinin (HA), neuraminidase (NA) and M2 that localize at the plasma membrane (PM) and form a lipid envelope supported by the matrix protein M1 that also interacts with and facilitates the assembly of
the internal capsid that encloses the virus’ nuclear material\textsuperscript{14}. Similar to the filoviruses, the matrix proteins of the type 1 human immunodeficiency virus (HIV) matrix protein form a structural matrix that is coated by a lipid envelope\textsuperscript{12, 15}. By wrapping itself in a coating of the host cell’s membrane, a virus is more likely to escape detection by the body’s molecular immune system.

Viruses are amazingly adaptable due to rapid mutations on the molecular level. The recent Sars-Cov2 outbreak has proved that there is always a looming threat of new viral outbreaks and the study of viruses is important for creating vaccines to fight existing outbreaks and prevent new outbreaks\textsuperscript{16}. My research work focuses on understanding on a molecular level the structural dynamics and functioning of three dangerous zoonotic viruses, namely Ebola, Marburg and Zika. Ebola and Marburg belong to the genus \textit{filovirus}, which is part of the family Filoviridae. Zika is a member of the \textit{flavivirus} genus, which is part of the Flaviviridae family. I focused my attention on viral proteins that interact with cellular membranes. Using molecular dynamics computational simulations, I investigated the atomic details of the conformational dynamics, flexibility, and membrane interaction mechanisms of the Ebola and Marburg virus’ VP40 proteins and the Zika virus NS1 protein.

1.2 VP40 PROTEINS IN THE \textit{FiloVirus} GENUS (EBOLA AND MARBURG VIRRUSES)

The \textit{filovirus} genus of viruses includes Ebola, Marburg, and Reston viruses. These are transmitted through bats, and cause hemorrhagic fever in humans and nonhuman primates that often results in high mortality rates\textsuperscript{3, 17}. The worm-like cylindrical structure of the virion is approximately 80 nm in diameter and can be as long as 14,000 nm in length.
The virion carries a negative sense RNA genome which codes for seven different proteins: nucleoprotein (NP), nucleocapsid protein VP35, major structural matrix protein VP40, glycoprotein GP, transcription factor VP30, VP24, and L (RNA polymerase). NP is responsible for covering the genetic material in a nucleocapsid, VP35 is also a nucleocapsid protein and binds with RNA in order to suppress cellular immunity, VP30 is a transcription factor, VP40 is the major structural matrix protein, GP facilitates viral entry through the host cell membrane, VP24 is minor matrix protein, and L protein, also known as RNA polymerase, plays a role in RNA replication\textsuperscript{18, 19}.

My research focused on the major structural matrix protein VP40 of the Ebola and Marburg viruses. The Ebola and Marburg VP40 proteins have small but important differences between them. The virus particle acquires a lipid coat from the plasma membrane (PM) of the host cell as the virus assembles and buds. The budding is driven by the association of VP40 on the inner surface of the cell’s PM lipid bilayer\textsuperscript{18, 19} followed by the new virus particle pushing out on the PM. By enveloping itself in the host cell’s PM, the virus is able to avoid detection by the organism’s immune defenses.

Because the binding of VP40 to the host cell’s PM is crucial in the life cycle of the virus, it is important to understand the structural dynamics of VP40 when it interacts with the lipid bilayer. The viral structural matrix layer underneath the lipid envelope is formed by the matrix protein VP40, and provides shape and stability to the viral particle\textsuperscript{12}. It has been shown in mammalian cells that the expression of only the VP40 protein but none of the other proteins in the EBOV and MARV genome is sufficient for forming non-infectious but authentic looking virus-like particles (VLPs)\textsuperscript{19}. 
In addition to being the major structural matrix protein, VP40 is involved in multiple other functions during the viral life cycle. Protein multi-functionality often requires proteins to undergo large structural conformational changes. It has been shown that, depending on the function, EBOV VP40 (eVP40) exists in different conformations such as a butterfly shaped dimer during transport of the protein inside the cell to the plasma membrane, a hexamer to form the viral structural matrix on the inner surface of the PM, and an octamer ring structure to bind RNA and regulate viral transcription\textsuperscript{12}. The X-ray crystal structure determination of the dimeric, hexameric, and octameric forms of eVP40 provides a basis for investigating the structural transformations of the protein into various oligomeric states for performing different functions\textsuperscript{12}. Structural information of such viral proteins also offers an opportunity to computationally explore potential inhibitors for treating the disease.

1.3 Flaviviruses (Zika virus)

Another genus of viruses with membrane-interacting proteins of interest is flavivirus, which includes the mosquito-borne Zika virus (ZIKV)\textsuperscript{20}. ZIKV has evolved as a major global health threat due to its link with severe clinical manifestations, including microcephaly in neonates and Guillain-Barre-syndrome in adults\textsuperscript{21-23}. The rapid spread of ZIKV around the globe has made it one of the most significant public health concerns.

ZIKV is a positive sense RNA virus, and its genome encodes ten different proteins, three of which are structural and seven are non-structural. The structural proteins are capsid protein (C), envelope protein (E), and precursor membrane prM protein\textsuperscript{24}. The M protein is responsible for monitoring the assembly of the envelope proteins. The seven non-
structural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) mainly function in RNA replication, immune evasion, and particle assembly. I studied the glycosylated NS1 protein which interacts with the host cell’s endoplasmic reticulum (ER) membrane where it clusters to form a dimeric structure inside the cell and assists with the viral replication process after infection. Studies suggested that it is the dimeric form of NS1 that interacts with the membrane, which allows the dimers to cluster together at the regions of lipid droplets in the membrane.

Figure 1.1 Crystal Structure of: a) Ebola VP40 where NTD and CTD are colored in silver and purple respectively, b) Zika NS1 protein where different domains are colored differently
CHAPTER 2: METHODS

2.1 MOLECULAR DYNAMICS (MD) SIMULATIONS:

Molecular dynamics (MD) simulation is widely used computational technique in order to study various biological processes\textsuperscript{28, 29}. It is computer simulation method which is used to study physical movement of atoms or molecules for a given period of time. MD is very useful tool to investigate conformational changes in the proteins, protein-protein interactions, protein-membrane interactions, drug discovery and many more\textsuperscript{30}. In my Ph.D., I employed MD simulation tools in my all projects. In MD simulations, large biomolecules are treated classically as a system of mass points with an energy function $U_{\text{total}}$ or $U(r)$ used to model the interactions between these atoms. This potential energy function as given in equation 2.1, is also often called the “force field” in MD investigations.

$$U_{\text{total}} = \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_{\theta} (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_{\phi} \left[1 + \cos(n\phi - \phi_0)\right]$$

$$+ \sum_{\text{van der Waals}} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \sum_{\text{electrostatic}} \frac{q_i q_j}{4 \pi \epsilon r_{ij}}$$

(2.1)

2.1.1 BONDED AND NON-BONDED INTERACTIONS

Figure 2.1 diagrams various interaction energies in the MD force field of eq. 2.1 involving electron bonds in a sequence of amino acids. These energy terms include bond, angle, and dihedral terms. The first term in eq. 2.1 shows that a harmonic potential is used to represent covalent bonds between two adjacent atoms. In similar manner, the bond angle interaction is also calculated by applying a harmonic potential for the angle defined by
three adjacent atoms. For four consecutive atoms covalently bonded in a linear sequence, which is commonly referred to as a dihedral interaction where, $\phi$ is the torsional angle of a central bond to two covalent bonds.

![Dihedral interaction](image)

Figure 2.1 Bonded interactions between atoms of a molecule.

Other interactions include “improper terms” and non-bonded interactions, where improper interactions are used to describe four planar atoms, and non-bonded interactions includes the van der Waals energy described by a Lenard-Jones potential, and electrostatic interaction energy represented by Coulomb’s law. These non-bonded interactions can occur between every pair of atoms in the entire protein and lipid molecular system, which can include thousands of atoms. However, non-bonded interactions weaken quickly as a function of distance and become negligible. Therefore, substantial computational time could be wasted by calculating insignificant non-bonded interactions between atoms that are well separated. To avoid wasting computational time, non-bonded interactions are cut-off with a switching function that is generally set between 12-14 Å in MD simulations.

2.1.2 Initial Atomic Coordinates

The initial co-ordinates of the atoms for the MD simulations are obtained from the Protein Data Bank computer server, commonly known as PDB, where thousands of crystal
structures from X-ray crystallography, NMR, and cryo-EM experiments of proteins are deposited. Given the initial positions and bonds of the atoms, MD computer applications assign initial velocities to the atoms using a Maxwell-Boltzmann distribution appropriate for a chosen temperature.

2.1.3 Simulation of Atomic Motion

The position and velocity of each atom are updated at user-defined time-steps using Newton’s equations of motion and sophisticated finite-difference mathematical techniques. The force is calculated from the gradient of the potential function given in equation 2.1 and used to calculate the acceleration for updating the position and velocity for the atoms. The time-step is often set to 2 fs, which captures tiny vibrational motions of atoms and approximates an infinitesimal time for use in $\frac{1}{2}at^2$. Shorter time-steps can be used but add to the computational time, add more steps that contribute to computational errors related to numerical round-off uncertainties, but do not contribute additional important details that might make the simulations more realistic.

The position and velocity of all the atoms at each time step can be saved to a trajectory file for analysis of the physical and chemical behavior of the system, as well for creating a movie for visualization of the evolving system. However, saving the position and velocity for thousands of atoms every 2 fs for microsecond simulations create data files that are unwieldy to analyze. In addition, femtosecond details usually do not contain important information. Therefore, the trajectory files usually contain frames of the position and velocity of the atoms at picosecond intervals (every 500 time-steps). Various statistical tools are used to analyze the data from this MD trajectory. As in MD simulation, 2 fs time step is mostly used to capture the atomic vibrational events it is very difficult to observe
significant re-arrangements in the proteins which generally requires large simulation time. Fortunately, due to development in GPUs, today it is possible run MD simulations for long period of time, up to millisecond time scale, which has been achieved by some of the research labs with very advanced super computers. Although it is impossible to run big systems even with GPUs, there are many other MD simulation techniques like Coarse Grained MD, Steered MD, Targeted MD that minimizes the amount the time. In my work, I have employed some of these tools to understand protein-protein interactions and dynamics.

2.1.4 ALL ATOM MOLECULAR DYNAMICS (AAMD) SIMULATION

All atom molecular dynamics simulations were performed using Charmm36 force fields and were run using NAMD 2.11 and NAMD 2.12 software. All the simulations were performed in explicit water environment with periodic boundary conditions. All protein only set-up and protein-membrane set-up, (both with and without membrane) were built using the Charmm-Gui solution builder and membrane Builder web interface. The standard plasma membrane composition was used for VP40-membrane construction which consists of phosphatidylcholine (POPC), phosphatidylethanolamine (POPE), phosphatidylserine (POPS), palmitoylsphingomyelin (PSM), phosphoinositol (POPI) and cholesterol (CHOL) molecules. The distribution of different lipid molecules in the lower leaflet of the plasma membrane were in the number ratio of 20:11:33:18:9:7 (CHOL:POPC:POPE:POPS:POPI:PSM) representing the high complexity of the plasma membrane composition. For NS1-membrane set-up endoplasmic reticulum membrane ratio was used to build the system. All the systems were solvated with TIP3 water model.
in cubic boxes and neutralized with 0.15 M KCl. In order to constraint hydrogen atoms, SHAKE algorithm\textsuperscript{37} was used and to calculate long range electrostatic interactions Particle Mesh ewald (PME) method was used\textsuperscript{38}. A Nose-Hoover Langevin piston method was used to control pressure with piston period\textsuperscript{39} of 50 fs and decay of 25fs. For temperature control, Langevin temperature coupling with friction coefficient of 1ps\(^{-1}\) was employed. All the simulations were performed at 300K temperature. Hydrogen bond analysis was performed using vmd h-bond plugin with 3.5 A distance cutoff and 30 degrees angle cutoff. For image rendering purposes, VMD was used\textsuperscript{40}.

### 2.2 Steered Molecular Dynamics (SMD)

SMD simulation was used to study the conformational flexibility of VP40 proteins. The basic idea of SMD simulation is to apply force to atom or group of atoms. During SMD simulation, the atoms where the external force is applied is refer as smd atom and another group of atoms are fixed which is refer as fixed atoms. During SMD, force is applied to dummy atom which is attached to SMD atoms through virtual spring\textsuperscript{41}. SMD is very useful tool to study biological processes on time scales accessible MD simulation techniques\textsuperscript{41, 42}. In this method, harmonic constrained with spring constant \(k\) is applied to SMD atom in a desired direction with velocity ‘\(v\)’ which changes the potential of the system. Equation 2 gives the potential energy of the system and \(r_0\) and \(r_1\) represents center of mass of smd atoms at \(t=0\) and \(t=n*\text{dt}\), where \(\text{dt}\) is the time step used in simulation\textsuperscript{32}. SMD was used to determine the bending flexibility for each of the three interfaces. For the SMD simulations, the equilibrated structures of the interfaces were used. We constrained the atoms at the outer edges of the structure and pulled on the \(C_\alpha\) atoms at an interface in a
direction that produces a bending motion at the interface. Later, in Figure 5, we provide a detailed description of the $C_\alpha$ atoms that were chosen. We used a spring constant $k$ of 1 kcal/molÅ$^2$ and pulled at a constant speed of 0.4 Å/ns with a time step of 1 fs for 10 ns for all three interfaces. These choices were made to create a bending stress at the interface with disturbing the internal structures of the domains, as explained in the Results section. The resistance offered to each interface was observed by plotting force-time graphs.

2.3 **Dynamic Network Analysis**

Dynamic network analysis is applied to understand the co-related atomic motions of the protein. It provides the detailed information in the form of community which highlights the important amino-acids responsible for connecting domains and protein-protein interface.$^{43}$ The NetworkView plugin$^{40,43,44}$ in VMD was employed to perform the dynamical network analysis for all interface systems. Carma$^{45}$ software is used to calculate covariance and correlations between pairs of amino-acids and Catdcd is used to break down the trajectory file for Carma calculations. In molecular dynamics simulations, the amino-acids co-relation motion are studied in order to generate communities. In each community, amino acids in the network are represented by a node centered at the $C_\alpha$ for each residue and co-relation values represent the weighted edge between the nodes. Nodes are connected by edges if they are within a cut-off distance of 4.5 Å for at least 75% of the simulation time. The shortest path between the nodes for communication are identified. The communities in the network are defined as the time-averaged connectivity of the nodes and were constructed using the Girvan-Newman algorithm$^{46}$. The thickness of edges represents the co-relations that define the probability of information transfer along the edge. The
strength and size of a community is determined based on many factors including the weight of edges joining nodes and communities, the number of connecting edges, etc. In order to understand the conformational flexibility of VP40 proteins, to explore VP40 and Sec24c protein dynamic network analysis was performed.

2.4 Principal Component Analysis (PCA)

PCA analysis was used to explore the coordinated motions of the atoms of the three interface structures of VP40 protein and were further examined with the use of the PCA function of the Bio3D software package. Default settings of the program were used for the PCA calculations, and we analyzed the last 200 ns of the MD trajectories. The least squares fitting method was used to remove the overall translational and rotational motions of the trajectories. Subsequently, the PCA function performs a diagonalization of the variance-covariance matrix of the data points of the system based on the mutual information between all Cα atoms in the interface structures. The diagonalization of the covariance matrix produces eigenvectors, each with its corresponding eigenvalue. The eigenvectors indicate the direction of the motion of atoms while the eigenvalues represent the magnitude of the motion.

2.5 Normal Mode Analysis and Dynamic Cross-Correlation:

The low-frequency collective motions of the three interface structures were analyzed by the elastic network model (ENM) of NMA with the use of the NMA function of the Bio3D software package. The default settings of the Bio3D program were used for the calculations, which compute the normal modes of the motions of the Cα atoms with an ENM model using the “Cα” force field. The dynamic cross-correlation analysis and map
(dccm) for interfaces trajectories obtained from the NMA were computed with the use of the dccm function of Bio3D. This function calculates the covariance matrix based on mutual information between all Ca atoms in the interface structures. This analysis was used to understand the conformational flexibility of different interfaces of VP40 protein.
CHAPTER 3: PLASMA MEMBRANE ASSOCIATION FACILITATES CONFORMATIONAL CHANGES IN THE mVP40 DIMER

(The content for this chapter is taken from my published paper\textsuperscript{48} with the permission from Royal Chemical Society journal RSC-Advances)

In this project, we report on molecular dynamics simulations that investigate the roles of various residues and lipid types in PM association as well as the conformational changes of the mVP40 dimer facilitated by the membrane association. We compared the structural changes of the mVP40 dimer with the mVP40 dimer in both lipid free and membrane associated conditions. Despite the significant structural differences in the crystal structure, the Marburg VP40 dimer is found to adopt a configuration very similar to the Ebola VP40 dimer after associating with the membrane. This conformational rearrangement upon lipid binding allows Marburg VP40 to localize and stabilize at the membrane surface in a manner to like the Ebola VP40 dimer. Consideration of the structural information in its lipid-interacting condition may be important in targeting mVP40 for novel drugs to inhibit viral budding from the plasma membrane.

3.1 INTRODUCTION

Filovirus infections cause hemorrhagic fever in human and non-human primates\textsuperscript{49} that often result in high fatality rates\textsuperscript{50}. The Marburg virus (MARV) is a lipid-enveloped virus from the Filoviridae family and is closely related to the Ebola virus (EBOV). The virus particle acquires a lipid coat from the plasma membrane (PM) of the host cell as the virus assembles and buds\textsuperscript{51}. The viral matrix layer underneath the lipid envelope is formed by the matrix protein VP40 and provides shape and stability to the viral particle. It has been
shown in mammalian cells that only the expression of the VP40 protein, among the EBOV and MARV genome, is sufficient for forming innocuous but authentic-looking virus-like particles (VLPs). VP40 is involved in multiple functions during the viral life-cycle and protein multifunctionality often requires proteins to undergo conformational changes. It has been shown that, depending on the function, EBOV VP40 (eVP40) exists in different conformations such as a butterfly shaped dimer involved in the transport of the protein to the membrane, a hexamer to form the viral matrix beneath the lipid-envelope, and an octamer ring structure to bind RNA and regulate viral transcription. The X-ray crystal structure determination of the dimeric, hexameric and octameric forms of eVP40 has provided a great deal of information about the structural transformation of the protein into various oligomeric states for performing different functions. Structural information of such viral proteins also offers an opportunity to explore potential inhibitors for treating the disease.

Many proteins are known to interact with lipids in a membrane, and lipid binding can induce protein conformational changes that modulate their functions. In addition to the direct role of the protein-phospholipid interactions on the conformational changes and allosteric modulations of integral membrane proteins such as ion channels and receptors, lipid interactions at the membrane surface can induce structural changes in peripheral proteins. A widely studied example is the human apolipoprotein E3, whose NTD has been shown to adopt an elongated globular four helix bundle structure in solution but undergo conformational changes upon lipid binding. It has been hypothesized that, upon membrane association,
the filovirus VP40 dimers undergo major structural rearrangements. This step is required for oligomerization into hexameric structures\textsuperscript{53,74} which further assemble to form filaments leading to the formation of the viral matrix\textsuperscript{75}. However, the mechanisms and consequences of the VP40-lipid interactions on VP40 dynamics and assembly are not well understood in either eVP40 or mVP40.

The crystal structure of the MARV VP40 (mVP40) dimer has recently been determined\textsuperscript{76}. As in eVP40, the structure of the mVP40 dimer features an α-helical dimer interface in the N-terminal domains (NTD) as well as a basic patch in the C-terminal domain (CTD) that mediates membrane binding\textsuperscript{76}. It was found that mutations in the basic patch residues greatly hindered the mVP40 assembly and adversely affected VLP budding due to the reduction in anionic lipid binding caused by these mutations\textsuperscript{76}. This shows that the mVP40 dimer, like eVP40, is involved in trafficking of the protein to the lower leaflet of the PM and membrane localization via the basic patch. Structural comparison shows that the NTD structures as well as the NTD-NTD interfaces in both mVP40 and eVP40 dimers are quite similar. Given the large sequence similarity (42%) between the NTDs of mVP40 and eVP40, this is not surprising. In contrast, the mVP40 CTD is significantly different from the eVP40 CTD, with only a 15% sequence similarity. The crystal structure shows that the mVP40 CTD basic patch is significantly flatter with a more extended surface than that of eVP40\textsuperscript{76}. This suggests that the mVP40 dimer could interact with the PM differently than the eVP40 dimer, leading to differences in the phospholipid specificity, oligomerization and budding of the VLPs\textsuperscript{5,27,77,78,79}.

Recently, Wijesinghe et al. used Hydrogen Exchange Mass Spectroscopy to investigate at various timescales, structural changes in mVP40 due to phospholipid
interactions. By determining the solvent accessibility of the mVP40 residues, important residues involved in binding to the membrane, as well as those residues at the oligomerization interface were identified. However, the structural details of the mVP40 conformation after membrane association is still lacking. In this paper, we investigated the PM association and the conformational changes of the mVP40 dimer induced by membrane association during the early stages of oligomerization at the plasma membrane. We compared the structural changes of the mVP40 dimer with the eVP40 dimer in both lipid free and membrane associated conditions. Despite the significant structural differences compared to the crystal structure of eVP40 dimer, the mVP40 dimer is found to adopt a configuration very similar to the eVP40 dimer after 200 ns of MD simulations. This conformational rearrangement upon lipid binding allows mVP40 to localize and stabilize at the membrane surface in a manner very similar to the eVP40 dimer. Once associated with the membrane, VP40 dimers assemble into higher oligomers and the oligomerization requires further large-scale conformational rearrangements. While the slippery CTD-CTD interface is likely to be the main contributor to the filovirus flexibility, our results provide insight on how the flexibility of the NTD-NTD interface can also play an important role on the overall virion flexibility.

3.2 MATERIALS AND METHODS

The mVP40 dimer structure was taken from the x-ray crystal structure in the Protein Data Bank (PDB ID: 5B0V ) and the missing residues were added with Modeller. The protein and plasma membrane systems (both with and without membrane) were built using the Charmm-Gui membrane Builder web interface. The plasma membrane consists of
phosphatidylcholine (POPC), phosphatidylethanolamine (POPE), phosphatidylserine (POPS), palmitoylsphingomyelin (PSM), phosphoinositol (POPI) and cholesterol (CHOL) molecules. The distribution of different lipid molecules in the lower leaflet of the plasma membrane were in the number ratio of 20:11:33:18:9:7 (CHOL:POPC:POPE:POPS:POPI:PSM) representing the high complexity of the plasma membrane composition\textsuperscript{83,84}. For the mVP40-membrane system, the membrane consists of 284 lipids in the upper leaflet and 290 in the lower leaflet. For comparison, a similar system was set up for the eVP40 dimer (PDB ID 4LDB). The eVP40-membrane system consists of 147 lipids in the upper leaflet and 156 in the lower leaflet. Both systems were solvated with TIP3 water in cubic boxes and neutralized with 0.15 M KCl. The solvated system (protein, membrane, water and the neutralizing ions) contained a total of 231,567 atoms for mVP40 and 121,417 for eVP40. The total number of lipid number used, and detailed equilibration steps are summarized in Table 3.1. The molecular dynamics simulation protocol is explained in section 2.

3.3 Results

Comparison of the crystal structures of mVP40 and eVP40 dimers shows a significant structural difference in the membrane-interacting interface. As shown in Figure 3.1, the CTDs on either side of the VP40 dimers contain a basic patch that interacts with the cytoplasmic leaflet of the plasma membrane. The basic patch residues reside in two different loop regions. In mVP40, loop 1 (residues 208-222) contains K210, K211, R215, and K218 and loop 2 (251-271) contains residues K259, K264, K265 and R266. Similarly, in eVP40, the basic loop 1 (residues 219-233) contains K221, K224, K225 and loop 2-
Table 3.1. Lipid bilayer setup a) mVP40 b) eVP40 and c) Equilibration steps (for both eVP40 and mVP40 with membrane)

(residues 274-283) contains K274, K275, and K279. In Figure 3.1, the basic loop 1 residues are highlighted in blue, and the basic loop 2 residues are highlighted in red. The slightly lower positioning of the CTDs in mVP40 (Figure 3.1a) allows a nearly flat and extended top surface that can interface with the membrane. This is in contrast to the crystal structure of the eVP40 dimer (Figure 3.1b), which shows that the CTDs are positioned to give a chevron-like shape to the overall dimer with the NTDs at the bottom of the V-shape so that only the CTDs on either end can interface with the membrane. Comparison of the monomer-monomer interactions at the dimer interfaces of eVP40 and mVP40 shows that the eVP40 dimer interface is much more robust than the mVP40 interface. As shown in 3.1(c), interfacial ionic/hydrogen bond interactions contribute significantly to the eVP40 dimer stability compared to the mVP40 dimer.
The dimer interface interactions in both eVP40 and mVP40 have similar van der Waals energy contribution but the interfacial electrostatic energy contribution is much larger in the eVP40 dimer interface compared to that in mVP40. In order to investigate the membrane localization process of the mVP40 dimer in the plasma membrane, we initially placed the dimer slightly below the lower leaflet of the membrane, with all protein atoms located at >5 Å below any lipid atoms. It is known that the membrane localization of eVP40 requires POPS, so we included this lipid in the system. The strong electrostatic interactions between the negatively charged PS head groups in the membrane and the positively
charged lysine residues in the CTDs allow the dimer to associate with the membrane. To understand the mVP40 dimer-membrane association and the resulting conformational changes, we performed a 300-ns all-atom MD simulation for the mVP40-membrane system described above.

3.3.1 Plasma membrane association of the mVP40 dimer

We monitored the lipid-protein interactions during the association of the mVP40 dimer to the lower leaflet of the plasma membrane. Figure 3.2a displays the initial configuration of the protein-membrane system after minimization and equilibration. Initially, the basic residues oriented towards the membrane were not too far away to interact strongly with the lipids. As the lysine residues in basic loops 1 and 2 start interacting with the anionic headgroups of lipids, the dimer gradually drifts towards the membrane, which is followed by further interactions with additional residues. By 50 ns, most of the basic patch residues have strongly interacted with the PS head groups as shown in Figure 2b. In Figure 3.2c, we show the final configuration of the mVP40 at the end of the 300-ns simulation. At this time, a significant number of lipids interact with the dimer (Figure 3.2d). Figure 3.3a shows the time evolution of the distance (dcm) along the z-axis between the center of mass of the protein and the lipid bilayer (calculated as the center of mass of the phosphorous atoms). The center of mass distance decreases gradually as the protein approaches the membrane. Movie S1 shows the overall drift of the mVP40 dimer towards the membrane and the resulting membrane association. After about 80 ns, a slight but interesting increase in dcm is observed. This is caused by conformational changes in the mVP40 dimer as explained later.
Figure 3.2. (a-c) Snapshots of the mVP40 dimer association with the plasma membrane at different times. d) Various lipid types interacting with the basic loop 1 and basic loop 2 residues at 300 ns. The lipids are colored as: POPS-cyan, POPI-green, POPC-gray, POPE-purple.

Figure 3.3b displays a plot of the hydrogen bonds for basic loops 1 and 2 and shows that the lipid interactions with the basic residues are important for the dimer’s membrane localization and stability. The hydrogen bonds were calculated with 3.5 Å distance cut-off and 30° angle cut-off. A significant number of hydrogen bonds are formed between the protein and the lipid head groups. We observed that the number of hydrogen bonds for loop 1 is in general more than that for loop 2, showing a greater role of the loop 1 residues in membrane association and stabilization of the protein at the lower leaflet of the plasma
membrane. Further details of the amino acids and lipids involved in the hydrogen bonding are discussed below.

![Figure 3.3](image)

**Figure 3.3** Time evolution of structural parameters during mVP40 dimer-membrane association. a) distance (along the z-axis) between the center of masses of the protein and the lipid bilayer (calculated using the center of mass of the phosphorous atoms). b) The number of hydrogen bonds between the lipid head groups and the protein. The red and blue curves represent the hydrogen bonds formed with loops 1 and 2, respectively.

### 3.3.2. LIPID-PROTEIN INTERACTIONS AND LIPID SPECIFICITY

The mVP40 dimer association shows preferential lipid selectivity. As shown in Figure 3.2d, almost all lipid types can be observed around the loop 1 and loop 2 regions. The electrostatic interactions with the basic residues are mostly made by POPS (colored cyan in Figure 3.2d). We calculated the number of lipid-protein contacts for various lipid types and plotted the results in Figure 3.4a as a function of time. The protein heavy atoms
were considered to be in contact with lipid heavy atoms if they were within 3.5 Å of each other. As the dimer approaches the membrane, the number of contacts for all lipid types increases, except for cholesterol, which is expected because the cholesterol head group is slightly buried inward compared to other lipid head groups in the bilayer. We find that mVP40 has more contacts with POPS than other lipid types. This agrees well with the experimental observation that POPS is important for plasma membrane localization for both eVP40 and mVP40\textsuperscript{77,78,79}. In addition to POPS, other lipids (POPE, POPC, and POPI) also have significant contacts with the protein atoms, mostly in the basic patch residues, which provide the electrostatic interactions for the mVP40 membrane association and stabilization at the lower leaflet.

Figure 3.4. E260 interactions with the PM. a) Lipid molecules clustered around E260 (purple: POPE, cyan: POPS). Extensive network of hydrogen bonds formed between E260, POPE, and POPS. C) Lipid-protein interactions between the basic patch residues in the mVP40-CTD and the plasma membrane lipids.

We calculated the hydrogen bonds between the lipid and protein atoms for the last 100 ns of the trajectory (200 to 300ns), with a distance cut-off of 3.5 Å and angle cut-off of 30° between the donor and acceptor heavy atoms. The mVP40 dimer is fully associated with the PM and structurally stable during this time. Multiple lipid molecules/types are found to interact with the basic patch residues (Figure 3.4c). Details of the lipid-protein
hydrogen bonds, including the information of the specific donor or acceptor atoms involved in hydrogen bonding, are given in Table 3.2. We calculated the relative contribution of each of the hydrogen-bonding residues to the total lipid-protein hydrogen bonds. As shown in Figure 3.5b, R215 is found to make the most hydrogen bonds with the lipids (20%). This is followed by K211 (15%), E260 (13%), and K264 (10%). Other major contributors in the lipid-protein hydrogen-bonding are K218, K259, Q216, K183, G261, and K210.

<table>
<thead>
<tr>
<th>Donor</th>
<th>Acceptor</th>
<th>Occupancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG215-Side</td>
<td>POPS398-Side</td>
<td>153%</td>
</tr>
<tr>
<td>ARG215-Side</td>
<td>POPS366-Side</td>
<td>144%</td>
</tr>
<tr>
<td>POPS356-Main</td>
<td>GLU260-Side</td>
<td>91%</td>
</tr>
<tr>
<td>POPE416-Main</td>
<td>GLU260-Side</td>
<td>88%</td>
</tr>
<tr>
<td>GLN216-Main</td>
<td>POPC376-Side</td>
<td>80%</td>
</tr>
<tr>
<td>POPE444-Main</td>
<td>GLU260-Side</td>
<td>75%</td>
</tr>
<tr>
<td>LYS211-Main</td>
<td>POPC349-Side</td>
<td>73%</td>
</tr>
<tr>
<td>LYS211-Side</td>
<td>POPS398-Side</td>
<td>70%</td>
</tr>
<tr>
<td>LYS259-Side</td>
<td>POPE416-Side</td>
<td>69%</td>
</tr>
<tr>
<td>LYS211-Side</td>
<td>POPE400-Side</td>
<td>68%</td>
</tr>
<tr>
<td>LYS264-Side</td>
<td>POPS350-Side</td>
<td>62%</td>
</tr>
<tr>
<td>LYS211-Side</td>
<td>POPS374-Side</td>
<td>58%</td>
</tr>
<tr>
<td>LYS185-Side</td>
<td>POPS374-Side</td>
<td>56%</td>
</tr>
<tr>
<td>ARG215-Side</td>
<td>POPE412-Side</td>
<td>51%</td>
</tr>
<tr>
<td>LYS264-Side</td>
<td>POPS450-Side</td>
<td>50%</td>
</tr>
<tr>
<td>LYS218-Side</td>
<td>POPS364-Side</td>
<td>49%</td>
</tr>
<tr>
<td>ARG215-Side</td>
<td>POPS386-Side</td>
<td>47%</td>
</tr>
<tr>
<td>LYS211-Side</td>
<td>POPC349-Side</td>
<td>46%</td>
</tr>
<tr>
<td>GLY261-Main</td>
<td>POPS350-Side</td>
<td>43%</td>
</tr>
<tr>
<td>LYS185-Side</td>
<td>POPS353-Side</td>
<td>40%</td>
</tr>
<tr>
<td>ARG215-Side</td>
<td>POPS364-Side</td>
<td>38%</td>
</tr>
<tr>
<td>LYS259-Side</td>
<td>PSM389-Side</td>
<td>37%</td>
</tr>
<tr>
<td>LYS183-Side</td>
<td>POPS372-Side</td>
<td>37%</td>
</tr>
<tr>
<td>LYS210-Side</td>
<td>POPC349-Side</td>
<td>34%</td>
</tr>
<tr>
<td>POPS350-Main</td>
<td>GLU260-Side</td>
<td>34%</td>
</tr>
<tr>
<td>LYS264-Side</td>
<td>POPS378-Side</td>
<td>32%</td>
</tr>
<tr>
<td>SER257-Side</td>
<td>POPS356-Side</td>
<td>30%</td>
</tr>
<tr>
<td>ARG215-Side</td>
<td>POPC376-Side</td>
<td>28%</td>
</tr>
<tr>
<td>SER182-Side</td>
<td>POPS350-Side</td>
<td>20%</td>
</tr>
</tbody>
</table>

Table 3.2. List of lipid-protein hydrogen bonds for the last 100 ns of simulation sorted by their relative occupancy (calculated by VMD with distance cutoff of 3.5 and angle cutoff of 30°). Only the hydrogen bonds with occupancy of >20% are listed.
The electrostatic interactions between the cationic arginine and lysine residues and the anionic lipid head groups provide the major stability for the mVP40 dimer at the plasma membrane. However, it is interesting to note that the negatively charged E260 is the third largest contributor to the overall hydrogen-bonding. This is also the only negatively charged residue making hydrogen-bonds with the membrane. To understand exactly how the E260 side chain interacts with the lipids, we explored the hydrogen bonds around this residue. As displayed in Figure 3.4b, multiple POPS and POPE can hydrogen bond with E260. Specifically, anionic oxygen of E260 is the acceptor and the serine amino group of POPS and the amine group of POPE are the donors. It is worth noting that the lipid head groups also make an extensive network of hydrogen bonds that shield the E260 anionic side chain from the negative phosphate head groups of lipids, akin to solvent screening of a charged group, and can affect the lateral fluidity of the membrane. Such lipid-protein interactions can also cause lipid clustering.

In Figure 3.5b (inset), we also display the contribution of each lipid type to the overall lipid-protein hydrogen bonding. The relative population of the lipid specific hydrogen bonds shows that POPS, which contains the negatively charged head group, provides the dominant contribution (55%) to the lipid-protein electrostatic interactions. This is followed by POPE (21%) and POPC (16%). POPI, PSM and CHOL participate in very little hydrogen bonding with protein.
3.3.3. CONFORMATIONAL CHANGES IN MVP40 DIMER

In order to determine any structural changes in the MVP40 dimer specifically due to membrane interactions, we compared various structural parameters for the MVP40 dimer simulated with or without membrane. Visualization of the MVP40-membrane trajectory clearly indicates that the two monomers in the dimer show a significant structural rearrangement due to lipid binding. As the two CTDs at either end of the MVP40 dimer
start interacting with the membrane, the relative orientations of the monomers start to change. The monomers can have twist, roll, and tilt motions about the monomer helices at the dimer interface. To monitor this conformational change, we calculated the angle between the best fit lines of the monomers and display the results in Figure 3.6. For each monomer, both the CTD and NTD residues were considered for the angle calculation.

In Figure 3.6, we plot the angle between the monomers for both mVP40 and eVP40 and compare how they evolve in the presence and absence of membrane. Initially, the angle between the monomers in mVP40 dimer is ~20° more than in eVP40, with an angle of ~125° for mVP40 and ~105° for eVP40. This difference in the monomer relative orientations contributes to a flatter interface in mVP40 as shown in Figure 3.1. As the basic residues in the CTDs of the mVP40 dimer interact with the membrane, the angle between the monomers gradually decreases due to reorientation of the monomers (Figure 3.6a). This contrasts with the behavior shown by eVP40 dimer. The angle between the monomers in eVP40 dimer starts out at a much lower angle but increases slightly with membrane interactions and settles to ~115° (Figure 3.6b). The decrease in the angle in mVP40 is quite significant, with as much as a 35° drop around 100 ns during the simulation. In fact, the angle in mVP40 is reduced to values lower than in eVP40 during the trajectory. Ultimately, the angles for both appear to converge to ~110°–115°. Still, these conformational changes in the mVP40 dimer as it associates with the membrane are more significant compared to those in the eVP40 dimer. To compare with the conformational changes in the absence of membrane interactions, we performed 100 ns all-atom simulations without the membrane as controls for both the mVP40 and eVP40 dimers. Except for the absence of a membrane, all other conditions were kept the same. As shown in Fig 3.6a and 3.6b, both mVP40 and
eVP40 dimers retain their overall shapes in the absence of the membrane interactions for the simulated timescales. The mVP40 without membrane curve (Figure 3.6a) remains around 128°, significantly higher than the eVP40 without membrane value of 110°. This large difference in the conformation of the mVP40 dimer compared to the eVP40 dimer decreases when both associate with a membrane.

Figure 3.6. Time evolution of the angle (in degrees) between the monomers in the presence (purple) and absence (green) of lipid interactions for the a) mVP40 dimer and b) eVP40 dimer.

Inspection of the dimeric interface during the conformational change revealed the dynamic nature of the antiparallel β-sheet formed from residue segment 40-52 on each monomer. As mVP40 interacts with the membrane and the angle between the monomers starts to decrease, the hydrogen bonds between the β-strands rearrange slightly. Interestingly, a complete dissociation of the strands is observed after around 80 ns. The lowest angle observed at ~100 ns is also marked by the largest separation between residues T40 and Y43, with no hydrogen bonds across the β-strands for about 20 ns. The β-strands then start to re-associate but with a slightly different H-bond pattern. The increase in the angle between monomers right around this time (120 ns) suggests a direct correlation between the monomer orientations and the β-strand dynamics at the interface. The dynamic
motion of the β-strands is consistent with the experimental observations of significant deuterium incorporation at longer timescales, which suggested structural fluctuations at this region. The conformational change in the mVP40 dimer also affects the lipid-protein interactions and therefore the flexibility of the residues at the membrane interface. As a measure of the residue flexibilities, we calculated the root-mean-squared fluctuations (rmsf) for all residues in the mVP40 dimer. Figure 3.7a shows the rmsf values for all residues in the dimer at various time windows, calculated from the simulation of the protein-membrane system as well as the rmsf of the residues in the protein-only system.

As the protein associates with the membrane, the flexibility of the residues starts to decrease in general. Compared to the protein-only system, a significant reduction in the rmsf values can be seen for most residues in the protein-membrane system at early time windows (e.g., 0-5 ns). Interestingly, the rmsf values are found to increase at later timescales (e.g., 95-100 ns), including for the residues in the basic loop regions that directly interact with the membrane. We hypothesize that this increase in the residue flexibility is due to the conformational changes induced by the lipid interactions. Specifically, the basic loop residues interacting with the membrane can show enhanced flexibility during monomer reorientations. Indeed, recent hydrogen-deuterium exchange experiments have found that the basic loop 1 showed reduced solvent accessibilities (reduced deuterium incorporation) after a 10 sec incubation with PS containing liposomes, compared to that in the absence of PS containing liposomes. The basic loop 1 interactions with POPS likely caused the reduction in deuterium incorporations. This study also suggested the mVP40 membrane association was a dynamic process with continuous association and dissociation.
events taking place\textsuperscript{80, 87}, which may lead to different relative flexibilities of basic loop 1 residues and varying levels of deuterium incorporation.

To understand the dynamic changes in the flexibility of the basic loop 1 due to the rearrangement of the mVP40 monomers, we calculated the rmsf values for every 5-ns window of the entire 0-300 ns trajectory, with a total of 60 windows. For each time window, we averaged the rmsf of the basic patch (K210, K211, R215, K218, K259, K264, K265, and R266) and plotted the average rmsf as a function of time in Figure 3.7b. The first data point in Figure 6 at $t=0$ was obtained from the 0-5 ns window of the protein-only system’s trajectory and represents the flexibility of the basic patch in the absence of any lipid interactions. The time dependence of the average rmsf clearly shows a decrease in basic patch flexibility as the dimer associates with the membrane. By $\sim 80$ ns, the average rmsf drops by more than 20%, after which it starts to increase. As can be seen in Figure 3.3a, the mVP40 dimer drifts steadily towards the membrane until $\sim 80$ ns, evidenced by the decrease in the protein-membrane center of mass distance, $d_{cm}$. During this time, the angle between the monomers also continues to decrease (Figure 3.6). As discussed before, contacts between the $\beta$-stands (segments 40-52 from each monomer) dissociate and $d_{cm}$ appears to reset its trend (Figure 3.3a) to a slightly higher value after 80 ns, suggesting a reorganization in the protein structure. This is also marked by noticeable rearrangements of the lipid contacts shown in Figure 3.5a between $\sim 80$-160 ns. Therefore, structural reorganizations and the changes in the lipid contacts seem to result in an increased flexibility of the basic loop 1 during this time window ($\sim 80$-160 ns). As the new contacts are formed and loop 1 is stabilized ($>160$ ns), the loop flexibility is reduced again. These
results provide more detailed insights on the dynamics of the mVP40 dimer at the PS containing membrane surface and help explain the deuterium exchange kinetics.

Figure 3.7a. Root-mean-squared fluctuations (rmsf) for all the residues in the mVP40 dimer at various time windows. As the protein associates with the membrane, the flexibility of all the residues decreases in general. The lipid interactions affect the CTD rmsf more than the NTD. b) Time course of the flexibility of the basic patch residues (K210, K211, R215, K218, K259, K264, K265, and R266). For each point, the rmsf values of the basic loop residues were calculated for a time window of 5 ns and averaged over the residues of the basic patch.

Although the membrane associated mVP40 dimer structure becomes more like the eVP40 dimer, some structural differences in the protein surfaces interfacing with the membrane still exist, which could lead to differences in the oligomerization and the budding of the VLPs. Indeed, recent studies examining lipid binding by eVP40 and mVP40 have found different degrees of anionic lipid selectivity for these two matrix proteins with eVP40 selectively associating with POPS and some phosphoinositides and mVP40 interacting with anionic lipids nonselectively based upon the anionic charge density of the membrane surface. Additionally, slight differences in eVP40 and mVP40 oligomerization may occur using a twisted hexameric filament at the plasma membrane interface for both filoviruses, but an alternative CTD oligomerization interface in the 4-helix has also recently been proposed for mVP40. The conformational flexibility of the NTD-NTD
interface observed in mVP40 dimer may have significance in the viral budding and virion flexibility. In contrast, due to the rigidity of the NTD-NTD interface in the eVP40, it is thought that only the CTD-CTD interface between two hexamers can provide the flexible surfaces necessary for forming the flexible and pleomorphic filovirus virion$^{53,88}$.

3.4 Conclusion

In this work, we investigated the lipid-protein interactions as well as the lipid-induced conformational changes in the mVP40 dimer as it associates with the lower leaflet of the plasma membrane. We performed all-atom molecular dynamics simulations and identified important residues that facilitate the membrane association of mVP40 dimer and stabilize it at the PM. Results show that the hydrogen bonds between POPS lipid and residues K211, R215, and E260 dominate the overall lipid-protein interactions. We compared the structural changes of the mVP40 dimer with the eVP40 dimer in the presence and absence of membrane interactions. Despite the significant structural differences in the crystal structure, the mVP40 dimer is found to adopt a very similar configuration to the eVP40 dimer after associating with the membrane. As the two CTDs at the end of the mVP40 dimer start interacting with the membrane, the relative orientation of the monomers that allows a nearly flatter top surface starts to change. Although the angle between the monomers in the mVP40 dimer is initially much wider than in the eVP40 dimer, the mVP40 angle decreases significantly due to lipid interactions. In contrast, the eVP40 dimer conformation does not show a significant change upon association with the membrane. Simulations of mVP40 and eVP40 in the absence of the membrane interactions showed that the dimers retain their overall different shapes, highlighting the role of the lipid
interactions in facilitating the conformational changes in the mVP40 dimer. These conformational changes upon lipid binding allow mVP40 to localize and stabilize at the membrane surface similarly to the eVP40 dimer but may give subtle differences in its function due to the differences in the solution conformations. In addition to providing a proper orientation for oligomerization into hexamers, the ability of mVP40 dimer to undergo conformational changes about the NTD-NTD interface may have significance in pleomorphic nature of the MARV virion. Finally, structural information in its lipid-interacting condition may prove useful in targeting mVP40 dimer for novel drugs to inhibit viral budding from the plasma membrane.
CHAPTER 4: CONFORMATIONAL FLEXIBILITY OF THE PROTEIN-PROTEIN INTERFACES OF THE EBOLA VIRUS VP40 STRUCTURAL MATRIX FILAMENT

(The content for this chapter is taken from my published paper with the permission from American Chemical Society)

In this project, we have performed explicit-solvent, all-atom molecular dynamic simulations to explore the conformational flexibility of the three different interface structures of the filament. Using dynamic network analysis and other calculation methods, we find that the CTD-CTD hexamer interface with weak inter-domain amino acid communities is the most flexible, and the NTD-NTD Oligomer interface with strong inter-domain communities is the least flexible. Our study suggests that the high flexibility of the CTD-CTD interface may be essential for the supple bending of the Ebola filovirus, and such flexibility may present a target for molecular interventions to disrupt the Ebola virus functioning.

4.1. INTRODUCTION

The pleomorphic Ebola virus (EBOV), a member of the Filoviridae family, is a deadly pathogen that causes a severe hemorrhagic fever with a high mortality rate in humans. This disease is especially dangerous because no approved vaccines or chemotherapeutics are currently available. Of the four known EBOV strains that are transmissible to humans, the Zaire and Sudan EBOVs have been responsible for most of the EBOV clinical cases. The elucidation of the underlying molecular mechanisms responsible for the EBOV replication and pathogenesis is required for the development of antiviral therapeutics. Of the proteins that are encoded by the seven genes of EBOV, the
40kDa VP40 is a major structural matrix protein that plays a key role in budding and assembly of the EBOV. The 326-amino acid long VP40 protein has an N-terminal domain (NTD; residues 1-195) and C-terminal domain (CTD; residues 196-326). Rearrangement of the NTD and CTD relative to each other allows VP40 to form different conformational states, permitting VP40 to display transformer-like protein characteristics and to perform different functions in the virus life cycle: a butterfly-shaped dimer structure is essential in the transport of the protein to a cellular membrane, a hexameric structure acts as a structural building block of the cylindrical viral matrix filament, and an octamer ring structure binds to RNA and regulates viral transcription.

Figure 4.1. (A) A section of the EBOV VP40 composing of two hexamers connected through end-to-end ‘unsprung’ CTDs. Three distinct interfaces are highlighted by dashed circles: (B) NTD-NTD oligomerization to connect dimers, (C) CTD-CTD interface to connect hexamers end-to-end, and (D) NTD-NTD dimer interface between monomers to form a dimer. For clarity, the ‘sprung CTDs’ are not shown in ‘B’ for the dimer-dimer oligomer interface, the NTDs are not shown in ‘C’ for the CTD-CTD hexamer interface, and the ‘unsprung’ CTDs are not shown in ‘D’ for the monomer-monomer dimer interface.
Before reaching the cellular membrane, the butterfly-shaped dimer is formed by interaction of the NTD of one protomer (NTD-NTD Dimer interface of Figure 4.1D) with the NTD of another VP40 protomer\textsuperscript{12,99}. The dimer further assembles into a hexamer through several steps as revealed by Bornholdt et al.\textsuperscript{12} First, the dimeric VP40 migrates to the plasma membrane and the basic patches of the CTDs interact with the membrane. Subsequently, the VP40 dimers assemble with each other through NTD-NTD oligomer interactions as an intermediate. Finally, the central CTDs spring away from the NTDs. At each end of the hexamer is a VP40 protomer with an unsprung CTD. The NTDs of four protomers form the central core of the hexamer and assemble through the NTD-NTD dimer-dimer oligomerization interface of Figure 4.1B, and their CTDs are sprung away. A linear VP40 filament is then created via the unsprung CTD-CTD interaction between hexamers (Figure 4.1C). The cylindrical structural matrix of the Ebola virus is formed by side-by-side interactions between the VP40 filaments as discussed in Pavadai et al.\textsuperscript{102}. Separately, a VP40 octameric ring can form by springing all CTDs away from their NTDs, followed by oligomerization through the NTD-NTD oligomerization interface\textsuperscript{12,101}. But recently, Wan et. al. has also reported that VP40 CTD-CTD interface of two dimers allows it to form the filament\textsuperscript{103}. However, information on the molecular mechanisms by which VP40 transforms from a dimeric to a hexameric or octameric structure is still limited. Understanding the dynamical molecular mechanisms of these structural conversion will shed light on how the VP40 protein performs multiple functions in the EBOV life cycle. Biochemical, biophysical, and structural studies using cryo-electron, tomography, mutagenesis, and mini-genome assays have provided information about the conformational flexibility of VP40\textsuperscript{12,101,48}. In this study, we have employed explicit-solvent atomistic
simulations to explore the dynamical flexibility of the three interfaces described above (Figure 4.1) in the VP40 filament. Our results from analyzing structural fluctuations, resistance to bending, and dynamic network connectivity implies a high degree of flexibility for CTD-CTD interface and low flexibility for the NTD-NTD dimer-dimer Oligomer interface. We have also used NMA and dynamic-cross correlation matrix to map which amino acid sections are rigid and flexible in the interface structures. Our results suggest that the high flexibility of the CTD-CTD interface may be essential for the supple bending of the Ebola virus, and the other filoviruses. The flexibility of this interface may also present a target for molecular interventions to disrupt the functioning of the Ebola virus.

4.2 MATERIALS AND METHODS

System preparation: The initial structures for the VP40 NTD-NTD dimerization, NTD-NTD oligomerization, and end-to-end CTD-CTD interfaces were extracted from the crystal structure of a linear VP40 filament (Protein Data Bank ID: 4LDD)\textsuperscript{12}. These interface structures are shown in Figure 4.1, and they are referred to as a protein-protein complex or a protein-protein interface throughout the manuscript. The missing residues from the PDB 3D structures of the interfaces are listed in the PDB file. These residues were added to the structure using the Modeller 9.17 software package\textsuperscript{104}. For all three interfaces, energy minimization was performed for 20 ps, followed by equilibration for 50 ps. The production runs were performed for 500 ns. The number of atoms for the oligomer was 68,000, for the CTD-CTD interface 62,000, and for the Dimer interface 80,000.
Steered Molecular Dynamics (SMD) simulations: SMD was used to determine the bending flexibility for each of the three interfaces. For the SMD simulations, the equilibrated structures of the interfaces were used. We constrained the atoms at the outer edges of the structure and pulled on the Cα atoms at an interface in a direction that produces a bending motion at the interface. Later, in Figure 4.5, we provide a detailed description of the Cα atoms that were chosen. We used a spring constant k of 1 kcal/molÅ² and pulled at a constant speed of 0.4 Å/ns with a time step of 1 fs for 10 ns for all three interfaces. These choices were made in order to create a bending stress at the interface with disturbing the internal structures of the domains, as explained in the Results section. The resistance offered to each interface was observed by plotting force-time graphs.

Normal mode and dynamic cross-correlation analyses: The low-frequency collective motions of the three interface structures were analyzed by the elastic network model of NMA with the use of the NMA function of the Bio3D software package. Default settings of the program were used for the calculations of the NMA. The dynamic cross-correlation analysis and map (dccm) for interfaces trajectories obtained from the NMA were computed with the use of the DCCM function of Bio3D. This function calculates the covariance matrix based on mutual information between all Cα atoms in the interface structures. The root mean square deviation (RMSD), root mean square deviation fluctuation (RMSF), center of mass distance, interaction energy, visualization of the trajectories and preparation of figures were performed with the use of VMD. The angle between the best-fit line of the Cα atom coordinates of domains was created utilizing the fit angle script of the VMD Script Library.
4.3 RESULTS AND DISCUSSION

Analysis of the VP40 linear filament (Figure 4.1A) shows three distinctive interfaces: NTD-NTD dimer-dimer oligomer interface (Figure 4.1B), NTD-NTD monomer-monomer dimer interface (Figure 4.1D), and hexamer-hexamer end-to-end CTD-CTD interfaces (Figure 1C). The dimer interface of Figure 4.1C is conserved among all EBOV strains and is formed by residues 52–65 on one NTD and 108–117 on the other NTD. This interface is dominated by hydrophobic interactions and has a limited number of hydrogen bonds. The NTD-NTD dimer-dimer oligomer interface (Figure 4.1B) is centered on the hydrophobic Trp-95 residue from each NTD domain when the CTDs are sprung away from the NTDs. This interface is the basis for creating the central core of four NTDs of the hexamer and is homologous to the interface found in the VP40 RNA-binding octameric ring structure. The hexamer-hexamer end-to-end CTD-CTD interface is conserved among all EBOV strains and formed by L203, I237, M241, M305, and I307 hydrophobic residues. To explore the flexibility of the filament that is necessary for the viral functions, the three representative interface structures were extracted from the linear VP40 filament. We performed 500 ns all-atom explicit-solvent MD simulations as well as NMA computations as described above. The ‘sprung’ CTDs for the oligomer interface were not considered as they were unstructured in the crystal structures. In addition, the ‘unsprung’ CTDs for the dimer NTD-NTD interface and the NTDs for the CTD-CTD interface were not included for the reasons of comparability and consistency.

4.3.1. FLEXIBILITY OF THE INTERFACE STRUCTURES

To quantitatively analyze the stability of the interface structures, we computed the root mean-square deviations (RMSD) summed over all of the backbone atoms of the
interface structures relative to their initial structures as a function of simulation time as shown in Figure 4.2. The CTD-CTD interface has an RMSD compared to its initial structure of approximately 8 Å, whereas the dimer and oligomer interface structures have an RMSD of approximately 2 Å and 3 Å, respectively, indicating that the CTD-CTD interface is significantly more flexible than other interface structures. This result agrees with the experimental result that the CTD-CTD interface is flexible, and the oligomer interface is rigid (Bornholdt et al. 2013).

Figure 4.2. The RMSD of the different interfaces as a function of MD simulation time.

To further understand the relative motion and dynamics of the interface structures, we calculated the angle between the best-fit line of the Cα atom coordinates of the two domains for each of the three interface structures as a function of time. The fluctuations in the curves of Figure 4.3 show that the NTD-NTD dimer interface is very rigid and the CTD-CTD interface is the most flexible. The RMSD and angle analyses indicate that all three systems, dimer, oligomer and CTD-CTD interfaces, reached equilibrium by 400 ns.
Figure 4.3. Fluctuations in the angle between the best-fit line of the Ca atom coordinates of the two domains for each interface as a function of time.

To obtain more detailed information about the structural elements that contribute to the flexibility of the interface structures, we examined the relative motion of small groups of amino acids on either side of each interface. For each interface structure, we chose two pairs of small structural elements. The small structural element is either an α-helix or a β-strand. Within each pair, the amino acids composing the structural element are the same on both domains. The groups of amino acids are listed in Table 4.1 and displayed in Figure 4.4.

Figure 4.4 Fluctuations in the distance between the structural elements of each group as a function of simulation time.
Table 4.1. For each interface, two small groups (Group I, Group II) of amino acids on each domain are chosen for detailed investigation of flexibility. For each group, the amino acids are the same on each of the two domains.

<table>
<thead>
<tr>
<th></th>
<th>Oligomer</th>
<th>CTD-CTD</th>
<th>Dimer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>β-strand</td>
<td>α-helix</td>
<td>β-strand</td>
</tr>
<tr>
<td>85-102</td>
<td>233-243</td>
<td>53-56</td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td>β-strand</td>
<td>β-strand</td>
<td>α-helix</td>
</tr>
<tr>
<td>153-159</td>
<td>303-310</td>
<td>107-118</td>
<td></td>
</tr>
</tbody>
</table>

We computed fluctuations in the angles between the best-fit line of the Cα atom coordinates of the two elements and center of mass distance (dCOM) in each group, e.g. for Group I of the oligomer interface, the angle was computed between the best fit lines of residues 85-102 of one NTD domain relative to the residues 85-102 of the other NTD domain. Figure 4.5 shows that the fluctuations in the angle are largest for the CTD-CTD interface, with the α-helix↔α-helix Group I display the largest fluctuations. As in the previous diagrams, the Dimer NTD-NTD interface is the most rigid with little fluctuation in the angle for either Group I or Group II. The Oligomer interface also shows little flexibility. Further examination of the flexible CTD-CTD interface reveals that the α-helix shows torsional-like motions until 300 ns and then reached equilibrium. The CTD-CTD β-strands are less flexible than the helices, but the twisted β-strands become parallel. This result is in line with the experimental observation that the CTD-CTD interface allows torsional motions of the domains. We also computed dCOM between the structural elements as explained above. The results are displayed Figure 4.4. The structural pairs
maintain the distances at both the dimeric and oligomeric interface structures throughout the simulation, indicating that both interfaces are stable. Nevertheless, the Group I (α-helixI-α-helixII) structural pair distance at the CTD-CTD interface displays much more motion, but a little motion was observed between β-sheetI and β-sheetII, Group II, signifying that CTD-CTD interface is flexible.

Figure 4.5. (A, B, C) Visualization of the structural elements located at the three interfaces; (D, E, F) fluctuations in the angle between the best-fit line of the Cα atom coordinates of the structural elements of each group as a function of simulation time. The group elements are highlighted in A, B and C, and the corresponding colors are used in the plot.

To further understand the fluctuation of individual residues at the interfaces, the root-mean square fluctuation (RMSF) of Cα atoms of interface structures from the last 100 ns simulation trajectory was analyzed, which reveals the flexibility of each residue of the interface structures, as shown in Figure 4.6.
Figure 4.6 Root-mean-square fluctuations (RMSF) computed from the last 100 ns trajectory for the Cα atoms of the interface structural elements.

In all interface structures, large fluctuations occur in the loop regions and low fluctuations occur in the α-helix and β-sheet regions. The high RMSF indicates loosely organized loop or terminal regions. As shown in Figure 4.6, the structural elements at the interface show lower RMSF values than other regions of the structures, indicating that the
domains interact strongly with each other. Interestingly, the CTD-CTD interface shows increased RMSF values compared to the other two interface structures, signifying that there is considerable fluctuation at the CTD-CTD interface.

4.3.2. **BENDING FLEXIBILITY OF INTERFACES**

The relative flexibility for bending of each interface was determined by measuring the force necessary to flex it. For each interfacial structure, we chose a transect line that started on the Cα atom of a residue near the middle of the outside edge of one domain, through the middle of the interface, to the Cα atom of a residue near the middle of the outside edge of the other domain. A schematic of this transect line is shown in Figure 4.7A. As shown in Figure 4.7B, for the SMD we immobilized the Cα on the transect line at the outer edge of each domain. We chose two Cα close to the transect line, one for each domain, that are directly at the interface and pulled on both amino acids in a direction perpendicular to the transect line of the domains so that the interface would flex. We pulled at a relatively slow constant flex speed of 0.4 Å/ns with an SMD virtual spring of $k=1$ Kcal/mol·Å$^2$ with a timestep of 1 fs for 10 ns. We chose relatively low values of pulling speed and spring constant to avoid disrupting the internal structure of the domains and non-equilibrium effects to focus on the bending motion. These low values of pulling speed and spring constant produces gradual changes on the system as compared to dynamics that would occur with higher values$^{106,107}$. Figure 4.7C shows the pulling force necessary for each interface to maintain the same speed of flex. Consistent with the high flexibility displayed in the previous figures, in Figure 4.7C, the CTD-CTD interface shows the least resistance to flexing.
4.3.3 Interaction of the Interface Structures

The flexibility analyses above show that the CTD-CTD interface is more flexible than the NTD-NTD Dimer and Oligomer interfaces. To understand which amino acids and molecular interactions are responsible for the differences, we calculated the energies for various types of interactions between the domains at the interface. Figure 4.8 shows the electrostatic interaction energy and van der Waals (vdW) contact energy as a function of simulation time. The plots of both the electrostatic inter-domain energy and the vDW inter-domain energy show that interactions between the domains of the CTD-CTD interface are intermediate in strength between the Dimer and the Oligomer interfaces. This shows that the CTD-CTD interface is a secure connection that does not easily break, but these bonding energies do not explain the enhanced flexibility of the CTD-CTD interface. Therefore, it is necessary to investigate the inter-domain interactions in more detail to explain the relative flexibility of the different interfaces. In the next section, we describe the inter-domain network analysis that we performed.
4.3.4 DYNAMICAL NETWORK ANALYSIS

The relative flexibility of a domain-domain interface is affected by the number and strength of the inter-domain interactions and depends on their topological arrangement. To visualize the topological networks formed by different groups (communities) of amino acids at the interface as a result of their correlated molecular motions, we calculated the dynamical networks of each of the three interface structures obtained from the last 100 ns of their separate MD trajectories using the NetworkView plugin in VMD. In Figure 4.9, amino acids are represented by a node centered at the Cα for each residue and co-relation values are represented by the weighted edge between the nodes. The shortest path between the nodes for communication are identified. The communities in the network are defined as the time-averaged connectivity of the nodes and were constructed using the Girvan-Newman algorithm. The thickness of edges represents the co-relations that define the probability of information transfer along the edge. We found a total of 18 communities for the CTD-CTD interface, 15 communities for the Oligomer interface, and 18 communities
for the Dimer interface as shown in Figs. 4.9(A-C). In Figs. 4.9, different communities are displayed in different colors.

Figure 4.9. Dynamic network community analysis from the last 100 ns of the MD trajectories. Different communities are shown in different colors. The line thickness between the amino acids represents how strongly the motions of residues within the community are correlated. (A) CTD-CTD, (B) dimer, (C) oligomer. Amino acid communities formed by residues that span both domains are shown in (D) CTD-CTD, (E) Dimer, (F) Oligomer.

Especially important for the flexibility of an interface are communities composed of amino acids from both domains and span the domain-domain interface. Since the amino acids within a community have highly correlated motions, the communities that span an interface reduce the flexibility of one domain relative to the other domain. The communities with amino acids that span each interface are numbered in Figs. 4.9(D-F). Communities that do not span the interface are not numbered in Figs. 4.9(D-F) and are shown in yellow or purple. There are five communities that span the oligomer interface, one community that spans the dimer interface, and three communities that span the CTD-
CTD interface. Most importantly for reducing domain-domain flexibility is the number of connections between amino acids on opposite domains. Examination of the domain-spanning communities in more detail shows that the number of inter-domain connections differs among the communities. Table 4.2 shows the amino acids that are connected across the interface for each community. Within the two columns for each interface structure, each row identifies the amino acids that are connected to each other across the interface from one domain to the other. Though the Dimer interface has only one community that spans both domains, this interface is especially dense with inter-domain connections. The bottom row of the table provides the sum of all inter-domain connections for each interface: Oligomer-33, Dimer-19, CTD-11. This ordering is the same in Figure 5 that displays the pulling force necessary to flex each interface.

Table 4.2. For each of the inter-domain communities pictured in Figs. 4.9 (D-F) for each interface structure, a list of amino acids in each inter-domain network community that are dynamically connected to amino acids on the other domain. The bottom row of the table provides the sum of all inter-domain amino acid connections for each interface.

<table>
<thead>
<tr>
<th>Community</th>
<th>Oligomer</th>
<th>Dimer</th>
<th>CTD-CTD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NTD-1</td>
<td>NTD-1</td>
<td>NTD-1</td>
</tr>
<tr>
<td>Comm-1</td>
<td>P93</td>
<td>P187</td>
<td>I54, A55, D56, D57</td>
</tr>
<tr>
<td></td>
<td>K90, I92</td>
<td>A189</td>
<td>A55, I59</td>
</tr>
</tbody>
</table>

51
Additional information about the contacts in Table 4.2 is provided in Table 4.3.
Table 4.3: Detailed interactions of amino acids involved in the community formation.

<table>
<thead>
<tr>
<th>Commuity</th>
<th>Oligomer</th>
<th>Interaction-type</th>
<th>Dimer</th>
<th>CTD-CTD</th>
<th>Interaction-type</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.5. NORMAL MODE ANALYSIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molecular dynamics (MD) simulations and normal mode analysis (NMA) have been successfully used to obtain information about the functionally relevant dynamic processes of proteins under their physiological-mimic environments (^{108-111}). To obtain additional information about the conformational flexibility, we performed normal mode analyses (NMA) of all three interfaces. The NMA can provide insight into the flexibility of interface structures in terms of collective motions of groups of atoms (^{112}). The analyses were performed using the NMA function of the Bio3D package(^ {113}) with default settings. Figure 4.10 displays one of the low frequency internal modes for each interface. Each figure is the superposition of the Ca NMA trajectory and the gray shading shows regions with large motion. As expected for flexing-type normal modes, residues execute larger motion the farther they are from the domain-domain interface. The CTD-CTD interface displays the largest flexibility.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To provide further insight into the flexibility of the interface structures, we plot in Figure 4.11 the correlation between the motions of the Cα atoms in the residues of one domain (horizontal axis) with respect to the residues in the other domain (vertical axis) for all three interface structures from their NMA trajectories. The correlation matrices were calculated using the dccm function of the Bio3D package. In order to capture a significant part of the motion, we plot the cross-correlation matrix using the combined motions from the 10 lowest frequency modes. The correlation is measured by calculating the cross-correlation coefficient (-1 to +1) between the Cα atomic motions. A value of +1 (red color) represents atoms moving in a similar direction, whereas -1 (blue color) represents atoms moving in the opposite direction. A value of zero (white color) represents no correlation in the motion of the atoms. Figure 4.11 shows that the interdomain motions are more positively correlated for the CTD-CTD interface and least correlated for the Oligomer interface. Further, the groups of amino acids across the two domains for each interface that are highlighted in Table 4.1 and Figure 4.5 as being especially important for interdomain flexibility display high correlation values in Figure 9 and are circled.
4.3.6 **Principal Component Analysis**

To further understand the distinct conformational transitions and motions of the interfaces, PCA was performed on the last 200ns of the MD trajectories for each of the three interfaces. The results are shown in Figs. 4.12-A, B, C for the CTD-CTD interface, Figs. 10-D, E, F for the Dimer interface, and Figs. 10-G, H, I for the Oligomer interface. The conformational motions are analyzed by projecting the trajectories onto two-dimensional subspaces spanned by the first PCA three eigenvectors (PC1, PC2, and PC3). Figures 10-A, D and G show the conformational spaces of the three interface structures: (PC1, PC2), (PC1, PC3), and (PC2, PC3). The different conformational states are shown in red and blue dots in the plot. The distributions displayed in Figure 10A show that the CTD-CTD interface has conformations that are clustered with long-distance jumps between conformations. This is also true for the Oligomer interface as shown in Figure 10G. However, Figure 10D shows that the conformational states of the Dimer are not arranged in separated clusters. In addition to the PC plots, we also display visualization of
the motions of the Cα atoms for PC1 and PC2 for each of the three interfaces: Figs. 10-B, C for the CTD-CTD interface, Figs. 10-E, F for the Dimer interface, and Figs. 10-H, I for the Oligomer interface. The analysis and visualization of the PC1 and PC2 for the interface structures in Figure 10 shows that the CTD-CTD is relatively more flexible compared to the other two interface structures. This result is consistent with the NMA and angle analyses.

Figure 4.12. From Principal Component Analysis, projections of the trajectories onto planes by the first three eigenvectors. (A) Projection of trajectories into PC1, PC2, and PC3 for the CTD-CTD interface, (D) Projection of trajectories into PC1, PC2, and PC3 for the Dimer interface and (G) Projection of trajectories into PC1, PC2, and PC3 for the Oligomeric interface. The color scale from blue to white to red shows that there are periodic jumps between these conformers through the trajectory. (B) Visualization of the molecular motions along (B) PC1 and (C) PC2 for CTD-CTD, (E) PC1 and (F) PC2 for the Dimer; and (H) PC1 and (I) PC2 for the Oligomeric interfaces. The gray shading shows regions with large motion.
4.4. CONCLUSIONS

The EBOV matrix protein VP40 is the main structural protein of the long, thin, flexible Ebola virion. VP40 hexamers attach end-to-end at a CTD-CTD interface into linear filaments. Analysis of the long VP40 filament shows three unique interfaces: Dimer, Oligomer, and the end-to-end CTD interfaces. The conformational flexibility of the VP40 structural matrix has been suggested to be essential for the pleomorphic nature of the Ebola virions\textsuperscript{12, 88, 114}. Here, we have performed MD simulations, normal mode analysis, principal component analysis, and other calculations to investigate the dynamical flexibility and stability of the three different structural interfaces to determine which contributes most to the flexibility of the virion. MD simulations and analysis show that the CTD-CTD interface is the most flexible and the oligomer interface is the most rigid. Though all three interfaces have similar numbers of strong domain-domain interactions, our dynamical network analysis shows that the rigidity of the oligomer interface is due to many weak interactions across the domain interface. In contrast, the flexible CTD-CTD interface has a relatively small number of this type of connections. The high flexibility of the CTD-CTD interface may be essential for the supple bending of the EBOV, while the rigidity of the Oligomer interface may be necessary for the structural stability of the EBOV. The flexibility and rigidity of the interfaces may present targets for molecular interventions to disrupt the functioning of the EBOV.
CHAPTER 5: EBOLA PROTEIN VP40 BINDING TO SEC24C FOR TRANSPORT TO THE PLASMA MEMBRANE

In this work, we use various molecular computational techniques to investigate the molecular details of how EBOV VP40 binds with the Sec24c complex of the ESCRT-I pathway. We employed different docking programs to identify the VP40 binding site on Sec24c. We performed molecular dynamics simulations for the highest ranked Sec24c-VP40 dock complex to determine the atomic details and binding interactions of the complex. We also investigated how the inter-protein interactions of the complex are affected upon mutations of VP40 amino acids in the binding region. In addition, we found that VP40 can bind to a site on Sec24c that can also bind Sec23 and suggests that VP40 may use the COPII transport mechanism in a manner that may not need the Sec23 protein in order for VP40 to be transported to the plasma membrane.

5. 1. INTRODUCTION

The Ebola virus (EBOV), a negative sense RNA virus, belongs to the Filoviridae family\textsuperscript{115, 116}. This virus causes hemorrhagic fever with up to 90% fatality rates\textsuperscript{117} and tragically, there are no approved drugs and vaccines readily available to treat Ebola diseases\textsuperscript{118, 119}. The 19-kb long EBOV genome encodes seven different proteins, including membrane associating proteins VP40 and VP24 that are involved in forming the viral filaments’ structural matrix\textsuperscript{120}. The other proteins, VP30, VP35, nucleoprotein (NP), glycoprotein (GP), and L protein, along with an oligomer of VP40\textsuperscript{121} are responsible for the transcription and replication processes of the virus. VP40 associates with the plasma membrane (PM) of the host cell and alone can form virus-like particles (VLPs), indicating
that VP40 is the most important protein for the structural assembly and budding of the virus\textsuperscript{122, 123}. The monomeric structure of VP40 consists of two domains: the C-terminal domain (CTD) that is involved in the lipid binding process, and the N-terminal domain (NTD) that is involved in the VP40 dimerization process\textsuperscript{12}. Soon after an infected cell produces a VP40 monomer, it dimerizes and the VP40 dimer migrates and associates with the PM molecules\textsuperscript{124-127} and forms oligomers\textsuperscript{81, 89, 128}. At the PM, three VP40 dimers associate into a linear hexameric structure and initiate the filamentous structure of a new VLP. Before interacting with the PM and forming the VLPs, the VP40 protein must be transported to the PM. As with other viral lifecycles, binding interactions between proteins play important roles\textsuperscript{129}. A crucial step in the transport process is the binding of VP40 to the Sec24c protein. VP40 is transported to the PM using ESCRT-I (endosomal sorting complexes required for transport). The late-domain PTAP and PPEY motifs of VP40 have been found to interact with the ESCRT-I complex\textsuperscript{130-132}. One of the pathways utilized by viruses to transport proteins within the host cell is the coat-proteins II (COPII) pathway\textsuperscript{133, 134}. In general, COPII forms coat vesicles on the ER membrane. These coat vesicles transport cargo molecules such as VP40 from the site where they are synthesized on the endoplasmic reticulum (ER) to other locations in the cell such as the PM or the Golgi complex in human cells\textsuperscript{135}. Various COPII proteins (Sec23, Sec24, Sec13, Sec31, Sar1, GTP) are required to form a coat vesicle. After insertion of Sar1 in the ER-membrane, it then interacts with GTP, which allows the recruitment of Sec23/Sec24 to form a Sar1-Sec23/Sec24 inner layer of the coat vesicle attached to the ER membrane. This inner layer of the coat vesicle binds with the cargo protein, and the transport vesicle is completed by the attachment of Sec13/Sec31 to form an outer layer. The coat vesicle forms curvature in
the ER membrane that then allows the budding of the coat vesicle and dissociation from the ER. The Sec24 protein plays a role in identifying the cargo proteins during coat vesicle formation. The Sec24 protein has four isoforms: Sec24a, Sec24b, Sec24c and Sec24d. These isoforms of Sec24 are responsible for packaging and interacting with various proteins. For example, the entry of the Hepatitis virus, which is regulated by Caludin-1 has been shown to be dependent on transport of Sec24c. Similarly, the Sec24c coat protein binds with other proteins, such as Ebola VP40 and transports them.

Yamayoshi et al. demonstrated experimentally that EBOV VP40 is transported to the PM using the COPII transport system. VP40 interacts with the Sec24c protein in the coat vesicle. The important binding residues on VP40 are known (residues 210, 211, and 303 to 308), but the binding site on the Sec24c protein is still not identified. Therefore, the important protein-protein residue interactions for the VP40-Sec24c complex also remain elusive. As the interaction of Sec24c with VP40 occurs in the initial phase of the EBOV life cycle, understanding the atomic level details of this interaction will be useful for determining targets for drug molecules on those binding regions of the complex. In this work, we investigated the molecular details of how EBOV VP40 binds with the COPII protein human Sec24c complex of the ESCRT-I pathway. We used various molecular computational techniques. We employed different docking programs to identify the binding site on Sec24c using the known binding site information of VP40. We performed molecular dynamics simulations for the highest ranked Sec24c-VP40 dock complex to determine the atomic details and binding interactions of the complex. We also investigated how the inter-protein atomic interactions of the complex are affected upon mutations of VP40 amino acids in the binding region, which has been performed experimentally. For
the L303A and I307A mutation, we observed a decrease in inter-protein interactions, as also observed experimentally. For the M305A mutation, we observed little change in the interactions. We describe the atomic level details that explain the changes in inter-protein interactions that we observe.

The sequence of molecular docking steps involving VP40, Sec24c and Sec23 that facilitates the transport of VP40 to the plasma membrane is not yet completely understood. To gain insight, we also performed docking of VP40 at the site on Sec24c that can also bind Sec23. The binding of VP40 in the trunk domain region of Sec24c that also binds Sec23 suggests that VP40 might use the COPII transport mechanism in a manner that may not need the Sec23 protein for VP40 to be transported to the plasma membrane.

5.2 Protein-protein docking

The human Sec24c (PDB ID: 3EH2)$^{139}$ and EBOV VP40 (PDB ID: 4LDB)$^{12}$ structures were obtained from the Protein Data Bank. The crystal structure of Sec24c is a trimer (Figure 1a) and that of VP40 is a dimer (Figure 1b). We used chain A of both crystal structures (Figure 1c and Figure 1d) for our study. The missing residues of the structures were added using the Modeller$^{140}$ software package. The protein structures used for docking contained residues 329–1094 for Sec24c chain A and 44–326 for VP40 chain A. The first 328 residues are missing in the crystal structure of Sec24c because they are disordered as well as hypervariable. To determine the reliability of our docking sites, we used three different docking software programs. Protein-protein docking was performed with the use of ZDOCK, CLUSPRO, and Patch Dock with the Fire Dock refinement.$^{141-144}$ Top ranked docking complexes were obtained from each docking program and analyzed
for the interactions between protein-protein complexes, and the best complex was selected for molecular dynamics simulations that is consistent with experimental information\textsuperscript{137}. Since the binding region of VP40 is in the CTD, we considered the CTD region of VP40 for our molecular dynamics (MD) simulations. This smaller system enables us to sample for longer time scales while performing MD simulations. We also performed MD simulations on three separate VP40 point-mutations that were experimentally investigated previously. The three different mutations are in the CTD domain of VP40: L303A, M305A, and I307A. Therefore, there are four different systems: the wild type and the three separate mutations.

5.3. RESULTS AND DISCUSSION

The crystal structure of Sec24c (PDB ID: 3EH2)\textsuperscript{139}, as shown in Figure 1a, has three identical chains colored differently. We performed domain prediction analysis with chain A of Sec24c (Figure 1c) using InterPro\textsuperscript{145}. Since the first 328 residues are missing in the crystal structure of Sec24c, for our computations we renumbered the residues so that residue 329 is now residue 1 and residue 1094 is now residue 767. The domain prediction identified five different domains in chain A of Sec24c: zinc finger (residues 94 to 132, yellow), trunk domain (171 to 417, gray), beta-sandwich (420 to 503, purple), helical domain (516 to 615, blue) and gelsolin domain (634 to 706, red). The crystal structure of the EBOV dimeric VP40 (PDB ID: 4LDB) is shown in Figure 1b and its monomeric structure is shown in Figure 1d. The monomer has two domains: N-terminal domain (residues 44 to 196, green) and C-terminal domain (197 to 326, pink)\textsuperscript{12}.
Fig. 5.1. (A) Crystal structure of the human Sec24c trimer (PDB: 3EH2); the three chains are colored differently. (B) Crystal structure of the Ebola virus VP40 dimer: NTDs (green) and CTDs (pink). (C) Monomer of Sec24c: zinc finger (yellow), trunk domain (gray), beta-sandwich (purple), helical-domain (blue), gelsolin-like domain (red). (D) Monomeric structure of the VP40 protein.

5.3.1 SEC24C AND VP40 DOCKING

Before performing docking of VP40 with Sec24c, we explored the different binding regions in the Sec24 protein (PDB ID: 1M2V) with various other proteins. Figure 2 shows the different binding sites in Sec24 (blue) and the corresponding region in Sec24c (red). Of the five different binding sites, sites A and B are the most studied. The A-site is responsible for binding the Sed5 protein, whereas the B-site binds Sys1, Bet1 and Yor1. Another studied site, which is also known as the IxM binding site, contains the LIL motif and is shown to be important for binding of the Hepatitis virus protein Claudin. Recently, Pagent et. al. found a new site in Sec24 which is referred to as the D-site and is responsible for binding Erv14. Apart from binding several other proteins,
Sec24 possess a trunk domain which is responsible for binding with the Sec23 protein\textsuperscript{39}, an essential interaction for COPII transport. To determine the docking site on Sec24c and understand the interactions between EBOV VP40 and Sec24c on the atomic level, we performed molecular docking using several docking programs, including ZDOCK, ClusPro, and Patch Dock with the Fire Dock refinement. It has been proposed that residues 196-326 in VP40 may be important for docking to Sec24c. However, mutagenesis studies showed that the region spanning residues 303 to 308 are essential for Sec24c binding\textsuperscript{24}. With this VP40 binding information, we performed docking of Sec24c to a VP40 CTD domain.

Figure 5.2. A) Sequence alignment of Sec24 (blue) and Sec24c (red) highlighting different binding sites. Sec23 (yellow) binding to Sec24 in the trunk domain region is also shown. (B) Docking complex of Sec24c (red) with VP40-CTD (brown).
We performed blind docking of chain A of Sec24c with the VP40-CTD. Each docking program ranked Sec24c-VP40 complexes according to their docking score. Top ranked docking complexes from each of the docking programs were visually inspected, and the orientation and interactions between Sec24c and VP40 were analyzed. We visually inspected the orientation of region 303 to 308 in VP40 with respect to Sec24c for all complexes. Based on the visualization and interaction of VP40 residues with Sec24c, the second complex from Patch Dock was selected for further analysis. As shown in Figure 5.2b, in this docked complex the VP40 binding site was observed to be near the A-site (Figure 5.2a) of Sec24c. The interactions between Sec24c and VP40 in the docked complex selected was further optimized using molecular dynamics simulations.

5.3.2 Molecular interactions determined from MD computational simulations

To investigate the structural stability of the VP40-Sec24c complex, we performed a 500 ns all-atom, explicit solvent molecular dynamics (MD) simulation for the selected complex. Figure 3A shows the structure used for the MD simulation, which includes the CTD domain of VP40 and the trunk domain of Sec24c. Figures 5.3A and 5.3B show the structure of the Sec24c-VP40 complex at the initial and final frames of the 500 ns MD simulation, respectively. Inter-protein interactions within 3.5 Å of the protein at the binding region are highlighted: yellow represents the residues in Sec24c and green represents the residues for the VP40 protein. The complex remained intact for the entire 500 ns MD simulation, indicating that the complex maintains inter-protein interactions throughout the simulation time. The binding regions in Sec24c that we identified are comprised of residues 331 to 343, 535 to 546, 580 to 603, and 729 to 767. Most of these segments are close to the
Sec24c A-site. Consistent with Yamayoshi et al.,137 we found that the binding region in VP40 consists mainly of residues 288 to 326. VP40 residues 196 to 224 also interact with Sec24c, which also supports the idea that the proline rich region 205 to 219 is important for binding with Sec24c as discussed by Reynard et al.138 Figure 5.3C shows the VP40 binding region in Sec24c in surface representation (yellow) at 500 ns.

Figure 5.3. For (A) and (B), residues involved in the inter-protein interactions of the complex are highlighted: Sec24c (yellow) and VP40 (green). (A) Docked Sec24c-VP40 complex at 0 ns. (B) Sec24c-VP40 complex after 500 ns of MD simulation. (C) Surface representation of (B).

To quantify the binding between Sec24c and VP40, we calculated the root mean square deviation (RMSD) of the Ca atoms of both proteins. Figure 4A shows the RMSD as a function of time. The significant number of fluctuations before 250 ns suggest that the complex is undergoing structural rearrangements, which is also reflected in the large increase in RMSD. After 250 ns, the smoothness in the RMSD indicates the stability of the Sec24c-VP40 complex. The system is now stabilized by different types of interactions, such as electrostatic, hydrophobic, and hydrogen bonding interactions. To gain further insight into the binding affinity of the complex, we plot in Figure 4B as a function of time the total interaction energy (black line) between Sec24c and VP40, as well as two types of
interaction energies: electrostatic (cyan) and Van der Waals (grey). Like the RMSD of Figure 5.4A, the large decreases in the Total Energy and Electrostatic Energy plots before 250 ns again signify that the complex is rearranging and equilibrating. After 250 ns, the system has settled into a stable configuration. After quantifying that the interaction between Sec24c and VP40 complex was stable after 250 ns of MD simulation, we begin more detailed analysis which involves characterizing the interactions between the proteins in the complex. As shown in Figure 5.4B, electrostatic interactions are the primary reason for the stabilization of the Sec24c and VP40 system. The electrostatic interactions are dominated by salt-bridges between amino acids. More than eight salt-bridges were formed between Sec24c and VP40. The most stable salt-bridges were: D296-R334, K326-D600, E325-K395 and K212-D340. Figure 4C plots the distance between the nitrogen and oxygen atoms on the residues for four important salt-bridges as a function of time. Figure 4C shows that the D296-R334 separation and the E325-K395 separation each become much closer after 240 ns, and therefore both salt-bridges become stronger and help to stabilize the complex. We also observed a significant number of hydrogen bonds between Sec24c and VP40 (Fig 5.4D). Figure 5.4D shows that after the system has stabilized around 250 ns, the number of hydrogen bonds continues to increase. This suggest that the complex was stabilized mainly due to salt-bridge interactions and that hydrogen bonds play a role in
fine-tuning the docking. A more detailed analysis for hydrogen bonds is given below.

![Graphs](image)

Figure 5.4. (A) Root mean square deviation of Cα atoms of Sec24c and VP40 with respect to time. (B) Interaction energy between Sec24c and VP40 as a function of time: Total Energy (black), Electrostatic (cyan), van der Waals (VDW) (grey). (C) The distance between the N and O atoms in the residues involved in four important salt-bridges. (D) Number of hydrogen bonds between Sec24c and VP40 continues to increase after the system has stabilized.

5.3.3 H-BOND ANALYSIS FOR THE SEC24C-VP40 WILD-TYPE COMPLEX

Figure 5.4 A-C shows that during the 500 ns MD simulation, the Sec24c and VP40 complex has stabilized its large-scale structural arrangement by 250 ns. However, Figure 5.4D shows that the hydrogen bond network continues to change after 250 ns, implying that small scale structural changes continue. We performed a detailed analysis of the inter-protein hydrogen bonds for the last 250 ns of the MD simulation. For the hydrogen bond
analysis, a distance cut-off of 3.5 Å and an angle cut-off of 30° was used. Additionally, only heavy atoms were considered for the calculation. Figure 5.5A represents the percentage of the time (percent occupancy) during the final 250 ns that a specific H-bond existed between the amino acids of Sec24c and VP40. The important interacting residue-pairs between VP40 and Sec24c include: D296-R334, K326-D600, E325-K395, K212-D340, E325-K271, L288-E595, K212-R334, K212-D332 and P196-E761. The inset of Figure 5.5A shows the hydrogen-bond occupancy percentage of individual amino acids on VP40 and Sec24c. We found VP40 residue K212 is especially important for hydrogen binding between VP40 and Sec24c binding. Interestingly, K212 has also been found experimentally to be important in the formation of VLPs.151

Figure 5.5. (A) Percentage of hydrogen-bond occupancy for different inter-protein amino acid pairs where charged interactions are colored in yellow. Inset: Hydrogen-bond occupancy of specific residues on VP40 (left column) and Sec24c (right column). (B) Residues involved in important hydrogen-bonding and salt-bridge interactions for the Sec24c-VP40-Wt complex. VP40 is on the left and Sec24c on the right.
5.4.4. Binding Region Dynamic Network Analysis:

The binding region for VP40 in Sec24c lies in a small pocket in Sec24c, as shown in Figure 5.6. In Fig 6A, the binding region in Sec24c is highlighted in white and the trunk domain of Sec24c is shown in red. The CTD region of VP40 that binds to Sec24c is displayed in brown. Figure 6B is an expanded version of the binding region and shows the VP40 residues that are within 3.5Å of Sec24c. These VP40 residues include I324, K326, L288, L320, K224, K212, P317, E325, V323, C311, D312, H210. As described above, the VP40 docking in the Sec24c pocket is stabilized by various electrostatic interactions.

To further understand the connections and contacts between Sec24c and VP40, we performed a dynamic network analysis (DNA) using the Network plugin in VMD. Network analysis is a tool which highlights communities of amino acids that display correlated motion. Analyzing the details of the communities provides information about amino acids on each protein that are interacting and the strength of their connections. The number of nodes (circles) in a community is proportional to the number of amino acids composing that community, and line thickness denotes the weight of the connection between the amino acids. For this analysis, we considered the last 100 ns of the MD simulation, thereby allowing the protein complex to be fully stabilized. Different communities were generated using Carma\textsuperscript{45} and Catcd software. We focused on the communities which were responsible for connecting the two proteins (Sec24c and VP40). We observe three communities, colored green, blue and yellow (Figure 5.6C) with a total of 20 inter-protein connections. Important connections were made by VP40 K212 with eight amino acids on Sec24c (N338, D333, K335, L336, I337, R334, D340, T339). Other important amino acids
on Sec24c are A749 (with VP40 D312, T313), G748 (with VP40 T313), M733 and L734 (with VP40 P317), H737 (with VP40 A318), and P592 and A594 (with VP40 R214). Additionally, strong network connections that stabilize the complex are made between VP40 residue M305 with Sec24c residues V555, V572, and between VP40 residue I307 with Sec24c residue V572. We also performed MD simulations with mutations of VP40 residues M305 and I307 and compare out computational results with experimental results, below.

Figure 5.6. (A): VP40-CTD (brown) in the binding pocket of Sec24c(white). The trunk domain of Sec24c is displayed in red. (B) Residues in VP40 interacting in the binding pocket of Sec24c. (C) DNA communities between Sec24c-VP40-WT, where each community is colored differently (green, blue, yellow).

5.5.5. COMPARISON OF SEC24C-VP40-WT COMPLEX VS VP40 POINT MUTATIONS

Experimental studies by Yamayoshi et al.\textsuperscript{137} examined the effect of mutations on several residues in the binding region (303-308) of the VP40 protein. Out of several mutations carried out, they found L303A and I307A decreases the binding affinity with Sec24c, with L303A showing the bigger decrease, while M305A increases the binding affinity. However, the atomic-level details of the changes of the inter-protein interactions of the mutated-complexes compared to VP40-WT are unknown. We therefore investigated
the effect of mutations (L303A, I307A and M305A) on the Sec24c-VP40 binding using molecular dynamics simulations. We set up three additional simulations using VP40 mutants Sec24c-VP40-L303A (L303A), Sec24c-VP40-I307A (I307A) and Sec24c-VP40-M305A (M305A) and performed 500 ns MD computational simulations on each complex for comparison with Sec24c-VP40-WT. Figures 5.7(A-D) represent the Sec24c-VP40-complex at the end of 500ns for all four complexes. Figure 5.7D shows that L303A has large structural differences in the binding region compared with the wild type. Structural differences in the binding region compared to the wild type are also displayed by I307A in Figure 5.7C, whereas Figure 5.7B shows that M305A looks most like the wild type. To further quantify the differences in the binding for all four complexes, we plotted inter-protein hydrogen bonds and interaction energy between Sec24c and VP40 for the last 250 ns of the MD simulation in Figs. 5.7E and 5.7F. As shown in Figs. 5.7E and 5.7F, the Sec24c-VP40-WT complex has the largest number of hydrogen bonds and strongest interaction energy of the four complexes. Consistent with the experimental results, we find that L303A has the largest decrease in the number of hydrogen bonds and in interaction binding energy. We also find that I307A has noticeable decreases in the number of hydrogen bonds and in interaction binding energy compared to the wild type. In contrast, we find the M305A complex to have similar characteristics to the wild type.
Figure 5.7. Sec24c and VP40 complex A) WT B) M305A C) I307A D) L303A where Sec4c is represented in red color and VP40 in brown. The mutated residue on VP40 (Figs. B-D) is colored differently. E) Number of inter-protein hydrogen bonds as a function of time for all four complexes. F) Total interaction energy between Sec24c and VP40 with respect to simulation time.

5.5.6. ADDITIONAL DOCKING COMPLEX

The sequence of molecular docking steps involving VP40, Sec24c and Sec23 that facilitates the transport of VP40 to the plasma membrane is not yet completely understood. To gain insight, we also performed docking of VP40 at the site on Sec24c that can also bind Sec23. This site is on the trunk domain of Sec24c (Figure 5.2). Multiple docking
complexes were generated using various docking servers like Patch dock, Z-dock and Cluspro. We observed that the VP40 region of amino acids 303-308 binds strongly in the trunk domain region (residues 171-417) of Sec24c. Specifically, the beta-sheet region 303 to 308 in VP40 interacts with the Sec24c trunk domain’s beta-sheet region 241 to 247 and helical region 276 to 286 as shown in Figs. 5.8A and 5.8B. The Sec24c region 241 to 247 is also found to be the binding region of Sec23. To optimize the interactions and to study the stability of the system, we performed simulations of this additional docked complex (Sec24c-VP40-AD) for 500 ns. Figures 5.8A and 5.8B show the Sec24c-VP40-AD complex at 0 ns and 500 ns, respectively and the regions of inter-protein interactions are highlighted. For the VP40 protein, the binding region of residues 303 to 308 determined experimentally and from the computational molecular docking simulations remained stable during the MD simulation. In addition to VP40 residues 303-308, VP40 residues 233-243 also interact with Sec24c. When the alpha-helix on VP40 (residues 233-240) aligns with the alpha-helix on Sec24c (residues 276 to 286), several favorable, stabilizing inter-protein interactions can form involving residues within these alpha-helices as well as residues close to the alpha-helices. The important interacting residue-pairs between VP40 and Sec24c include T232-Q242, I307-V245, T232-M243, T242-D284, S233-Q242, D302-K227, A229-Q242, T304-S247, Q238-E280, S233-P241, K212-D251. Figure 8C displays a view that includes some of these interprotein bonds.

Among the amino acids involved in inter-protein interactions in the complex, 63% of the Sec24c residues are in the trunk domain’s beta-sheet region 241 to 247. For the VP40 CTD, 50% of the contributing residues lie within the 303-308 region and the other 50% are in an alpha-helical region (233 to 240). This result is consistent with the experimental study.
that showed that VP40 residues 303-308 are important for the binding to Sec24c. Additionally, the interfacial region of the complex contains a hydrophobic core involving residues VP40 residues M305, V306, I307 and L203, and Sec24c residues V228, M243, M244, V245, V246. These hydrophobic residues in the interface are shown in surface representation (gray) in Figure 8D. The binding of VP40 in the trunk domain region of Sec24c that also binds Sec23 suggests that VP40 might use the COPII transport mechanism in a manner that may not need the Sec23 protein for VP40 to be transported to the plasma membrane. However, how the VP40 utilizes the COPII mechanisms and whether it binds with Sec24c before Sec23 binds with Sec24c still remains elusive.

Figure 5.8. Sec24c-VP40-AD complex with highlighting of the VP40 binding region (green) and the Sec24c binding region (yellow). MD structure of the complex at: A) 0 ns, B) 500 ns. C) Important residues at the protein-protein interface. Hydrophobic residues (gray) form a strong hydrophobic core that contributes to the stability of the complex. The polar, negative, and positive residues at the interface are shown in green, red, and blue surface representation, respectively. D) View showing some interprotein residue-residue interactions between VP40 and Sec24c.
5.4. CONCLUSION

The Ebola virus life cycle requires transport of VP40 proteins to the plasma membrane and the COPII transport system plays a key role in this process. Sec24c is one of the proteins of the COPII transport system. Sec24c binds with VP40 and assists in the transport process to the intracellular leaflet of the PM where VP40 interacts with lipid molecules and oligomerizes into hexamers and forms the structural matrix filaments of the virus’ cylindrical lattice. Experimentally, it has been shown that Sec24c binds with the VP40 C-terminal domain (CTD), particularly in the region spanning VP40 amino acids 303 to 308. In this study, we investigated which regions of Sec24c are most likely to be involved in the binding process using various computational techniques. The binding regions in Sec24c that we identified are comprised of residues 331 to 343, 535 to 546, 580 to 603, and 729 to 767. There is also another possible VP40 docking site in the trunk domain of Sec24c that can also bind Sec23. The site on the trunk domain of Sec24c involves a beta-sheet region composed of amino acids 241 to 247 and an alpha-helical region consisting of amino acids 276 to 286. Interestingly, in addition to the known VP40 binding region including the amino acids 303-308, we also observed binding interactions involving VP40 residues 232-243, which are part of an alpha-helical secondary structure. Our MD computations showed that after the initial docking of VP40 at this alternative site on Sec24c, the secondary structures (beta-sheet and alpha-helix) on both proteins at the interface undergo structural rearrangements which allow the Sec24c-VP40 complex to optimize the interactions and make the complex stable via inter-protein hydrogen bonds, hydrophobic interactions, and salt-bridges. We also examined the effects of three VP40 mutations on the inter-protein interactions of the complex. We investigated the Sec24c-
VP40-L303, Sec24c-VP40-M305A, and the Sec24c-VP40-I307A complexes. Consistent with experimental results on binding affinity, we find that L303A has the largest decrease in the number of inter-protein hydrogen bonds and in interaction binding energy. We also find that I307A has noticeable decreases in the number of hydrogen bonds and in interaction binding energy compared to the wild type. In contrast, we find the M305A complex to have similar characteristics to the wild type. The sequence of molecular docking steps involving VP40, Sec24c and Sec23 that facilitates the transport of VP40 to the plasma membrane is not yet completely understood. The atomic-level findings provide insight into the mechanism of how Sec24c binds with VP40, a necessary step for the transport of VP40 to the cell’s PM. These molecular details about a critical step in the EBOV life cycle provides information that can be helpful in developing drug molecules that target the EBOV. The binding of VP40 in the trunk domain region of Sec24c that also binds Sec23 suggests that VP40 might use the COPII transport mechanism in a manner that may not need the Sec23 protein for VP40 to be transported to the plasma membrane. However, how the VP40 utilizes the COPII mechanisms and whether it binds with Sec24c before Sec23 binds with Sec24c still remains elusive.
CHAPTER 6: DYNAMICS AND MEMBRANE ASSOCIATION OF ZIKA VIRUS PROTEIN NS1: A MOLECULAR DYNAMICS STUDY

6.1 INTRODUCTION

Zika virus (ZIKV) is a human pathogen transmitted primarily through mosquitoes and is closely related to other flaviviruses such as Dengue (DENV), West-Nile, and Encephalitis\textsuperscript{152}. First discovered in 1947 in Uganda, ZIKV has evolved as a major global health threat due to its link with severe clinical manifestations, including microcephaly in neonates and Guillain-Barre-syndrome in adults\textsuperscript{22,152-155}. Like other flaviviruses, ZIKV is transmitted through mosquitoes. It is primarily transmitted through the Aedes aegypti mosquito, but it can be transmitted by sexual activity as high viral loads have been detected in semen from infected patients\textsuperscript{156}. ZIKV infection to a pregnant woman may cause miscarriage and preterm birth\textsuperscript{154} and babies born to ZIKV infected mother may have birth defects and develop microcephaly and brain disorders\textsuperscript{157}. The rapid spread of ZIKV around the globe has made it one of the most significant public health concerns\textsuperscript{158}. Due to the increased threat of ZIKV infections and the lack of effective antiviral therapeutics or vaccines, there is an urgent need to develop novel preventive and therapeutic modalities based on an understanding of the molecular mechanisms underlying the disease. Unfortunately, very little is understood about the molecular level details of the pathophysiology of ZIKV. Like DENV, ZIKV genome encodes ten different proteins, three of which are structural and seven are non-structural. The structural proteins are Capsid protein (C), Envelope protein (E), and M protein. The M protein is responsible for monitoring the assembly of the envelope proteins The seven non-structural (NS) proteins
(NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) mainly function in RNA replication, immune evasion and particle assembly. Except NS1, rest of the non-structural proteins are responsible for forming replication complex on the cytoplasmic side of the ER, whereas glycosylated NS1 protein localized with the ER-membrane where it clusters to from dimeric structure inside the cells and execute the viral replication process after the infection. Gutsche et al. studies suggested that dimeric form of NS1 interacts with the membrane, which allows the dimers to cluster together at the regions of lipid droplets in the membrane. This activity disrupts the membrane which results in formation of NS1 hexamers structure with central lipid core. The hydrophobic surface of each dimer points inward in hexameric structure, forming a lipid cargo complex which is refer as “hexameric lipoprotein particle”. The hexameric lipoprotein function in immune circumvention and pathogenesis.

Many major proteins of virus are known to change their structure and form higher oligomeric states after it binds with lipid molecules. To gain insights of viral life cycle and its other aspects, the membrane-protein interaction is very important and has been studied for various viruses. The structural rearrangement upon phospholipids binding are important for forming different oligomer states and also for ion-channels. Experimentally, it has been studied that ZIKV NS1 dimer associates with the endoplasmic reticulum (ER)-bilayer and this step is required in forming hexameric lipoprotein. Till date, there has been limited number of studies which explains the molecular level details of how the different proteins in Zika virus perform their roles during their viral life cycle. The mechanism and molecular level details of NS1-membrane interaction remains still elusive. In this work, we investigated the molecular level mechanism of how NS1-
membrane interaction occurs and identified the important amino-acids and lipids for the association by performing one microsecond of all-atom molecular dynamics simulations with explicit water molecules.

6.2 System Set-up

The NS1 crystal structure of ZIKV-virus from 2015 Brazil strain was extracted from Protein Data-Bank (PDB ID: 5GS6). The missing residues and missing N-linked glycosylation at two of the sites N207 and N130 were modelled. The protein-membrane system was made using Charmm-gui and VMD software. As, ZIKV NS1 dimer interacts with lower endoplasmic reticulum membrane, we constructed ER-membrane with a symmetric lipid composition on lower and upper leaflet of membrane, to investigate the lipid selectivity in the NS1 dimer-ER membrane, different lipid compositions of the membrane were considered. To achieve fully equilibrated protein membrane systems, we employed the highly mobile membrane-mimetic (HMMM) model. The HMMM membrane uses lipids with shortened tails, floating on a layer of 1,1-dicholoroetahne representing the hydrophobic inner core lipid bilayer. The fluidity of the system allows significantly enhanced lipid diffusion that can provide unbiased lipid-binding of the protein. We placed the membrane interacting interface of ZIKV NS1 dimer facing the complex membrane containing phosphatidylcholine (POPC), phosphatidyethanolamine (POPE), phosphatidylserine (POPS), phosphoinositol (POPI), phosphatidylinositol 4,5-biphosphate (PIP2), cholesterol (CHOL and single chain fatty-acid (TRI)) in both the leaflets of the membrane. The complex was solvated with TIP3 water molecules in a cubic box and was neutralized with counter ions. The total system consists of 240,000 atoms.
6.3 Structure of NS1:

The crystal structure of NS1 Zika virus exists in dimeric form. It consists of 352 residues where two residues N130 and N207 on each monomer consists of N-linked glycosylation with one sugar residue. Being like Dengue virus, Zika NS1 each monomer consists of central beta-sheet domain, wing domain, beta-hairpin domain and beta-ladder domain. The beta-hairpin domain consisting of residues 1 to 30 twist with another monomer and make “roll” completing the dimer structure. This interface is also referred as beta-roll dimerization domain. The wing domain (31 to 181) continues from beta-ladder domain which also contain N-linked glycosylation sites at N130 and N207 is also critical for membrane localization. A wing domain consists of various charged residues and hydrophobic residues Y122, F163 which are suggested to be important for membrane association. The beta-ladder domain (consists of ladder of beta-sheets connecting wing domain and to beta-roll domain (Figure 6.1). The NS1 dimer from beta-ladder which includes in total of 20 beta-strands arranged in the form of ladder.

Figure 6.1: Crystal structure of NS1 ZIKV dimer, beta-ladder(gray), wing domain(yellow), intertwined loop (green) and beta-roll (magenta). The residues N130 and N207 are glycosylated with sugar group (BGL group). Important hydrophobic residues for membrane localization are highlighted.
6.4. Result

6.4.1. NS1 - Membrane Localization:

To investigate the lipid protein interaction and localization phenomenon, (as the system was quite large and to save some computational time) we placed dimer slightly below the lower leaflet of ER-membrane, with protein atoms being < 3-4 Å below the atoms of the lipid molecules. We incorporated all the lipid molecule types and also used 10% of fatty acid. We observed the ZIKV dimer and membrane interactions on lower leaflet of membrane. Fig 6.2a shows the arrangement of protein-lipid at around 5 ns where of the residues on wing domain and beta-roll were slightly interacting with head groups of lipid molecules. Initially, the basic residues started interacting with bilayer which then allowed protein to be more stabilized which was then followed by interaction of hydrophobic and polar residues which were present in beta-roll and wing domain. Y122, H164, C4, F8 from both monomers of NS1 are the important hydrophobic residues penetrating the membrane (Fig 6.2c). The positively charged lysine and arginine residues form electrostatic interaction with the head groups of negatively charged lipid molecules PS, PIP2, POPI (fig 6.2d) and facilitated the localization of dimer on ER-membrane. Specifically, the residues in the wing domain, intertwined loop and some from beta-roll were responsible for protein membrane interaction and allow the ZIKV dimer to associate fully with lipid bilayer.
Figure 6.2: Snapshots of protein association with membrane at different trajectory, where different regions of NS1 are shown; wing domain (yellow and green), beta ladder domain (white), beta roll of two monomers (purple and blue) for: (a) 5 ns (b) 1000ns. c) Few residues inserted in lipid bilayer after protein-membrane localization d) Lipids interacting with different residues of protein within 3.5 Å of protein.

6.4.2. INSERTION OF HYDROPHOBIC RESIDUES

Figure 6.3: Amino acids important for membrane stabilization and association in (a) Wing domain of NS1 with intertwined loop (green), (b) NS1 beta roll.

Recent studies have suggested the possibility of the role of intertwined loop and beta-roll of NS1 protein is important for membrane association. The intertwined loop contains several hydrophobic residues, mainly Y122, F123, F163, and H164 (Figure 6.3).
This hydrophobic surface acts as membrane interacting interface for NS1 protein. We placed the NS1 dimer below the membrane so that this hydrophobic surface is well exposed for membrane interaction and insertion. We performed all-atom simulation for NS1-membrane system for 1000ns and observed the process of insertion of hydrophobic residues in the lipid bilayer. Initially, the residue Y122 from one of the chains of NS1 begins penetrating the hmmm membrane which facilitates the environment for charged residues R40, R125, R69, K120, R31, R29 present in beta roll and wing domain to interact with lipid molecules. This interaction of charged residues with membrane allows the dimer to tilt one side. Around 400ns, after the interaction were strong enough, Y122 from second chain begins penetrating the membrane and start anchoring the dimer. In the meantime, interaction with Y122 and the charged residues with the lipid molecules was also continuously strengthening. Due to continuous lipid molecules interacting with charged residues, they attract the more lipid molecules which results in the lipid accumulation around the charged residues in beta roll. Interestingly, we even observed the slight insertion of positively charge residue K11 in membrane. As a result, the significant void is created in lipid bilayer around the beta roll region which allowed the interaction of hydrophobic residues with the membrane (Figure 6.4b and 6.4c). The anchoring by Y122 and slight protrusion of F8, C4, F163, K11, H164 allows the dimer to be fully stabilized with lipid
Figure 6.4: (a) Residues penetrating in the lipid bilayer. Orientation of NS1 where cyan represents phosphate atoms (b) before lipid interaction begins and (c) after lipid interaction where, dotted circle indicates the region of void formation where residues begin penetrating are also shown.

We found that residue Y122 plays an important role for the stabilization of the protein. Initially, Y122 interacts with the phospholipids and in both monomers, Y122 acts as anchoring residue for NS1 in both monomers and facilitates the interaction of charged residues with phospholipids bilayer. Apart from Y122, F8 and H164 also well associates with lipid molecules and promotes the ZIKV-membrane localization. As previously suggested by experimentalist, these hydrophobic protrusions basically on wing domain are critically important for dimer functioning on membrane.

6.4.3. ROOT MEAN SQUARE DEVIATION

To investigate the stability of association of NS1 with ER-membrane we calculated root mean square deviation for protein averaged for all carbon-alpha (CA) atoms in dimer with respect to time (Fig 6.5b). For 0-200 ns, the rmsd changes rapidly and after 200 ns as the protein begins to interact with the membrane the protein begins to stabilize. The ZIKV NS1 dimer localization was mostly mediated through electrostatic interaction between
positively charged residues and negatively charged head groups of lipid molecules and hydrophobic interactions. To quantify the interaction, we plotted hydrogen bond between NS1 dimer and membrane for last 500ns of the simulation. We plotted number of hydrogen bonds for two monomers separately (Chain A and Chain B) and plotted for overall protein with the membrane. Fig 6.5a displays the plot of hydrogen bonds between heavy atoms of dimer and membrane. The distance cut of 3.5 Å and angle cut-off of 30 degrees was used for the calculation. The protein–membrane interaction starts with Chain A. Initially, the charged residues from Chain A begins associating with lipid molecules and this association is followed by interaction of more residues form Chain A. This results in slight tilt for protein, specifically for Chain B. After localization of Chain A, it pulls back the Chain B towards the membrane. Hence, Chain A was found to be making more hydrogen bonds initially with the membrane. We observed significant increase in hydrogen bonds till 650-700 ns and a slight decrement begins after 700 ns, which was at lowest point at around 780 ns. The first 500 ns of simulation allowed dimer to be fully associate with the bilayer which results in significant increase in hydrogen bonds between protein and membrane. After 600ns, the proper orientation of dimer due to anchoring of hydrophobic residues in wing domain and beta roll may be responsible for slight change in hydrogen bonds pattern. At later time in simulation, the most contacts with the NS1 were maintained by hydrophobic residues. This contacts with hydrophobic residues made system more stable. The more detail of hydrogen bonds and contribution of each amino acids involved in bonding will be discussed in later sections.
Figure 6.5: (a) Time evolution of number of hydrogen bonds for Chain A (cyan), Chain B (gray) and Protein (red). (b) Root mean square deviation (rmsd) of ZIKV dimer with respect to time.

6.4.4 Lipid-count and Relative Percentage of H-bonds

Figure 6.6a displays the normalized number of counts of different lipid atoms with the atoms of NS1 dimer as a function of time. For this calculation, the distance cut off of 3.5 Å was used. Initially, there were few contacts made by head groups of lipid molecules with the protein. As the protein begins associates with the membrane, the contacts begin to increase significantly for all the lipid types except for POPS and Cholesterol. All the lipid types were found to be making significant contact with the protein. Out of seven different lipid types, the most contacts were made by POPC, which was then followed by POPI, POPE and TRI (type of fatty acid). These contacts were mainly made by charged residues present in wing domain and beta-roll.

To further understand the contribution of different amino acids for association with membrane, we plotted the histogram showing the relative percentage of hydrogen bond occupancy for top residues for the protein (fig 6.6b) and compared values for two monomers the hydrogen bond occupancy of less than 15% was not taken into account for this calculation. Important residues for hydrogen bonding with lipid molecules include...
amino acids mainly from wing domain and beta roll. Out of all, R40 was the top contributor which was then followed by amino-acids R125, K116, R31, S121, K120, R69, T165, R306, H164 and so on. For Chain A, R40, R125, R31, S121, R306, K120, G114 were the top contributing residues whereas for Chain B, R40, R69, K116, T165, H164, K120, Y32 were the important residues for lipid binding. As suggested from experimental studies, we observed most residues from wing domain (R40, R125, K116, S121, R69, T165, H164, Y32, G114, G119) and R29 from beta roll. The beta roll region consists mostly hydrophobic residues which might be reason for less hydrogen bonding from beta-roll region. Our results are in well agreement with experimental findings.

Figure 6.6: (a) Normalized lipid contacts between ZIKV dimer and membrane with respect to simulation time. (b) Histogram representing relative percentage of hydrogen bonds formed by different residues of ZIKV dimer with lipid molecules. c) Different residues in beta roll region interacting with lipids d) Amino-acids interacting with lipids in wing domain e) Hydrogen-bond formed by Y122 with lipid.
6.4.5 ROOT MEAN SQUARE FLUCTUATION

The protein lipid association results in the slight conformational changes in protein. To monitor the flexibility of the residues we plotted the root mean square fluctuation (rmsf) of the protein. We plotted rmsf for two monomers separately for different time period, first 200 ns and last 200 ns of simulation (Fig 6.7). The plot clearly shows the decrease in flexibility for most of the residues in both monomers as they interact with the membrane. Basically, the residues on wing loop domain (residues 30 to 180) and beta roll (1 to 30) domain have large decrease in flexibility. For first 200 ns the residues on wing domain for monomer A has comparatively more flexibility than monomer B. For last 200 ns, as the NS1 localized with the membrane by the rmsf drops significantly for monomer A. At the end of simulation, both the monomers have similar trend of flexibility which is as expected as the protein now is completely associated with membrane (fig above).

Figure 6.7. Root mean square fluctuation (rmsf) for a) Chain A b) Chain B and c) Wing domain of NS1 protein dimer.

6.4.6. INTERACTION ENERGY BETWEEN PROTEIN AND MEMBRANE:

To quantify the stability of protein and membrane interactions we calculated the total interaction energy between the protein and membrane. The NAMD Energy plugin was used for the calculation. Fig 6.8 shows the plot of electrostatic, Vander Waal’s and
total energy (electrostatic+ vdw + non-bonded) between NS1 dimer and membrane as a function of time. Initially, the total energy of the system was significantly higher (-5 kcal/mol), but, as the protein begins stabilizing on the bilayer, the energy of the system lowered to around -30 kcal/mol and this energy was consistent throughout the simulation time. The top energy contribution was due to electrostatic energy, which was as expected, as the most interactions with lipid molecules were mediated through charged residues of the protein.

![Interaction energy between protein and membrane with a function of time.](image)

Figure 6.8. Interaction energy between protein and membrane with a function of time.

6.4.7 Change in electrostatic potential distribution

Previous findings from experimental study indicates that Zika NS1 protein exhibits unique electrostatic potential distribution on comparing with other Flavivirus species. We investigated the difference in electrostatic potential before NS1 fully interacts with the membrane and after NS1 is completely localized with the membrane with its few hydrophobic residues inserting in the bilayer. Fig 6.9 represents the electrostatic potential
of ZIKV NS1 at 0 ns and at around 800 ns. We highlighted both top view and lipid binding view. For the top view, the significant difference was observed as indicated by dotted black circle. The charge distribution was quite symmetric for the crystal structure of the protein while after associating with membrane (at 800 ns) the positive charges were seems to be accumulated at wing domain region. In the lipid interacting zone of NS1 dimer, we observed some important changes in the pattern of positive charges. Before interacting with lipid molecules (at 0 ns) the positive charges were not forming the clusters. After the dimer was localized with the membrane (at 1 us), we clearly observed the clusters of positively charge residues in beta roll and wing domain region which is also zone for lipid interaction. Additionally, at the end of simulation, the beta roll seems to have nice arrangement of hydrophobic residues and positive residues at the end of 1 us simulation which also characterize the insertion of hydrophobic residues into the lipid bilayer.

Figure 6.9. Electrostatic potential distribution of NS1 protein at 0 ns and at around 800 ns with top and bottom view shown.
6.4.8 Conformational Changes in ZIKV Dimer

There has been lot of studies which focuses on the how the protein undergoes conformational changes after associating with membrane. This type of conformational changes is important for the protein to perform its function. For some virus, the conformational changes upon lipid binding are crucial to undergo various higher order oligomeric structures. We observed the slight conformational changes on ZIKV dimer upon lipid association, which was mostly dominant in beta-ladder region. This conformational change not only facilitated the lipid binding but also, made the protein more organized with respect to mass density. We noticed the slight extension of protein, specifically on beta-ladder region after 1000 ns of simulation. Visualizing the trajectory more carefully, we noticed the changes in salt-bridge pattern formed between residues K191-D157 and D234-K227 between two chains of dimer. Initially, both salt-bridges are quiet strong and later after membrane association, the salt bridge begins to weaken and breaks apart completely. Fig 6.10a shows the plot of distance between residue pair K191-D157 and D234-K227 as a function of time. This weakening of salt bridges between the domains mainly in center region of protein (Fig 6.10a) resulted in the overall extension of beta-ladder and conformational change of the protein.
To quantify more in detail, we mapped the mass density of the protein using Volmap plugin form VMD software. We took the averaged structure for first 10 ns and last 10 ns of simulation to compare them. Fig 6.11 shows the mass density profile for three different views of protein along X, Y and Z-direction. As shown in figure, the mapped mass density profile for last 10 ns averaged structure seems more organized between the two domains of protein. From this map, we also can visualize the significant changes in the structure of the protein after it interacts with the membrane along all three axes. Along, X-axis which is also membrane-binding view, we can notice the cluster of blue, specifically in beta-roll region which also indicates the collection of residues after it interacts with lipid-bilayer. The fading of blue color can be seen in beta-ladder region (along Y) in the center, which also further supports the extension of the protein by breaking of the salt bridges. Along z-direction, we can see that the mass density of the protein is more uniformly distributed, and protein is more stretched after 1000 ns of simulation.
In this work, we investigated the molecular details and mechanism of how ZIKV NS1 dimer associates with a lipid bilayer. We employed all-atom molecular dynamic simulations using highly mobile membrane system for the complex of NS1-membrane for one microsecond and identified the important residues for the localization of NS1 in the membrane. We studied the insertion of hydrophobic residues mainly Y122, F8, C4, F163 in beta roll and wing domain region which facilitated the stabilization of protein in the membrane. The anchoring of Y122 residue present in wing domain initiates the localization process for the protein, which then facilitates localization by interaction of various charged residues. This results in the insertion of hydrophobic residues into the membrane. The residues important for hydrogen bonding were R40, R125, K116, R31, S121, K120, R69,
T165, R306, H164 which were majorly present in wing domain. We studied the electrostatic potential map and observed the significant differences between the initial and final configuration of protein. The protein forms more cluster of positive charge residues after it associates with the lipid-bilayer. We studied the conformational changes in ZIKV protein after it interacts with the lipid molecules. These conformational changes were mostly due to breaking of salt bridges which were present between the domains. The weakening of the salt-bridge allowed the protein to stretch slightly mainly in beta-ladder domain region. These residues might also be important for stabilizing the dimer as it is between the monomer-monomer interface. Further mutational study is needed to identify the role of these residues. To further understand the structural re-arrangement, we studied the mapped mass-density of the protein for averaged structure of first 10 and last 10 ns of simulation and observed the significant differences along three axes X, Y and Z. These findings conclude that the protein undergoes the conformational changes between the domains allowing the protein to be more organized in terms of mass-density, extending the protein slightly. The membrane-protein interaction mechanism study will be important to identify the role of various residues and lipid molecules for viral life cycle and to understand the structural re-arrangements, how the protein stabilizes itself in lipid-bilayer. This study gives insight into molecular level details of mechanism of how NS1 ZIKV dimer localizes itself in membrane and how it undergoes structural changes. This detail might be important for targeting the drug molecules for NS1 protein.
CHAPTER 7: OVERALL SUMMARY

The recent Sars-Cov2 pandemic, along with continued outbreaks of Ebola and Zika shows that there is always a looming threat of new viral outbreaks. Molecular-level investigations of viruses are very important to treating existing outbreaks and preventing new viral outbreaks. Viruses are amazingly adaptable due to rapid mutations on the molecular level. My research work focuses on understanding on a molecular level the structural dynamics and functioning of three dangerous zoonotic viruses, namely Ebola, Marburg and Zika. Different types of proteins in viruses perform various functions. An important type of viral protein is membrane interacting viral proteins that perform the interactions between the virus and the host cell membrane. These membrane-interacting proteins are crucial in allowing a virus to enter a host cell, replicate the viral genetic material, and create new virus particles that evade the immune system of the body. Understanding the dynamics of viral membrane-interacting proteins on an atomic level offers an opportunity to develop potential inhibitors for treating the disease. These investigations are on the forefront of bio-nanophysics and molecular medicine. I find that the elegant simplicity of the laws of physics when combined with the random events of evolutionary natural selection allows viruses to use a range of different molecular interactions to create proteins that facilitate viral life cycles. This is a specific example of the underlying tension of life: organismal strength versus adaptability, molecular stability versus flexibility, thermodynamic enthalpy versus entropy.

Filovirus infections cause hemorrhagic fever in human and non-human primates\textsuperscript{49} that often result in high fatality rates\textsuperscript{50}. The Ebola and Marburg viruses are lipid-enveloped viruses from the *Filoviridae* family. The virus particle acquires a lipid coat from the plasma
membrane (PM) of the host cell as the virus assembles and buds. The viral structural matrix layer underneath the lipid envelope is formed by the matrix protein VP40 and provides shape and stability to the viral particle. It has been hypothesized that, upon binding to the host cell’s plasma membrane (PM), the filovirus VP40 dimers undergo major structural rearrangements. This step is required for oligomerization into hexameric structures which further assemble to form filaments leading to the formation of the viral matrix. In my work, the mechanisms, and consequences of the VP40-lipid interactions on VP40 dynamics and assembly are investigated. My study focuses on molecular details of how VP40 associates with the host cell’s plasma membrane, the conformational flexibility of different regions of VP40, and how VP40 interacts with the human Sec24c protein to enable transport within the cell to the PM. I explored the binding regions and interactions with the Sec24c protein and identified the important residue pairs for Ebola virus VP40 and Sec24c interaction using molecular dynamics simulations. As VP40 is transported to the plasma membrane it undergoes conformational changes which was observed prominently in the Marburg virus protein (mVP40). Despite the significant structural differences with the Ebola eVP40, the mVP40 dimer is found to adopt a very similar configuration to eVP40 dimer after associating with the membrane. These structural changes were not observed in the absence of a membrane where each dimer retains its noticeably different shape. The structural change upon lipid-interaction may prove useful in targeting and disrupting the mVP40 dimer.

I also highlighted the important residues and lipids for membrane association of VP40. Membrane interactions also enable oligomerization of VP40, but oligomerization also requires conformational flexibility within VP40. On exploring different domain
interfaces of the eVP40 filament forming oligomer, the CTD-CTD interface was found to be most flexible and the oligomeric NTD-NTD interface was the most stable interface. The high flexibility of the CTD-CTD interface in the eVP40 filament may be essential for the supple bending of the Ebola virus, while the rigidity of the oligomer interface may be necessary for the structural stability of the Ebola virus. The flexibility and rigidity of the interfaces may present targets for molecular interventions to disrupt the functioning of the Ebola virus.

I also studied proteins found in the Zika virus. Zika is a member of the flavivirus genus, which is part of the Flaviviridae family. The genus flavivirus also includes the Dengue virus (DENV), West-Nile, and Encephalitis. Zika virus (ZIKV) is a human pathogen transmitted primarily through mosquitoes and is closely related to other flaviviruses such as Dengue (DENV), West-Nile, and Encephalitis. Due to the increased threat of ZIKV infections and the lack of effective antiviral therapeutics or vaccines, there is an urgent need to develop novel preventive and therapeutic modalities based on an understanding of the molecular mechanisms underlying the disease. My research focuses on one of the Zika virus proteins NS1 which associates with the host cell’s endoplasmic reticulum membrane. When bound to the ER membrane, the NS1 protein forms dimeric structures and assists in viral replication. I found computationally that the insertion of hydrophobic residues (Y122, F8, C4, F163) in a beta-roll and wing domain region of NS1 facilitates the stabilization of the protein in the membrane. The anchoring of the Y122 residue present in the wing domain initiates the localization process for the protein, which is then enhanced by interaction of various charged residues in the protein. I also highlighted residues important for hydrogen bonding with lipids. This study gives insight into
molecular level details of the mechanism of how Zika’s NS1 dimer binds to membranes and how NS1 undergoes structural changes. These details might be useful in designing drug molecules that disrupt the functions of the NS1 protein.
REFERENCES


40. Humphrey, W.; Dalke, A.; Schulten, K. J. o. m. g., VMD: visual molecular dynamics. 1996, 14 (1), 33-38.


106. Huang, H.; Ji, H.; Li, H.; Jing, Q.; Labby, K. J.; Martasek, P.; Roman, L. J.; Poulos, T. L.; Silverman, R. B., Selective monocationic inhibitors of neuronal nitric oxide...


159. Muller, D. A.; Young, P. R., The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral research* **2013**, *98*(2), 192-208.


VITA

NISHA BHATTARAI

Education

2016-2021     Ph.D., Physics, Florida International University, Miami, FL
2016-2019     MS, Physics, Florida International University, Miami, FL
2012-2014     MS, Physics, Tribhuvan University, Kathmandu, Nepal
2009-2012     BS, Physics, Tribhuvan University, Kathmandu, Nepal

Publications and Presentations


10. Marburg VP24 Protein K-loop Cysteine Interactions with the Human Keap1 Protein, American Physical Society Meeting, Denver, 2020. [Link]

11. Molecular Basis of the Marburg Virus Protein VP24 interactions with Human Keap1, Biophysical Society Meeting, Baltimore, Nisha Bhattarai, 2019. [Link]