Molecular Simulations of HIV-1 gp41 Protein-Membrane Interactions

A Thesis
Submitted to the Faculty
of
Drexel University
by
Vamshi K. Gangupomu
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical and Biological Engineering
October 2010
Dedications

This thesis is dedicated to my parents, Uma Sundari Gangupomu and Uma Ramana Gangupomu. Mom and Dad, I am grateful for everything you have done and continue to do for me.
Acknowledgments

I would first like to thank my advisors Prof. Franco M. Capaldi and Prof. Cameron F. Abrams for the opportunity and for their invaluable guidance and motivation throughout my doctoral studies. I appreciate their respect and trust towards their students’ ability and for providing me the facilities and freedom to perform my work. Both my advisors have been incredible sources of knowledge and support and have profoundly influenced my scientific thinking and research acumen. I am fortunate to have had an opportunity to work with two true scholars and excellent researchers and will look up to them as role models both professionally and personally.

I would like to acknowledge the invaluable suggestions and feedback of my thesis advisory committee Prof. Raj Mutharasan, Prof. Nily Dan and Prof. Irwin Chaiken. Special thanks to Prof. Chaiken for many insightful discussions and for sharing his vast knowledge.

Many thanks to my lab partners in the Abrams group, Dr. Harish Vashisth, Dr. Debashish Mukherji, Ali Emileh, Spencer Stober, Shyno Mathew and Michelle Baker for their fruitful discussions and help in many ways. Also, I would like to thank Dr. Mark Contarino and Karyn McFadden from the Chaiken group for the invaluable insight into in vitro studies. I would also like to thank my fellow graduate students at Drexel John Fitzpatrick, Roza Mahmoodian, Ranjita Bose, Harsh Sharma and Jackie Yim for providing a fun and stimulatory environment to work.

My eternal gratitude to my friends Naveen Kammili, Madhuri and Krishna Pinnamaraju, Swapna and Vijay Harsha for helping me take my mind away from the challenges of my research. Special thanks to Hari Kamaraju for his interesting discussions and Chetana Sunkari for her wonderful support and help in many ways. I am especially grateful to Dr. Sally Nyquist for being an incredible source of inspiration. I also thank Prof. Kevin Scoles of the Department of Electrical and Computer
Engineering for being an excellent mentor and for making my teaching experience at Drexel thoroughly enjoyable.

I would like to acknowledge the National Institutes of Health (grant No. R01AI-084117-01), the National Science Foundation (grant No. DMR-0427643), the Department of Chemical and Biological Engineering and the Department of Mechanical Engineering and Materials at Drexel University for their financial support and the National Science Foundation for the computational grants TG-MCB070073N (CFA) and TG-MCB090139 (FMC) through TeraGrid.

I am highly grateful to my brother Ajay Gangupomu and his family, my sister-in-law Sri Vidya Gangupomu and my little niece Tanvi Suma Gangupomu, for their love and emotional support. I am especially thankful to my brother for his encouragement and for being an excellent role model. I have always looked up to him both professionally and personally and will continue to do so. Finally, I thank my parents Uma Sundari Gangupomu and Uma Ramana Gangupomu for their unconditional love and support and for cultivating a general interest in science.
# Table of Contents

List of Tables ................................................................. ix
List of Figures ............................................................... x
Abstract ................................................................. xiii
1. Introduction .......................................................... 1
  1.1 Background and Motivation ........................................ 1
  1.2 Lipid Membranes .................................................. 2
    1.2.1 Phospholipids ................................................. 4
    1.2.2 Cholesterol .................................................... 5
  1.3 Membrane fusion .................................................. 6
    1.3.1 Fusion proteins ............................................... 6
    1.3.2 Fusion mechanism ............................................. 7
  1.4 The human immunodeficiency virus type 1 (HIV-1) ............... 8
    1.4.1 Glycoprotein-41 ............................................. 9
  1.5 Objectives and specific aims ..................................... 10
    1.5.1 Specific Aim 1: Compute the forces and free-energies of membrane rupture ............................................. 10
    1.5.2 Specific Aim 2: Structure and orientation of HIV-1 gp41 MSD and effects on lipid membrane structure ......................... 11
    1.5.3 Specific Aim 3: Folding of gp41 ectodomain from a pre-fusion intermediate (PFI) to a six-helix bundle (6HB) ..................... 11
2. Simulation Methods .................................................. 12
  2.1 Introduction ....................................................... 12
    2.1.1 Statistical mechanics and ensembles ......................... 13
  2.2 Molecular Dynamics .............................................. 15
    2.2.1 Langevin Thermostat ........................................... 15
2.2.2 Numerical Integration ........................................... 16
2.2.3 Force Field ..................................................... 17
2.2.4 Limitations of molecular dynamics ............................. 18
2.3 Steered molecular dynamics ....................................... 19
2.4 Metadynamics ..................................................... 20
2.5 Targeted molecular dynamics (TMD) .............................. 21
2.6 Software and resources used ...................................... 22
3. Mechanical destabilization of a cholesterol containing lipid bilayer .......................... 23
  3.1 Introduction ..................................................... 23
  3.2 Simulation methods and models .................................. 24
     3.2.1 Order parameter ........................................... 25
     3.2.2 Membrane destabilization simulations ...................... 26
  3.3 Results ......................................................... 26
     3.3.1 Membrane properties ...................................... 26
     3.3.2 Membrane rupture force and effects of cholesterol ........ 27
     3.3.3 Free energy of membrane rupture .......................... 29
  3.4 Discussion ...................................................... 29
  3.5 Conclusions ..................................................... 33
4. HIV-1 gp41 membrane spanning domain structure in water ................................. 36
  4.1 Introduction ..................................................... 36
  4.2 Simulation description .......................................... 38
     4.2.1 General .................................................. 38
     4.2.2 Gp41 MSD model ......................................... 38
     4.2.3 Metadynamics ............................................. 38
  4.3 Results ......................................................... 39
4.3.1 HIV-1 gp41 MSD structure in water from molecular dynamics simulations .................................................. 40

4.3.2 MSD structure in water from metadynamics ...................... 40

4.4 Discussion ........................................................................ 42

4.5 Conclusions................................................................. 44

5. HIV-1 gp41 membrane spanning domain structure in cholesterol containing lipid membranes .................................................. 45

5.1 Introduction ..................................................................... 45

5.2 Simulation description .................................................... 46

5.2.1 General ................................................................. 46

5.2.2 Gp41-membrane model ............................................ 47

5.2.3 Metadynamics ........................................................... 49

5.3 Results ........................................................................... 50

5.3.1 Effects of Exoplasmic snorkeling of Arg694 on MSD conformational ensembles .................................................. 50

5.3.2 Effects of Endoplasmic snorkeling of Arg694 on MSD conformational ensembles .................................................. 52

5.3.3 Effects of Arg694 to Leu mutation on MSD conformational ensembles .................................................. 55

5.3.4 Effects of MSD conformation on membrane thickness ...... 58

5.4 Discussion ..................................................................... 60

5.5 Conclusions................................................................. 63

6. Model building HIV-1 gp41 ectodomain pre-fusion intermediate ............... 65

6.1 Introduction ..................................................................... 65

6.2 Simulation methods and models ....................................... 67

6.2.1 General methods....................................................... 67
List of Tables

3.1 Membrane structural properties, surface area per lipid and membrane thickness, obtained from 3 ns MD runs. *Surface area per lipid compared with experimentally observed values [1] ........................................ 27

3.2 Free energy of insertion of nanotube into bilayers with and without cholesterol. % change in surface area per lipid in term of number of lipid molecules deleted by the nanotube ........................................ 29
## List of Figures

1.1 A schematic cross-section of a lipid membrane with 10% cholesterol. The lipid head groups are shown in van der Waals representation. Cholesterol is shown in cyan and POPC in tan. Water is shown as points in red on both the exoplasmic and endoplasmic sides of the membrane ................................................................. 3

1.2 Representation of lipid molecules. 1-palmitoyl-2-oleoyl-phophatidylcholine (POPC) with monounsaturated sn-2 chain, 1,2-dipalmitoylphosphatidylcholine (DPPC) and cholesterol ................................................................. 5

3.1 Representative system of lipid membrane, carbon nanotube and water system prior to nanotube insertion. ................................................................. 25

3.2 Order parameter profiles for the sn-1 (left) and sn-2 (right) acyl chains of POPC with varying cholesterol content. Pure POPC (black), POPC:10%Chol(red), POPC:30%Chol(green),POPC:50%Chol(blue) ................................................................. 34

3.3 Force-distance profiles of pure POPC (blue) and POPC:Cholesterol bilayer with 30% cholesterol (red) obtained at a nanotube velocity of $v = 0.01 \, \text{Å/ps}$ ................. 35

3.4 Force-distance profiles of pure POPC (blue) and POPC:Cholesterol bilayer with 30% cholesterol (red) obtained at a nanotube velocity of $v = 0.01 \, \text{Å}$ ......................... 35

4.1 Schematic of HIV-1 gp41 membrane spanning domain. The basic residues are shown in blue, polar residues in green and non-polar residues in silver ................................. 37

4.2 RMSD$_\alpha$ of the MSD in pure water with respect to the initial $\alpha$-helical conformation vs. simulation time during 20 ns of MD ................................................................. 39

4.3 CV profiles of two 120 ns metadynamics runs with gp41 MSD in pure water starting from a complete $\alpha$-helical conformation (red) and a kinked conformation (green) ...... 41

4.4 Free energy profiles from two metadynamics runs with gp41 MSD solvated in pure water starting from an $\alpha$-helical conformation (red) and a kinked conformation (green). Average free energy from both the runs is shown in blue. ................................................................. 41

4.5 Representations of stable kinked conformation (left) and metastable loop conformation of MSD (right) of gp41 MSD in pure water observed in 20 ns MD simulation ......... 42
5.1 Membrane-peptide models of MSD α-helix with Arg_{694} snorkeling towards exoplasmic head groups (A) and endoplasmic head groups (B). (C) Structural conservation between the two MSD models and the exoplasmic (red) and endoplasmic (blue) snorkeling positions of Arg_{694} after 20 ns of equilibration. The residues with the highest structural conservation are shown in blue and the lowest are shown in red. The Leu mutant was prepared using the endoplasmic snorkeling model. ...................... 47

5.2 CV trace during metadynamics calculations for MSD in lipid bilayer with Arg_{694} snorkeling towards the exoplasmic leaflet (red), the endoplasmic (green) leaflet and the Leu mutant as a function of the simulation time .................. 51

5.3 Free energy profiles of MSD in a lipid membrane with Arg_{694} snorkeling towards exoplasmic head groups sampled during 260 ns metadynamics (red), Arg_{694} snorkeling towards endoplasmic head-groups sampled during 140 ns metadynamics run (blue) and Leu_{694} mutant (cyan). Schematic illustration of MSD with basic charged residues in blue and polar residues in green. Hydrophobic residues are shown in silver and Arg_{694} side-chain is shown in exoplasmic (red) and endoplasmic (blue) snorkeling positions. ... 52

5.4 Representation of stable and metastable conformations of MSD in a lipid bilayer with Arg_{694} snorkeling towards the exoplasmic, endoplasmic head groups and the Leu mutant. ...................................................... 57

5.5 (Upper panels) Map of mean membrane thickness (measured as a distance between centers of mass of lipid head groups in 10 Å square sections of each leaflet); color bar units in Å. Left, stable; right, metastable. (Lower panels) Side-view snapshots of stable (left) and metastable (right) conformations. .................................................. 58

5.6 (Upper panels) Map of mean membrane thickness (measured as a distance between centers of mass of lipid head groups in 10 Å square sections of each leaflet); color bar units in Å. Left, stable; right, metastable. (Lower panels) Side-view snapshots of stable (left) and metastable (right) conformations. .................................................. 59

5.7 (Upper panels) Map of mean membrane thickness (measure as a distance between centers of mass of lipid head groups in 10 Å square sections of each leaflet); color bar units in Å. Left, stable; right, metastable. (Lower panels) Side-view snapshots of stable (left) and metastable (right) conformations. .................................................. 61

6.1 A schematic cross-section of the model build six-helix bundle. The NHR(blue) and CHR(red) structure were obtained from available crystallographic structural information (PDB). The sequence of the linker(grey) was used to model build the complete structure .................................................. 75
6.2 A schematic cross-sections of the 6HB representing the hydrophobic groove (white) and the deep hydrophobic cavity in the NHR trimeric core. Trp\textsuperscript{628} can be seen to interact with residues in the hydrophobic cavity. .......................... 76

6.3 A schematic cross-section of the model build pre-fusion intermediate. The PFI is model built using the 6HB structure. ................................................................. 77

6.4 The RMSD of the 6HB (blue), RMSD of the NHR/CHR domains (green) and the RMSD of the linker (red) during 10 ns of MD run. The NHR/CHR domains can be observed to be extremely stable. ........................................ 78

6.5 Target RMSD (red) and instantaneous RMSD (green) of the TMD atoms during the course of 10 ns TMD simulations. The stiff spring constant of 200 kcal mol\textsuperscript{-1}Å\textsuperscript{-2} resulted in the instantaneous RMSD closely following the target RMSD. ............... 78
Cellular and viral membrane fusion is a critical process during viral infection. The HIV-1 surface envelope glycoprotein (gp41) anchored in the cholesterol-rich host cell derived viral membrane is responsible for mediating membrane fusion between the viral and target-cell membranes. The objective of this thesis is to understand the role of cholesterol and gp41, specifically the membrane spanning domain (MSD) and the two ectodomain heptad repeats (N-terminal heptad repeat and C-terminal heptad repeat), during membrane fusion using all-atom molecular modeling.

The 27-residue membrane-spanning domain (MSD) of the HIV-1 glycoprotein gp41 bears conserved sequence elements crucial to the biological function of the virus, in particular a conserved mid-span arginine Arg$^{694}$. We found that the MSD assumes a stable α-helical conformation and metastable kinked conformations that differ depending upon whether Arg$^{694}$ side-chain snorkels towards the exoplasmic or endoplasmic side of the viral membrane. It was observed that when the Arg$^{694}$ side-chain snorkels towards the endoplasmic headgroups the MSD assumes a kinked conformation that was hypothesized from experimental studies. Moreover, we found significant local thinning of the membrane and penetration of water molecules into the membrane core in the presence of mid-span Arg$^{694}$. The presence of cholesterol in the lipid membrane was found to assist in better anchoring of a non-polar structure in the membrane while increasing the free-energy of membrane destabilization. Mechanically destabilizing a cholesterol-rich lipid membrane also resulted in the formation of a larger rupture pore compared to a pure POPC lipid bilayer. Model built structures
of the gp41 ectodomain were then used to simulate folding of the gp41 pre-fusion intermediate into a stable six-helix bundle. These models provided important insights into the mechanism of folding and the specific interactions involved during folding.

Membrane fusion requires that the opposing membranes are brought into proximity and destabilized to facilitate lipid mixing. The membrane lipids and fusion proteins in both the membranes play important roles in facilitating membrane fusion. In this work we showed specific roles cholesterol and HIV-1 gp41 play in priming the viral membrane to promote fusion.
Chapter 1: Introduction

1.1 Background and Motivation

The acquired immunodeficiency syndrome (AIDS) pandemic caused by the human immunodeficiency virus type 1 (HIV-1) currently affects an estimated 33.4 million people worldwide and infects 2.5 million people each year. Twenty seven years after the first publication reported the pathogen and its role in the etiology of AIDS, a vaccine to prevent HIV-1 infection is still elusive [2–6]. HIV compromises the human immune system, specifically targeting the CD4+ T-cells, and quickly reverse-transcribes to form a latent reservoir of infected T cells resulting in a depletion of memory T-cells making it a much more formidable adversary compared to viruses such as influenza. The recent Thailand trails involving a combination of a recombinant canarypox vector vaccine (ALVAC-HIV [vCP1521]) and a recombinant glycoprotein 120 subunit booster vaccine provided promising results [7]. However, the sequence and strain diversity of HIV along with its ability to mutate and adapt presses upon the need to find new vaccine mechanism and targets [3, 5]. Considering the prevalence of AIDS in the low-income group countries a major challenge is to provide inexpensive vaccines. Therefore research is now being focused to develop an understating of HIV and targets to develop safe and efficient vaccines.

One of the most basic yet important targets are the surface envelope glycoproteins of HIV. Env is a trimeric complex heterodimers including receptor-binding glycoprotein 120 (gp120) and membrane-spanning glycoprotein 41 (gp41). The env complex
is responsible for viral tropism and fusion of the viral and target cell membranes leading to infection [8, 9]. Some early efforts included using monomeric gp120 to induce humoral immune responses. However, env gp120 immunogens failed to induce broadly neutralizing antibodies and production of cytotoxic T cells. Another promising strategy to hinder viral infection is to prevent the fusion of the viral and target cell membranes by inhibiting the conformational changes in gp41 that facilitate membrane fusion. The peptide T20, a 36 amino acid long peptide from the C-terminal heptad repeat of gp41, and the monoclonal antibodies 2F5 and 4E10, that bind to specific linear binding sites in the membrane proximal external region (MPER) belong to this category [10–15]. The ability of HIV-1 to mutate and the rapid folding of gp41 into a very stable six-helix bundle might however make these ineffective. Understanding the structure and function of env glycoproteins and their effects on viral and target cell membranes provide a first step towards developing safe and effective fusion inhibiting vaccines.

1.2 Lipid Membranes

Lipids are amphiphilic biomolecules that have diverse biological functions as energy storage components in the form of fats and oils, enzyme cofactors, signaling molecules, pigments and as biological membranes [16]. In aqueous media, lipids spontaneously form into macromolecular structures such as micelles and bilayers. Lipid bilayers are planar structures that are biologically important since they form the basis for cell plasma membranes that protect the cell from its environment and organelle membranes that compartmentalize cellular contents. Plasma membranes function as selective permeability barriers while also acting as a solvent for membrane proteins assisting in various functions such as cell signaling, ion transport and cell adhesion. The intricate interactions between the lipid molecules and the membrane proteins are
critical for their biological function. Therefore, the cell plasma membranes are complex structures containing a vast array of lipids that reflect their biological function.

Figure 1.1: A schematic cross-section of a lipid membrane with 10% cholesterol. The lipid head groups are shown in van der Waals representation. Cholesterol is shown in cyan and POPC in tan. Water is shown as points in red on both the exoplasmic and endoplasmic sides of the membrane.

The fluid mosaic model, by Singer and Nicholson in 1972, proposed that lipid membranes are complexes of proteins diffusing freely in a fluid lipid bilayer [17]. The amount of protein/lipid ratio and the lipid composition of plasma membranes vary widely not only between cell types but also between the inner and outer leaflets of the plasma membranes. However, the two predominant lipids in the plasma membranes
of eukaryotic cells are phospholipids and cholesterol. Carbohydrates are covalently attached to either the proteins or lipids and make up about 10% weight in plasma membranes [18]. Research into specific interactions between cholesterol and phospholipids in plasma membranes resulted in the discovery of phase separation in lipid bilayers [19]. Cholesterol rich lipid domains that formed a rigid liquid-ordered phase ($L_o$) were observed to coexist within a more fluid liquid-disordered phase ($L_d$). These cholesterol rich domains or “rafts” gathered immense research interest and were found to sequester a variety of membrane proteins and play an important role in cell signaling, viral entry, immune response and cell adhesion [20–27].

1.2.1 Phospholipids

Phospholipids are the most abundantly found lipid molecules in biological membranes [16]. The basic structure of most phospholipids consists of a glycerol backbone, a phosphorylated alcohol polar “head” group and two fatty acid chains as the non-polar “tail” groups. The carboxyl groups of the two fatty acids chains are esterified to the C-1 and C-2 hydroxyl groups of glycerol and are called the sn-1 and sn-2 chains respectively (Fig. 1.2). The head group is esterified to the C-3 hydroxyl group of glycerol backbone. The most common alcohol moieties present in the mammalian plasma membrane lipid head groups are choline (PC) and ethanolamine (PE) followed by serine (PS) and inositol (PI) to a much lesser extent [18]. The fatty acid chains are typically 14 to 22 carbon atoms long and for the hydrophobic core of lipid bilayers. These acyl chains can be derived from saturated fatty acids as in palmitic acid (C16), or unsaturated fatty acids like oleic acid (C18-1). The unsaturated fatty acids chains are usually found in the sn-2 position. The length of the acyl chains and the degree of unsaturation has a profound effect on membrane structure and fluidity.
1.2.2 Cholesterol

Cholesterol is a major component of eukaryotic cell plasma membranes and is thought to be important for cell growth [18]. The structure of cholesterol consists of a small polar 3\(\beta\)-hydroxyl group followed by a very hydrophobic planar fused ring structure and a hydrocarbon side chain(Fig. 1.2). The 3\(\beta\)-hydroxyl group allows cholesterol to position itself parallel to the acyl chains of the lipid molecules. The planar ring structure of cholesterol then increases the motional order of the lipid tails above the phase transition temperature of the lipid molecules resulting in a densely packed lipid molecules and a rigid membrane structure as observed in a liquid-ordered (Lo) state [28]. The rigid fused ring structure enables cholesterol to interact more favorably with lipids containing straighter saturated acyl chains than the kinked unsaturated tail groups. Below the lipid phase transition temperature, i.e. with the lipid membrane in the gel phase, cholesterol increases the free space between lipid molecules thereby increasing hydrocarbon chain disorder and making the membrane
more fluid [29–32]. The ability of cholesterol to manipulate membrane fluidity and properties likely make it important for membrane protein function and explains the highly energy intensive synthesis and specificity of the sterol in eukaryotic cells [18]. Moreover, cholesterol-rich domains in plasma membranes are known to localize certain proteins. These proteins have specific structural domains that interact with cholesterol. Examples of these structural domains are the cholesterol recognition amino acid consensus (CRAC) domain and the sterol-sensing domain (SSD) [33].

1.3 Membrane fusion

Membrane fusion is important for various processes in nature including endocytosis, neurotransmission, organelle formation and viral infection and typically consists of a two-step process [34, 35]. In step one, the opposing membranes are brought into close proximity in the aqueous environment against the counteracting electrostatic forces. In the second step the membranes are destabilized to facilitate lipid mixing and fusion [35]. A variety of proteins and lipids are involved to control and coordinate the fusion process. The most studied of these fusion proteins are the viral envelope proteins and the SNARE proteins that mediate intracellular fusion events [34, 35].

1.3.1 Fusion proteins

Viral fusion proteins can be divided into two major classes based on their structural features [36]. Class I proteins consist of paramyxov-, filo-, retro- and orthomyxoviruses and include HIV-1 (Env) and influenza (HA) fusion proteins [34]. Class I proteins typically have a fusion-inactive precursor that is proteolytically cleaved into a fusion active receptor-binding protein and a membrane anchored protein with an amino-terminal fusion peptide. These proteins assume a characteristic stable triple stranded post-fusion coiled-coil structure. Class II viral proteins are not proteolyt-
ically cleaved and have an internal fusion peptide. The flavi- and alphaviruses are thought to have class II fusion peptides. The structure and function of Class I fusion peptides, especially HA, are the most studied and excellent review articles on both Class I and II proteins are available [36]. A detailed description of the HIV-1 Env is provided in section 1.4.

SNARE protein complexes mediate the synaptic vesicle fusion with the plasma membrane and have been proposed to be essential for hundreds of intracellular membrane fusion events [37–39]. The SNARE complex involves the formation of a stable four-helix bundle where the vesicle associated synaptobrevin contributes one helix and the target-membrane associated SNAP-25 and syntaxin contribute three helices. A common structural feature between the SNARE complexes and class I viral proteins are the conserved heptad repeats that are responsible for the formation of stable helical bundles while pulling the opposing membrane to close proximity [38].

1.3.2 Fusion mechanism

It was postulated that the fusion process involves at least two steps [18, 34, 35]. In the first step the proximal layers of the opposing membranes merge to form a “stalk”. In the second step the distal monolayers merge to form a hemi-fusion intermediate the eventually grows into a fusion pore. The merging of the bilayer requires that the membranes are destabilized to facilitate interaction between the hydrophobic tail groups of the lipids. However, there are still large gaps in understanding the mechanism of membrane fusion and the specific way the fusion proteins destabilize the membranes to catalyze the fusion process [18].
1.4 The human immunodeficiency virus type 1 (HIV-1)

HIV-1 is a retrovirus and has an RNA genome that is reverse transcribed and integrated into the host DNA during infection. The life cycle of HIV-1 starts as a mature virion approaches a target cell for infection [40]. Viral entry is mediated by the surface envelope glycoproteins (Env). The envelope glycoprotein gp160 is post-transitionally cleaved into two non-covalently linked subunits, gp120 and gp41. Viral tropism is directed by gp120 that binds to the cell receptor CD4 and co-receptors CCR5/CXCR4. Gp120 receptor binding induces rapid but poorly understood conformational changes in the Env gp41 transmembrane protein which initiates fusion [8, 9, 41, 42]. These conformational changes result in the exposure followed by insertion of a hydrophobic fusion peptide (FP) of gp41 into the target cell membrane. Following FP insertion another cascade of conformational changes results into the gp41 folding into a very stable six-helix bundle thereby bringing the viral and target-cell membranes within close proximity inducing membrane fusion and the delivery of the viral capsid into the target cell. The viral capsid consists of the enzymes reverse transcriptase and integrase, the viral RNA and the viral protein Vpr that assists in transporting the capsid contents into the host cell nucleus.

In the nucleus, the RNA is reverse transcribed by reverse transcriptase and is incorporated in the host-cell chromosome. The synthesized RNA is transported back into the cytoplasm where the viral core is assembled. Gp160 is cleaved into gp120 and gp41 and will be in a complex with CD4 in the endoplasmic reticulum. The viral protein Vpu degrades CD4 and enables Env transport to the plasma membrane in the form of trimeric spikes of gp120/gp41 dimers. The env spikes are assembled into the plasma membrane and the virion buds using the host-cell derived membrane. During this process HIV-1 is believed to assemble the env spikes and bud from cholesterol rich domains present in the host-cell membrane. Therefore, the HIV-1 viral lipid coat
is rich in saturated lipids and cholesterol. After the virion is released from the host-cell membrane it matures into an approximately 100 nm diameter virus and ready to repeat the life cycle.

1.4.1 Glycoprotein-41

Gp41 is often described structurally in terms of five major regions: 1. a glycine rich fusion peptide (FP) that is exposed upon the conformational changes following gp120 binding to CD4 and CCR5/CXCR4 and inserts into the target cell membrane, 2. a core consisting of the N- and C-terminal heptad repeats (NHR and CHR), that following FP insertion form a six-helix bundle bringing the target membrane and viral membrane together, 3. a membrane proximal external region (MPER) following the CHR, that contains a CRAC segment that is thought to be responsible for sequestering the peptide in cholesterol-rich domains, 4. a membrane spanning domain (MSD), that is thought to anchor gp41 to the viral membrane and 5. a long cytoplasmic tail that extends into the endoplasmic side of the virion [43–45]. Although their triggering mechanisms can be different, the ectodomain core complexes of various viral membrane fusion proteins have remarkably similar structures [8, 46]. The role of the fusion peptide of viral fusion proteins, including influenza(HA) and HIV-1(Env), in initiating viral infection by inserting itself into the target cell membrane and the formation of the six-helix bundle to bring the target cell membrane and the viral membrane closer to each other has been well studied [47, 48]. Since, proximity of the membranes is not sufficient for membrane fusion there should be a mechanism for the virus to destabilize its own membrane. Therefore recent experimental studies were focused on understanding the conformation of the MSD and its role in viral membrane destabilization [49–51]. Although molecular simulations were proved to be an invaluable tool in studying and visualizing protein structure and dynamics, the
absence of crystallographic or nuclear magnetic resonance (NMR) structure of the HIV-1 MSD hindered molecular simulation studies of HIV-1 MSD [52].

The complete mechanism of HIV-1 infection, the membrane fusion process and the function of various viral proteins during infection are provided in some excellent reviews [8, 40, 53, 54]. The main goal of this work is to explore a structure-function relationship of gp41, specifically the membrane spanning domain (MSD) that anchors the env to the viral membrane and its effect on the viral membrane using molecular simulations.

1.5 Objectives and specific aims

To achieve the objective of understanding gp41 MSD effects on viral membrane, we started by studying the effects of cholesterol on the forces and free-energies required to rupture lipid bilayer.

1.5.1 Specific Aim 1: Compute the forces and free-energies of membrane rupture

An important component of molecular simulations is to validate the results generated by the simulations by comparing reliable observable properties with experimental data. In Chapter 3 we validate the results by comparing the surface area per lipid obtained from our simulations of pure POPC and POPC bilayers with 10%, 30% and 50% cholesterol. We also study the effect of cholesterol content on membrane properties, specifically the membrane thickness and the lipid order parameters. We then simulated mechanical rupture of pure POPC and POPC bilayers containing 30% cholesterol using a pristine carbon nanotube. The dimension of the nanotubes were chosen to replicate a viral fusion peptide and were approximately 20 Å in length and 12 Å in diameter. We used these simulations to compute the forces and free-energies
required to penetrate a lipid bilayer with a predominantly non-polar structure and the effects of cholesterol on these force and free-energies.

1.5.2 Specific Aim 2: Structure and orientation of HIV-1 gp41 MSD and effects on lipid membrane structure

The HIV-1 gp41 MSD likely plays an important role in the destabilization of the viral membrane during infection. In Chapters 4 and 5 we study the structure of gp41 MSD in water and a cholesterol-containing lipid bilayer respectively. The MSD has a mid-span Arg\textsuperscript{694} that is important for its function. Specifically, we study the importance of the mid-span Arg\textsuperscript{694} by testing (i) the effect of Arg\textsuperscript{694} snorkeling direction on the MSD structure in a membrane, and (ii) the Arg\textsuperscript{694}→Leu point mutation to replace Arg\textsuperscript{694} which abrogates infectivity.

1.5.3 Specific Aim 3: Folding of gp41 ectodomain from a pre-fusion intermediate (PFI) to a six-helix bundle (6HB)

Folding of the gp41 ectodomain from a PFI conformation to the stable 6HB results in bringing the viral and target-cell membrane into close proximity leading to membrane fusion. These series of conformational changes likely assist in viral membrane destabilization of the MSD. We simulate the folding of the ectodomain into a 6HB from a pre-fusion intermediate to provide a first step towards understanding the interactions involved during the process. In Chapter 6 we provide the simulation details of the folding of a PFI into a 6HB using TMD.
Chapter 2: Simulation Methods

2.1 Introduction

Molecular simulations are the tools for study and analysis of molecular structure and function. Molecular simulations are now being used to study various biological systems, most importantly proteins, lipids and nucleic acids [52, 55, 56]. These simulations can be especially powerful to explore, observe and understand experimentally inaccessible phenomena in a cheap, safe and efficient way [57].

Molecular simulations need models with atomic level detail, the interactions of which are then computed to study their macroscopic properties. These models are initial configurations of atoms in space that are typically obtained from experimental data such as nuclear magnetic resonance (NMR) or X-ray crystallography. In the absence of this data, a thorough knowledge of the system can used to model build structures [55]. The interactions between these atoms are then described using an empirical potential energy function. Although quantum mechanics can be used to describe these interactions, for many atomic systems quantum mechanical calculations are computationally expensive. Therefore classical mechanics, in which the energy of the system is calculated only based on nuclear positions ignoring electron motion, is best suitable for large biological macromolecules. Force fields have been developed with the help of experiments and quantum mechanical calculations [58, 59]. Using these initial structures and force fields, molecular simulations generate information at the atomic level in the form of a series of snapshots of the system.
2.1.1 Statistical mechanics and ensembles

Experimentally observable macroscopic properties are calculated from the microscopic information using statistical mechanics. Each particular snapshot with unique internal coordinates and momenta of particles can be described as a microscopic state. In a system with \( N \) particles, where \( N \) is large, there can be a large number of microscopic states that collectively conform to a macroscopic state characterized by a handful of thermodynamic variables. A collection of microscopic states that have an identical macroscopic state is called an ensemble. Different kinds of statistical ensembles exist that are characterized by fixed values of thermodynamic values such as the total number of particles \( N \), volume \( V \), temperature \( T \), total energy \( E \), or chemical potential \( \mu \). The canonical ensemble is characterized by constant \( NVT \), the microcanonical ensemble by constant \( NVE \), the grand canonical ensemble by \( \mu VT \) and the isobaric-isothermal ensemble by \( NPT \) [58–60].

The value of an observable property of a system depends on the positions and momenta of the \( N \) particles of the system. Therefore, the instantaneous values of a property \( A \) can be described as \( A(p(t), r(t)) \) where \( r \) and \( p \) are the positions and momenta of the \( N \) particles respectively at time \( t \). The instantaneous value of \( A \) fluctuates in time because of the chaotic motion of the particles. To obtain an experimentally observable value we must average \( A \) over time. As the time over which the measurement is made increases to infinity the time average value reaches the “true” average of the property as given by:

\[
\bar{A} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau A(r(t), p(t)) \, dt,
\]

(2.1)

Since, most macroscopic systems are large (in the order of \( 10^{23} \) particles), it is impractical to simulate the dynamic behavior of the particles as expected by the time averaging. In such cases, the time average can be replaced by an ensemble average.
given by:

\[
\langle A \rangle_{\text{ensemble}} = \int \int d\mathbf{p} d\mathbf{r} \ A(\mathbf{p}, \mathbf{r}) P(\mathbf{p}, \mathbf{r}),
\]  

(2.2)

where the \( \langle \rangle \) brackets indicate an ensemble average and \( P \) is the probability density of the ensemble, defined as the probability of observing a microstate \((\mathbf{p}, \mathbf{r})\). The ensemble average of a property \( A \) can therefore be determined by averaging over all possible configuration of the system and its use is facilitated by the ergodic hypothesis, which states that the ensemble average is equal to the time average.

In the canonical ensemble the probability density is given by

\[
P(\mathbf{p}, \mathbf{r}) = \frac{1}{Z} \exp \left( -\frac{H(\mathbf{p}, \mathbf{r})}{k_B T} \right),
\]  

(2.3)

where \( H \) is the Hamiltonian of the system, defined as \( H = K + U \), where \( K \) is the kinetic energy of the system and \( U \) is the potential energy of the system. \( k_B \) is the Boltzmann constant, \( T \) is the temperature of the system and \( Z \) is called the partition function given by

\[
Z = \int \int d\mathbf{p} d\mathbf{r} \ \exp \left( -\frac{H(\mathbf{p}, \mathbf{r})}{k_B T} \right),
\]  

(2.4)

If we know the partition function of the system, we can derive all other thermodynamics properties of the system from it. For example:

\[
U = k_B T^2 \left( \frac{\partial \ln Z}{\partial T} \right)_{V,N},
\]  

(2.5)

\[
S = k_B \ln Z + \frac{U}{T},
\]  

(2.6)

\[
F = -k_B T \ln Z.
\]  

(2.7)
There are two major approaches in molecular simulations to derive these properties: 1) Molecular dynamics (MD), which is a deterministic approach and studies the dynamics by simulating the time evolution of the system and 2) Monte Carlo (MC) which is a stochastic approach. Molecular dynamics can therefore be used to extract both the kinetic and thermodynamics properties of the system while MC can be used to calculate thermodynamic properties of the system. Some excellent texts and reviews can be found on both these methods and the statistical mechanics [55, 56, 58–60]. However, in this work only MD was used and therefore a brief introduction of MD is provided in Section 2.2. This is followed by the force fields used in this work, the limitations of MD and other simulation techniques used in this thesis.

2.2 Molecular Dynamics

Molecular dynamics simulates the time dependent dynamics of a system of interacting particles and is a major simulation technique used to study biological macromolecules. MD can be used to extract kinetic and thermodynamic information, to refine macromolecular structures based on NMR and X-ray crystallographic data, and to study the free-energy changes and requirements in biological systems. To access these properties using MD simulation requires knowledge of the empirical interatomic potential that describes the interactions between various atoms. The forces acting on these atoms are gradients of this potential, and can be used in the integration of Newtonian equations of motion to predict the evolution of particle positions and momenta.

2.2.1 Langevin Thermostat

Since the total energy $E$ is not conserved in an NVT ensemble a thermostat is needed to constantly add to remove energy from the system. In the Langevin
thermostat each particle is coupled with a heat bath and the fundamental equation of motion $F = ma$ is modified into the Langevin equation

$$m_i \ddot{r}_i = -\left( \frac{\partial U}{\partial r_i} \right) - \gamma m_i \dot{r}_i + \eta_i(t), \quad (2.8)$$

where $m$ is the mass, beta, $k_B$ is the Boltzmann’s constant and $T$ is the temperature. $U$ is the interatomic potential, $\gamma$ is the friction constant, $\eta$ is the random force or the Gaussian white noise, which obeys a fluctuation-dissipation theorem

$$\langle \eta_i(t)\eta_j(t') \rangle = \delta_{ij} \delta(t - t')6k_B T \gamma m_i \quad (2.9)$$

### 2.2.2 Numerical Integration

During an MD run the Langevin equation is numerically integrated to propagates the system in time. One the most widely used and efficient integrating schemes is the velocity-Verlet algorithm. The velocity-Verlet algorithm is accurate, fast and time reversible [58, 61]. This algorithm updates the coordinates and velocities of the particles at the next time step ($r_{(n+1)}$) from the current one ($r_n$) using

$$v_{n+1/2} = v_n + M^{-1}F_n \Delta t/2, \quad (2.10)$$

$$r_{n+1} = r_n + v_{n+1/2} \Delta t, \quad (2.11)$$

Compute, $F_{n+1} = F(r_{n+1}),$

$$v_{n+1} = v_{n+1/2} + M^{-1}F_{n+1} \Delta t/2. \quad (2.12)$$

where $\Delta t$ is the time step and $M$ is the mass matrix.

To start an MD simulation, apart from assigning the atomic coordinates of the system, the initial velocities should also assigned. The velocities are randomly assigned...
for each particles from a Maxwell-Boltzmann distribution for the desired temperature:

\[
P(v) = \left( \frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[ -\frac{1}{2} \frac{m_i v^2}{k_B T} \right].
\] (2.13)

The Maxwell-Boltzmann equation gives the probability that atom \(i\) of mass \(m_i\) has a velocity \(v\) at a temperature \(T\). The instantaneous value of the temperature is related to the kinetic energy via the particles’ momenta \(p_i\) by:

\[
H = \sum_{i=1}^{N} \frac{|p_i|^2}{2m_i} = \frac{k_B T}{2} (3N - 6).
\] (2.14)

where \(H\) is the Hamiltonian of the system, and \(3N-6\) is the total number of degrees of freedom for a 3 dimensional system. The temperature control for MD simulations in this work was achieved via Langevin thermostat (Eq 2.8). Other techniques for temperature control include Andersen thermostat [62], and Berendsen thermostat [63]. The pressure control in all MD simulations presented in this work was achieved by the Nosé-Hoover barostat [64–68].

2.2.3 Force Field

Molecular dynamics simulations start with the knowledge of empirical potential functions or force fields that describe the interactions between atoms. These potential functions contain information about the system components such as the bond lengths, bond angles, torsion angles and the non-bonded interactions that treat the Van der Waals and electrostatic interactions. The empirical potential used in this thesis is the Chemistry at HARvard Macromolecular Mechanics (CHARMM) potential and has the following contributions [69–71]:

\[
U_{\text{total}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}} + U_{\text{vdW}} + U_{\text{Coulomb}} + U_{\text{Urey–Bradley}}.
\] (2.15)
Where the Urey-Bradley term accounts for the angle bending using 1,3 non-bonded interactions. Each of these interactions are given by

\[
U_{\text{bond}} = \sum_{\text{bonds}} K_b (b - b_0)^2, \tag{2.16}
\]

\[
U_{\text{angle}} = \sum_{\text{angle}} K_\theta (\theta - \theta_0)^2, \tag{2.17}
\]

\[
U_{\text{Urey-Bradley}} = \sum_{\text{Urey-Bradley}} K_U (u - u_0)^2, \tag{2.18}
\]

\[
U_{\text{dihedral}} = \sum_{\text{dihedral}} K_\chi (1 + \cos(n\chi - \delta)), \tag{2.19}
\]

\[
U_{\text{impropers}} = \sum_{\text{impropers}} K_\varphi (\varphi - \varphi_0)^2, \tag{2.20}
\]

\[
U_{\text{vdW}} = \sum_{i} \sum_{j > i} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{2.21}
\]

\[
U_{\text{Coulomb}} = \sum_{i} \sum_{j > i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \tag{2.22}
\]

### 2.2.4 Limitations of molecular dynamics

The quality of MD simulation most importantly depend on the accuracy of the interatomic potential and the extent of phase space sampled. The interatomic potential can be modified to obtain accurate results. However, in general the first step is probably to validate the accuracy of the simulation by comparing reliable properties or behavior of the system either with experimental, theoretical or previous simulation values. For membrane simulations the surface area per lipid is one such reliable property. In this work, Chapter 3 starts with validation of the simulation results by comparing surface area per lipid values with previous experimentally obtained values.

The extent of phase space sampling depends on the length of the simulation. The two major factors that affect simulation length are the size of simulation system and the availability of the computational resources i.e. the physiological time. Although
computers are becoming cheaper and faster there is a need for accelerated MD techniques to efficiently sample the phase space, especially for large systems. Some of these accelerated methods, steered molecular dynamics (SMD), metadynamics and targeted molecular dynamics (TMD) are used in this work.

2.3 Steered molecular dynamics

In steered molecular dynamics the SMD atom or the center-of-mass of the atoms to be steered is attached to a dummy atom via a virtual spring. The dummy atom is then moved at a constant velocity (cv-SMD) or by applying a constant force (cf-SMD). SMD can therefore be used to apply external steering forces to drive the reaction coordinate along a predetermined or hypothesized reaction path. The force between the SMD atom and the dummy atom is the negative gradient of the potential energy given by

\[
U = \frac{1}{2} k [\nu t - (\vec{r} - \vec{r}_0) \cdot \vec{n}]^2
\]  

(2.23)

The potential of mean force (PMF) is the free energy profile along the reaction coordinate and is important to determine conformational equilibrium properties of proteins or transition rate of dynamical processes. The PMF is an equilibrium property while steering in an SMD simulation is a non-equilibrium process. Therefore the Jarzynski’s equality, which is an exact relation between the free energy differences and the work done through non-equilibrium process is often used.

\[
e^{-\beta \delta G} = \langle e^{-\beta W} \rangle
\]

(2.24)

Where, \( \beta = \frac{1}{k_B T} \), \( G \) is the Gibbs free energy, \( T \) is the temperature and \( W \) is the total
work done to pull the SMD atom. The work $W$ can be calculated by integrating the total force over the SMD trajectory obtained using a sufficiently stiff spring. Potential of mean force (PMF) calculations were made using the Jarzynski’s equation that was generalized for isobaric-isothermal systems. These trajectories can be repeated to reduce the statistical error of the calculated PMF. In this work we used the second order cumulant expansion of the Jarzynski’s equality to calculate the PMF. The second order cumulant is given by:

$$\Delta F = \langle W \rangle - \frac{1}{2}\beta \left( \langle W^2 \rangle - \langle W \rangle^2 \right)$$  \hspace{1cm} (2.25)

where $W$ is the work performed, $\beta = 1/k_B T$, $k_B$ is Boltzmann’s constant, $T$ is the bulk temperature, and $\Delta F$ is the free energy difference.

SMD is used in the Chapter 3 to steer a carbon nanotube through a lipid bilayer to simulate lipid mechanical destabilization of lipid membranes. In Chapter 5, SMD is used to steer a membrane spanning domain of HIV-1 gp41 into lipid bilayers to model membrane spanning domain-lipid bilayer systems. SMD is also used in Chapter 6 to model the pre-fusion intermediate of the gp41 ectodomain from a model of a six-helix bundle, where the C-terminal heptad repeats of the ectodomain were steered to create an extended ectodomain.

### 2.4 Metadynamics

In metadynamics, system forces in the MD simulation are augmented by those arising from a history-dependent potential, which induces uniform sampling (i.e, overcomes free energy barriers) along a specified collective variable (CV) \[72–77\].
Following Laio et. al., we used the following external potential:

\[ V_{\text{meta}}(\xi) = \sum_{t'<t} W \prod_{i=1}^{N_{\text{c}}} \exp \left\{ -\frac{[\xi_i - \xi_i(t')]^2}{2\delta^2_{\xi}} \right\} \]  

(2.26)

where \( \xi_i \) is the value of the CV, \( \xi_i(t') \) is the value of \( \xi_i \) at time \( t' \) from the atomic coordinates, \( W \) is the Gaussian height, \( \delta_{\xi} \) is the Gaussian width and \( t' \) are the number of MD steps between every Gaussian deposited. The metadynamics potential, when converged, also provides a good estimate of the inverse free energy as a function of the CV. Metadynamics in various forms has recently been applied for efficient conformational sampling and developing free energy profiles in various protein systems [77–82]. Metadynamics is used in Chapters 4 and 5 to study conformations of the membrane spanning domain of HIV-1 gp41 in water and lipid environments respectively.

### 2.5 Targeted molecular dynamics (TMD)

In targeted molecular dynamics, an initial conformation of a system can be driven to a target conformation through the application of a time-dependent restraining force [83]. The method requires identification of TMD atoms in the system. At each timestep, a steering force is applied on each atom by the potential

\[ U_{\text{TMD}} = \frac{1}{2N} k [\text{RMSD}(t) - \text{RMSD}^*(t)]^2 \]  

(2.27)

Where, \( \text{RMSD}(t) \) is the instantaneous best fit RMSD of the current structure from the target structure and \( \text{RMSD}^*(t) \) evolves linearly from the initial RMSD to the target RMSD. Also, \( k \) is the spring constant and \( N \) is the number of TMD atoms in the system.

TMD is used in Chapter 6 to drive a model built initial pre-fusion intermediate structure of the HIV-1 gp41 ectodomain to a target six-helix bundle.
2.6 Software and resources used

All trajectories in this work were performed using NA nonscale Molecular Dynamics (NAMD) versions (2.6 - 2.7b2) [84, 85]. Visual Molecular Dynamics (VMD) (versions 18.6 - 1.8.7) software was used to visualize and analyze the trajectories. VMD was also used to solvate membranes and proteins and to add counter ions to the system.

The Protein Data Bank (PDB) was used to obtain some of the protein structures used in this work. The PDB is available as http://www.pdb.org/pdb/home/home.do and is repository for several thousands of biomolecular structures.

The CHARMM-GUI membrane builder plugin was used to generate initial structures of the lipid bilayers used in this work. The obtained structures were then minimized and equilibrated using NAMD.
Chapter 3: Mechanical destabilization of a cholesterol containing lipid bilayer

3.1 Introduction

Cholesterol rich detergent resistant domains (DRMs) present in the plasma membrane of eukaryotic cells have been the focus of various research studies recently [20]. One of the important observations about these DRMs is that they likely act as platforms for viral entry [86–89]. Enveloped viruses are also thought to bud from the cholesterol rich domains present in the host-cell plasma membrane. Viral membranes, including that of HIV-1, were observed to contain a higher cholesterol/phospholipid ratio than that present in the plasma membrane of the host-cell. Although the exact mechanism of viral entry is still unclear, it is understood that destabilization of both the viral and target-cell membranes is necessary for successful infection. After binding to the receptor CD4 and the coreceptors CCR5/CXCR4 the hydrophobic glycine rich fusion peptide of HIV-1 gp41 is exposed and is inserted into the target cell membranes initiating membrane destabilization while anchoring gp41 in the target-cell membrane. Therefore, studying the forces and free energies of rupture of lipid membranes and the effect of cholesterol on them provides an understanding towards their role in viral entry and as a source for viral envelope.

The purpose of this chapter is (i) to compare and validate the various membrane properties calculated from molecular dynamics with those observed in experiments and to study the effects of cholesterol content on those properties, and (ii) to calculate
and compare the forces and free-energies of mechanical rupture, by inserting a highly non-polar pristine carbon nanotube, of pure phospholipid bilayer with that of cholesterol/phospholipid complexes. We calculated the surface area per lipid, membrane thickness and membrane fluidity as given by the lipid order parameter ($S_{cd}$), membrane properties that are sensitive towards simulation parameters and are generally thought to provide a reliable measure of the accuracy of lipid membrane simulations [90]. We observed a good agreement between the values calculated from our simulations and those obtained from experiments and the condensing and thickening effects of cholesterol on a lipid membrane. Calculating the forces and free-energies of insertion also suggested that the presence of cholesterol does not influence the force of insertion, however bilayers containing 30% cholesterol had a higher free-energy of rupture and likely assisted in membrane anchoring of the nanotube and membrane destabilization.

3.2 Simulation methods and models

The simulation system consisted of a lipid bilayer, water and a carbon nanotube. Lipid bilayers consisting of pure 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) and with 10%, 30% and 50% cholesterol were prepared using the CHARMM-GUI membrane builder. POPC was used because of it ubiquitous presence in host-cell membranes [18, 91]. The bilayers consisted of a total of 200 lipid molecules and were fully hydrated with approximately 39 water molecules per lipid. All simulations were performed using NAMD version 2.6 and visualized using VMD v1.8.6. The CHARMM force field and explicit TIP3P water was used. The simulation systems were initially minimized for 10000 steps. Following minimization the systems were equilibrated in the NPT ensemble for at least 5 ns. Surface area per lipid, membrane thickness and the lipid order parameters were calculated for the last 3 ns of the simulation.
3.2.1 Order parameter

Lipid order parameters provide a measure of the fluidity of the lipid tails [18]. It is generally given by a second order Legendre polynomial function expressed as

\[ S_{CD} = \frac{1}{2} (3 \cos^2(\theta_i) - 1) \]  

(3.1)

where \( \theta_i \) is the angle between the membrane normal and the bond between \( C_i \) and \( C_{(i-1)} \). The order parameter gives values of \( S_{CD} = 1 \) when the tails are parallel to the membrane normal and \( S_{CD} = -0.5 \) when the tails are perpendicular to the membrane normal. However, \( S_{CD} \) can be 0 both when there is a very fluid random motion of the lipid tails groups or when they are all aligned at an angle of 54.7°.
3.2.2 Membrane destabilization simulations

Carbon nanotube insertion simulations were performed only on pure POPC bilayer and bilayers with POPC and 30% cholesterol based on the composition of target-cell plasma membranes. The carbon nanotube was prepared to represent the dimensions of a viral fusion peptide and was 20 Å length and approximately 12 Å in diameter. Constant velocity steered molecular dynamics was used to steer the nanotube along the membrane normal. The center-of-mass of the nanotube was used as the SMD atom and the lateral degree of freedom of the nanotube were restricted using a force of 10 kcal/(mol*A) to avoid nanotube rotation or diffusion along the plane perpendicular to the membrane normal. A sufficiently stiff spring constant of 7 kcal/mol/A was used. The work done and the PMF were calculated as described in section 2.3.

3.3 Results

3.3.1 Membrane properties

Membrane structural properties for pure POPC and POPC/cholesterol bilayers are given in Table 3.1. The surface area per lipid was found to be 62.46±2.03 Å² for pure POPC. The area per lipid for pure POPC was found to be 63 Å² experimentally [1]. The area per lipid and the % area condensation values obtained at different cholesterol content also compare well with previously reported experimental values [1]. The surface area per lipid decreased with increasing cholesterol content likely due to the condensing effect of cholesterol. Conversely, the membrane thickness, calculated as the average distance between the phosphorous atoms of the exoplasmic and the endoplasmic lipid head groups, increases with increasing cholesterol content. Thickness of a pure POPC bilayer was found to be 38.07±0.87 Å. Addition of 10% cholesterol showed the maximum increase in membrane thickness.
Table 3.1: Membrane structural properties, surface area per lipid and membrane thickness, obtained from 3 ns MD runs. *Surface area per lipid compared with experimentally observed values [1]

<table>
<thead>
<tr>
<th>%Chol</th>
<th>Sim. area (Å²)</th>
<th>Expt. area (Å²)*</th>
<th>% Area. condensation</th>
<th>Thickness Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>62.46±2.03</td>
<td>63</td>
<td>0</td>
<td>38.07±0.87</td>
</tr>
<tr>
<td>10</td>
<td>58.22±2.65</td>
<td>58</td>
<td>6.78</td>
<td>40.04±1.34</td>
</tr>
<tr>
<td>30</td>
<td>51.33±1.90</td>
<td>48</td>
<td>11.83</td>
<td>40.70±1.35</td>
</tr>
<tr>
<td>50</td>
<td>48.35±1.94</td>
<td>44</td>
<td>5.81</td>
<td>41.20±1.19</td>
</tr>
</tbody>
</table>

The lipid order parameters (scd) for the sn-1 and sn-2 chains are given in Fig. 3.2. The sn-1 chain in general had a high order parameter in the upper chain region and had a higher disorder lower in the acyl chains due to availability of higher free space for motion. The addition of cholesterol increased the overall order of the sn-1 chain, although the increase is minimal between 30% and 50% cholesterol content. The ordering effect of cholesterol is higher between C5 and C14 the region where the fused rings of cholesterol are present. The sn-2 chain in POPC is monounsaturated and the double-bond is present between C9 and C10. The portion of the sn-2 chain close to the glycerol backbone is rigid and almost parallel to the membrane surface and therefore has a lower order parameter. Cholesterol had minimal effect on the sn-2 chain except at a high concentration of 50%. In general the order parameter values are in good agreement with experimental values in the lower portion of the acyl chains. The upper chain region of the sn-1 chain had a dip near C3 in the experimental values which was not reproduced in our simulations [92].

3.3.2 Membrane rupture force and effects of cholesterol

To study the effects of mechanical rupture of a pure POPC bilayer with a carbon nanotube we initially pulled the nanotube at three different velocities of \( v = 0.1 \)
Å/ps, 0.05 Å/ps and 0.01 Å/ps using constant velocity SMD and calculated the forces required to pull the nanotube using equation 2.23. Forces-Distance curves were plotted as a function of the distance traveled by the dummy atom(-z) as the nanotube traverses across the membrane. The magnitude and the profiles of the force-distance curves depended upon the CNT velocity (Fig. 3.3).

The highest insertion forces decreased from 3612.26 pN to 1428.12 pN between nanotube velocities of 0.1 Å/ps and 0.01 Å/ps respectively. Immediately after penetration of the lipid head groups the insertion force decrease as the nanotube enters into the hydrophobic core of the membrane. At the two higher velocities the force continued to decrease as the nanotube enters the endoplasmic leaflet of the membrane before exiting the membrane. However at the lowest velocity of 0.01 Å/ps the force remained constant at approximately 800 pN following penetration.

To test the effect of cholesterol on membrane rupture the insertion simulations were performed at the lowest velocity of v = 0.01 Å/ps on a bilayer with 30% cholesterol content. The highest force of insertion was observed to be 1376.74 pN and is lesser that that observed for a pure POPC bilayer. Comparing the two force-distance profiles at v=0.01 Å/ps for pure POPC and POPC with 30% cholesterol, the force of insertion decreased with increasing cholesterol content. However, as the nanotube exits the endoplasmic leaflet of the bilayer a second peak of 1268.44 pN was observed (Fig. 3.4). The nanotube was also observed to destabilize the bilayer by pulling the lipid molecules out of the endoplasmic leaflet of the membrane that increased in the presence of cholesterol. In a pure POPC bilayer the nanotube removed approximately 2 lipids when pulled out of the membrane, while in the presence of cholesterol it increased to approximately 11 lipids molecules. The change in surface area values presented in table 3.2 were computed as a function of the amount of free space on the bilayer surface due to the removed molecules.
Table 3.2: Free energy of insertion of nanotube into bilayers with and without cholesterol. % change in surface area per lipid in term of number of lipid molecules deleted by the nanotube

<table>
<thead>
<tr>
<th>%Lipids</th>
<th>CNT vel. (Å/ps)*</th>
<th>Free-energy (kcal/mol)</th>
<th>% change in surface area</th>
</tr>
</thead>
<tbody>
<tr>
<td>POPC</td>
<td>0.01</td>
<td>35.74±3.54</td>
<td>2.07</td>
</tr>
<tr>
<td>POPC: 30%Chol</td>
<td>0.01</td>
<td>50.02±4.73</td>
<td>13.50</td>
</tr>
</tbody>
</table>

3.3.3 Free energy of membrane rupture

Since the applicability of Jarzynski’s equality requires reversibility of the process, the free-energy calculations were performed only at the lowest velocity of \( v = 0.01 \text{Å/ps} \). Since the nanotube disrupts the membrane as it exits the endoplasmic leaflet it disrupts the membrane by pulling lipid molecules out of the membrane, we calculated the rupture free-energies as a function of the work required for the nanotube to be embedded in the membrane core. At the lowest velocity of 0.01 Å/ps the free-energy of insertion into a pure POPC bilayer was observed to be 35.74±3.54 kcal/mol. For the bilayer with 30% cholesterol the free-energy of insertion increased to 50.02±4.73 kcal/mol.

3.4 Discussion

The surface area per lipid is a membrane property that can be used to validate the accuracy of membrane simulations [90]. The surface area per lipid values for both pure POPC and cholesterol containing bilayers obtained from our simulations compare well with experimental values. The effect of cholesterol on the area condensation and the thickening of the lipid bilayer was obvious with the obtained area per lipid and membrane thickness values. The area per lipid decreased almost linearly
with the cholesterol concentration until 30% cholesterol. However, the membrane thickness had maximum increase at 10% cholesterol content and the percent change in thickness decreased with increasing cholesterol between 10% and 50% cholesterol content. The stiffening effect of cholesterol was also observed on the lipid order in the bilayers. The order of both the sn-1 and sn-2 chains increased with increasing cholesterol content. The interatomic potential could accurately reproduce the order parameters in the lower portion of the lipids tails. However, the CHARMM potential was slightly inaccurate in reproducing the experimentally observed order parameter in the upper portion of the lipid acyl chains and it was corrected in a recent version of the potential [93].

We further computed the forces and free-energies of insertion of a non-polar carbon nanotube into a pure POPC and POPC with 30% cholesterol bilayers. To penetrate the lipid bilayer it should create hydrophobic defect in the bilayer thereby exposing the hydrophobic tail groups to water. The energy required to porate the membrane depends on the interplay between the curvature and hydrophobic defects of the membrane [53]. Previous coarse grained simulations have shown that as the nanotube approaches the lipid bilayer the applied force becomes negative [94]. This could be because of the “smooth” potentials used in coarse graining techniques. However, it should also be noted that the velocity of nanotube penetration in this study is higher and induces a negative curvature in the lipid bilayer thereby increasing the applied force. Therefore the mechanism of nanotube penetration into the exoplasmic monolayer probably depends on the velocity of CNT. At lower velocities the lipid head groups have more time for lateral movement and let the nanotube into the membrane core and relieve the hydrophobic mismatch. This is probably also the reason for observable higher forces of penetration in simulations than in experiments. Similarly, the penetration force was also observed to increase with the velocity in our simula-
tions. However, in all cases the highest force was observed as the nanotube penetrates the lipid head groups. This phase was followed by a steep lowering of the applied force as the nanotube enters the hydrophobic core of the membrane. After the nanotube enters the membrane the hydrophobic mismatch between the lipid molecules and the nanotube results in the lipid tails wrapping the nanotube to reduce the unfavorable exposure of the hydrophobic core to water thereby reducing the pulling forces.

The force-distance profiles are however different following internalization for different pulling velocities of the nanotube. A non-polar nanotube would partition into the hydrophobic core of the lipid bilayer under no applied external force [95]. Therefore, a second peak could be observed between 60 Å and 70 Å as the nanotube exists the membrane at a velocity of 0.05 Å/ps as it penetrates the head groups of the lower lipid monolayer. Also, blocking of the lipid molecules and the strong interactions of lipids in close proximity of the nanotube probably resulted in an increased force. However, at the highest velocity of 0.1 Å/ps although a second peak can be observed the applied force is too high for these interactions and therefore less membrane deformation was observed. At the lowest velocity of 0.01 Å/ps more lipid molecules were interacting with the CNT and were removed along with the nanotube and therefore there was no observable increase in force. Blocking of nanotube by lipids, especially those from the lower leaflet, was previously observed in simulation studies [94]. Therefore, poration forces depend on the energies involved in the membrane bending, breaking of the hydrogen bonds between the head groups and water molecules, the hydrophilic interactions between the lipid head groups, the hydrophobic interactions between the nanotube and the lipids and the energy required for the lateral movement of the lipid molecules. Besides the friction between the nanotube and the membrane lipids also plays an important role in the total force.

To estimate the effect of cholesterol on the forces and required for nanotube pen-
etration, we tested our method on lipid bilayer with 70% POPC and 30% cholesterol with a nanotube velocity of 0.01 Å/ps. A comparison of the obtained force-distance profile with that of pure POPC bilayers (Fig. 3.4) showed a much reduced observed membrane bending. However, the force required to puncture the membrane remained similar to pure POPC bilayers. Although the presence of cholesterol stabilizes the bilayer, cholesterol also partitions into the spaces between the lipid molecules and therefore induces a spontaneous negative curvature on the bilayer. Further an observable second peak was observed as the nanotube exits the bilayer and the force required for the nanotube to exit the bilayer was also higher than for pure POPC. It was also observed that as the nanotube exits the bilayer, higher number of lipids were extracted from the bilayer containing 30% cholesterol than pure POPC thereby disrupting the bilayer. Therefore, the presence of cholesterol likely stiffens the membrane and makes it relatively “brittle” compared to a pure POPC bilayer, leading to greater destabilization of the membrane.

The free energy required for nanotube insertion into the bilayer was found to be $35.74\pm3.54$ kcal/mol. This estimate is slightly lesser than to the amount required to porate a lipid monolayer which was expected to be 40-55 kcal/mol [53]. The energy released in relieving the membrane curvature defect probably facilitates nanotube internalization into the polar head groups of the membrane. After the nanotube across the energy barrier, the non-polar nanotube is spontaneously internalized by the hydrophobic lipid tails and the PMF decreases. Following this phase the PMF should be expected to have a minimum at the center of the bilayer.

In the presence of 30% cholesterol, the free energy of rupture increased to approximately 50 kcal/mol. The formation of two to three extended coiled coils required to porate a cell membrane by viruses was estimated to provide 60-90 kcal/mol of energy. However, it should be noted that the amino acid sequence of the viral fusion
peptides and the composition of the lipid bilayers also play an important role in determining the forces required and the energetics of membrane rupture during viral infection. In the absence of external applied force on the nanotube there would probably be minimal membrane bending and the mechanism of internalization depends only on the hydrophobic mismatch. This would probably explain the spontaneous internalization of carbon nanotube with small diameters. However, for the nanotube to traverse across the bilayer a two-phase internalization mechanism involving an energy requiring step is needed to cross the saddle point at the endoplasmic head groups.

3.5 Conclusions

In this work we calculated membrane structural properties of pure POPC bilayers and POPC/cholesterol complexes using molecular dynamics. The obtained surface area per lipid, membrane thickness and lipid order parameter compared well with experimentally observed results [1, 92]. The lipid order parameter calculations, however, could only reproduce the order parameters for the lower portion of the acyl chains. We further used a non-polar carbon nanotube as a nanosyringe to test the rupture forces and free-energies of the bilayers. These results suggested similar rupture forces for bilayer with and without cholesterol. However, the free-energy required to cross the lipid head groups increased in the presence of cholesterol. Also, the presence of cholesterol resulted in higher destabilization of the membrane as the nanotube is pulled out of the hydrophobic core of the membrane.
Figure 3.2: Order parameter profiles for the sn-1 (left) and sn-2 (right) acyl chains of POPC with varying cholesterol content. Pure POPC (black), POPC:10%Chol(red), POPC:30%Chol(green), POPC:50%Chol(blue)
Figure 3.3: Force-distance profiles of pure POPC (blue) and POPC:Cholesterol bilayer with 30% cholesterol (red) obtained at a nanotube velocity of $v = 0.01 \text{ Å/ps}$.

Figure 3.4: Force-distance profiles of pure POPC (blue) and POPC:Cholesterol bilayer with 30% cholesterol (red) obtained at a nanotube velocity of $v = 0.01 \text{ Å/ps}$. 
Chapter 4: HIV-1 gp41 membrane spanning domain structure in water

4.1 Introduction

The HIV-1 membrane spanning domain anchors the Env in the viral membrane. The MSD is a highly conserved region of gp41, spanning 27 residues from Lys\(^{681}\) to Arg\(^{707}\) and contains a mid-span charged residue Arg\(^{694}\) [50], a common feature of lentiviral MSD’s (Fig. 4.1) [96, 97]. The MSD also contains a GXXXG motif between residues 688 and 692. Although GXXXG motifs are thought to be involved in inter-helical interactions [98, 99], the loss of these residues does not affect the trimerization of the Env monomers required during transport from the endoplasmic reticulum to the Golgi apparatus [50]. However, loss of Arg\(^{694}\), the GXXXG motif or the charged residues Lys\(^{681}\) and Arg\(^{705}\) affect membrane fusion and viral infectivity [49, 50, 97, 100–102].

Despite the MSD’s importance in viral infectivity, the absence of X-ray crystallographic or NMR structural information resulted in minimal information about its structure and function at the atomic level. However, based on it sequence the MSD likely assumes an \(\alpha\)-helical conformation in the viral membrane. In this chapter we model build MSD based on its sequence using a \(\alpha\)-helical subdomain of viral protein Vpu. We further investigate the structure of MSD in water using all-atom simulations. However, classical molecular dynamics simulation is likely insufficient for sampling all relevant protein conformations due to computational limitations. We therefore used metadynamics to accelerate conformational sampling of MSD and we predict con-
Figure 4.1: Schematic of HIV-1 gp41 membrane spanning domain. The basic residues are shown in blue, polar residues in green and non-polar residues in silver.

Formations lowest in free energy. We demonstrate that the MSD shuttles between a stable kinked conformation and a metastable “loop” conformation in solvent water.
4.2 Simulation description

4.2.1 General

All simulations were performed using NAMD 2.7b1 with the CHARMM force field and explicit TIP3P water \([84, 103, 104]\). VMD 1.8.7 was used for visualization and preparation of the simulation systems \([85]\). All production runs were performed in the NVT ensemble unless explicitly specified using a Langevin thermostat at a temperature of 310K with a coupling constant of 5 ps\(^{-1}\).

4.2.2 Gp41 MSD model

HIV-1 gp41 MSD (681-KLFIMIVGGLVGLRVFAVLSIVNRVR-707) is a 27 amino acid peptide thought to span the bilayer as an \(\alpha\)-helix based on sequence [49–51]. We modeled gp41 MSD from a 27 amino acid helical subdomain of the HIV-1 Vpu protein. The side-chain atoms of the Vpu subdomain were deleted and replaced with those of the HIV-1 gp41 MSD. The obtained MSD structure was then solvated with 14,647 water molecules, neutralized and brought to an ionic strength of 0.1 M by adding \(\text{Na}^+\) and \(\text{Cl}^-\) ions and was volume equilibrated for 100 ps in the NPT ensemble. This system was then equilibrated for a total of 20 ns in the NVT ensemble.

4.2.3 Metadynamics

A detailed description of metadynamics is given in section 2.4. For this study the root-mean-squared displacement of the backbone (C, C\(_\alpha\), N, O) atoms with respect to those in a perfect \(\alpha\)-helix (RMSD\(_{\alpha}\)) was chosen as the CV. We used \(W = 0.15\) kcal/mol, \(\delta_\xi = 0.01\) Å, and deposited a new Gaussian every 1 ps. A low RMSD\(_{\alpha}\) on the free energy profile indicates a helical conformation, while kinks and unfolding of the peptide occur at relatively larger RMSD\(_{\alpha}\).
We performed two 120 ns metadynamics simulations starting from two different initial conformations obtained from the 20 ns MD simulation with the MSD solvated in water, (i) an $\alpha$-helical conformation with RMSD$\alpha$ less than 1 Å and (ii) a stable kinked conformation with RMSD$\alpha$ 6 Å. A repulsive wall with a force constant of 500 kcal/mol was used at 14 Å for the run starting from $\alpha$-helical conformation and at 12 Å for the run starting from the kinked conformation. The CV profile during the course of the two simulations shows sampling of the CV space between 0 Å and the upper wall (Fig. 4.3). The free energy profiles between CV space of 0 and 10 Å were considered converged and the RMSD$\alpha$ CV extensively sampled conformations during these simulations. Metadynamics calculations at RMSD$\alpha$ higher than 10 Å resulted in unfolding of the MSD peptide and are not thoroughly sampled by the RMSD$\alpha$ CV with respect to a perfect $\alpha$-helix and therefore are not reported in this work.

4.3 Results

Figure 4.2: RMSD$_\alpha$ of the MSD in pure water with respect to the initial $\alpha$-helical conformation vs. simulation time during 20 ns of MD
4.3.1 HIV-1 gp41 MSD structure in water from molecular dynamics simulations

We tested the stability of the α-helical structure of gp41 MSD solvated in water using molecular dynamics. The RMSD$_\alpha$ evolution from a 20 ns MD is shown in Fig. 4.2. These data suggest a stable conformation of the peptide between RMSD$_\alpha$ 5 and 6 Å. However, after 5 ns the MSD assumes a conformation with a higher RMSD$_\alpha$ between 7-9 Å. Similar behavior was observed after about 18 ns of MD simulation for the final 2 ns. The α-helical structure was observed to be highly unstable and therefore the MSD immediately kinked within 1 ns at Val$^{691}$ during the MD run. This kink is a result of water molecules interacting with the glycine residues and destabilizing the backbone hydrogen bonds while the peptide minimizes the exposure of hydrophobic surface area to water. However, the hydrogen bond between Arg$^{694}$ and Leu$^{690}$ was intact even after helix bending. Although the GXXXG motif unfolded to facilitate Gly interactions with water molecules, the rest of the MSD remained helical. The bulky arginine side chains and the phenylalanine side chain along with the kink minimize the protein backbone atom interactions with water molecules between residues 692 and 707. The two unstable high-RMSD$_\alpha$ structures occur due to the occasional breaking of intrahelical hydrogen bonds between Leu$^{690}$-Arg$^{694}$ and Arg$^{694}$-Ala$^{698}$. This results in a loop-like conformation with RMSD$_\alpha$ between 7-9 Å. However, a 20 ns MD run is probably insufficient to unfold a helical peptide and sample all relevant conformations.

4.3.2 MSD structure in water from metadynamics

To accelerate conformational sampling of the MSD two 120 ns metadynamics trajectories were started from different conformations of the peptide randomly selected from the initial MD run, one from a α-helical structure and another from the kinked structure. The RMSD$_\alpha$ was the CV. The CV trace during the course of the simulation
Figure 4.3: CV profiles of two 120 ns metadynamics runs with gp41 MSD in pure water starting from a complete $\alpha$-helical conformation (red) and a kinked conformation (green) suggests a thorough sampling of CV space (Fig. 4.3).

Figure 4.4: Free energy profiles from two metadynamics runs with gp41 MSD solvated in pure water starting from an $\alpha$-helical conformation (red) and a kinked conformation (green). Average free energy from both the runs is shown in blue.
The free energy profiles from the two runs and the average are presented in Fig. 4.4. This free energy profile suggests a very stable conformation that has a single kink at the GXXXG motif between 4-6 Å RMSD\(\alpha\). The free energy profile also shows that the conformations between 8-10 Å RMSD\(\alpha\), although metastable, can be accessed only by overcoming a few kcal/mol free energy barrier. These results support the conformations observed during our initial MD run as in a stable kinked helix at the GXXXG motif and a metastable structure with a second kink at Phe\(^{697}\). Furthermore, the free energy profile also suggests a metastable \(\alpha\)-helical structure in water. These results indicate that the residues C-terminal to Arg\(^{694}\) can form a stable \(\alpha\)-helical structure even in water and that the residues N-terminal to Arg\(^{694}\) kink at the GXXXG motif. However, it should be noted that with water as solvent and the polar residues accessible, most of the helical structure remained intact, despite the kink.

### 4.4 Discussion

![Diagram](image)

Figure 4.5: Representations of stable kinked conformation (left) and metastable loop conformation of MSD (right) of gp41 MSD in pure water observed in 20 ns MD simulation
During the 20 ns MD run the $\alpha$-helical MSD immediately ($<2$ ns) assumed a kinked conformation in water. The RMSD$_{\alpha}$ evolution of the MD simulation suggested that the MSD shuttles between a stable conformation between 4-6 Å RMSD$_{\alpha}$ and a metastable conformation (Fig. 4.2). The kink in the stable conformation was observed at the GXXXG motif. In the metastable state the MSD assumed a “loop” structure with one kink at the GXXXG motif and a second kink at Phe$^{697}$ (Fig. 4.5). The stable kinked conformation was previously observed during short ($<4$ ns) MD simulations of MSD. However, to our knowledge we are first to report the “loop” conformation of the MSD in water. Therefore, we performed metadynamics simulations starting from two different initial configurations to obtain the PMF profile of the MSD in water.

The average PMF profile obtained from the two 120 ns metadynamics simulations corroborated the observations from the MD simulations. The PMF profile had a minimum between 4-6 Å RMSD$_{\alpha}$ suggesting a stable conformation in that region. The PMF profile also suggested multiple metastable conformations between 8-10 Å RMSD$_{\alpha}$ that are accessible across a low energy barrier from the stable conformation. The structures of MSD observed in the stable and metastable regions of the PMF profile are similar to the kinked and loop structures observed during the MD run respectively. The obtained PMF profile also suggested highly metastable completely $\alpha$-helical conformation that explains the fast conformational change from an $\alpha$-helix to the kinked conformation during the MD run. It should however be noted that in a MD simulation the MSD could still sample in the $\alpha$-helical conformation for longer time but it is unlikely that a kinked or loop conformation would assume a $\alpha$-helical conformation. The CV profiles obtained from metadynamics however showed a thorough and extensive sampling of the conformational space.
4.5 Conclusions

In this chapter, we model built an all-atom structure of MSD using its sequence and an α-helical subdomain of the viral protein Vpu. We further sampled conformations of MSD in water using both molecular dynamics and metadynamics and reasoned that although the MSD assumes a stable kinked structure in water, it can shuttle among the stable conformation and several metastable conformations.
Chapter 5: HIV-1 gp41 membrane spanning domain structure in cholesterol containing lipid membranes

5.1 Introduction

Human immunodeficiency virus (HIV) entry into the target cells is mediated by trimeric complexes or “spikes”, comprised of glycoprotein heterodimers. Each heterodimer is a non-covalently linked complex of receptor-targeting gp120 and membrane-anchoring gp41. Beginning at its N-terminus, gp41 is comprised of a glycine rich fusion peptide (FP), the so-called N- and C-terminal heptad repeats (NHR and CHR respectively), a membrane-proximal external region (MPER), a membrane-spanning domain (MSD) and a long cytoplasmic tail that extends into the endoplasmic side of the viral lipid bilayer envelope. When spike-associated gp120 binds to its target CD4 receptor, it undergoes conformational changes that (i) decrypt a binding site for CCR5/CXCR4 receptors and (ii) initiate a cascade of conformational changes in gp41 [8, 9, 41–45]. These changes are thought to include insertion of previously spike-sequestered FP’s into the target cell membrane and eventual refolding of the NHR/CHR’s to form a so-called six-helix bundle (6HB) which pulls the viral and cell membranes together in preparation for fusion [48]. Although this “pull-model” of membrane fusion is theoretically appealing, it does not address the experimentally observed behavior of certain gp41 mutants, especially those that strengthen the interaction of gp41’s MSD with the viral membrane while compromising infectivity [50]. We speculate that a major functional role of gp41 is to destabilize the viral membrane
as a necessary part of the fusion mechanism.

For the entire MSD to span the viral membrane, Arg$^{694}$ localizes to the hydrophobic core, which is energetically unfavorable compared to being solvated by water. However, residues such as arginine, when required by sequence to be inside a lipid membrane, “snorkel”; i.e. the charged end groups of the side chains localize to the hydrophilic head-groups of one membrane leaflet [105].

The purpose of this study is to understand the structure and orientation of MSD in a cholesterol containing lipid membrane and the effects of MSD conformation on the viral membrane structure in order to provide clues toward an explanation of the functional roles of conserved MSD features. Recent simulation studies using molecular dynamics suggested that the MSD assumes an $\alpha$-helical conformation in a lipid membrane [51]. We used metadynamics to accelerate conformational sampling of MSD, and we predict conformations in membrane-spanning states lowest in free energy. In particular, we demonstrate the effects of Arg$^{694}$ and snorkel direction of the Arg$^{694}$ side-chain on water penetration to the membrane core and local membrane thinning.

5.2 Simulation description

5.2.1 General

All simulations were performed using NAMD 2.7b1 with the CHARMM force field and explicit TIP3P water [84, 103, 104]. VMD 1.8.7 was used for visualization and preparation of the simulation systems [85]. All production runs were performed in the NVT ensemble unless explicitly specified using a Langevin thermostat at a temperature of 310K with a coupling constant of 5 ps$^{-1}$. 
5.2.2 Gp41-membrane model

We used steered molecular dynamics (SMD) to insert water-equilibrated α-helical MSD into solvated lipid bilayers comprised of 50 mol% dipalmitoylphosphatidylcholine (DPPC)/cholesterol [87–89]. Two MSD-membrane simulation systems were prepared (Fig. 5.1): (1) MSD inserted with Arg<sup>694</sup> snorkeling towards the exoplasmic head groups and (2) MSD inserted with Arg<sup>694</sup> snorkeling towards the endoplasmic head groups. The exoplasmic (N-terminal) and endoplasmic (C-terminal) sides of the membrane are only based on the orientation of the MSD peptide in the bilayer and the lipid and solution composition of the leaflets are similar.

![Membrane-peptide models of MSD α-helix with Arg<sup>694</sup> snorkeling towards exoplasmic head groups (A) and endoplasmic head groups (B). (C) Structural conservation between the two MSD models and the exoplasmic (red) and endoplasmic (blue) snorkeling positions of Arg<sup>694</sup> after 20 ns of equilibration. The residues with the highest structural conservation are shown in blue and the lowest are shown in red. The Leu mutant was prepared using the endoplasmic snorkeling model.](image-url)
The CHARMM-GUI membrane builder was used to generate initial bilayer structures containing 512 lipid molecules [106]. An initial MSD-membrane system was prepared by placing MSD 35 Å above the lipid bilayer and the system was fully solvated by adding explicit TIP3P water molecules. The simulation box measured 105 Å × 98 Å × 127 Å and contained a total of 141642 atoms with 512 lipid molecules, 29561 water molecules for complete solvation of the bilayer and MSD, and 280 Na⁺ and Cl⁻ ions. The system was then equilibrated in the NPT ensemble for 6 ns. Following equilibration, the MSD peptide was inserted into the lipid bilayer using constant velocity SMD. To obtain the exoplasmic Arg⁶⁹⁴ snorkeling model, we used the center-of-mass of Arg⁷⁰⁷ as the SMD atom and steered the peptide in the membrane from the exoplasmic side of the bilayer with Arg⁶⁹⁴ interacting with the water molecules and lipid head groups towards the exoplasmic leaflet. Similarly, to obtain the endoplasmic Arg⁶⁹⁴ snorkeling model we used the center-of-mass of Lys⁶⁸¹ as the SMD atom and inserted the peptide from the endoplasmic side of the bilayer. For both the models we used a steering velocity of 10 Å/ns and the lipid head groups were restrained in the z-direction to prevent bilayer distortion. Insertion of MSD into the membrane to obtain the exoplasmic snorkeling model resulted in slight unfolding of the helix near 696-698 and 704-707. Therefore, in the endoplasmic snorkeling model the MSD secondary structure was restrained during insertion. However, both the models showed a good structural conservation following a 20 ns MD equilibration (Fig. S1C).

Following MSD insertion the simulation system was cropped by removing the excess water required to solvate MSD prior to insertion. Moreover, the lipid bilayer was also cropped to reduce the system size while maintaining at least twice the theoretical length of MSD (40 Å) in the lateral dimensions. The simulation boxes measured 80 Å × 84 Å × 80 Å and 85 Å × 81 Å × 77 Å and the total number of atoms were
58524 and 58185 for the exoplasmic and endoplasmic snorkeling models, respectively. The final gp41-membrane models contained approximately 320 lipid molecules and a 32:1 water to lipid ratio to ensure complete solvation. The systems were then neutralized and brought to an ionic strength of 0.1 M by adding Na$^+$ and Cl$^-$ ions and were equilibrated in the NPT ensemble for 3 ns followed by equilibration in the NVT ensemble for at least 5 ns prior to the metadynamics calculations.

To create the Arg$^{694}$ to Leu mutant, the membrane inserted and equilibrated peptide from the endoplasmic snorkeling model was used. Arg$^{694}$ was mutated to Leu$^{694}$ using Mutate Residue plugin in VMD. The obtained system was then further equilibrated in the NPT ensemble for 3 ns and NVT ensemble for at least 5 ns.

### 5.2.3 Metadynamics

For this study the root-mean-squared displacement of the backbone (C, C$_\alpha$, N, O) atoms with respect to those in a perfect $\alpha$-helix (RMSD$_\alpha$) was chosen as the CV. For all metadynamics simulations, we used $W = 0.15$ kcal/mol, $\delta_\xi = 0.01$ Å, and deposited a new Gaussian every 1 ps. A low RMSD$_\alpha$ on the free energy profile indicates a helical conformation, while kinks and unfolding of the peptide occur at relatively larger RMSD$_\alpha$. The arithmetic average of the free energy profiles in two 10 ns intervals after a certain filling time ($t_f$) were compared to judge the convergence of metadynamics calculations (Fig. S8,S9). The filling time $t_f = 240$ ns for the exoplasmic model and $t_f = 120$ ns for the endoplasmic model after which the calculation sampled the CV range diffusively. The free energy profiles between CV space of 0 and 10 Å were considered converged and the RMSD$_\alpha$ CV extensively sampled conformations within this region. Metadynamics calculations at RMSD$_\alpha$ higher than 10 Å resulted in unfolding of the MSD peptide and are not thoroughly sampled by the RMSD$_\alpha$ CV with respect to a perfect $\alpha$-helix and therefore are not reported in this
5.3 Results

5.3.1 Effects of Exoplasmic snorkeling of Arg\textsuperscript{694} on MSD conformational ensembles

A single 260 ns metadynamics simulation resulted in uniform exploration of CV-space between 0-13 Å RMSD\textsubscript{α}, when initiated from a helical conformation in which the Arg\textsuperscript{694} side chain snorkels towards the exoplasmic head-groups. The averaged free energy profile for the last 10 ns computed using metadynamics is shown in Fig. 5.3. Conformations with RMSD\textsubscript{α} between 1 and 2 Å display the lowest free energy. The RMSD\textsubscript{α} visited helical and non-helical states several times throughout the 260 ns metadynamics calculation (Fig. 5.2).

The most stable conformation in this model is a complete α-helical conformation with Arg\textsuperscript{694} sidechain snorkeling to interact with the water molecules and lipid head groups towards the exoplasmic leaflet. In this conformation, the peptide tilts with respect to the membrane normal to facilitate Arg\textsuperscript{694} snorkeling while preventing water molecules to interact with the hydrophobic residues between 681 and 692 thereby stabilizing the α-helix. However, the proximity of Arg\textsuperscript{694} to the glycines in the GXXXG motif results in water molecules interacting Gly\textsuperscript{688}, Gly\textsuperscript{689} and Gly\textsuperscript{692}. Thermal fluctuations result in opening and closing of the backbone hydrogen bonds in the α-helix and water molecules competing for the hydrogen bonds destabilize the α-helical structure. Furthermore, the propensity of hydrophobic residues to localize in the membrane core results in peptide bending at Val\textsuperscript{691}. However, water interactions with glycines result in the MSD eventually kinking at Gly\textsuperscript{692} and also unfolding of the GXXXG motif, with the hydrophobic Leu\textsuperscript{690} and Val\textsuperscript{691} buried in the membrane core and Gly\textsuperscript{688}, Gly\textsuperscript{689} and Gly\textsuperscript{692} interacting with water. The free energy profile
suggests metastable kinked conformations with RMSD$\alpha$ between 4 and 6 Å.

The metastable state has the Arg$^{694}$ side-chain interacting with the exoplasmic head groups and water molecules, with the MSD kinked at Gly$^{692}$ and unfolding of the GXXXG motif (Fig. 5.4). The rest of the peptide between residues 694 and 702 retains an $\alpha$-helical structure in this conformation. However, the C-terminal residues from 703 to 707 unfold to facilitate Arg$^{694}$ snorkeling and thereby loss of a helical turn at the C-terminus. Also, peptide conformations with a complete helical structure but kinked at Arg$^{694}$ were sampled at RMSD$\alpha$ of 4-8 Å.

Unfolding of the subdomain between residues 681 and 694 is the combined result of hydrophobic localization in the membrane core, Lys$^{681}$ snorkeling towards the exoplasmic lipid head groups, and the presence of water molecules destabilizing the helical structure. However, these conformations had a relatively high RMSD$\alpha$ (8-10 Å) and as can be seen from the free energy profile are relatively unstable. Finally,
we note that the Arg$_{694}$ side-chain remained in interaction with the exoplasmic head-groups throughout the calculation.

### 5.3.2 Effects of Endoplasmic snorkeling of Arg$_{694}$ on MSD conformational ensembles

![Free energy profiles of MSD in a lipid membrane with Arg$_{694}$ snorkeling](image)

Figure 5.3: Free energy profiles of MSD in a lipid membrane with Arg$_{694}$ snorkeling towards exoplasmic head groups sampled during 260 ns metadynamics (red), Arg$_{694}$ snorkeling towards endoplasmic head-groups sampled during 140 ns metadynamics run (blue) and Leu$_{694}$ mutant (cyan). Schematic illustration of MSD with basic charged residues in blue and polar residues in green. Hydrophobic residues are shown in silver and Arg$_{694}$ side-chain is shown in exoplasmic (red) and endoplasmic (blue) snorkeling positions.

We conducted a second metadynamics calculation initiated from an α-helix with the Arg$_{694}$ side-chain snorkeling toward the endoplasmic head-groups. During this run metadynamics sampled the CV space between 0 and 15 Å RMSD$_{α}$ again uniformly and the simulation was stopped at 140 ns when the peptide completely unfolded.
The two predominant conformations in this model are (i) a completely helical conformation, with RMSD$_{\alpha}$ less than 2 Å and the axis of the $\alpha$-helix tilting with respect to the membrane normal to facilitate snorkeling of the Arg$^{694}$ side chain and (ii) a kinked helix, RMSD$_{\alpha}$ 4-6 Å, with the kink at Phe$^{697}$ and residues 681 through 697 spanning the bilayer along with the snorkeling side chains of Lys$^{681}$ and Arg$^{694}$ (Fig. 5.4). Of these, the tilted $\alpha$-helical conformation between RMSD$_{\alpha}$ 0 to 2 Å is the more stable structure. In this conformation, Arg$^{694}$ snorkels towards the endoplasmic head groups and interacts with water molecules penetrating the lipid bilayer. Arg$^{705}$ positioned on the same side of the $\alpha$-helix, near the C-terminus of the MSD, assists in $\alpha$-helix tilting by interacting with water molecules and lipid head groups at the membrane-water interface. Moreover, Arg$^{705}$ also interacts with water molecules penetrating the membrane core towards Arg$^{694}$. Interaction of Arg$^{705}$ with both molecules at the membrane water interface and water molecules in the membrane core with a “paddle” motion was also observed in the 20 ns MD simulation. Therefore, in a completely $\alpha$-helical conformation the presence of a charged residue with a relatively large side chain assists in snorkeling of Arg$^{694}$ by inducing helix tilt and also acting as a mediator for water molecule penetration into the membrane core. Furthermore, with two long side chains of Arg$^{694}$ and Arg$^{705}$ interacting with the water molecules in the membrane core, the backbone hydrogen bonds between the residues 694 to 705 were shielded from competition from water. A tilted, completely $\alpha$-helical MSD conformation was also favorable to the two charged residues, Lys$^{681}$ and Arg$^{707}$, on either terminus of the MSD. The long side chains of lysine and arginine are also predominantly hydrophobic except for the charged moieties at the end. We observed that in a tilted $\alpha$-helix the N-terminus of the MSD was in the hydrophobic membrane environment while the tip of the side chain snorkels and interacts with the lipid head groups and water. Similarly, the tilt also pulled Arg$^{707}$ relatively further
into the membrane core where it could then snorkel and interact with molecules at the membrane-water interface.

The set of conformations that were extensively sampled with approximately 4 Å RMSD_α were very similar to the hypothesized snorkeling model proposed by Hunter et al. (Fig. 5.4) [49]. In this set, the presence of water molecules near the C-terminal residues of the MSD, especially when Arg^{705} interacts with molecules at the membrane-water interface, results in destabilization of the helix with loss of a hydrogen bond between Phe^{697} and Ser^{701}. This results in the MSD α-helix bending at Phe^{697}, while the rest of the subdomain from 698 through 707 maintains an α-helical conformation. Additionally, the hydrogen bond between Arg^{694} and Ala^{698} was maintained even after helix bending. The partitioning of the charged groups towards the membrane-water interface and the hydrophobic groups towards the membrane core assists in maintaining the α-helical conformation by protecting vulnerable backbone hydrogen bonds from attack by water. Residues 681 through 697 spanned the bilayer in this model with the two long side chains of Lys^{681} and Arg^{694} snorkeling towards the exoplasmic and endoplasmic head groups respectively. These residues maintained a helical structure with the hydrophobic part of the Arg^{694} side chain shielding the hydrogen bonds between residues 694 and 697 from water molecules in the membrane core. Although these conformations were sampled extensively by metadynamics, the final free energy profile suggests that these conformations are metastable.

As seen in the case of exoplasmic Arg^{694} snorkeling, metadynamics sampled various unstable and short lived conformations during endoplasmic Arg^{694} snorkeling simulation. However, most of these conformations retained the α-helical structure between residues 681 and 697. The presence of lipid head groups and water molecules at the membrane water interface resulted in destabilization of the helical subdomain between 698 and 707. Conformations with a completely unfolded C-terminal subdomain
were also observed. During this run we also observed a completely unfolded MSD peptide. In these conformations the C-terminal charged residues and water molecules associated with these residues entered deep into the lipid bilayer core resulting in the interaction with GXXXG motif and destabilization of the backbone hydrogen bonds.

### 5.3.3 Effects of Arg$^{694}$ to Leu mutation on MSD conformational ensembles

Metadynamics calculations with the Leu mutant were performed for a total of 170 ns and was initiated from an α-helix. Metadynamics extensively sampled the CV space between 0 and 13 Å during this simulation. The simulation was stopped when the peptide completely unfolded. The two major conformations sampled during this run were a completely α-helical conformation with RMSD$^\alpha$ less than 2 Å and a kinked conformation with RMSD$^\alpha$ between 4 and 6 Å and similar to the metastable conformation observed in the endoplasmic snorkeling model. Similar to the other two models the α-helical conformation is the stable conformation and the kinked conformation is the metastable state as observed from the PMF profile (Fig. 5.3). Although the conformations observed with the Leu mutant are similar to the endoplasmic snorkeling model, water penetration into the membrane was not observed due to the lack of a polar side chain the hydrophobic core of the membrane. Moreover, in the kinked conformation the axis of the subdomain between Lys$^{681}$ and Phe$^{697}$ was also not always parallel to the membrane normal as observed in the endoplasmic snorkeling model. This is likely due to the absence of Arg$^{694}$ stabilizing the subdomain by interacting with the endoplasmic head groups and water molecules. Water molecules competing for the hydrogen bonds with the backbone were also observed in the kinked conformation near Phe$^{697}$. In the stable completely α-helical conformation the helix tilt was observed likely due to the presence of Arg$^{705}$ and also the hydrophobic part of the Lys$^{681}$ side-chain trying to be buried in the hydrophobic core.
while the tip interacts with the polar groups.

Metadynamics calculation also sampled several short lived conformations during this run including a conformation with a kink in the GXXXG motif similar to the exoplasmic snorkeling model. However, water molecules present in the vicinity prior to the mutation and interacting with the backbone atoms during the metadynamics run likely contributed towards the kink that was unstable and short lived. The completely unfolded peptide at the end of the metadynamics run resulted in water molecules entering into the core of the membrane as observed in the other two models.
Figure 5.4: Representation of stable and metastable conformations of MSD in a lipid bilayer with Arg<sup>694</sup> snorkeling towards the exoplasmic, endoplasmic head groups and the Leu mutant.
5.3.4 Effects of MSD conformation on membrane thickness

Figure 5.5: (Upper panels) Map of mean membrane thickness (measured as a distance between centers of mass of lipid head groups in 10 Å square sections of each leaflet); color bar units in Å. Left, stable; right, metastable. (Lower panels) Side-view snapshots of stable (left) and metastable (right) conformations.

In Fig. 5.5, Fig. 5.6 and Fig. 5.7 we show plots of average membrane thickness for the exoplasmic snorkeling model, endoplasmic snorkeling model and the Leu mutant respectively of MSD in their stable and metastable states. The average bilayer thickness of DPPC-cholesterol (50 mol%) was observed to be 45 Å. Membrane thickness
for the exoplasmic and endoplasmic snorkeling models was found to be significantly lower (approximately 35 Å) than pure DPPC-cholesterol bilayer thickness in the proximity of the peptide in both the stable and metastable state. Exoplasmic snorkeling Arg$^{694}$ resulted in a localized membrane defect in the upper leaflet while endoplasmic snorkeling of Arg$^{694}$ results in a localized membrane defect in the lower leaflet due to water penetration and lipid head groups interacting with Arg$^{694}$. Although, the stable and

Figure 5.6: (Upper panels) Map of mean membrane thickness (measured as a distance between centers of mass of lipid head groups in 10 Å square sections of each leaflet); color bar units in Å. Left, stable; right, metastable. (Lower panels) Side-view snapshots of stable (left) and metastable (right) conformations.
metastable state of the endoplasmic snorkeling model displayed similar membrane thickness, it was observed that the stable conformation of the exoplasmic snorkeling model had lower average membrane thickness than the metastable state. It should be noted that in the metastable state of the exoplasmic model the MSD unfolded near the C-terminus to facilitate Arg$^{694}$ snorkeling towards the exoplasmic head groups. This condition is true for both the tilted completely $\alpha$-helical conformations that places Arg$^{694}$ near the center of the bilayer and the metastable kinked conformations in which the membrane spanning subdomains are relatively short. For conformations with higher RMSD$_{\alpha}$ in both the models, unfolding of the peptide resulted in charged groups at the termini burying themselves in the hydrophobic core of the bilayer. These conformations resulted in significant water penetration into the membrane.

Conversely, the Arg$^{694}$ to Leu mutant displayed a membrane thickness similar to pure DPPC bilayers in the stable $\alpha$-helical conformation. In the metastable state the kink at Phe$^{697}$ combined with the polar residue interacting with water and head groups at the terminals resulted in minimal decrease in membrane thickness.

5.4 Discussion

In our model with Arg$^{694}$ snorkeling towards the exoplasmic head groups, the free energy profile suggested a stable completely $\alpha$-helical conformation. The MSD peptide tilts to facilitate Arg$^{694}$ snorkeling while preventing water molecules to interact with the hydrophobic residues to maintain the $\alpha$-helical conformation. A metastable conformation kinked at Gly$^{692}$ with an unfolded GXXXG motif was also observed with RMSD$_{\alpha}$ between 4 and 6 Å. This is likely because an $\alpha$-helical MSD with Arg$^{694}$ snorkeling towards the exoplasmic head groups resulted in water molecules in the proximity of the glycines, which then can attack backbone hydrogen bonds involving glycines. This model requires that the MSD subdomain between residues 694 through
707 anchor the “spike” to the viral membrane in the metastable state. However, recent experimental results based on truncation mutants suggested that addition of just three residues C-terminal to residue 694, up to Phe$^{697}$, is enough to stably anchor the Env glycoprotein in the membrane [49, 100]. Therefore, we believe that a MSD model with Arg$^{694}$ snorkeling towards the exoplasmic head groups is improbable in the native spike, although it should not be discounted that these conformations could still play a role in disruption of the viral membrane.

The free energy profile for the model with Arg$^{694}$ snorkeling towards the endo-
plasmic head groups also suggested that the completely α-helical conformation is the most stable conformation. The possibility of the α-helix to tilt and Lys$^{681}$ and Arg$^{694}$ to snorkel towards either leaflet’s head groups along with the observed shielding of the backbone hydrogen bonds likely resulted in this stable conformation. A similar phenomenon of shielding of backbone hydrogen bonds thereby stabilizing the α-helical structure by the arginine side chain was previously observed in alanine rich peptides [107]. This conformation is consistent with results from truncation studies that show that addition of residues C-terminal to Arg$^{694}$ increase surface expression by better anchoring of the peptide with increasing number of residues in the α-helix [49]. Although this conformation explains the importance of the length of the subdomain between 694 and 707 for the incorporation of the spike and the role of Arg$^{705}$ in destabilization of viral membrane and therefore fusion, it was also important that this simulation predicted as metastable the experimentally hypothesized snorkeling model of the MSD [49]. We also observed that the MSD peptide from Lys$^{681}$ to Phe$^{697}$ is enough to span the lipid bilayer and anchor the MSD.

The stable and metastable conformation observed in the Leu mutant model are similar to the endoplasmic snorkeling model. This suggests that the MSD peptide has a natural tendency to kink at Phe$^{697}$ and the presence of a polar side chain in the hydrophobic core and snorkeling towards the endoplasmic head groups facilitates the kink. It should be noted that penetration of water molecules to the extent observed in the endoplasmic snorkeling model was not observed in the Leu mutant model.

Both the Arg$^{694}$ snorkeling models showed a consistent thinning of the lipid membrane due to water penetration and lipid head groups interacting with Arg$^{694}$. Cryo-electron tomographic images of HIV-1 Env spike show an obvious decrease in membrane thickness in the lower leaflet of the bilayer [108] leading us to speculate that Arg$^{694}$ snorkels towards the endoplasmic head groups in the native spike. Conversely,
the Leu mutant model showed minimal membrane thinning in both the stable and metastable state. The presence of water molecules in the hydrophobic core and membrane thinning due to the presence of the polar sidechain of Arg\textsuperscript{694} likely have a role in membrane destabilization during membrane fusion. These observations may therefore explain the loss of infectivity observed from experimental results when Arg\textsuperscript{694} was mutated into a Leu [100].

Based on these observations we hypothesize that the MSD exists in the metastable kinked snorkeling model in the native spike. Conformational changes in the gp41 ectodomain during the formation of the pre-fusion intermediate perhaps result in the MSD assuming a tilted, completely $\alpha$-helical conformation. A tilted $\alpha$-helix, continuous from Lys\textsuperscript{681} to Arg\textsuperscript{707}, further assists the penetration of water into the membrane core to interact with Arg\textsuperscript{694}. The GXXXG motif, although insignificant for the incorporation of the native spike [50], is perhaps important to induce inter-helical interactions and in maintaining gp41 MSD trimers during these conformational changes [98, 99].

5.5 Conclusions

We sampled various conformations of the HIV-1 gp41 MSD in a cholesterol-rich lipid bilayer based on two starting models, (i) with Arg\textsuperscript{694} snorkeling towards the exoplasmic head groups and (ii) with Arg\textsuperscript{694} snorkeling towards the endoplasmic head groups, using all-atom metadynamics calculations. We further prepare a third model with Arg\textsuperscript{694} mutated into a Leucine. Based on our observations and previous experimental results, we hypothesize that the MSD assumes a conformation with the mid-span Arg\textsuperscript{694} snorkeling towards the endoplasmic head groups in the native spike. Furthermore, there are two conformations, a stable tilted $\alpha$-helical conformation and a metastable kinked snorkeling conformation, that the MSD likely can shuttle be-
tween during conformational changes of the gp41 ectodomain. It was also observed that the MSD residues from Lys$^{681}$ to Phe$^{697}$ are enough to span the bilayer and effectively anchor gp41, consistent with experimental observations. MSD with Arg$^{694}$ snorkeling towards endoplasmic head groups also resulted in significant thinning of the membrane.
Chapter 6: Model building HIV-1 gp41 ectodomain pre-fusion intermediate

6.1 Introduction

The envelope (Env) glycoproteins of HIV-1 are comprised of non-covalently interacting subunits of surface glycoprotein gp120 and transmembrane glycoprotein gp41 that are assembled in the form of trimeric spikes on the viral membrane. Gp120 is responsible for viral tropism and binds to the cell surface receptor CD4 and coreceptor (CCR5/CXCR4). Gp120 binding results in the exposure of the hydrophobic N-terminal fusion peptide of gp41 that is inserted into the target-cell membrane. Following FP insertion, the N- and C- heptad repeats (NHR and CHR respectively) of gp41 form into an extended pre-fusion intermediate (PFI) conformation bridging the viral and target-cell membranes. The NHR and CHR then fold into a thermally stable six-helix bundle (6HB) bringing the viral and target-cell membranes into proximity initiating fusion (Fig. 6.1). The structural information of HIV-1 gp41 in the 6HB form has been solved using X-ray crystallography [109–111]. According to these structures, the NHRs form into a parallel, trimeric coiled coil of α-helices. The three CHRs then wrap antiparallel into the three interhelical grooves of the NHRs as α-helices. The important features of 6HB involve a hydrophobic cavity in the interhelical groves of the NHR core that includes residues Lys$^{565}$, Leu$^{566}$, Leu$^{568}$, Thr$^{569}$, Ile$^{573}$, Leu$^{576}$, Val$^{570}$, Trp$^{571}$, Gly$^{572}$, Lys$^{574}$ and Gln$^{577}$ from the NHR [109, 110, 112]. The hydrophobic residues in the CHR especially Trp$^{628}$ and Trp$^{631}$ are known to collapse
into the hydrophobic cavity in the 6HB (Fig. 6.2). Further away from the flexible linker hydrophobic interactions between residues \( \text{Ile}^{559} \) and \( \text{Ile}^{646} \) are thought to stabilize the 6HB. However, the point mutation studies suggested that in general the residues closer to the flexible linker on both the NHR and CHR are more important for the formation and stability of the 6HB that those further away from it [113]. The flexible linker is known to consist of three glycosylation sites at Asn\(^{611}\), Asn\(^{616}\) and Asn\(^{624}\).

Several other enveloped viruses, including influenza, share a similar fusion mechanism. However, the native and final structures of these viral proteins are available while that of HIV-1 gp41 is still elusive. With the gp41 ectodomain being the most conserved region of the HIV-1 Env, research is now focused to understand the structure of the ectodomain in its native and pre-fusion intermediates. The purpose of this chapter is to model build a complete 6HB and an extended state representing the PFI of HIV-1 gp41 with NHR, the flexible linker connecting the two heptad repeats and the CHR from existing structure (PDB access code: 1AIK) [109, 110]. The PFI was then successfully folded into a stable 6HB using a combination of molecular dynamics and targeted molecular dynamics (TMD). The interactions between the amino acids in NHR and CHR and the final state were consistent with existing literature. The mechanism of folding of PFI into 6HB is still speculative. However topological features of the ectodomain, specifically the position of the CHRs in the extended state and the flexible linker are at variance with current knowledge. It should be noted that the goals of this exercise were to develop a basic knowledge of the HIV-1 ectodomain and the specific interactions involved during the cascade of conformational changes from PFI to 6HB, attempt at model building 6HB and PFI and test the ability of molecular simulations to replicate the mechanism of folding.
6.2 Simulation methods and models

6.2.1 General methods

All simulations were performed using NAMD2.7b2 with the CHARMM force field and explicit TIP3P water. The trajectories were visualized using VMD 1.8.7. Secondary structure restraints were applied on the α-helical subdomains of the ectodomain during all vacuum simulations.

6.2.2 Model building gp41 ectodomain six-helix bundle (6HB)

The initial structure of 6HB was obtained from the crystallographic data obtained by Chan et al. (PDB access code: 1AIK) [109]. The structures consists of one NHR and one CHR. The complete trimer to represent the known biologically significant state can be obtained by using the coordinates provided in the structure information for the other two NHR/CHR pairs. However, the flexible linker connecting the C-terminus of the NHR to the N-terminus of the corresponding CHR is absent. The complete sequence of the linker is given by 581 SGIVQQQNLRAIEAQQHLLQLTVWGKQLQARIL 628 (HXB2R) [110]. The linker was applied as patch to the C-terminus of the NHR in CHARMM. The disulfide link was applied between Cys$^{598}$ and Cys$^{604}$ (Fig. 6.1). The system with the NHR connected to the linker and the CHR in vacuum were minimized for 1000 steps. The entire system was then equilibrated using the NVT ensemble in vacuum for 100 ps at a temperature 310K. The linker collapses at the disulfide link and forms a loop. The C-terminus of the linker was then steered towards the N-terminus of the CHR using steered molecular dynamics. cv-SMD was performed using a velocity of 0.1 Å/ps in vacuum for 1 ns in the NVT ensemble at a temperature of 310 K and the α-carbon of Tyr$^{627}$ was used as the SMD atom. The C-terminal of the linker was then applied as a patch to the N-terminus of the CHR. The system containing one NHR and one CHR connected
by the flexible linker was then used to generate a complete 6HB structure using the coordinates provided in the structure. The system was then inspected for overlapping atoms and minimized for 1000 steps. Equilibration of the system was performed in the NVT ensemble at a temperature of 310 K for a total of 1 ns in vacuum.

The ectodomain system was then solvated with approximately 100,000 water molecules, neutralized and brought to an ionic strength of 0.1 M by adding Na$^+$ and Cl$^-$ ions. The solvated system was then volume equilibrated for 100 ps in the NPT ensemble with a pressure of 1 bar before performing a 10 ns MD run in the NVT ensemble.

6.2.3 Model building of gp41 ectodomain pre-fusion intermediate (PFI)

To model build the gp41 PFI, the monomeric NHR-linker-CHR complex described in the previous section was used. SMD was used to steer the C-terminus of the CHR along the axis of the 6HB in vacuum while the backbone $\alpha$-carbons of NHR are restrained. A steering velocity of 0.1 Å/ps was used and the center-of-mass of Glu$^{661}$ was used as the SMD atom. The simulations were performed at a temperature of 310 K in the NVT ensemble. Based on the estimate from influenza fusion proteins (HA), gp41 was also assumed to span approximately 100 Å from the viral membrane to the target-cell membrane in the extended state of the PFI. Since the 33 amino acid long fusion peptide is absent in our structure, the distance between the N-terminus of the NHR and the C-terminus of the CHR was targeted to be at least 80 Å. The complete trimer of extended ectodomains was prepared from the crystallographic coordinates as explained in the previous section to obtain a model of the PFI in vacuum(Fig. 6.3). The PFI system was then minimized for 1000 steps, equilibrated in the NVT ensemble for 100 ps while the NHRs and the CHRs are restrained.

The PFI system was then solvated by adding water molecules and neutralized to
0.1 M using Na\(^+\) and Cl\(^-\) ions. Folding of PFI into 6HB was performed using MD and TMD. TMD requires that the initial and final systems have the same number of particles. Therefore, the final PFI system was ensured to have the similar number of water molecules and ions as the 6HB system. The entire system was then subjected to equilibration in the NPT ensemble for 100 ps at a pressure of 1 bar and temperature of 310 K followed by equilibration the NVT ensemble at a temperature of 310 K for a total of 10 ns.

### 6.2.4 Targeted molecular dynamics (TMD)

TMD was used to fold the PFI of the ectodomain into 6HB and was explained in Section 2.5. The initial state was obtained after the 10 ns equilibration of PFI model. The 6HB model equilibrated for 10 ns was used as the final state. All restraints were released during TMD and the simulations were performed in the NVT ensemble. The backbone \(\alpha\)-carbons of the NHRs and CHRs were subjected to TMD forces. A spring constant \((k)\) of 200 kcal mol\(^{-1}\) Å\(^{-2}\) was used and the TMD simulations were performed for 10 ns.

### 6.3 Results

#### 6.3.1 Molecular dynamics simulations of six-helix bundle and pre-fusion intermediate

The RMSD profiles for the backbone C, C\(^\alpha\) and N atoms of the 6HB, the linker and the NHR/CHR subdomains during 10 ns of MD is provided in Fig. 6.4. As expected the NHR/CHR subdomains are extremely stable in the 6HB. At approximately 7 ns the C-terminus of one of the CHRs in the 6HB was observed to unfold from the stable state that resulted in a peak in the RMSD profile. However, the interactions between Ile\(^{599}\) and Ile\(^{646}\) stabilized the 6HB structure from further unfolding. Conversely, the
linker is highly flexible and the RMSD profile of the 6HB overall was influenced by the fluctuations in the linker. The linker collapsed during the vacuum equilibration that resulted in the glycosylation site being buried. In water, during the initial 10 ns some of these glycosylation sites especially Asn$^{611}$ and Asn$^{616}$ were observed become solvent accessible.

The PFI was extended using SMD and was distance between the N-terminal of NHR and the C-terminal of the CHR was at least 80 Å in the extended conformation. The NHR and CHR are approximately 50 Å each and there was minimal loss in the interactions between residues of the CHR and NHR near the hydrophobic cavity. Lys$^{628}$ interactions with the residues in the hydrophobic cavity were still intact. Since, secondary structure restraints were used the NHRs and the CHRs are intact α-helices. The helical secondary structure was completely retained even after the 10 ns of MD run.

During the 10 ns run the CHR α-helices were observed to “flip” and initiate folding back into a stable six-helix bundle. The interactions between Trp$^{628}$ and Trp$^{631}$ and the residues in the hydrophobic cavity were sufficient to initiate formation of 6HB. However, classical molecular dynamics is likely insufficient to observe complete folding of PFI into 6HB. Therefore, we used TMD to simulate complete folding of PFI into 6HB.

6.3.2 Folding of pre-fusion intermediate into six-helix bundle using TMD

TMD run was started from the final configuration of the molecular dynamics simulation of the PFI. Generally TMD drives the system from an initial conformation to a target conformation linearly. The equilibrated six-helix bundle was provided as the target conformation. Even during the TMD run the the CHRs were observed to maintain intact α-helical conformation. The RMSD profile during the TMD run is
provided in Fig. 6.5. The target RMSD varies evolves during the simulation while the current RMSD is the instantaneous RMSD that follows target RMSD depending upon the spring constant. Since, we used a “stiff” spring constant of 200 mol$^{-1}$ Å$^{-2}$, the current RMSD followed the target RMSD closely.

During the TMD run the secondary structure restraints and the spatial constraints on the NHRs were removed. However, the trimeric NHR core was observed to be intact throughout the TMD simulation. In spite of the rigid spring constant used in the TMD run the NHRs were observed to kink while residues Gln$^{562}$ and Gln$^{563}$ and the hydrogen bonding maintain the NHR core to facilitate interactions of Trp$^{628}$, Trp$^{631}$ and Ile$^{635}$ with the residues in the hydrophobic cavity and the subsequent folding of CHRs into the hydrophobic grooves.

### 6.4 Discussion

The intent in this chapter is to model build a complete 6HB from the available structural information, model build a pre-fusion intermediate from the 6HB using SMD and existing knowledge of the PFI and to test the ability of molecular simulation techniques to fold PFI into 6HB. The available 6HB structure contains the $\alpha$-helical NHR and CHR subdomains. Therefore, we used the available sequence of HIV-1 HXB2R to model build the flexible linker connecting the NHR and the CHR. The 10 ns MD simulation of the 6HB structure was highly stable with the observed fluctuations dominated by the flexible linker. In the modeled linker prior to the water-solvated simulations the glycosylation sites at Asn$^{611}$ were Asn$^{616}$ buried in the bulky linker. During the 10 ns run it was observed that the 4 of the 6 residues were solvent accessible. Following the MD we used SMD to extend the ectodomain and model the PFI.

The existing knowledge about influence (HA) transmembrane protein ectodomain
structure in its low-pH conformation where the fusion peptide moves approximately 100 Å is used to model the PFI of HIV-1 gp41. The distance between the C-terminus of the CHR and the N-terminus of the NHR without the 33 amino acid long fusion peptide was 80 Å. The obtained PFI structure retained the important interactions between Trp$^{628}$ and the residues in the hydrophobic cavity in the NHR core. These interactions resulted in the PFI partially folding back towards a 6HB in a 10 ns MD simulation. To accelerate folding into 6HB we used TMD.

TMD successfully facilitated folding of PFI into 6HB by linearly driving the system towards the target 6HB structure. Since, the restraints on the system were removed the ability of the NHRs and CHRs to kink in order to fold into a stable 6HB was observed. Also, the importance of the highly conserved glutamine residues Gln$^{562}$ and Gln$^{563}$ to hold the NHR trimer was observed. Previous experimental point mutations of these residues resulted in a complete loss of viral infectivity. The final conformation restored the important structural features of 6HB, Trp$^{628}$, Trp$^{631}$ and Ile$^{634}$ on the CHRs collapsed into the hydrophobic pocket in the NHR trimer. The salt-bridge between Lys$^{574}$ and Asp$^{632}$ was also restored. These interactions between residues close to the flexible linker were also found to be important in the formation of 6HB and viral infectivity. Further away from the linker hydrophobic interactions between Ile$^{559}$ in the NHR and Ile$^{646}$ and Glu$^{647}$ on the CHR that are important for the stability of 6HB were also restored.

In spite of successfully modeling and folding PFI into a stable 6HB it should be noted that the system is partially inconsistent with existing literature about native spike and PFI. There are two competing hypotheses about the structure of gp41 in the native spike based on spike structure obtained from cryo-electron tomography images [108, 114–116]. Liu et. al suggested that the spike is connected to the viral membrane in the form of a “stalk” that is approximately 35 Å wide [108]. In our PFI
model the CHRs were at a distance between 38 to 41 Å from each other. Moreover, the spike was observed to be in the form of a mushroom and the flexible linker was thought to protrude to the side of the spike while the linker in our model would essentially be hidden in the electron-tomography structure. Conversely, Zhu et. al suggested that the base of the native spike is splayed apart in the form of a tripod that compares well with our model of CHRs [115, 116]. A better model would likely position the CHRs in the grooves between the flexible linkers.

Moreover, synthetic peptides prepared to replicate CHRs were observed to be unstructured in isolation in water experimentally. Monoclonal antibodies could only bind to CHRs following formation of 6HB suggesting conformational changes during the formation of 6HB. Therefore, it was hypothesized that CHRs are unstructured in water and require interactions with NHRs to obtain the α-helical conformation. However, in our TMD simulation the CHR retained its α-helical conformation even in the extended conformation and while folding into a 6HB. Apart from the relatively short simulation time, the high spring constant used during TMD likely restricted the simulation from sampling conformations farther away from the 6HB. Nevertheless, these simulation proved that TMD can be successfully used to simulate folding of PFI to 6HB and can be extended to study the the effect of the formation of 6HB on viral and target-cell membranes.

6.5 Conclusions

In this chapter, we successfully model built a complete six-helix bundle and pre-fusion intermediates of the HIV-1 gp41 transmembrane proteins. The models were then used to study the interactions and mechanism of the conformational changes from 6HB to PFI using TMD. Although some topological features of the models are at a variance with existing literature, the specific interactions between the residues
important for the formation and stability of 6HB were consistent with present knowledge. Furthermore, TMD was successfully used to simulate the conformational change from PFI to 6HB and the pros and cons of the simulation were discussed.
Figure 6.1: A schematic cross-section of the model build six-helix bundle. The NHR (blue) and CHR (red) structure were obtained from available crystallographic structural information (PDB). The sequence of the linker (grey) was used to model build the complete structure.
Figure 6.2: A schematic cross-sections of the 6HB representing the hydrophobic groove (white) and the deep hydrophobic cavity in the NHR trimeric core. Trp^{628} can be seen to interact with residues in the hydrophobic cavity.
Figure 6.3: A schematic cross-section of the model build pre-fusion intermediate. The PFI is model built using the 6HB structure.
Figure 6.4: The RMSD of the 6HB (blue), RMSD of the NHR/CHR domains (green) and the RMSD of the linker (red) during 10 ns of MD run. The NHR/CHR domains can be observed to be extremely stable.

Figure 6.5: Target RMSD (red) and instantaneous RMSD (green) of the TMD atoms during the course of 10 ns TMD simulations. The stiff spring constant of $200 \text{ kcal mol}^{-1}\AA^{-2}$ resulted in the instantaneous RMSD closely following the target RMSD.
Chapter 7: Findings and Future Directions

7.1 Findings

Surface envelope glycoproteins of viruses have always been a major source of vaccines and targets for drug molecules. With research now being focused to develop cheap and potent vaccines and drug molecules to help prevent and cure the global epidemic of AIDS, there is a renewed drive to understand the structure and function of the HIV-1 surface envelope glycoproteins, gp120 and gp41. Although molecular simulations are an excellent tool to study, visualize and analyze protein structure, dynamics and function, the absence of complete crystallographic or NMR structural information of gp41 resulted in inadequate simulation studies of the protein. In this work, we studied the HIV-1 transmembrane protein (gp41), specifically two highly conserved regions of gp41 (ectodomain and membrane-spanning domain), and we also tried to understand its natural preference for cholesterol-rich domains in the cell membranes to enter as well as derive its own envelope. The main goals of this work are therefore, (i) to study the forces and free-energies of rupture of cholesterol-containing bilayers, (ii) to explore the structure-function relationship of the membrane-spanning domain of HIV-1 gp41 destabilization of lipid membranes and (iii) to simulate the folding of a pre-fusion intermediate of gp41 into a stable six-helix bundle. In order to circumvent the absence of structural information of these systems we model built (i) the membrane-spanning domain using the available sequence and the $\alpha$-helical structure of the viral protein Vpu, (ii) the complete 6HB with the flexible linker from
the available linker sequence and the ectodomain structure (PDB access code: 1AIK) and (iii) the PFI from the 6HB.

In Chapter 3, we computed the structural and dynamical effects of cholesterol on lipid membrane properties and validated our results with existing knowledge. We further calculated the forces and free-energies of rupture of pure POPC and POPC/cholesterol membranes but simulating mechanical insertion of a pristine carbon nanotube into a lipid bilayer. We found that while the force of insertion is similar for both pure POPC and POPC/cholesterol bilayers the free-energy of insertion is significantly higher for cholesterol-containing bilayers. Since, cholesterol is known to increase the rigidity of the bilayers by straightening the lipid tails the increased free-energy is likely the results of increased interactions between the lipid tails and the nanotube. Considering these observations, it can be hypothesized that cholesterol-containing lipid bilayers increase the membrane anchoring of non-polar structures. The higher free-energy required for destabilization also likely explains the formation of oligomers of fusion protein in viruses. Moreover, it was also found that higher-cholesterol content in membranes resulted in higher mechanical destabilization and likely a larger fusion pore.

In Chapters 4 and 5, we explored the structure of the MSD in water and cholesterol-containing lipid membranes. We found that, in water the MSD exits in the form of a stable kinked helix with the kink at the GXXXG motif. However, it can shuttle between the stable kinked structure and a metastable loop structure with a second kink at Phe697 across a low energy barrier and that it is highly metastable as a complete \(\alpha\)-helix. In a lipid membranes, we used two models based on the snorkeling direction (Exoplasmic or Endoplasmic) of the mid-span Arg694 side-chain and a third model with Arg694\(\rightarrow\)Leu point mutation to replace Arg694 which abrogates infectivity. In all the models, we found that the MSD assumes a stable completely \(\alpha\)-helical con-
formations and a metastable kinked conformations. In the endoplasmic snorkeling model, which is likely the naturally occurring direction of the Arg\textsuperscript{694} sidechain, the metastable state observed in our simulations compared well with existing hypothesis based on experimental observations. Moreover, we also found that in the presence of mid-span Arg\textsuperscript{694} the MSD significantly decreases membrane thickness in the proximity of the peptide in both the stable and metastable conformations. Mutation of Arg\textsuperscript{694} into a relatively hydrophobic Leu resulted in minimal effect of the MSD on the lipid bilayer thickness. These observations likely explain the presence of a polar residue in the hydrophobic environments of the viral membrane core and the loss of viral infectivity upon mutation.

In Chapter 6, we model built a complete 6HB and PFI and studied the folding of PFI into 6HB. These simulations suggested the importance of the presence of highly conserved residues in the gp41 ectodomain. Although, the models provided insights into the structure and folding of 6HB, it was concluded that the models can be improved to better represent current knowledge of gp41 in the native spike and the pre-fusion intermediate.

At a fundamental level, membrane fusion requires two steps, (i) opposing membranes are brought into close proximity and (ii) membranes are destabilized to facilitate lipid mixing. We used model built structures to study the structure-function relationship of HIV-1 gp41 during membrane fusion using all-atom simulations. HIV-1 consists of a viral envelope rich in cholesterol. The presence of cholesterol increases the rigidity of lipid membranes that is likely required to maintain the high internal pressure of a mature virion. In this work we showed that the presence of a mid-span Arg\textsuperscript{694} results in significant thinning of the lipid membrane and penetration of water molecules into the membrane core thereby weakening the membrane and priming it for fusion.
7.2 Future directions

We propose the following suggestions to motivate further research in protein-membrane interactions and HIV-1 gp41 structure and function in membrane fusion:

(i) Biological membranes are complex macromolecular structures with a variety of lipid molecules that influence membrane and protein structure and function. With force field parameters available for a variety of lipids and becoming increasing accurate complex bilayers can be used to study both the structure of proteins in a relatively realistic membranes and also the role each of these lipids play in the membrane and protein function.

(ii) The MSD is a highly conserved peptide of HIV-1 gp41 and contains a GXXXG motif and charged residues at Lys$^{681}$, Arg$^{705}$ and Arg$^{707}$. Each of these residues were found to be important for viral infectivity from recovery-of-function and point mutations. Studying the role of these residues in MSD and membrane structure by simulating point mutations can offer excellent insight into their role in infectivity.

(iii) Understanding the mechanism by which various other fusion proteins facilitate membrane fusion is also important. For example, SNARE proteins also consist of a transmembrane domain that anchors the protein to both the vesicle and the target membranes. However, a midspan charged residue that weakens the membranes is absent. Therefore, understanding the mechanism by which various fusion proteins destabilize lipid membranes using both experimental and computational methods is important.

(iv) The CHRs were observed to be $\alpha$-helical in our simulations in both the 6HB and PFI. However, previous experimental results suggested that the CHRs are unstructured when isolated form the NHR trimer. The structure of CHR can therefore be studied and the knowledge can be used to understand the structure of the viral spike, the mechanism of formation of 6HB and to build better models of PFI.
(v) Model built structures of 6HB and PFI can be fit to cryo-electron tomography images and improved by comparing them with the unknown densities.

(vi) The role of the linker in membrane fusion is still not completely studied. It was observed from previous experimental studies that the entire flexible linker is not essential for the formation of a stable 6HB. However, the role of the linker in membrane destabilization and fusion is still not completely understood. Further studies, both experimental and computational, are likely required to study the role of the flexible linker and its sequence in membrane fusion.

(vii) Finally, the model build structures of 6HB and PFI although not completely accurate can be used to understand their role in membrane destabilization. Ideally, a complete viral membrane to target-cell membrane system with gp41 bridging the two membranes in the PFI state can be used as a starting state and the effects of formation of 6HB on the MSD, fusion peptide, the viral and target-cell membrane could be studied. However, the system will be extremely large and computationally expensive and hard to simulate.


Gallo, S., C. Finnegan, M. Viard, Y. Raviv, A. Dimitrov, S. Rawat, A. Puri,


ensemble and folding mechanism of the C-terminal GB1 beta-hairpin. *JACS* 130:13938.


Vamshi K. Gangupomu was born in Ichchapuram, India, on Aug 14, 1981 and grew up in Visakhapatnam, India. Vamshi received his Bachelor of Technology in Chemical Engineering from the Maharajah Vijayram Gajapathiraj College of Engineering of the Jawaharlal Nehru Technological University, Hyderabad in 2002. He then attended Bucknell University, Lewisburg, Pennsylvania and received his Master of Science in Chemical Engineering in 2004. Vamshi worked as a process and validation engineering consultant at Wyeth Pharmaceuticals from 2004 to 2005. He joined the PhD program in the Department of Chemical and Biological Engineering at Drexel University in Fall 2005, where he worked as a graduate student under Prof. Franco M. Capaldi (MEM) and Prof. Cameron F. Abrams (CBE). During his graduate study, Vamshi was a teaching assistant for 5 years. He plans to pursue post-doctoral training to gain a broader expertise and perspective in the field of molecular simulations.