3-22-2017

Molecular Dynamics Investigations of Structural Conversions in Transformer Proteins

Jeevan GC
Florida International University, jgc001@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

MOLECULAR DYNAMICS INVESTIGATIONS OF STRUCTURAL CONVERSIONS

IN TRANSFORMER PROTEINS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Jeevan B. GC

2017

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

This dissertation, written by Jeevan B. GC, and entitled Molecular Dynamics Investigations of Structural Conversions in Transformer Proteins, having been approved in respect to style and intellectual content, is referred to you for judgment.

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

_____
Jessica Liberles

_____
Jin He

_____
Bernard S. Gerstman, Co-Major Professor

_____
Prem P. Chapagain, Co-Major Professor

Date of Defense: March 22, 2017

The dissertation of Jeevan B. GC is approved.

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

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

Florida International University, 2017

ACKNOWLEDGMENTS

ABSTRACT OF THE DISSERTATION

MOLECULAR DYNAMICS INVESTIGATIONS OF STRUCTURAL CONVERSIONS

IN TRANSFORMER PROTEINS

by

Jeevan B. GC

Florida International University, 2017

Miami, Florida

Professor Bernard S. Gerstman, Co-Major Professor

Professor Prem P. Chapagain, Co-Major Professor

Multifunctional proteins that undergo major structural changes to perform different functions are known as "Transformer Proteins", which is a recently identified class of proteins. One such protein that shows a remarkable structural plasticity and has two distinct functions is the transcription antiterminator, RfaH. Depending on the interactions between its N-terminal domain and its C-terminal domain, the RfaH CTD exists as either an all-α-helix bundle or all-β-barrel structure. Another example of a transformer protein is the Ebola virus protein VP40 (eVP40), which exists in different conformations and oligomeric states (dimer, hexamer, and octamer), depending on the required function.

I performed Molecular Dynamics (MD) computations to investigate the structural conversion of RfaH-CTD from its all-α to all-β form. I used various structural and statistical mechanics tools to identify important residues involved in controlling the conformational changes. In the full-length RfaH, the interdomain interactions were found to present the major barrier in the structural conversion of RfaH-CTD from all-α to all-β

form. I mapped the energy landscape for the conformational changes by calculating the potential of mean force using the Adaptive Biasing Force and Jarzynski Equality methods. Similarly, the interdomain salt-bridges in the eVP40 protomer were found to play a critical role in domain association and plasma membrane (PM) assembly. This molecular dynamic simulation study is supported by virus like particle budding assays investigated by using live cell imaging that highlighted the important role of these saltbridges. I also investigated the plasma membrane association of the eVP40 dimer in various PM compositions and found that the eVP40 dimer readily associates with the PM containing POPS and PIP_2 lipids. Also, the CTD helices were observed to be important in stabilizing the dimer-membrane complex. Coarse-grained MD simulations of the eVP40 hexamer and PM system revealed that the hexamer enhances the PIP_2 lipid clustering at the lower leaflet of the PM. These results provide insight on the critical steps in the Ebola virus life cycle.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

ABF	Adaptive Biasing Force
AMD	Accelerated Molecular Dynamics
CHARMM	Chemistry at Harvard Molecular Mechanics
CGMD	Coarse Grained Molecular Dynamics
CTD	Carboxy Terminal Domain
EEF1	Effective energy function-1
GROMACS	Gronigen Machine for Chemical Simulations
JE	Jarzynski Equality
LJ	Lennard-Jones
MD	Molecular Dynamics
MMTSB	Multiscale Modeling Tools for Structural Biology
NAMD	Not Another Molecular Dynamics
NMR	Nuclear Magnetic Resonance
NTD	Amino Terminal Domain
PDB	Protein Data Bank
PM	Plasma Membrane
PMF	Potential of Mean Force
REMD	Replica Exchange Molecular Dynamics
SASA	Solvent Accessible Surface Area

1. INTRODUCTION

1.1 Transformer proteins

Proteins are molecular machines that provide the high physical organization and low entropy required for life. Proteins perform a myriad of biophysical tasks in cells such as molecular transportation, enzymatic activities, energy transfer, and signaling. In order to perform their physiological functions, most proteins have to fold to their corresponding native state structures, which are highly specific three-dimensional conformations (1). Depending on the biochemical environment, some proteins are observed to have multiple functions. A protein that exhibits multiple functions is categorized as an interesting subset of moonlighting proteins (2). In performing alternate functions, a moonlighting protein can utilize different protein interfaces (3) of the same configuration. However, a recent development has added a new twist to the relationship between protein structure and function.

Protein conformational dynamics in the functional state includes local, small-scale flexibility observed in sidechains, loops, or at a binding site, as well as larger, global conformational rearrangements such as domain separation or oligomerization (4). Intrinsically unstructured proteins may have sufficient structural flexibility to interact with multiple partners, whereas globular proteins can undergo structural changes (5) that allow them to perform different functions. Globular proteins such as lymphotactin (6) and Mad2 (7) undergo small structural rearrangements that lead to different functions and are referred to as metamorphic proteins (8).

Recently, a new class of multifunctional proteins has been identified as "transformer proteins" (9). Transformer proteins show a remarkable ability to rearrange their structure to significantly different conformations that perform completely different functions. An example of such a protein is the transcription antiterminator RfaH. Depending on the interactions between its N-terminal domain (10) and its C-terminal domain (10), the CTD of RfaH exists as either an all-α-helix bundle or all-β-barrel structure. The RfaH-CTD in the all-α topology is involved in regulating transcription, whereas in the all-β topology it is involved in stimulating translation by recruiting a ribosome to an mRNA (11). Another example of a transformer protein is the Ebola virus protein VP40, which exists in different conformations depending on the required function: a butterfly shaped dimer is involved in the transport of the protein to a membrane, a hexamer to form the viral matrix, and an octamer ring structure to bind to RNA and regulate viral transcription (12).

I performed Molecular Dynamics (13) computations to investigate the dynamic behavior of transformer proteins RfaH and Ebola virus protein VP40 and identified the interaction mechanisms and pathways of conformational changes in these proteins. The large scale structural changes in RfaH-CTD were observed using various enhanced sampling simulation methods as described in Chapter 2. The MD studies performed on RfaH and the Ebola virus protein VP40 highlighted the important salt-bridges that play crucial roles in domain-domain association. My research also involves the study of the interactions between the protein and a membrane. I investigated the association of the VP40 dimer with the plasma membrane using all-atom molecular dynamics simulations as well as coarse-grained simulations. Through all-atom simulations, I have investigated

the mechanisms of plasma membrane localizations of VP40 and identified residues that are important for stabilizing the VP40 dimer at the lower leaflet of the plasma membrane. The coarse grained simulations allowed me to investigate the dynamics of lipids and proteins for the larger hexamer-membrane system for much longer time-scales. This study revealed an enhanced clustering of a specific lipid type (14) in the lower leaflet of the plasma membrane in the presence of a VP40 hexamer.

1.1.1 RfaH Transcription Factor

I studied the RfaH protein because it is one of the smaller proteins that exhibit transformer properties. A recent study (11) on the two-domain transcription factor RfaH revealed that its function is very different depending on whether its CTD is in an α-helix or β-sheet conformation (15). Magnetic resonance (13) measurements revealed that the C-terminal domain (CTD) of the transcription anti-terminator RfaH (PDB-ID: 2OUG) exists as an α-helix bundle when it interacts with the N-terminal domain (NTD) of the same protein, but folds to a five-stranded all-β conformation (PDB-ID: 2LCL) when it is not interacting with the NTD. These two structures are shown in figure 1.1. The RfaH protein conformation is closely related to another two-domain transcription factor NusG(16). Although the NTD conformations of both RfaH and NusG are similar, the CTD conformations differ significantly. Unlike the CTD of RafH that has an all-α native fold when it interacts with its NTD, the CTD of NusG (PDB-ID: 2JVV) has an all-β fold. The similar NTD conformations of both RfaH and NusG have identical roles in RNA Polymerase (RNAP) binding. However, the existence of two different conformations of RfaH-CTD allows an additional function for this domain. The RfaH-CTD in the all-α

topology is involved in transcription regulation by masking the RNAP binding site in the absence of a DNA recruitment signal, whereas in the all-β topology it is involved in stimulating translation by recruiting a ribosome to an mRNA that lacks a ribosome-binding site. Just as in NusG, binding of RfaH to the RNAP and protein S10 occurs when the CTD is in its β-form, which requires breaking of the NTD and CTD contacts. The domain separation is believed to be triggered by RfaH interactions with an *ops* element (17), (18) which then leads to spontaneous refolding of the CTD into the β-barrel. Figure 1.1 shows the structural conversion of the RfaH-CTD. The N-terminal domain (NTD) is colored in green and the C-terminal domain (CTD) is colored in yellow.
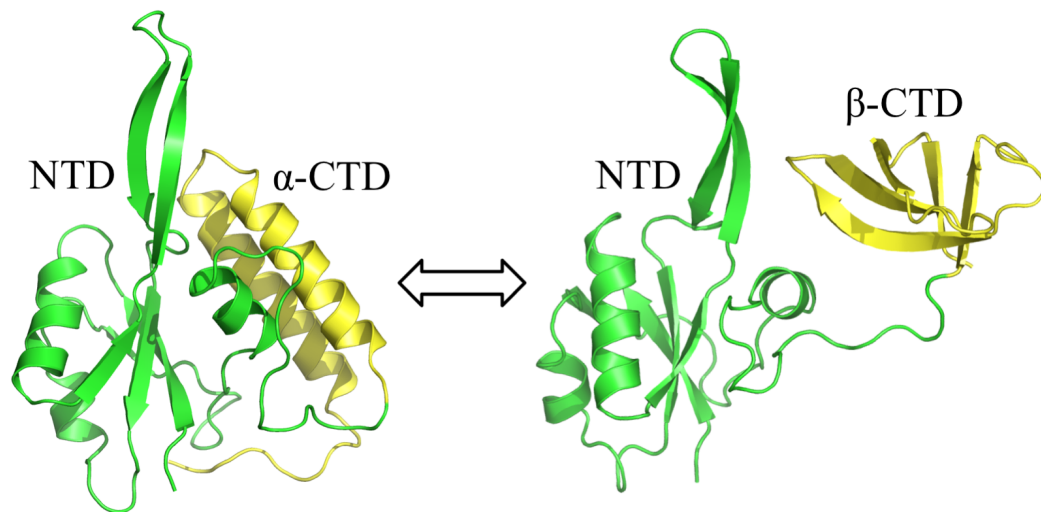
Figure 1.1 Structural conversion of RfaH-CTD. The N-terminal domain (NTD) is colored in green and C-terminal domain (CTD) is colored in yellow.

The conformational transformations in RfaH-CTD that allow it to perform different functions places it in the category of a transformer protein (TFP) (19), a class which may eventually include many other functional proteins yet to be identified. The details of the folding mechanism and structural transitions in RfaH are of great interest. It

has recently been shown (13) that after denaturation, the inter-domain contacts with the NTD are sufficient to fold the CTD into the α-helical form from the denatured state. Further, when only the CTD is expressed, it folds completely to a β-barrel structure. Interestingly, when the domains are swapped to reverse the order of NTD and CTD translation, functional activity of the α-helical form is still observed *in-vitro*, but with the observed activity being significantly reduced, suggesting the possibility that a fraction of chains formed β-structure(13).

The α-helix to β-sheet structural transition in the CTD of RfaH is an example of a biophysical process that is frequently observed in the aggregation and fibrilization of amyloidogenic proteins. Such protein structural transitions are thought to be critical processes in the development of a variety of neurodegenerative diseases such as Alzheimer's, Parkinson, Huntington, and prion diseases, which are considered to be the result of protein misfolding (20). The amyloid fibrils that are implicated in these diseases are composed of proteins that transform from innocuous, soluble configurations such as α-helices, into β-strands that aggregate into sheets (21), (22). A well studied example (23), (24) is the structural conversion of the native α-helical form of the prion protein PrP^C into the pathogenic β-sheet-rich form PrP^{Sc} that causes transmissible spongiform encephalopathy (scrapie). Recent investigations have shown that the α-to-β conversion in PrP can occur even in the absence of the infectious molecular species (25). This α-to-β conversion also holds true for the structural transitions of the amyloid beta (Aβ) precursor peptide system, which is responsible for amyloid deposits in Alzheimer's disease (26). Although such peptides can undergo structural transitions, the presence of an oligomer seed can accelerate the oligomerization and fibril formation processes (27), (28). Another

example of spontaneous structural conversion occurs in a group of cc-beta peptides, which form coiled-coil helical structures at low temperatures but convert to β-sheet rich amyloid fibrils at elevated temperatures (29), (30). In addition to the role played in many neurological disorders (31), the α-helix to β-sheet transition in amyloidogenic peptides is of great interest for the engineering of novel biomaterials from synthetic peptides (32).

1.1.2 Ebola virus protein VP40

The Ebola virus is a lipid-enveloped virus that causes hemorrhagic fever with a high mortality rate. The virus uses its negative sense RNA genome to replicate its seven genes in the host cell. To overcome the limitations posed by its small genome, the Ebola virus utilizes proteins with astonishing abilities to change structures and perform multiple functions (33). The Ebola virus protein VP40 shows remarkable conformational plasticity in forming various oligomeric structures that perform different functions and therefore is a transformer protein. The Ebola virus protein VP40 consists of an N-terminal domain (NTD) that is involved in dimerization and a C-terminal domain (CTD) that mediates membrane binding and oligomerization (12). It exists in various conformations depending on the required function: a butterfly shaped dimer is involved in the transport of the protein to a membrane, a hexamer to form the viral matrix, and an octamer ring structure to bind to RNA and regulate viral transcription (12). As with other recently identified transformer proteins such as the RfaH transcription factor (11), (13), (34) the structural basis and mechanisms of large-scale structural transformation in VP40 are not well understood.

6

The Ebola virus protein VP40 is a peripheral protein that localizes in the inner leaflet of the plasma membrane (PM) (35) and forms the viral matrix which provides the major structure for the virus-like particles (VLP). Formation of the matrix from the assembly and oligomerization of VP40 dimers involves a series of steps. First, a VP40 dimer is trafficked to the plasma membrane where electrostatic interactions occur between the anionic lipids and the basic patch of the VP40 C-terminal domain (CTD). This electrostatic interactions was determined with experiments in which deletion of several lysine residues in CTD loops prevented VLP formation, even though the dimeric structure was unaltered. Once at the plasma membrane, it has been hypothesized that the VP40 dimers undergo major structural rearrangements and oligomerize into hexameric structures (12), (36) which oligomerize to form filaments that associate further to form the viral matrix (37). Although the x-ray crystal structures of the dimeric and hexameric forms provide information about structural transformations that must occur, the mechanisms of VP40-membrane interactions leading to the structural transformations from dimers to hexamers and hexamers to larger filamentous structure are not well known.

Figure 1.2 presents the structures of the VP40 dimer, hexamer and octamer. Because of its low molecular weight, the dimer is trafficked to the lower leaflet of the plasma membrane, whereas the hexamer is the major structural oligomer and the octamer binds RNA to control transcription.
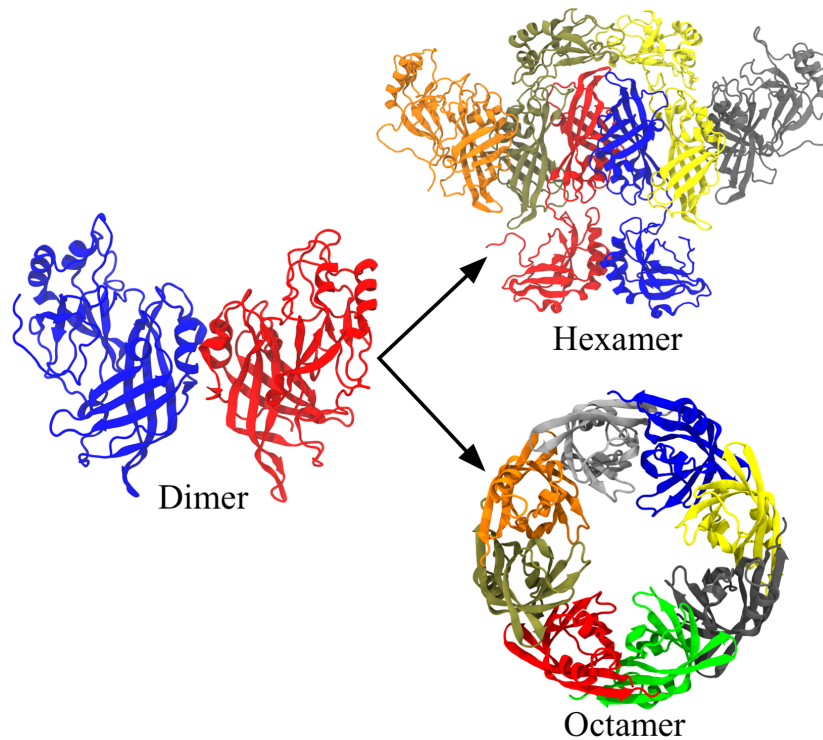
Figure 1.2 Different Structures of VP40.

1.2 Molecular Dynamics computer simulation

Most of the work reported here used molecular dynamics (MD) (13) computer simulations to uncover molecular biophysics principles concerning the structural transformations. The MD method evolved from theoretical physics in the late 1950s and is now used in a variety of fields, including drug discovery (38). The MD computer simulations use basic laws of physics. A biomolecule such as a protein is treated as a long polymer chain consisting of many atoms that are represented as classical particles. The forces between interacting atoms and the overall energy of the system is calculated by a force field. The equation below represents the overall energy of the system.

$$U_{\text{total}} = \sum_{\text{bonds}} K_b \left(r - r_0\right)^2 + \sum_{\text{angles}} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} K_\varphi \left[1 + \cos\left(n\varphi - \varphi_0\right)\right]$$

$$+ \sum_{\substack{\text{van der Waals} \\ i,j \text{ pairs}}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \sum_{\substack{\text{electrostatic} \\ i,j \text{ pairs}}} \frac{q_i q_j}{\varepsilon r_{ij}}$$

(1.2.1)

Figure 1.3 below shows various interactions used in MD simulations. A chemical bond between two adjacent atoms is modeled as a virtual spring with a harmonic potential. The angular term measures the energy from the angles between each pair of covalent bonds sharing the same atom at the vertex. A dihedral term describes a torsional angle φ of a central bond connected to two covalent bonds. For the four covalently bonded atoms i,j,k,l, the torsional angle is the angle between plane i,j,k and j,k,l. This interaction is also periodic in nature. The geometry or chirality of four planar covalently bonded atoms can defined by "improper terms" that can be added to eq. (1.2.1), and are also harmonic in nature. The non-bonded interaction consists of Van der Waal's forces defined by Lenard Jones 6-12 potentials and the pairwise electrostatic interaction given by Coulomb's Law with the screening constant ε. For a system of N atoms there are $N^2/2$ non-bonded distinct atom pairs, with most of the computation time in MD used in calculating these interactions. Therefore these interactions are truncated using switching function in MD simulations that smoothly cutoff the non-bonded interactions at user chosen cutoff distances. Typically, non-bonded interactions between atom pairs beyond a cutoff distance of 12Å are set to zero.
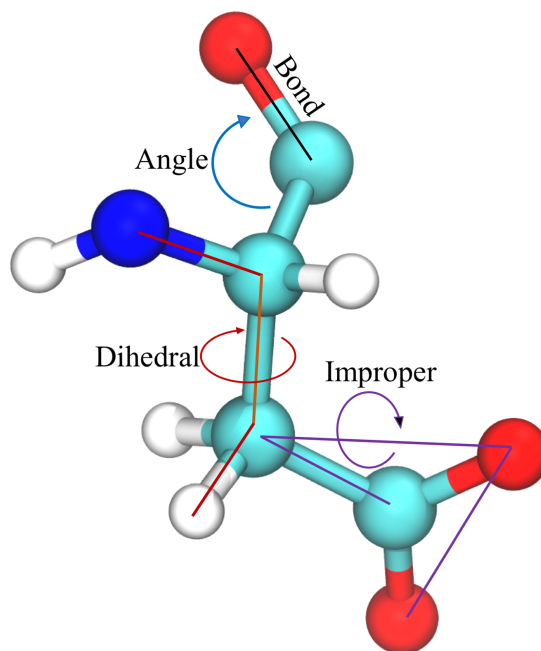
Figure 1.3 Illustration of all bonded terms of potential energy used in MD

The initial position of a biomolecule is assigned from the x-ray crystallographic structure, NMR or homology modelling. The initial velocity is assigned using Maxwell Boltzmann distribution of velocities at certain temperature T. Then the force is calculated using the gradient of potential energy function. The position and velocities are updated in a short time step and are saved. The position and velocity of each atom at each time step is the information contained in a file called the MD trajectory. The MD trajectory is post processed using concepts of statistical physics to calculate a variety of thermodynamic and kinetic properties to understand the dynamics of complex biomolecules, which can be compared with the experimental observables.

The MD time steps are on the order of a femtosecond, but the large scale structural changes in biomolecules occurs in time scales from millisecond, seconds, or hours. Conventional MD is unfeasible to apply for these long time scales because of

computer limitations. Enhanced sampling simulation methods like Steered MD, Targeted MD, Coarse Grained MD etc. close this gap by using statistical physics algorithms. Also, the use of GPUs has significantly accelerated the MD. The current study employed most of these techniques to study protein folding, large structural transformations, and membrane protein interactions.

2. METHODS

2.1 Replica Exchange Molecular Dynamics (REMD)

In order to explore the protein configurational space, I performed replica exchange molecular dynamics (REMD) (39) simulations using the MMTSB toolset (40). To overcome the multiple minima problem, REMD is considered as an efficient and powerful sampling technique. In the REMD method, a number of replicas of the system, each at a different temperature, are simultaneously simulated independently for a certain number of time steps, τ. After τ time steps, a pair of replicas i and j at different temperatures, T_i and T_j, that have evolved into different conformations are tested for a temperature swap. The probability for temperature swapping is given by a Metropolis type criterion. In this criterion, the probability of exchange is given by $P_{ij} = \min\{1, \exp(-\Delta)\}$, where $\Delta = (E_j - E_i) \times (1/kT_i - 1/kT_j)$ and E_i is the potential energy of the replica i. The swapping criterion satisfies the condition of detailed balance and conformations generated at each temperature are a Boltzmann distributed canonical ensemble. As a result of this swapping, a random walk in temperature space is performed for each replica, which enhances the replica's ability to explore the protein's configuration space. During stages when its temperature is high, it is more easily able to exit from configurational kinetic traps and globally sample configuration space, and when its temperature is low a replica can explore the details of the local region of configuration space. Although the time sequence of configurations for any replica is not realistic because of periodic changes in temperature and information on kinetics is lost, the simulations contain information that can be used in calculations of various thermodynamic quantities.

Before commencing the replica exchange procedure, I started the MD simulation of the α-helix bundle of RfaH-CTD by performing 20 ps of equilibration with a time step of 2 fs. The system was confined by a spherical quartic potential with a radius of 60 Å, which is large enough to accommodate fully extended helices. I then created 16 replicas with exponentially spaced temperatures between 300 and 500 K. A 50-ns production run was obtained using the implicit solvent EEF1 force field and a 2 fs integration time step, with a replica swap attempted every 500 steps (1 ps). The atomic dynamics were propagated via the Langevin method with a reduced friction coefficient of 1 ps^{-1}. The trajectories for each replica were used to calculate the root-mean-square deviation (rmsd) from the initial PDB structure, the radius of gyration (rgyr), as well as the number of hydrogen bonds. VMD (41) was used to visualize and further analyze the trajectories.

The configurational trajectories of all 16 replicas were examined and the structure closest to the experimentally reported all-β CTD configuration was extracted from one of the replicas and was relaxed for 50 ns using explicit solvent all-atom MD simulations. This detailed, explicit solvent all-atom simulation was performed with the NAMD molecular dynamics package (42) using the CHARMM27 force field (43). The structure was solvated with the TIP3P water model by using the *solvate* plugin in VMD. With a box cut-off of 10 Å away from the protein, the dimensions of the simulation box were 51 x 45 x 43 Å^3. The solvated system was electrically neutral. The final system contained a total of 10,246 atoms. The particle mesh Ewald method (44) was used to treat long range interactions with a 12 Å non-bonded cutoff. Energy minimization was performed using the conjugate gradient and line search algorithm. The system was then heated with a linear gradient of 20 K/ps from 20K to 300K. At 300 K, the system was equilibrated for

10 ps, followed by a 50-ns production run with a 2 fs integration time step in the NVT (constant number, volume, and temperature) ensemble. Langevin dynamics with a damping constant of 1 ps^{-1} was used to maintain the temperature at 300 K.

2.2 Steered Molecular Dynamics Simulations (SMD)

In order to explore the all-α to all-β transformation of the CTD in the full RfaH, I performed Steered Molecular Dynamics (SMD) simulations [45], [42] in explicit water.

The full RfaH was solvated with TIP3P water in rectangular boxes. Pulling occurred along the x-axis. A cut-off of 10 Å was used for the y- and z-directions, whereas a large padding was added in the x-direction to allow enough space for protein extension due to pulling. The resulting box dimensions were 198×56×81 Å^3 with a total of 85,567 atoms. The system was neutralized by adding two chloride ions. The particle mesh Ewald method was used to treat the long-range interactions with a 12 Å non-bonded cut-off. Simulations were performed with the CHARMM27 force field using NAMD.

For SMD, the dummy atom attached to the C-terminal C$_\alpha$ SMD atom (residue L162) via a virtual spring (k=3 kcal/mol/Å^2) was pulled in the negative x-direction with a constant speed of 2 Å/ns while keeping the C$_\alpha$ atom of the N-terminal residue (M1) fixed. SMD was performed with a 1 fs time step and the steering forces were saved every 1 ps. Trajectories were obtained by alternating pulling and relaxation, i.e., 10 ns of SMD followed by 10 ns of relaxation. During relaxation, the extension was preserved by keeping the first C$_\alpha$ in the NTD and the final SMD C$_\alpha$ in the CTD fixed (i.e., setting the velocity of each to zero), but the protein's internal structure was allowed to relax after the forced stretching [46], [47]. The average work done by pulling was calculated using nine

independent simulations, and the potential of mean force (PMF) was calculated using the Jarzynski equality (48), using an ensemble average of the nine SMD trajectories. Although PMF calculations with ABF have been shown to represent the reaction more accurately than with constant velocity SMD (49), the SMD simulations give important insights into the barriers along the pulling reaction coordinates.

Steered molecular dynamics (SMD) simulations were also performed on Ebola virus protein VP40 protomer by pulling on the center-of-mass of the CTD (residues 202-326) along the negative x-axis. The system consisted of 61,320 atoms in a box of size $117 \times 84 \times 66 Å^3$. All C_α atoms were harmonically restrained during heating and equilibration. The dummy atom attached to the center-of-mass of the CTD via a virtual spring ($k=3$ kcal/(mol/ $Å^2$)) was pulled at a constant speed of 2 Å/ns with respect to the harmonically restrained NTD (residues 44−194).

2.3 Targeted Molecular Dynamics Simulations (TMD)

In order to further understand the all-α to all-β transformation of the CTD in the full RfaH, I performed Targeted Molecular Dynamics (TMD) simulations in explicit water. Targeted MD is especially useful for investigating conformational changes or large molecular motions that are otherwise unfeasible to access in reasonable computational times. Targeted MD method has been extensively used in various molecular systems that undergo large-scale conformational changes that are relevant to protein function. For example, Targeted MD was used to elucidate the ATP binding mechanism, and to determine the transition pathways between the open and closed states of GroEL (50). Similarly, Targeted MD has been used to investigate transitions between open and closed

states of ion channels and their gating mechanisms in various systems such as a nicotinic receptor (51), KcsA potassium channels (52), (53), mechanosensitive channels (54), and the AMPA receptor (55). More relevant to the transitions in RfaH, the dynamical transition between two distinct states in the Mad2 metamorphic protein was recently studied using a combination of conventional and Targeted MD and the structural characteristics as well as dynamic transition mechanisms were explored (56).

The full RfaH with the helical CTD conformation was solvated with TIP3P water and the system was electrically neutralized by adding counter ions using VMD (41). The resulting dimensions of the simulation box were $56 \times 54 \times 78$ Å^3 and the final system contained a total of 24,082 atoms. The MD simulations were performed with the CHARMM36 (57) force-field using NAMD (42). The particle mesh Ewald method (44) was used to treat the long-range interactions with a 12 Å non-bonded cut-off. Energy minimization was performed using the conjugate gradient and line search algorithm for 5 ps. The system was then heated with a linear gradient of 20 K/ps from 20 to 300 K. At 300 K, each system was equilibrated with a 2 ns NPT run using a 1 fs integration time step, followed by a 5 ns NVT run also using a 1 fs integration time step. Langevin dynamics with a damping constant of 1 ps^{-1} was used to maintain the temperature at 300 K. The RATTLE algorithm was used to constrain protein bonds involving hydrogen. SETTLE was used to maintain water geometry. Multiple time-stepping algorithms (1-2-4) were used to compute interactions between atoms. Interactions through covalent bonds were calculated at every time step, short range non-bonded interactions were calculated every other step, and long range electrostatic forces at every fourth step.

A 10 ns equilibration run was performed before commencing each targeted MD simulation. The target RfaH structure with the beta conformation of CTD was created using Modeller by connecting the RfaH-NTD and the β-CTD (PDB ID: 2LCL) with a linker. For α → β conversion, force was applied only on the N=777 heavy (non hydrogen) atoms of the helical CTD (residues 115-162). For β → α conversion, force was applied on N=1322 heavy atoms which include all of the CTD (residues 115-162) as well as an additional 30 residues of NTD that participate in interdomain interaction. The targeted MD calculation uses the potential energy function $U = \frac{k}{2N} [rmsd(t) - RMSD(t)]^2$, where rmsd(t) is the root mean squared deviation (rmsd) of the simulated structure with respect to the target structure and RMSD(t) is the assigned, target rmsd, which decrease linearly over time. A force constant per atom of k_{atom}=k/N=0.4 kcal/mol/Å^2 was used for all TMD simulations.

2.4 Coarse Grained Molecular Dynamics Simulations (CGMD)

To understand the roles of various phospholipids on VP40 hexamer binding in the PM, I performed coarse-grained molecular dynamics simulations (CGMD) on the hexamer-membrane system. While all-atom simulations are useful in understanding the temporal and spatial evolution of a system in the nanosecond timescales, they are computationally demanding for investigating physical properties such as lipid diffusion and clustering and protein-membrane interactions that occur on microsecond timescales. The CGMD (58), (59), (60), (61) model, which maps four heavy atoms on average to a single interactions site, provides an excellent alternative. The CGMD model simplifies

the complex energy landscape and enhances the kinetics of the system (61) and has been extensively used in recent investigations on membrane protein and lipid systems (62), (63), (64), (65). In one study (66), coarse-grained simulations were found to better represent the protein-membrane binding as compared to other extended-lipid protein-membrane models (67). The CG simulations of the highly asymmetric plasma membrane showed the clustering of the Glycolipids GM3 in the outer leaflet and PI(4,5)P$_2$ in the inner leaflet. The different types of lipid clustering affect the membrane bending differently and the correlation of such lipid clustering with the membrane curvature has important biological implications (68). For example, Basu et al. investigated the clustering of GM1 with or without PSM using CGMD (69). It was found that in the presence of PSM, the GM1 domains formed loosely bound clusters, whereas in the absence of PSM, they were more strongly bound in the clusters. Similarly, Koldso et al. showed that local enrichment of PI(4,5)P$_2$ in the inner leaflet can induce concave curvature (viewed from intracellular side) to the plasma membrane (68). This computaional result is directly relevant to our present work because such a membrane curvature is needed for the budding of the Ebola virus particle.

For CG simulations, I used the Martini force field in which approximately four heavy atoms are mapped onto a single interaction site (bead), except for the residues with aromatic side chains (59). Specifically, Ala and Gly residues are each represented by a single bead; Cys, Asn, Asp, Gln, Glu, Ile, Leu, Met, Pro, Ser, Thr, and Val are each represented by two beads; Arg and Lys are each represented by three beads; His, Phe, and Tyr are each represented by four beads, and Trp is represented by five beads. Similarly for lipids POPC, POPE and POPS each have 12 beads, PIP$_2$ (70) has 16 beads, PSM has

11 beads, and CHOL has 8 CG beads. These beads are divided into four interaction types: Polar (P), nonpolar (N), apolar (C), and charged (Q). There are subtypes d (donor in H-bond), a (acceptor in H-bond), n (none), da (both) and P1 (lower polarity)-P5 (highest polarity) (59).

The *Martini Maker* plugin (71) in the Charmm-Gui webserver was used to prepare both the protein-membrane and membrane-only systems. The Martini 2.0 force field was used for lipids with a non-polarizable water model, combined with the elastic network model for the protein (VP40 hexamer). All systems were equilibrated using the six-step equilibration process generated by the Charmm-Gui (72). Each system was solvated using standard MARTINI water beads and was neutralized in 0.15 M NaCl. The final protein-membrane system consisted of 52, 827 particles and the membrane-only system consisted of 27, 512 beads. The number of lipids used for the CGMD simulations is given in Table S1 (14). For one of the simulation setups, the elastic network constraints were removed for all the atoms in the top two CTDs in the middle by setting the spring constant to zero. Various simulations performed in this work are summarized in Table S2 (14).

All CGMD simulations were run using Gromacs5.1.1 (73), (74) (gpu-version). A time step of 15 fs was used and the Lennard-Jones and coulomb potential were cutoff at 11 Å. The Verlet neighbor scheme was used with a straight cutoff to keep track of particles within specified distances (75). Coulomb interactions were treated using a reaction-field. Ring systems and stiff bonds were controlled using the LINCS algorithm. The pressure was maintained at 1 bar using the Berendsen barostat during equilibration runs and the Parrinello-Rahman barostat during production runs. Pressure coupling

between protein and membrane groups was semi-isotropic with a compressibility of $3x10^{-4}$ bar^{-1}. The temperature was maintained at 303.15 K and the temperature coupling was done using a velocity rescale algorithm.

2.5 Accelerated MD (aMD)

Accelerated molecular dynamics (aMD) simulation is an enhanced sampling method that adds a non-negative boost potential and can accelerate transitions between the low energy states (76), (77). This aMD method is superior to other because they donot require prior knowledge of reaction coordinates. The aMD has been successfully applied various biomoelcules such as folding of Trp-cage, villin headpiece, WW domain and chignolin (78). The equations below defines general idea of the aMD. The bias potential is given by

$$\Delta V(r) = \frac{[E-V(r)]^2}{\alpha + [E-V(r)]} \qquad (2.5.1)$$

Where V(r) is the true potential, E is the reference (boost) energy and α is the tuning parameter that controls the depth of modified potential energy as shown in Figure 2.1. The modified potential is

$$V^*(r) = V(r), \ V(r) \geq E$$

$$V(r) + \Delta V(r), \ V(r) < E \qquad (2.5.2)$$
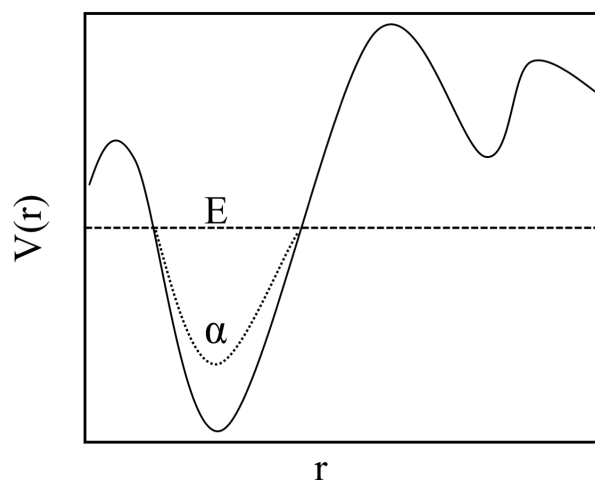
Figure 2.1 Schematic diagram of aMD

We performed aMD simulations on RfaH-CTD to exlore the confromational transition from α-form to β-form. We were able to explore the conversion of α-form RfaH CTD to random coiled state.

2.6 Jarzynski Equality (JE)

I performed nine independent 60 ns SMD simulations (pulling only) on RfaH to calculate the potential of mean force (PMF) using the Jarzynski equality (48), which relates the non-equilibrium work done (W) with the PMF as $e^{-\beta\Delta F} = \langle e^{-\beta W} \rangle$ where, $\langle \ \rangle$ refers to the ensemble average. The free-energy profile can be obtained from

$$\Delta F = \langle W \rangle - \frac{\langle W^2 \rangle - \langle W \rangle^2}{2kT} \qquad (2.6.1)$$

where k is the Boltzmann constant. The harmonic force applied to the virtual spring attached to the C-terminal C_α was integrated over the pulled distance to calculate the work done over the simulation time using $W = \int_0^T v\, f(t)\, dt$, where v is the pulling

speed, $f(t)$ is the force as a function of time, and $v\ dt=dx$ is a small increment in the pulling distance. All nine simulations were used in eq. 2.6.1 to calculate the average and the standard deviation for calculating the free-energy profile.

2.7 Adaptive Biasing Force (ABF)

To further investigate the α→β structural transformation of the CTD in the full RfaH, I calculated the Potential of Mean Force (PMF) using the Adaptive Biasing Force (ABF) method implemented in NAMD. In ABF, a free-energy profile is computed using a thermodynamic integration scheme in which the biasing force is adapted continuously in the Hamiltonian until the system overcomes an energy barrier. ABF has proven to be a very useful tool to investigate and explain protein interactions as well as dynamic pathways in a complex energy landscape (79), (80).

The PMF was calculated using 20 ns ABF simulations with a time step of 0.5 fs using rmsd as the reaction coordinate in the NAMD colvar module. The upper and lower boundary wall constant was 310 kcal/mol and a bin width of 0.001 Å. Force samples were accrued for 5000 steps until the adaptive biasing force was applied. The Langevin damping constant was 1 ps^{-1}.

2.8 Transfer Entropy (TE)

Transfer Entropy (TE) can be a valuable tool to obtain information about the dynamics of proteins, especially in allosteric conformational changes. In addition to quantifying important residue-residue correlated motions, TE provides a directionality of the drive-response relationship between residues i and j through the directionality index

$D_{j \to i}$. TE elucidates which part of the protein drives structural transitions and which other parts respond. For example, TE calculations illustrated the roles of helix motions on how the information of DNA binding is transmitted in the folded protein Ets-1 (81). Recently, TE calculations have been applied to protein folding and unfolding processes. Using long simulation trajectories, including a 200 μs trajectory of Trp-cage that contained multiple folding and unfolding events, Qi and Im (82) calculated a time dependent directional index $D_{j \to i}$ which showed remarkable correspondence with the overall folding score and further suggested that the main contribution to the information transfer entropy reflected the folding and unfolding events. The authors performed the TE analysis for single folding events in the trajectory and showed that the drive-response relationship is still preserved in separate non-equilibrium events such as folding and unfolding. During such non-equilibrium events, residues may interact differently at different times so that the drive-response relationship is altered. However, the calculations can still identify the residue-residue pairs that contribute most to the information transfer during the overall folding or unfolding process.

I performed transfer entropy calculations as described in ref. (81). In this method, the transfer entropy $T_{j \to i}$ is calculated as conditional mutual information; $T_{j \to i} = I(I_{k+1}; J_k^{\mu_j} | I_k^{\mu_i})$ where $I_k^{\mu_i}$ and $J_k^{\mu_j}$ are state vectors representing positional fluctuations of atoms i and j. The value of $T_{j \to i}$ indicates how strongly the fluctuations of atom j are driving the fluctuations of atom i. Directionality of the drive-response relationship can be obtained by further calculating the normalized directional index $D_{j \to i}$

$$D_{j \to i} = \frac{T_{j \to i}}{h(I_{k+1} | I_k^{\mu_i})} - \frac{T_{i \to j}}{h(J_{k+1} | J_k^{\mu_j})} \qquad (2.7.1)$$

If $D_{j \to i} > 0$, the fluctuations of atom j drive the fluctuations of atom i, and vice versa if $D_{j \to i} < 0$.

3. STRUCTURAL CONVERSION FROM α-HELIX TO β-BARREL IN THE RFAH TRANSCRPTION FACTOR

3.1 RfaH Protein Structure and Function

In the RfaH CTD, the protein segment from residues 115 to 156 exists as a helix-turn-helix motif when it is interacting with the RfaH NTD as shown in Figure 3.2a, but the CTD prefers a β structure when not interacting with the NTD. Its NusG paralog Figure 3.1b has a similar NTD structure but the CTD is never interfaced with the NTD, and the CTD is found only in the β-barrel structure. In both NusG and RfaH, RNAP and protein S10 binding occurs when the CTD is in its β-form, which requires breaking of the interfacial contacts between the NTD and the CTD in RfaH. Since the α-form of the RfaH-CTD is stabilized by interdomain interactions, it is important to examine these interactions for understanding the structural transformation of the RfaH-CTD.

Initially, we performed 850 ns all-atom MD (also referred as conventional MD (cMD)) and 10 ns accelerated MD (aMD) to elucidate the structural conversion of RfaH-CTD from α-form to β-form. During 850 ns simulation, such structural conversion of CTD was not observed. The shorter helix loses its helicity while the other remains intact. Since this shorter helix was in close contact with NTD, and loss of the contact in isolated CTD allowed greater flexibility and reduction in the secondary structure. There was loss in secondary structure of both helices during 10 ns aMD simulations.

I then performed REMD simulations for the CTD helix-turn-helix motif without including the NTD. Since the motif is simulated in the absence of the NTD, it is expected to convert to the β scaffold, as observed experimentally. Analysis of MD trajectories showed that the implicit solvent EEF1 force field allowed a fast structural conversion from α-helix →random coil →β-conformation as well as a through sampling of conformational space with an acceptance ratio between 41-48%. I note that given the relatively fast structure conversion and high acceptance ratio, additional lower temperature replicas would improve the stability of the preferred conformations. Since both the replica exchange procedure and choice of force field affect the kinetics information, the timescales of the observed structural conversion are not exact. However, relative information on kinetics such as identifying rate limiting steps can be obtained.
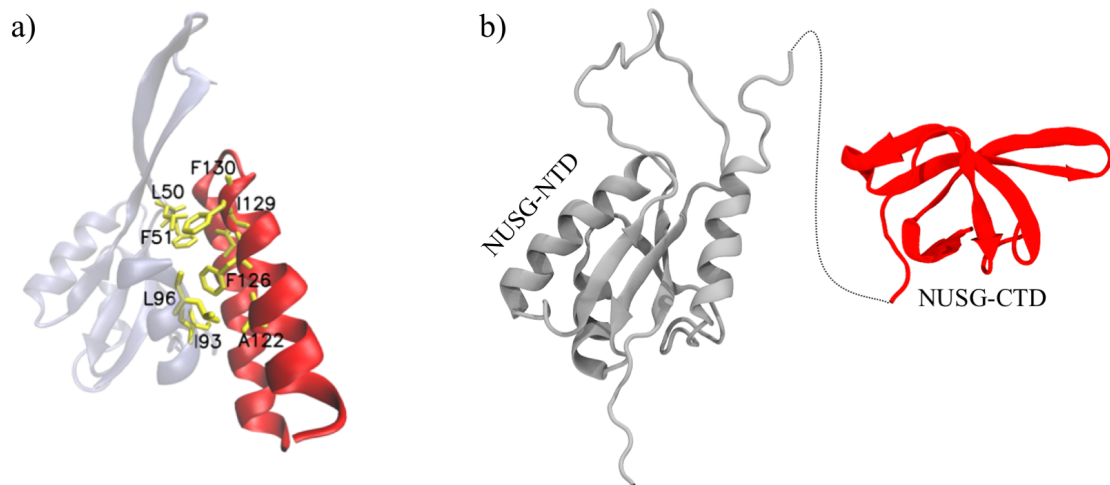
Figure 3.1 a) RfaH structure showing both the NTD (gray) and the CTD (red). Hydrophobic interactions at the interface between NTD and CTD are highlighted in yellow. b) NUSG-CTD and NUSG-NTD.

To understand the role of the interfacial interactions in the structural rearrangements of the RfaH-CTD, I also used SMD and calculated the Potential of Mean

Force (PMF) from the ensemble average of the SMD trajectories using the Jarzynski equality (48). SMD simulations have been valuable in understanding many biophysical processes such as mechanical responses to biomolecules (83), (84), (47), (85), (86) protein dissociation and conformational changes (87), (88), (89), stability of protein-protein or domain interfaces (90), (91), (92), and ligand binding (45), (93). Through the TMD and SMD simulations, I found that the interdomain interactions constitute the main barrier in the α-helix to β-barrel structural conversion. Once the interfacial interactions are broken, the structural conversion of the CTD occurs relatively easily. I identified key residues that are involved in the interdomain motion and structural transformation.

3.2 Structural Transitions in the RfaH Carboxy Terminal Domain (CTD)

3.2.1 Conformational Sampling of RfaH-CTD

In order to elucidate the conformational space of RfaH-CTD, I calculated and plotted in Figure 3.2 the free-energy landscape as a function of rmsd from the initial structure, and end-to-end distance (94) at 310 K. I used the 25-ns implicit solvent EEF1 REMD simulation to obtain the probability $P(rmsd, d_{ee})$ for the chain to have specific values of rmsd and d_{ee} and used $F(rmsd, d_{ee}; T) = -RT \ln P(rmsd, d_{ee})$ to calculate the free-energy. In order to plot F at 310K, I gather the 310K sequences that are distributed throughout various replicas. Multiple free-energy minima corresponding to various conformational states of RfaH-CTD are observed. As indicated by low free-energy regions, both α-helix as well as β-sheet structures are sampled at T=310 K. Since the α-helix to β-structure conversion occurs relatively quickly, the β-structures are sampled more frequently than α-helix structures for this temperature and time-window.

The free-energy minima corresponding to α-helix structures are observed mostly in the range of 5<rmsd<10 Å and 5<d_{ee}<30 Å. The α-helix structures with the two helices positioned approximately at right angle to each other occur at very high d_{ee} (~48 Å). A region with relatively high d_{ee} (~31 Å) and intermediate rmsd (~10 Å) consists of mixed α-helix and β-structure conformations. Similarly, several free-energy minima are observed for β-structures, which mostly have d_{ee}<30 Å and rmsd >10 Å. Deep free-energy minima are observed for the β-sheet structures with antiparallel stacking of β-strands. As expected in the absence of the NTD, the β-structures of the CTD are found to be relatively more stable than the α-helix structures. Also, in examining the MD time trajectories, once β-structure is attained, conversion back to α-helix is not observed at this temperature.
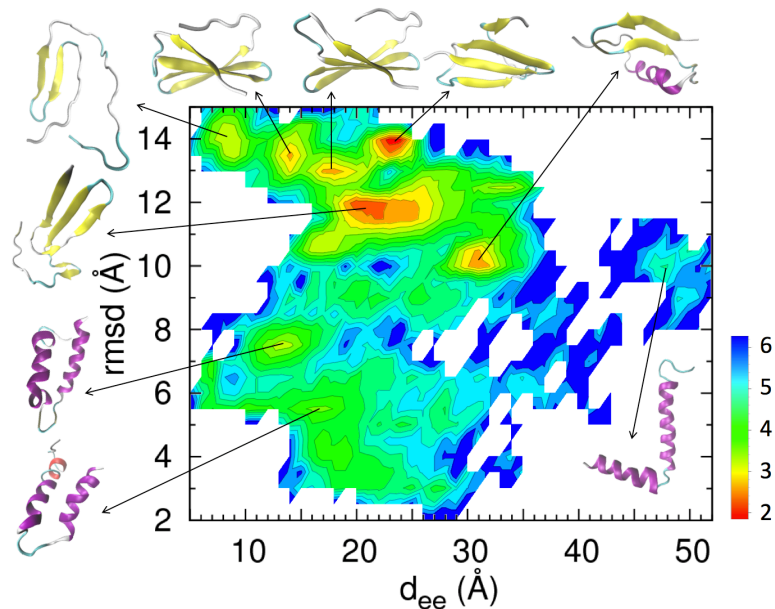
Figure 3.2 Free-energy surface for RfaH-CTD at 310 K in terms of rmsd and end-to-end distance. The energy scale is given in kcal mol-1 with red color indicating low free-energy regions and blue indicating the high free-energy regions. Typical structures for the free-energy minima are shown with arrows.

3.2.2 Time evolution of structural transitions in RfaH-CTD

During the 25-ns window that I used for my implicit solvent EEF1 simulations, α-to-β conversion was observed in all 16 replicas, but with only one replica finishing in a conformation close to the experimentally observed RfaH-CTD β-structure. I selected the replica that best displayed the experimentally observed CTD structure for further analysis to investigate the details of the structural transition from α-helix → random coil →β-conformation.

For the selected replica, Figure 3.3a shows the time evolution of the root-mean-square deviation (rmsd) from the initial α-helical configuration averaged over all the C_α atoms during the structural conversion obtained from the implicit solvent REMD simulations. The rmsd graph shows that the major structural conversion starts around 13

ns. Once relatively stable β structure is formed, the rmsd plateaus after 20 ns. I plot in

Figure 3.3b the time evolution of the temperature showing the walk in the 16 different

temperatures for this replica. Sampling at the lowest temperature mostly occurs early,

before the helix unfolding, and late, after the β-structure formation. During the helix to

coil and coil to β-structural transition, the sampling is at higher temperatures and this

allowed a faster structural conversion. An additional 25 ns of REMD simulation with

EEF1 revealed that this β-structure spends the most time at the lowest temperature,

making it the preferred structure.
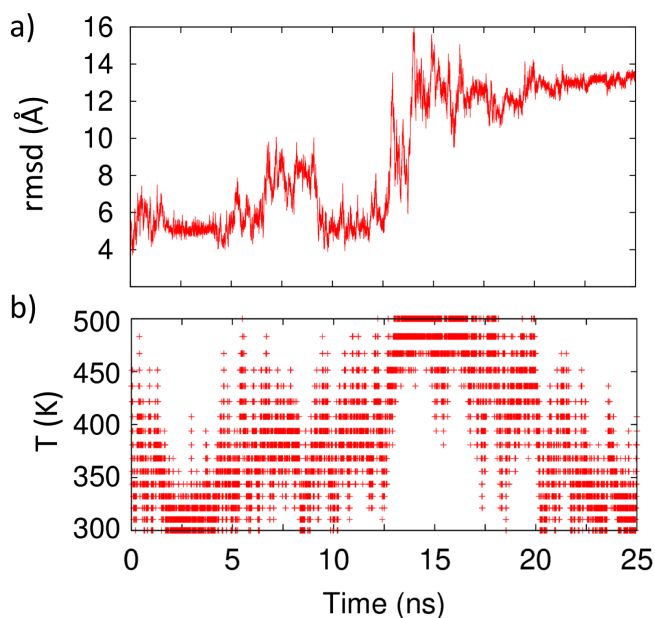
Figure 3.3 a) Root-mean-squared deviation (rmsd) from the initial structure averaged
over all the Cα atoms of the CTD during the structural conversion from α-helix structure
to β-scaffold. b) Time series of temperature exchange showing replica's walk throughout
the temperature space.

The stages of the structural conversion of RfaH-CTD are shown in Figure 3.4, in

which I display snapshots of representative structures at various times during the replica

exchange simulation. The supplementary movie S1 (95) gives a more detailed view of the dynamics of the structural conversion.
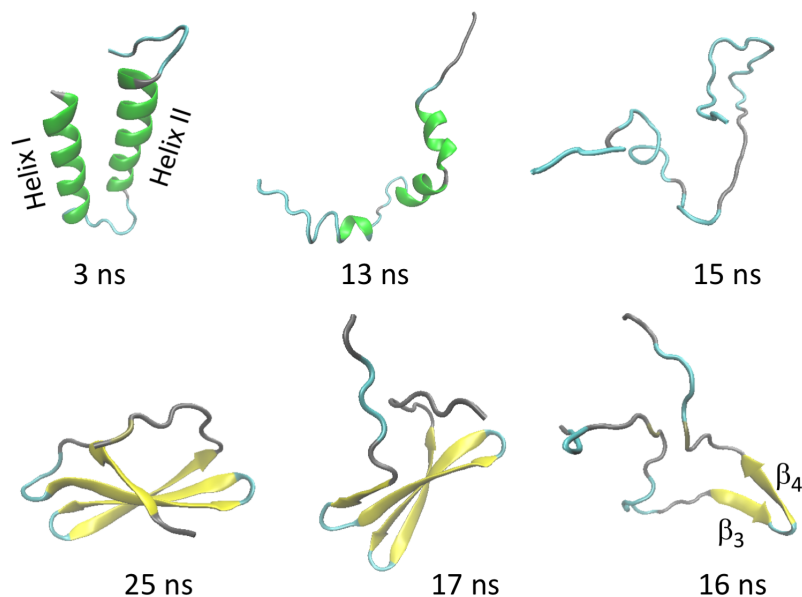
Figure 3.4 Snapshots of the evolving CTD configuration at various stages during the all-α to all-β structure conversion process.

In order to examine the changes in the protein structure for the selected replica in more detail, I plot the evolution of the secondary structure of each amino acid in the CTD in Figure 3.5a. In the helix-turn-helix motif, Helix I consists of residues 115-130 and Helix II consists of residues 136-153. Helix I loses most of its helicity around 8 ns, as seen in Figure 3.5a, whereas Helix II takes almost twice as much time to lose its helical structure. This shows that Helix I is relatively unstable under the conditions in which it does not experience tertiary interactions with the NTD. In the native state, Helix I interacts with a NTD helix via hydrophobic interactions. Specifically, as shown in Figure 3.1, residues A122, F126, I129, F130 in the CTD make hydrophobic interactions with

residues L50, F51, L96, I93 in the NTD. These inter-domain contacts are sufficient to refold the CTD into the α-helical form after denaturation (13).
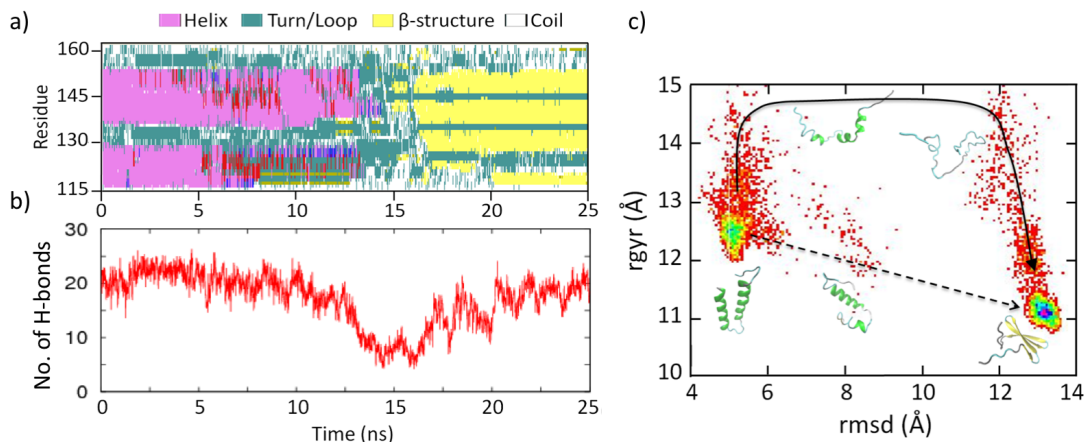
Figure 3.5 a) Time evolution of the secondary structure of each amino acid in the CTD. In addition to the color legends, noticeable π-helix (red) and 3-10 helix (blue) are also observed. b) Time evolution of the number of hydrogen bonds. c) Distribution of the conformations in a single replica as a function of rgyr and rmsd values. The solid line represents the MD pathway for the structural conversion and the dashed line represents another possible pathway.

Figure 3.5a shows that Helix II starts to unravel around 12 ns, and by 15 ns both Helix I and Helix II have completely unfolded. During the large structural change between 12-15ns, this replica is sampling conformations at high temperatures as shown in Figure 3.3b. The protein spends relatively less time in the intermediate unfolded state as compared to the α-helical structure. Large-scale β-structural arrangements appear between 15-20 ns, followed by the organization of the β-strands into a stable configuration by 20 ns. After 20 ns, smaller fluctuations occur which are reflected in the small fluctuations after 20 ns in the number of hydrogen bonds displayed in Figure 3.5b. The loop region in the helix bundle (residues 131-135) remains flexible until the β-structure starts to form. Residue Gly135 is in the loop conformation in both the helix

bundle as well as in the final β-scaffold. During helix unfolding, a noticeable fraction of π- as well as 3-10 helical structures also transiently appear.

I also analyzed changes in the hydrogen bond pattern due to the structural transformation by calculating the total number of hydrogen bonds between protein backbone atoms, which is plotted in Figure 3.5b. I used an H-bond distance cut-off of 3.5 Å and angle cut-off of 30°. The time evolution of the number of hydrogen bonds is similar to the evolution of the changes in secondary structure and also correlates with the temperatures at which the conformations are sampled. The number of hydrogen bonds decreases as the helices unfold. The unfolded random coil structure around 15 ns shows the least number of hydrogen bonds, before it increases as a result of formation of hydrogen bonds between β-strands. In the unfolded state around 15 ns, a slight increase in hydrogen bonding occurs because of a collapsed random structure. These non-native contacts break within a few nanoseconds and allow the β-strands to align and nucleate into the first β-hairpin. As discussed later, from this time onwards, the hydrogen bonding pattern trends upwards but with oscillations during structural rearrangement of β-strands until the final β-scaffold is reached.

Figure 3.5c displays the distribution of the structures having specific radius of gyration (rgyr) and rmsd values. In the MD simulations, the pathway for structural conversion from all α-helix to an all β-conformation that is most similar to the experimentally observed structure proceeds via a completely unfolded intermediate structure, as shown by the solid line in Figure 3.5c. However, other possible pathways may exist (dashed line in Figure 3.5c) in which structural conversion to β-conformation can occur concurrently with unraveling of α-helix structure. Such a process in a different

protein has been shown to be orders of magnitude faster than the structure conversion that involves a completely unfolded intermediate (96). Similarly, a recent investigation of the α-helix to β-structure transition in the human islet amyloid peptide fragment with REMD simulations at atomistic detail also showed two distinct mechanisms, one without unraveling the helix into random coil and the other with a random coil intermediate (97). On the basis of the observed pathway alone, the sampled structures that have intermediate rmsd but low rgyr in Figure 3.5c are off-pathway structures. However, they may belong to the on-pathway structures for the alternate pathway represented by the dotted line. A plot of the free energy profiles of the observed pathway and a possible alternate pathway based on the free-energy landscape of Figure 2 is given in the supplementary figure, Figure S1 (95). In our simulations with the EEF1 implicit solvent model, I did not observe the structural conversion via the faster concurrent β formation pathway, which can potentially be revealed by more detailed and long time-scale all-atom simulations.

3.2.3 Folding into a β-barrel scaffold

Starting from the unfolded state configuration around 15 ns, folding into β-structure starts by nucleating a β3-loop-β4 hairpin structure as shown at 16 ns in Figure 3.4. Here, β3 consists of residues 137-144, whereas β4 consists of residues 147-154. Residues 145 and 146 form the loop segment that connects the β3 and β4 strands. Once this hairpin structure is formed, another β-strand, β2 formed by residues 127-134, aligns with β3 very quickly (~1 ns) as seen in Figure 3.4 at 17 ns. The number of hydrogen bonds increases steadily during the formation of β-sheet structure involving β2, β3, and β4 that occurs around 17 ns. The number of hydrogen bonds continues to increase due to

partial formation and alignment of β1 formed by residues 116-119. However, the initial hydrogen bonding to β1 is transient as seen in Figure 3.5b. After peaking at 17 ns, these hydrogen bonds break and rearrange. Therefore, formation and proper alignment of β1 takes longer (~3 ns). The longer time for proper alignment of β1 is partially due to the higher entropy of the loop segment connecting β1 to β2 which consists of eight residues, as compared to only two residues in the loop segment connecting β2 to β3.

Oscillations with a period of a few nanoseconds in the hydrogen bond pattern provide additional insight into the mechanism for this protein to fold into a β-motif. The antiparallel β-sheet structure formed with β1 through β4 is slightly curved and makes it possible for β5 to wrap around the β-sheet structure and align with β1 in an antiparallel fashion. This wrapping gives an additional twist to the β-scaffold, resulting in a barrel-like shape. With the implicit solvent EEF1 model, the final β-structure is still relatively flat as compared to the experimentally determined β-barrel structure, and the structure did not show improvement upon further REMD simulation with EEF1. Therefore, I selected a structure that is closest to the native β-barrel structure for further relaxation with an all-atom, explicit solvent MD simulation.

3.2.4 β-structure refinement with all-atom simulation

When relaxed using the all-atom CHARMM 27 force field with explicit solvent, the β-structure tends to be organized closer to the native state of the experimentally determined all-β configuration with an rmsd of ~5 Å and Q-score of ~0.4. Figure 3.6 displays the structural overlap between the experimentally determined NMR structure (red) of the β-barrel and the MD simulated structure (blue). The order and orientations of

the β-strands, as well as the overall barrel shape from the simulation match relatively well with the NMR structure. An additional 50 ns simulation at 300 K shows very small change in the rmsd, suggesting that the obtained structure is quite stable. I also compare the specific amino acid residues making up the β-strands. The β-strand residues match reasonably well, but with some differences that prevent the MD β-strands from aligning perfectly with the experimental structure. The β2 strand from the simulation (amino acids 127-134) extends longer than the experimental structure (127-130) and the MD β3 strand starts one residue earlier at 137, leaving only two amino acids for the MD connecting loop as compared to seven amino acid residues in the experimentally observed structure.
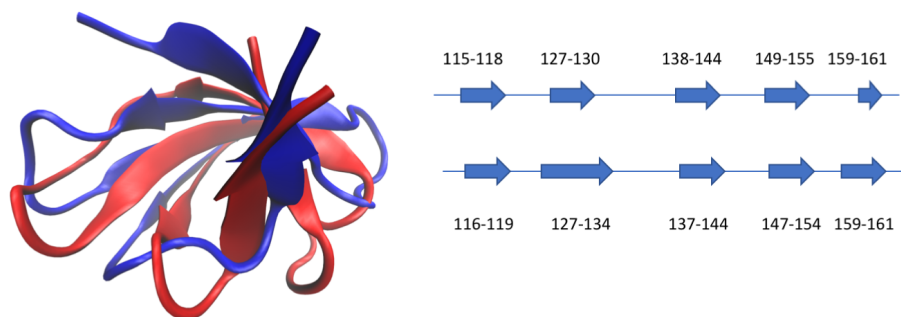
Figure 3.6 Left: Comparison between the β-barrel structure from the MD simulation (blue) and the NMR configuration (red; NMR structure 2LCL from ref. [10]). Right: Comparison of amino acid residues that form various β-strands in the β-scaffold.

3.3 TE Calculations on the α-helix unfolding and β-barrel formation

I performed information transfer entropy (TE) analysis for our selected REMD trajectory using C_α backbone information. Since the trajectory consists of two major events, unfolding of α-helix to random coil and folding to a β-structure as shown in Figures 3.3 and 3.5, I calculated $D_{j \to i}$ of eq. (2.7.1) for the unfolding event and the folding

event separately. By calculating TE separately for helix unfolding and formation of beta structures, the role of each residue in each of these processes can be examined. The trajectory for the TE analysis for helix unfolding was chosen to be the 0-14 ns window. Since the protein configuration was saved every 2 ps, there were 7,000 frames in the trajectory for this process. I display the $D_{j \to i}$ values for all i,j residue pairs for the helix unfolding process in Figure 3.7a. The values were color coded as shown in the color bar, with positive values in red, brown or yellow and negative values in blue. Red means that the residue on the y-axis drives the residue from the x-axis. For the helix unfolding, residues on Helix II appear to be a strong driver for several residues in both Helices I and II. Specifically, Helix I residues Ile117, Ile118, and Helix II residues Pro133, Asp134 appear to respond strongly to the residues between Leu142 to Glu158 which are mostly in Helix II. In general Helix I experiences much larger structural fluctuations as compared to Helix II. Consistent with the transfer entropy analysis in ref. (82) in which driving residues were found to be in segments with more robust secondary structure, I found that residues in Helix II which has a more stable secondary structure drive the residues in more floppy regions in the chain. Experiments show that in the absence of appropriate CTD/NTD contacts, the preferred conformation of the CTD is the β-barrel structure (13). Our results show that in the absence of NTD tertiary contacts, Helix-I is not stable and therefore unravels relatively quickly. This is reflected by the transfer entropy calculations which show that Helix-I is primarily a responder. These observations suggest that the helix-I is most likely formed by nucleation at the CTD/NTD interface, instead of first forming the helix and then making the tertiary contacts with NTD.
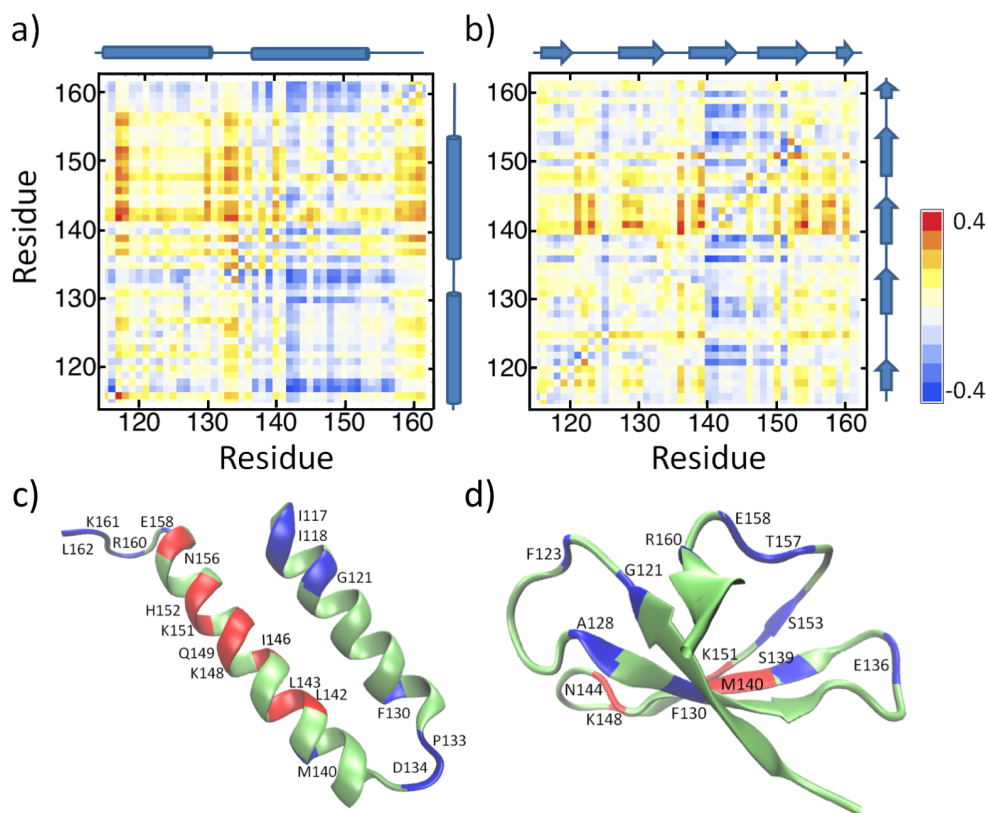
Figure 3.7 Normalized transfer entropy directional index $D_{j \to i}$ for two different stages in the α→β conversion process: a) α-helix unfolding and b) β-scaffold formation. Red means that the amino acid on the y-axis drives the response of the amino acid on the x-axis. c) major driver (red) and responder (blue) amino acids in helix unfolding and d) in β-barrel folding.

The second part of the trajectory was taken from 14 ns to 25 ns of the REMD trajectory and included a total of 5,500 frames. As displayed in Figure 3.7b, the driving amino acid residues are different for β-structure formation from those in helix unfolding. However, the responding residues fall in the same general segment of the protein. Major drivers M140, N144, K148, and M140 are in β3 and β4, which form the β-hairpin structure early on during the β-structure folding. Residues between Met140 and Leu145 in β3 are strong drivers for residues Gly121, Phe123, Glu136, Ser139, Val154, Glu158, and Arg160. Similarly, residues Asp147, Lys148, Glu149, and Lys151 are weak drivers.

Residues Ile146 and Ile150 are strong responders to many other residues. Isoleucine 146 is in the flexible loop region between β3 and β4 and Ile150 is in β4. In both cases, the Ile responders are exposed to the solvent, which is consistent with the observation that driving residues are mostly buried in the hydrophobic core (82). During the β-folding, information transfer occurs among amino acid residues that are not necessarily in close proximity. In addition, many amino acids have different transfer entropy identities, drivers for some residues, responders for others. In Figure 3.7c and 3.7d, I color code only the major drivers and the associated responders from Figure 3.7a and 3.7b, respectively.

3.4 Full RfaH Structural Fluctuations and Correlated Motions

In the RfaH-CTD, the protein segment from residues 115 to 156 exists as a helix-turn-helix motif when it is interfaced with the RfaH-NTD, but the CTD prefers a β-barrel structure when not interfaced with the NTD. Its NusG paralog has a similar NTD structure but the CTD is never interfaced with the NTD, and the CTD is found only in the β-barrel structure. In both NusG and RfaH, RNAP and protein S10 binding occurs when the CTD is in its β-form, which requires breaking of the interfacial contacts between the NTD and the CTD (Table S1) in RfaH. Since the α-form of the RfaH-CTD is stabilized by interdomain interactions, it is important to examine these interactions for understanding the structural transformation of the RfaH-CTD.

I performed a 100 ns NVT simulation on the native state structure of the full RfaH with the CTD in the α-helix form. Figure 3.8a shows overlays of 100 sampled conformations taken from the 100 ns trajectory. These, together with the root mean

squared fluctuations (rmsf) for all C_α plotted in supplemental Figure S1 (34) provide insight on the conformational fluctuations during the simulation. The rmsf figure (Figure S1 a) shows that both the NTD and the CTD are stable but large fluctuations occur in the linker and loop segments, as expected. Specifically, the segment composed of residues 35-45, corresponding to the loop region of the β-hairpin motif in the NTD shows particularly large fluctuations. Also, the long NTD-CTD linker segment (residues 101-114) shows large flexibility. Interestingly, the NTD helix (90-100) at the domain interface is slightly more flexible than the CTD helices (Helix I and Helix II).

Dynamic network analysis can be used to understand the correlated molecular motions (98), (99), (100). Dynamic analysis from the 100 ns NVT simulation of the full RfaH resulted in 11 communities that are highlighted in color in Figure 3.8a and the constituent amino acids for each community are listed in Table S2 (34). A community is composed of amino acids that are in close proximity and have highly correlated motions. The size of the circle for each node is proportional to the number of residues in each community, and the line thickness represents the inter-community edge weights. As shown in Figure 3.8b, community H (lime green) in the NTD β-sheet core region seems to have the most correlated motion with other communities. I found three communities (C, D, G) that each are composed of residues from both the NTD and CTD. The amino acids that comprise these three communities are highlighted in Figure 3.8c. By examining the composition of each community, I was able to locate the amino acids pairs that most strongly bridge the NTD to the CTD within each community. These pairs are shown in Figure 3.8c: L96-F126 (community C), F33-F130 (community D), and F81-I118 (community E). In all of these cases, phenylalanine residues seem to play important

roles in modulating the interdomain motions. F126 plays a major role in interdomain communication and therefore it is a functionally important residue for modulating the domain separation. Alanine substitution using FoldX (101) revealed that mutating F126 with alanine resulted in an increase in $\Delta\Delta G$ by ~3 kcal/mol (Figure S1 (b) (34), indicating that the F126A substitution destabilizes the domain interface, and is one of the largest $\Delta\Delta G$ among all residue positions upon alanine mutation. Figure S1 (34) (b) also shows that more than 75% of the interface residues do not favor alanine substitution. I performed a 100 ns MD simulation on the RfaH-F126A mutant and investigated its kinetic/dynamic behavior. Figure S2 shows that the C_α rmsf for the WT and the RfaH-F126A are comparable. The F126A mutation caused slightly larger fluctuations in the loop/turn region (residues 35-50) in the β-turn-β motif of the NTD, but the linker segment connecting the two domains, as well as the some other loop segments are more restricted in the mutant. The restriction of loop segments in the mutant are due to the interactions with the CTD tail segment (residues 158-162), which became flexible in the WT after ~50 ns, whereas it remained intact for the entire 100 ns in F126A. The network analysis displayed in Figure S2 (34) shows that the F126A mutation causes changes in the community structure. Specifically, the interdomain correlated motion through F126 disappears in the mutant.
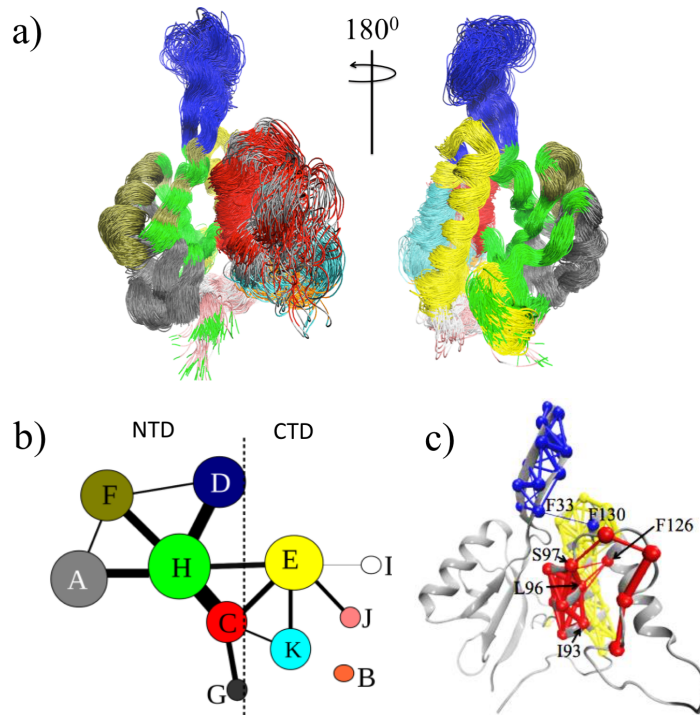
Figure 3.8 RfaH structural fluctuations and correlated motions. a) Structural fluctuations of amino acids in the full RfaH determined from MD simulations. b) Community map of RfaH showing 11 different communities. The size of the circle for each node is proportional to the number of residues in each community, and the line thickness represents the inter-community edge weights. The communities are colored as in Figure 3.8a. c) Three communities especially relevant for the interdomain interactions are highlighted.

3.5 Structural transformation with TMD:

The α-helical conformation of the CTD (α-CTD) in the full RfaH was targeted to map onto the β-barrel CTD (β-CTD) and vice versa using targeted MD. Several targeted MD simulations of length 10 ns were performed. Each simulation resulted in conversion to the desired target structure. A typical α-helix to β-barrel structural transformation process is shown in Figure 3.9, which shows snapshots of the protein conformation at various stages of the structural transformation. In general, the α-helix secondary structure

is lost by the mid-part of the simulation where the target RMSD has decreased by 50%. For the CTD, once the tertiary contacts with the NTD are lost as well as the CTD helicity, folding into the CTD β-structure begins to occur. By the end of the simulation, CTD β-sheets are fully formed and arranged into the final β-barrel structure. The final structure was within 0.94 Å rmsd of the experimental structure (11). To confirm the stability of the transformed structure, I performed a 100 ns conventional MD and found that the β-structure was very stable at 300K.
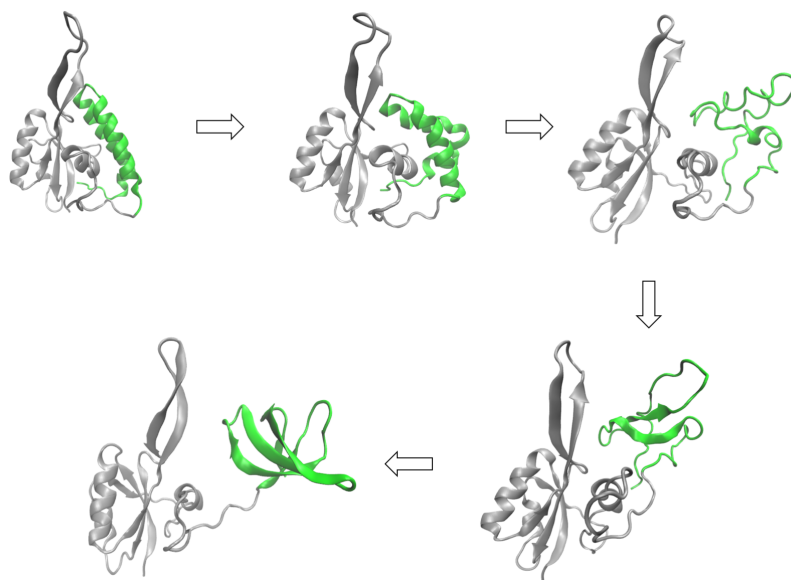
Figure 3.9 Snapshots of RfaH conformations during TMD simulation for the α-CTD to β-CTD conversion in the full RfaH. The N-terminal domain is shown as gray and the transforming C-terminal domain is highlighted in green.

I plot the time evolution of the secondary structure of the full RfaH in Figure 3.10a from the TMD. The structural transformation of the CTD (residues 115-162) can be divided into two parts: helix-to-coil transition (approximately 2–5 ns), and coil-to-beta transition (approximately 6-9 ns). The total number of intradomain CTD hydrogen bonds

also follows this pattern, as shown in Figure 3.10b. Hydrogen bonds are calculated with a

distance cut-off of 3.5 Å and an angle cut-off of 30°. As discussed later, the coil structure

that is intermediate between the α-helix and β-barrel structure is in fact a collapsed

globule state of the protein.
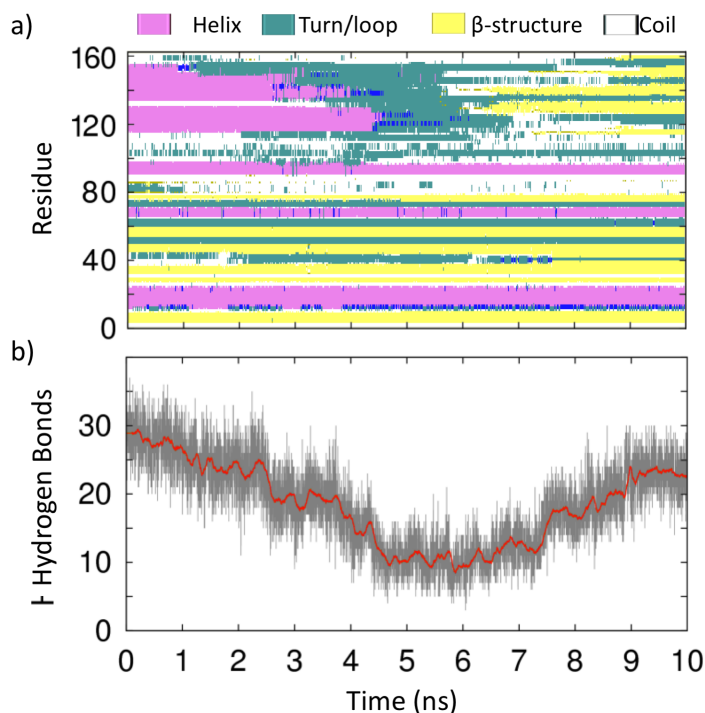
Figure 3.10 Results from RfaH TMD simulations for the α-CTD→ β-CTD structural conversion. The CTD is composed of residues 115-162. a) Time evolution of the secondary structure b) Time evolution of the number of intradomain hydrogen bonds in the CTD. The number of hydrogen bonds decreases as the CTD helices unfold (2-5 ns) but increases with the formation of the CTD β-structure (6-9 ns).

3.5.1 Unfolding of the helix to the collapsed coil state

As shown in Figure 3.10a, the CTD begins to lose its Helix II helicity around 2

ns. Interestingly, Helix I (residues 116-131) retains its structural integrity better than

Helix II (residues 135-156). This result using the full RfaH is in contrast to previous

computational results (95), (102) obtained for the isolated CTD. Simulations for the

isolated CTD showed faster unfolding of Helix I because of the absence of the interdomain tertiary interactions. On the basis of amino acid secondary structural propensity calculations, Balasco et al. (102) showed that Helix I has weaker stability as compared to Helix II, whereas the interdomain tertiary interactions with the NTD for both helices are comparable. Targeted MD results from the simulations of the full RfaH show that in the presence of the NTD interactions, the structural transformation proceeds through the pathway that rearranges the amino acids in Helix II first. I further explored this process with steered MD simulations, as discussed later.

As seen in Figure 3.10b, during the helix unfolding the number of hydrogen bonds decreases by ~70%. Although some residual hydrogen bonds are still present in the coil state (5-6 ns), the CTD helix secondary structure is completely lost by 5 ns (Figure 3.10a). Initially, Helix II is stabilized by both CTD Helix I-Helix II interactions as well as the NTD-CTD inter-domain interactions during the structural conversion. Part of Helix II unravels to form a loop/turn at 2 ns, and then a transient 3_{10}-helix at 3 ns. The CTD inter-helical contact is lost once Helix II is sufficiently destabilized. Key stages during the helix to coil conversion of the CTD are shown in Figure 3.11.
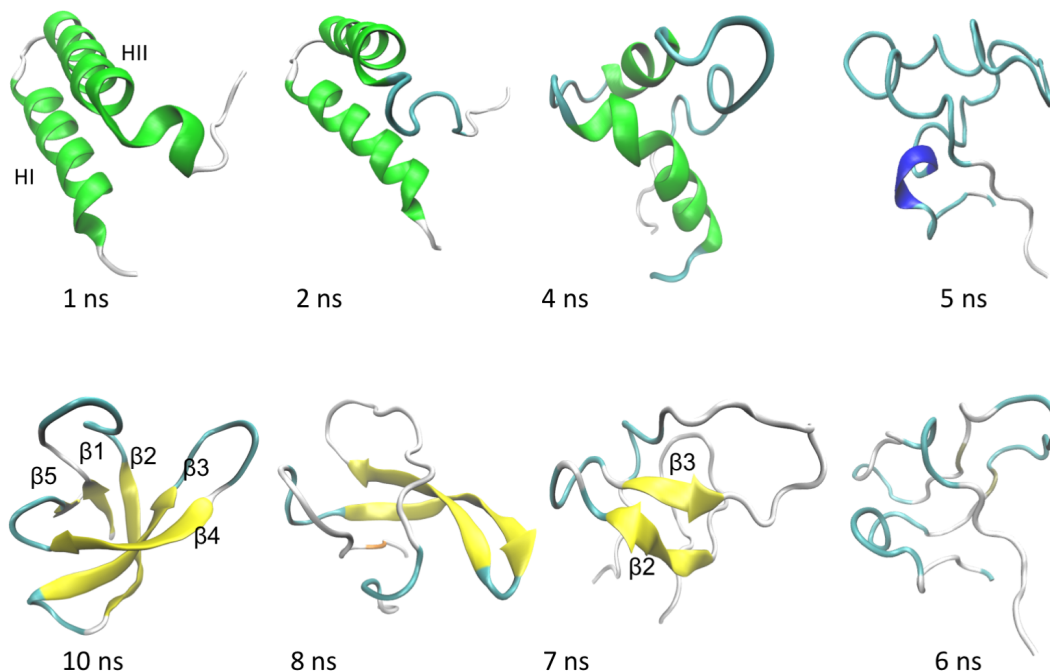
Figure 3.11 Conformations of the CTD at various stages of structural transformation during helix unfolding (top, left to right) and beta-folding (bottom, right to left).

A closer examination of the CTD coil structure (5-6 ns) reveals that it has significant internal hydrophobic interactions, making it more like a molten-globule state than a random coil structure. This is also evidenced by only a small change in the solvent accessible surface area (SASA) calculated for the hydrophobic amino acids in the CTD (Figure 3.12, top). This hydrophobic collapsed state constitutes the intermediate structure that connects the CTD α-helix and β-barrel basins in the energy landscape. This result is in contrast to the observed mechanism for the structural transformation in the isolated CTD (95) in which the helices first unfolded to a random coil structure, followed by the folding into the β-barrel.
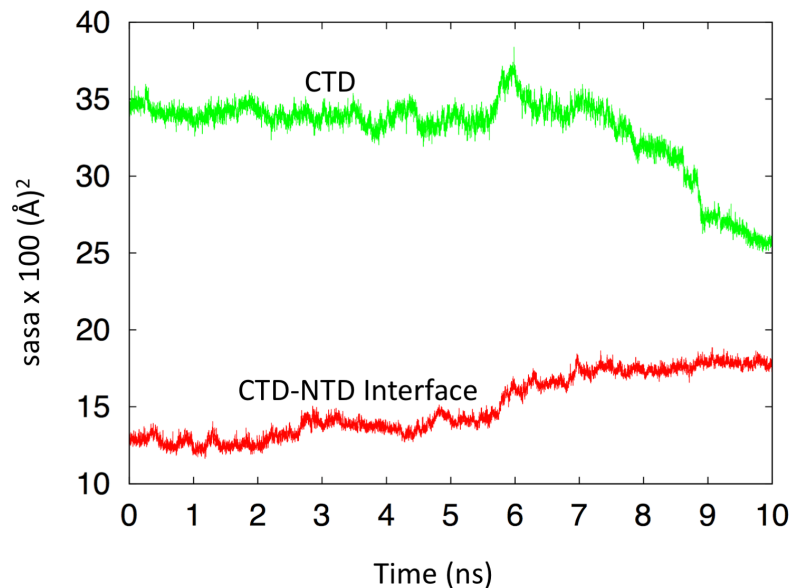
Figure 3.12 Solvent accessible surface area (SASA) calculated for the hydrophobic amino acids in the CTD (top-green) and at the CTD-NTD interface (bottom-red).

3.5.2 Folding into the β-barrel

The structural rearrangement of the collapsed globule state of the CTD around 6 ns in Figure 3.10a allows the formation of the first anti-parallel β-strands. At this point, the SASA for the CTD hydrophobic residues (Figure 3.12, top) increases slightly and then steadily decreases as the hydrophobic residues bury due to reorganization. The formation of the first anti-parallel β-strands also causes the SASA for the NTD-CTD interface residues (Figure 3.12, bottom) to increase due to increased exposure of the hydrophobic residues. During 7-8 ns, the middle two β-strands of the CTD, β2 (residues 127-133) and β3 (residues 138-143) form first. This is followed by the alignment of β1 (residues 115-117) with β2 as well as the alignment of β4 (residues 150-155) with β3. Finally, β5 (residues 160-161) forms and aligns with β1 to complete the β-barrel

structure. The above mentioned sequence of β-structure formation is consistent with observations by Li et al. (103) and Xiong et al. (10) through MD simulations.

To investigate the dynamics of the resulting RfaH β-CTD, I performed a 100-ns conventional MD simulation and found the β-structure of the CTD is stable at 300K. However, the absence of the interdomain contacts, combined with the flexible nature of the linker, allows large-scale diffusive motions of the β-CTD, which can be important for its function. In Figure S3, I plot the time evolution of the CTD-NTD center of mass separation distance (d_{cm}), as well as the root-mean-squared distance (rmsd) from the starting structure during the 100 ns simulation and compared these with the results for the RfaH α-CTD. The rmsd and d_{cm} for the RfaH β-CTD shows significant fluctuations as compared to those of the RfaH α-CTD. As the linker segment undergoes large fluctuations, d_{cm} tends to increase in the β-CTD whereas the d_{cm} for the closed form α-CTD remains constant due to CTD-NTD interactions. In addition, the rmsf for the NTD show that the residue fluctuations are comparable in the beta and alpha forms, suggesting that the NTD integrity is not significantly perturbed by domain dissociation. However, the intradomain network connectivity in the NTD changes significantly upon domain separation (Figure S2) (34) I also observed that when the E48-R138 interdomain salt-bridge interaction breaks, E48 makes a different ionic interaction with R11 which is within the NTD, giving additional structural stability. This R11-E48 interaction was also observed in the 100-ns simulation of the RfaH in the α-CTD form when the E48-R138 interaction is intermittently broken.

3.5.3 Dynamics of the interdomain contacts

The RfaH-CTD in the α-helical state is stabilized by interdomain contacts. I calculated the time-evolution of the fraction of native contacts (Q) of the NTD-CTD interfacial residues and CTD intradomain contacts and plotted the results in Figure 3.13a. Native contacts were defined as contacts between any heavy-atoms within 4 Å of distance in the native state configuration. The NTD-CTD interfacial native contacts are almost completely lost by 4 ns. Intradomain alpha CTD contacts are lost by 6 ns, at which time the native contacts for the CTD β-structure begin to form. Interestingly, some residue pairs that were part of the native intradomian contacts in the α-CTD were also part of the native intradomian contacts in β-CTD. Specifically, residue pairs G135-K139 and A128-L141 were within a cutoff distance of ~4 Å in both the α-helix and β-structure of the CTD. Persistence of the hydrophobic contact between residues A128 and L141 is an interesting feature of the structure conversion.

One of the major interdomain interactions in the RfaH with α-CTD is the salt-bridge interaction between E48 and R138. As shown in Figure S1 (34), alanine substitution at either E48 or R138 destabilizes this interdomain interaction. Similarly, E48S mutation increased the $\Delta\Delta G$ by 0.82 kcal/mol. The alanine substitution is consistent with the experimental observation (11) that E48S mutation favored the dissociated state leading to the β-barrel structure of the CTD. Figure 3.13b shows the distance (Δr) between the anionic carboxylate oxygen of the E48 side chain and the cationic ammonium nitrogen of the R138 side chain. The distance Δr remains around 5Å until 3 ns but increases steadily once the salt-bridge is broken. The time evolution of the interdomain contacts in Figure 3.13a shows a sharp decrease in Q around 4 ns, which is

when the salt-bridge interaction is completely broken. Therefore, the E48-R138 salt-bridge is one of the last interdomain contacts that are still present during the structural transformation. This is explored further with SMD in the next section.
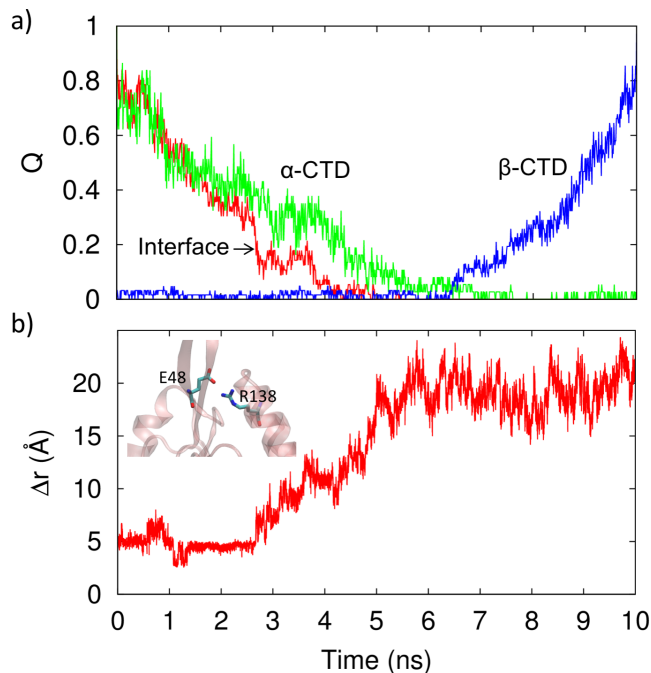
Figure 3.13 a) Fraction of the native contacts Q at the NTD-CTD the interface (red), and intradomain for the CTD for α-CTD (green) and β-CTD (blue). b) Distance (Δr) between the E48 oxygen and the R138 nitrogen that are involved in the interdomain salt- bridge interaction.

3.6 Potential of Mean Force Using ABF

To further investigate the α→β structural transformation of the CTD in the full RfaH, I calculated the Potential of Mean Force (PMF) using the Adaptive Biasing Force (ABF) method implemented in NAMD, and display the results in Figure 3.14. In ABF method, a free-energy profile is computed using a thermodynamic integration scheme in which the biasing force is adapted continuously in the Hamiltonian until the system overcomes an energy barrier. The calculation of PMF using ABF method has proven to

be a very useful tool to investigate and explain protein interactions as well as dynamic pathways in a complex energy landscape (79), (80).
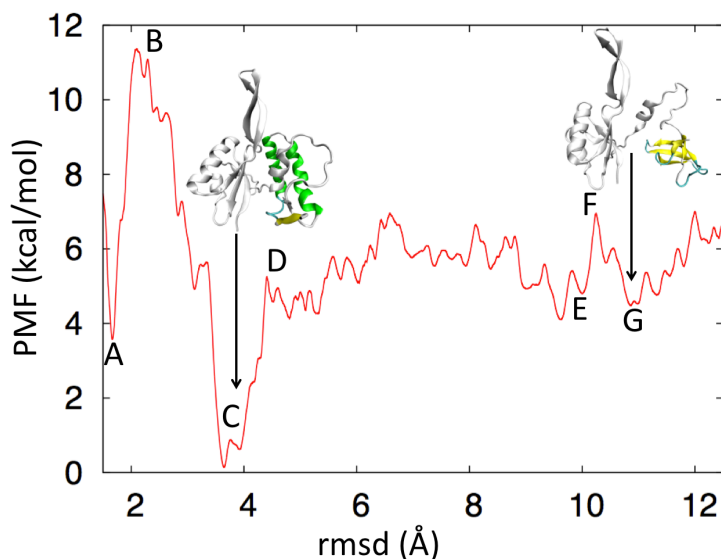
Figure 3.14 PMF as a function of rmsd for the structural transformation of the RfaH-CTD.

Figure 3.14 shows the calculated PMF with the root-mean-squared-deviation (rmsd) as the reaction co-ordinate. The rmsd is calculated using a starting reference structure that is obtained after a 100 ns equilibration simulation of the full RfaH with the alpha form of the CTD. Figure 3.14 shows that readjustment (A→C) of the residues at the CTD-NTD interface requires overcoming a large barrier B and ultimately leads the system to an especially low free-energy configuration at C. The deep basin at C corresponds to a structure that has subtle, but important changes in amino acid positions and interactions compared to the initial structure in A, which represents the native state basin. Specifically, a slight bending in the CTD Helix I allows its amino acid I118 to make a hydrophobic contact with the CTD tail residue V154 as shown in Figure 3.15a. Similarly, the NTD residue A91 makes a hydrophobic contact with the CTD tail residue

F159. This interaction also allows the CTD tail segment (residues 158-161) to align with the NTD-CTD linker segment (residues 110-113) in an antiparallel fashion. Both the linker segment and the CTD tail segment were parts of the segments that were added using Modeller and their interactions in this region seem to yield a more compact and stable structure.
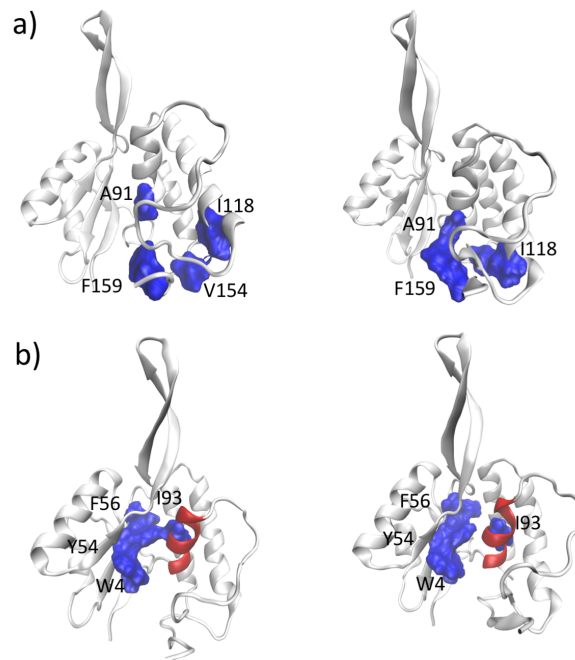
Figure 3.15 a) Snapshot of the structures corresponding to state A (left) and state C (right) showing the rearrangement of the hydrophobic residues in the CTD. b) Displacement of I93 of the NTD helix in state A (left) slightly exposes the hydrophobic core (right, state C).

The changes in structure from A to C in Figure 3.14, due to small rearrangements of the CTD may be functionally important. For RNAP binding to RfaH, the clamp helix known as β' CH of the RNAP β-subunit binds to the RfaH NTD hydrophobic region composed of residues W4, Y54 and F56 (17) (Figure 3.15b). This binding of β' CH is important for reducing the transcriptional pausing and excluding the RNAP binding of

NusG-NTD (11). In the native state (A of Figure 3.14 and the left side of Figure 3.15b), these hydrophobic residues interact with I93 of the NTD helix and form a completely buried hydrophobic core, which may not allow the β-subunit binding. However, in structure C (right side of Figure 3.15b), this NTD hydrophobic core is slightly exposed as the Y54-I93 interdomain hydrophobic contact is broken as I93 is pulled away by the CTD Helix II through a hydrophobic interaction with F126.

Figure 3.14 shows a large barrier separating states A and C and overcoming the large barrier at B is not possible merely because of thermal fluctuations, but requires assistance. It was suggested (17) that the *ops* element binding at another binding site (residues R16 and R73) in RfaH triggers the domain separation and subsequent binding of the RNAP. Consistent with the suggested mechanism, the minimum at C might constitute the structure that is accessible for RNAP binding. These structural changes present correlations with the functionally relevant structures, though I note that the allosteric changes caused by the *ops* binding may be different from the structural changes observed here under biasing forces.

The transition from C to D in Figure 3.14 involves another large barrier in the free-energy profile. The structure corresponding to the state D exhibits significant solvent exposure of the hydrophobic core residues at the RNAP binding site suggests that the RNAP β-subunit binding may facilitate the C→D transition and allow further domain separation. The complete domain separation occurs when the rmsd exceeds 6 Å, with a corresponding domain-domain distance of ~16 Å. Once the domains are separated, significant structural changes (Figure 3.14, Structures E, F, G) take place during the helix unfolding and beta-structure formation. The free-energy plot shows a rugged landscape,

as expected for the calculations with rmsd as a reaction coordinate (80). The β-structure formation occurs between rmsd 9 to 11 Å. The state E in Figure 3.14 corresponds to a structure that shows complete loss of the CTD helicity as well as the initiation of the antiparallel β-sheet. After overcoming a small free-energy barrier, β-structure formation is completed at state G. Most structures beyond state G are part of the native state basin of the β-CTD.

3.7 Interdomain and CTD Interhelical Interactions Using SMD

To further investigate the NTD-CTD interdomain, as well as the CTD intradomain interactions, I performed Steered Molecular Dynamics (SMD) simulations on the full RfaH. Before applying the pulling force, the system was first equilibrated for 5 ns in an NPT simulation in a long rectangular box with enough size to accommodate the pulling. The N-terminal C_α atom was held fixed and a dummy atom pulls the C-terminal atom C_α as the SMD atom for 10 ns, followed by another 10 ns without pulling to allow the molecular system to relax from mechanical stretching. The process was repeated until the SMD atom was close to the box edge, with a final end-to-end distance (94) of ~130 Å. Thus, SMD simulations were performed for intervals (0-10, 20–30, 40–50, 60–70, 80–90, 100–110, 120–130) ns with relaxation simulations in-between. The pulling/relaxation simulation is displayed in Movie S1 (34).

Figure 3.16 shows representative snapshots of various intermediate stages during the mechanical stretching. After the initial rearrangement of the CTD loop segment, the positions of the CTD helices as well as the NTD helix at the interface are affected around 40 ns. Consistent with our description given above with respect to Figs. 3.14 and 3.15

53

and the ABF simulations, Figure 3.16 shows (40 ns) that the NTD helix initially moves together with the CTD helices, opening up the NTD hydrophobic pocket at the RNAP binding site (residues W4, Y54 and F56). After 45 ns of pulling (80 ns frame in Figure 3.16), contacts between the CTD helices and the NTD helix weaken, which ultimately lead to the separation of the CTD from the NTD helix. As shown in Figure S5 (34), the SASA for the hydrophobic core opens up around 100 ns due to a slight movement of I93.
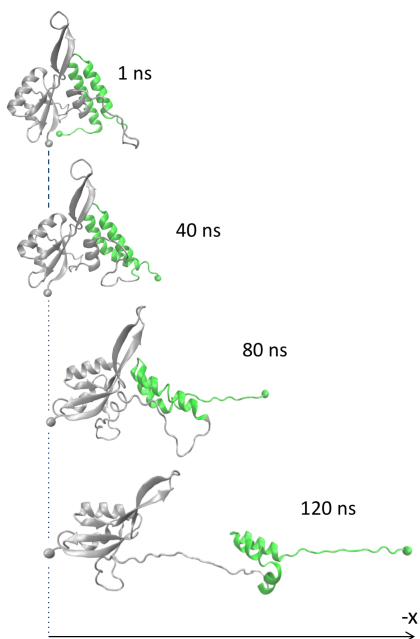
Figure 3.16 Snapshots of protein conformations at 1 ns, 40 ns, 80 ns, and 120 ns from the SMD pulling and relaxation trajectory.
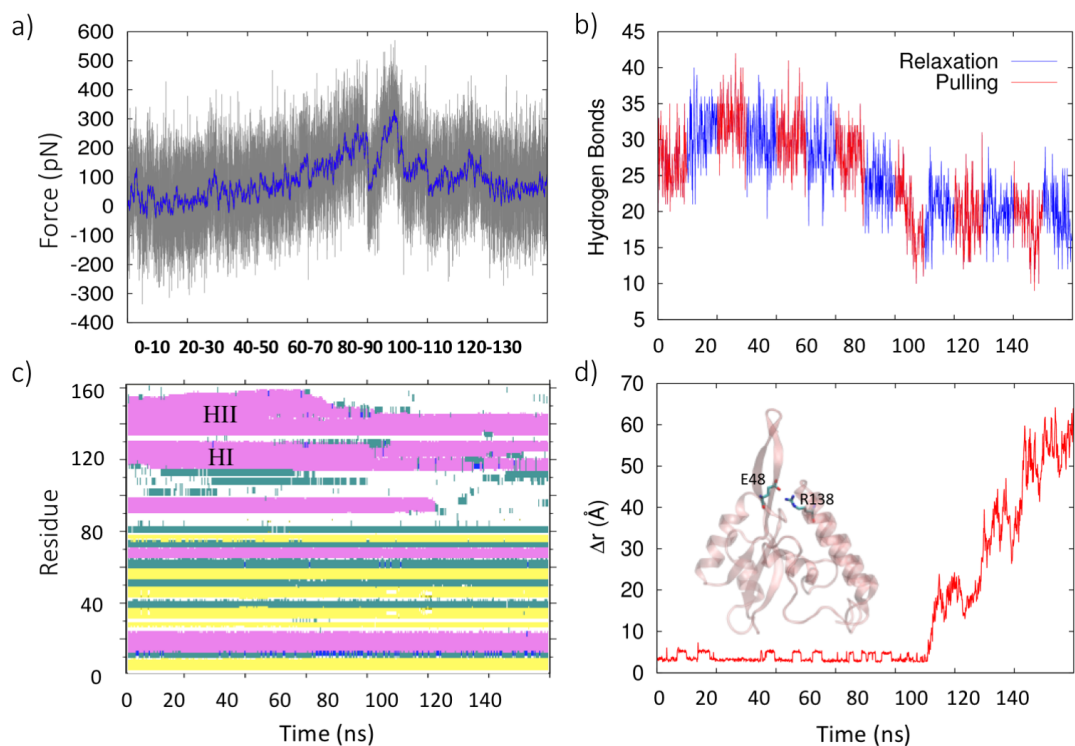
Figure 3.17 Various dynamic and structural parameters as a function of time during SMD: a) force profiles during pulling intervals, b) number of hydrogen bonds during SMD (red) and relaxation (blue) c) secondary structural conversion, and d) Distance (Δr) between the E48 oxygen and the R138 nitrogen.

The SMD simulation is set up so that the dummy atom (attached to the CTD C_α SMD atom) travels at constant velocity while being pulled requires a pulling force that varies in strength in response to the protein's resistance. The helical form of the CTD is stabilized by interdomain interactions with the NTD, which include the E48-R138 salt-bridge (involving the CTD Helix II) and other hydrophobic and hydrogen bond (involving both CTD helices). The force profile displayed in Figure 3.17a shows a general increase in the pulling force with time until a large drop at 80 ns. The force rises again, and another significant drop in force is observed around 90 ns that coincides with breaking of major interdomain interactions, including the E48-R138 salt-bridge and some

hydrogen bonds. Hydrogen bond analysis (Figure 3.17b) shows an initial increase in hydrogen bonds due to rearrangement of the CTD-tail segment (residues 158-162). Figure 3.17c displays the changes in secondary structure of all residues during the pulling and relaxation. At around 60 ns, a significant reduction in helicity is observed for the CTD Helix II, whereas Helix I is relatively unperturbed and retains most of its secondary structure. This is consistent with our results presented above in Figure 3.10. As shown in Figure S6 (34), the interdomain interacting hydrophobic residues F33-F130 separate in two stages. They begin to separate and become more exposed to the solvent (increasing solvent accessible surface area - SASA) around 80ns, which allows solvent molecules to penetrate the interfacial region, weakening the interdomain interactions (104) facilitates breaking of the E48-R138 salt-bridge interaction around 90 ns (increasing Δr in Figure 3.17d) and increases the SASA of E48-R138 (Figure S6) (34). As the E48-R138 salt-bridge breaks, the CTD moves far away from the NTD and the F33-F130 separate further, resulting in a larger SASA.

3.8 Calculation of Potential Mean Force Using JE

I performed nine independent 60 ns SMD simulations (pulling only) to calculate the potential of mean force (PMF) using the Jarzynski equality, which relates the non-equilibrium work done (W) with the PMF as $e^{-\beta\Delta F} = \langle e^{-\beta W}\rangle$ where, $\langle\ \rangle$ refers to the ensemble average. The free-energy profile can be obtained from

$$\Delta F = \langle W\rangle - \frac{\langle W^2\rangle - \langle W\rangle^2}{2kT} \qquad (3.8.1)$$

where k is the Boltzmann constant. The harmonic force applied to the virtual spring attached to the C-terminal C_α was integrated over the pulled distance to calculate the work done over the simulation time using $W = \int_0^T v\, f(t)\, dt$, where v is the pulling speed, $f(t)$ is the force as a function of time, and $v\, dt = dx$ is a small increment in the pulling distance. All nine simulations were used in eq. 3.8.1 to calculate the average and the standard deviation for calculating the free-energy profile. Figure 3.18 shows the resulting free-energy profile as a function of M1-$C_\alpha \rightarrow$L162-C_α end-to-end distance, d_{ee}. State B at 40 Å (~15 ns) represents the native state minimum after the rearrangement of the 158-162 CTD tail segment residues from structure A, and before any structural changes occurred at the NTD-CTD interface. From B to C, the initial transient interactions of the tail segment with the NTD are broken, followed by the rugged energy landscape from state C to D representing multiple structural transitions. The minimum at D around 110 Å (~ 56 ns) corresponds to the structure in which the CTD is mostly detached but the E48-R138 salt-bridge is still intact. The final, major barrier between D and E corresponds to breaking of the salt-bridge interaction. This is followed by the complete detachment of the CTD at E, where all interdomain contacts are lost. At this stage, the CTD-helices still retain their secondary structure as a helix-bundle. The PMF in Figure 3.18 calculated from SMD describes the results of the pulling simulations that are relevant for AFM experiments.
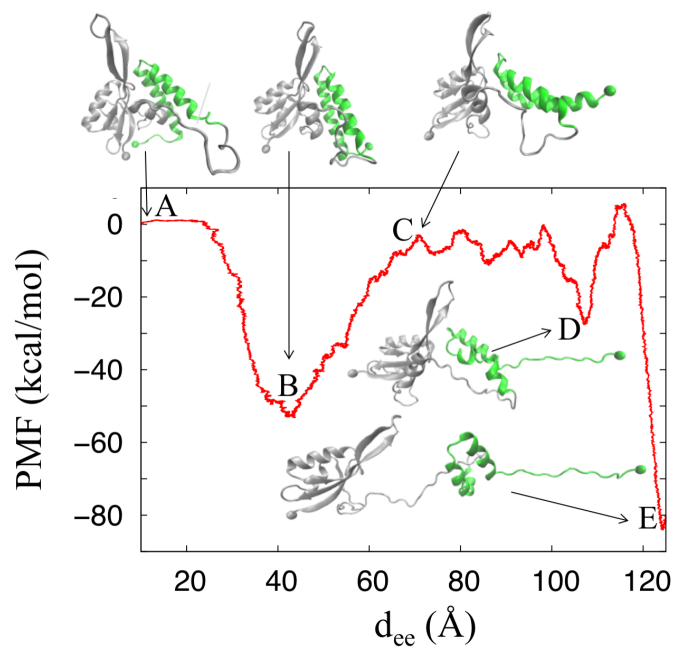
Figure 3.18 PMF as a function of end-to-end (Cα-Cα) distance between M1 and L162.

4. INTERDOMAIN SALT-BRIDGES IN THE EBOLA VIRUS PROTEIN VP40 AND THEIR IMPORTANT ROLES

4.1 Dynamical Analysis of the Closed-form Structure

MD simulations on the wild-type (WT) VP40 as well as on several mutants were performed to investigate the interdomain interactions that are important in stabilizing the closed-form structure, the sequence of steps that allow the CTD to disengage from the NTD, and the interactions that facilitate domain-domain association into the closed form. The effects of the removal of some important salt-bridges that were identified computationally were further investigated experimentally by live cell imaging.

In many proteins, the CTD-tail segment interactions affect the protein conformational dynamics and play important functional roles (105-108). Though not resolved in the x-ray crystal structure, the VP40 CTD-tail segment (residues 320–326) has been shown to be important for regulating oligomerization. This is evidenced by the fact that the removal of the CTD-tail segment leads to spontaneous oligomerization of VP40 and membrane binding *in vitro* (109). This is further supported by the observation that the VP40 in its closed-form is stabilized by latch-like interactions between an unstructured region of the N-terminal domain and the C-terminal tail region(110). In order to investigate the NTD-CTD interactions involving the CTD-tail segment, I modeled this segment by inserting the missing residues 320–326. The modeled CTD-tail segment in the closed-form structure was a random coil structure with no contacts with the NTD. Starting from this conformation, I performed eight independent MD simulations for 50 ns each (total of 400 ns) and monitored the CTD-tail interactions with the NTD. I observed the latch-like interactions with the NTD via a hydrophobic patch as

well as the D45-K326 salt-bridge interaction in four of the eight simulations. Figure S1 (111) displays the representative trajectories that show the formation of the latch. The interdomain contacts involved in the latch are highlighted in Figure S2 (111). Once formed, the latch is stable for several nanoseconds (~3 to 30 ns) before breaking. The flexibility of the CTD tail segment resulting from the transient stability allows the separation of the CTD from the NTD, as required for higher order oligomerization of the VP40 dimers. In addition, membrane association may make it easier for the latch to release(110) to facilitate the VP40 hexamerization in the membrane.

In order to explore the interdomain as well as intra-domain interactions and residue-residue contacts, I used 10 ns of the trajectory from the region with the intact latch and calculated a contact map for the VP40 structure using Carma (112). Figure 4.1a shows the resulting contact map obtained with a distance cut-off of 5 Å. Residues 44–201 belong to the NTD and residues 202–326 belong to the CTD. The lower right block displays the interdomain contacts. This contact map allows us to identify all interdomain interactions and in this paper, I focus on interdomain salt-bridges and hydrophobic interactions. Salt-bridge interactions were identified with a cut-off distance of 4 Å and are shown in Figure 4.1b. A total of five interdomain salt-bridge interactions were obtained: D45-K326, D56-K256, E76-K291, R148-D312, and E160-K212. As discussed later, the D45-K326 salt-bridge is important in guiding domain association for the open- to closed-form structure.
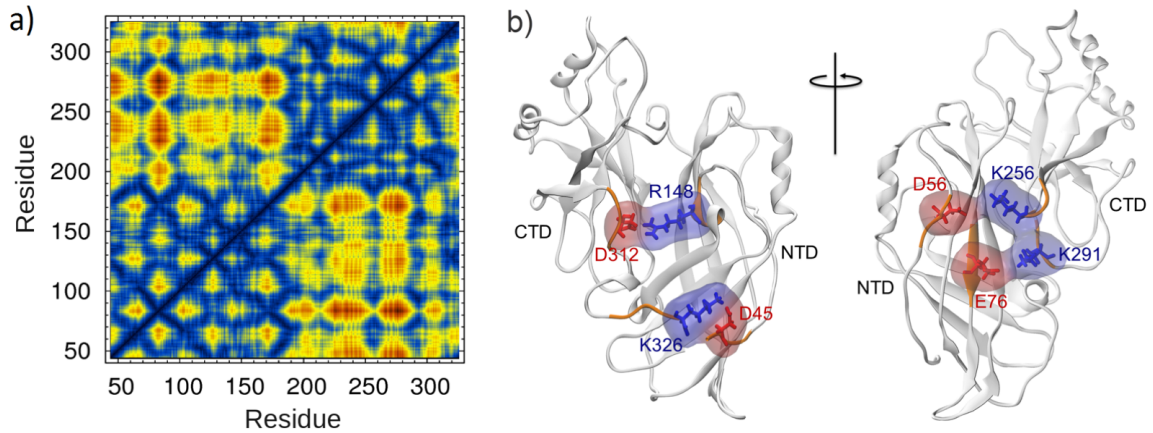
Figure 4.1 (a) Amino acid contact map of the VP40 protomer. The blue regions (<5Å) in the lower right block display the NTD-CTD interdomain contacts within 5Å distance (b) VP40 inter-domain salt-bridges. Red are negatively charged amino acids and blue are positively charged.

4.2 Forced pulling and Dissociation of Domains

I used Steered Molecular Dynamics (SMD) to investigate the relative strengths of interdomain interactions by pulling the center of mass of the C-terminal domain (CTD) away from the harmonically restrained N-terminal domain. For SMD simulations, the last frame of the 5-ns equilibrium NVT simulation was taken as the initial structure and solvated again in a larger rectangular box to allow room for extension. In order to focus on the relative motion of the NTD and the CTD, translation and rotation was prevented by harmonically restraining the protein during heating and equilibration. After that, the NTD was kept harmonically restrained and the center of mass of the CTD was pulled away from the NTD using SMD. The protein was aligned so that the NTD-CTD interface is approximately perpendicular to the pulling direction (−x). Stages in the resulting domain separation due to the SMD pulling are shown in Figure 4.2.
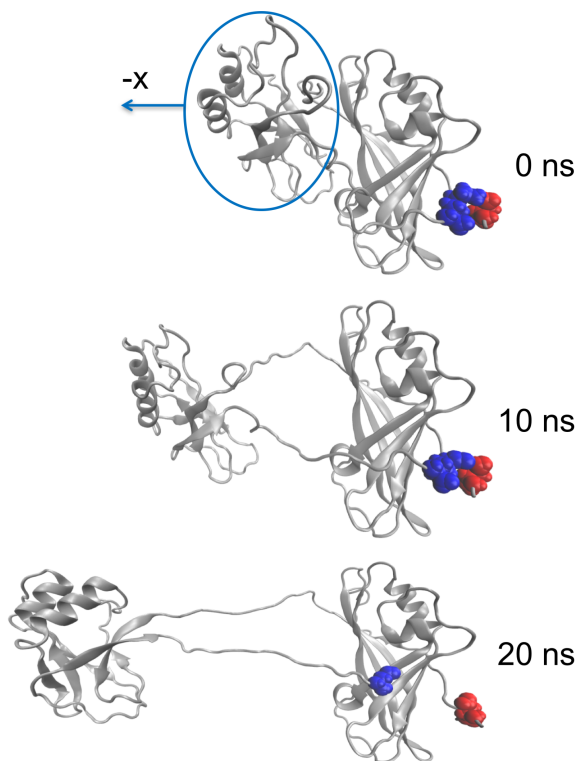
Figure 4.2 Domain separation during the SMD simulation in which the CTD (circled) is pulled away from the NTD. Residues D45 and K326 that participate in the interdomain salt-bridge are highlighted. This salt-bridge that participates in the latch remains intact for 20 ns and restricts CTD–NTD relative reorientation even when the domains are well separated.

I performed seven 25-ns SMD simulations and averaged the force profiles for pulling the CTD center of mass away from the harmonically restrained NTD. During the initial stage of pulling, the SMD force continuously increase because all NTD–CTD interactions are intact and undergoing stress as shown in Figure S3 (111) . I monitored the interdomain interactions during the pulling for a representative trajectory and the results are displayed in Figure 4.3. Figure 4.3a shows the SMD force necessary to move the center of mass of the CTD at a constant speed of 2 Å/ns as it disengages from the NTD. The SMD force curve initially increases and reaches a peak of 700 pN around 4 ns

(region A in Figure 4.3a) and then suddenly drops to approximately 300 pN. Just before

this sharp drop in the SMD force, the solvent accessible surface area (SASA) at the

interface of the NTD and CTD suddenly increases (Figure 4.3c) by 800 $Å^2$, suggesting

the separation of the interdomain hydrophobic interface and the exposure of hydrophobic

residues to water. This may be precipitated by the breaking of hydrogen bonds (Figure

4.3b) and the breaking of the R148–D312 salt-bridge (Figure 4.3d) that occur at

approximately 1.5 ns. In addition, two other salt-bridges, D56-K256 and E76-K291, are

each stretched and weakened before 1.5 ns. Thus, it appears that the weakening of

specific salt-bridges and hydrogen bonds permits enough flexibility to allow water to

enter into the hydrophobic domain-domain interface.
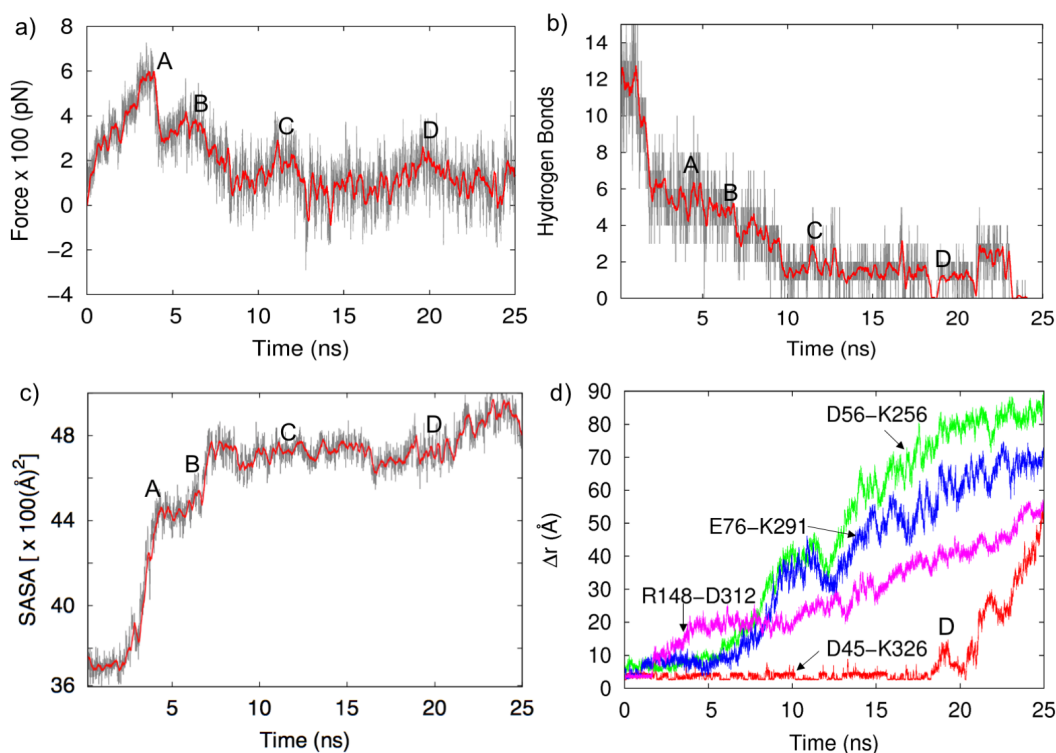
Figure 4.3 a) SMD force necessary to keep the CTD moving at constant speed away from
the NTD (gray curve: full data, red curve: time averaged data with 200 ps windows) b)
Time evolution of the number of hydrogen bonds in the interface, c) solvent accessible

surface area (SASA) of hydrophobic residues at the NTD-CTD interface, d) various interfacial salt-bridges distance.

Once the hydrophobic interactions begin to be disrupted, the SASA increases rapidly and the rest of the hydrophobic interactions are disrupted. This allows the domains to further separate, and additional hydrogen bonds are broken (region B in Figure 4.3b). Separation of the domains becomes easier, the SASA increases and the D56-K256 and E76-K291 salt-bridges in the interface are fully broken by 8 ns. This sequence of bond disruption is consistent with the work of Bhaskara (113) and Waldburger (114) who emphasize the importance of hydrophobic interactions in stabilizing proteins. The NTD residues 72, 74, 95, 97, 125, 132, 156, 161, 162, 186, 187, 189, 191 and the CTD residues 283–286, 288–290, 293, 295, 317, 322, 323 form a strong hydrophobic interface. As shown in Figure 4.4a, at 0 ns the interface is tightly packed and provides little access to water, By 4 ns (Figure 4.4b), the domains have begun to separate and gaps open at the hydrophobic interface, exposing SASA as quantified in Figure 4.3c. The SASA averaged over seven different trajectories in Figure S3 (111) also shows the same general trend.
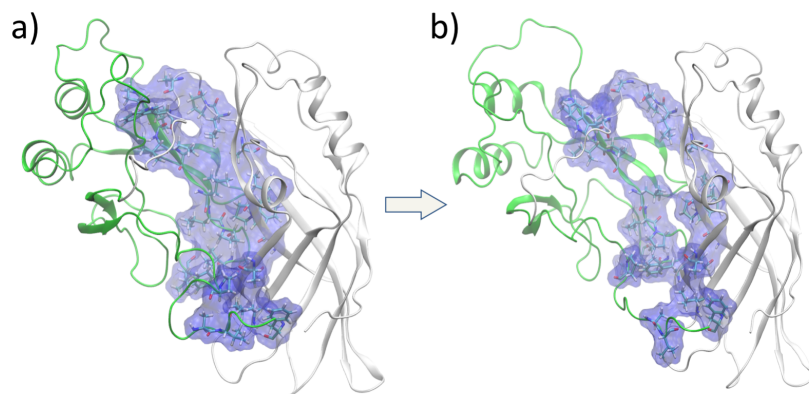
Figure 4.4 Hydrophobic residues at the domain-domain interface. (a) Initial closed-form. (b) After the domains begin to separate (~4 ns), gaps open at the hydrophobic interface, exposing SASA as quantified in Figure 4.3c.

After 8 ns, the few remaining interdomain hydrogen bonds are broken (B→C in Figure 4.3b). Though the domains are always connected by the linker segment, after 10 ns, the only other interaction between the domains is the D45-K326 salt-bridge. Residue K326 is at the end of the flexible tail of the CTD. As the CTD separates from the NTD, this salt-bridge remains fully intact and acts as a connecting latch until 20 ns (region D in Figure 4.3b) as displayed in Figure 4.2. I will show later that the extended latch-like D45-K326 salt-bridge facilitates association of the separated domains into the closed-form structure.

4.3 Salt-bridges and the interdomain stability

In order to understand the role of specific interaction in stabilizing the closed-form structure, I mutated several important residues and pulled the CTD until the CTD is disengaged from the NTD. Figure 4.5 shows for different mutants, the force necessary to pull the CTD away from the NTD at constant speed. Since I was only interested in

interdomain interactions, I pulled the CTD at a speed of 5 Å/ns for 4 ns and stopped after the interface begins to open. For each mutant, 10 similar runs were performed and the average force profile was calculated and displayed in Figure 4.5a. Although the differences are small, the force profile for the WT generally shows more resistance to pulling compared to the mutants. The K291E mutation removes the 76-291 salt-bridge that is present in the WT, and as seen in Figure 4.5a, results in less force required to separate the domains. This shows that the K291 salt-bridge is important in stabilizing the closed-form structure. The comparison of hydrogen bonds in the WT and K291E as a function of time during SMD is displayed in Figure 4.5b. The number of hydrogen bonds is reduced by half in the initial state of the mutant, also weakening the domain-domain interactions and contributing to the destabilization of the closed-form structure.
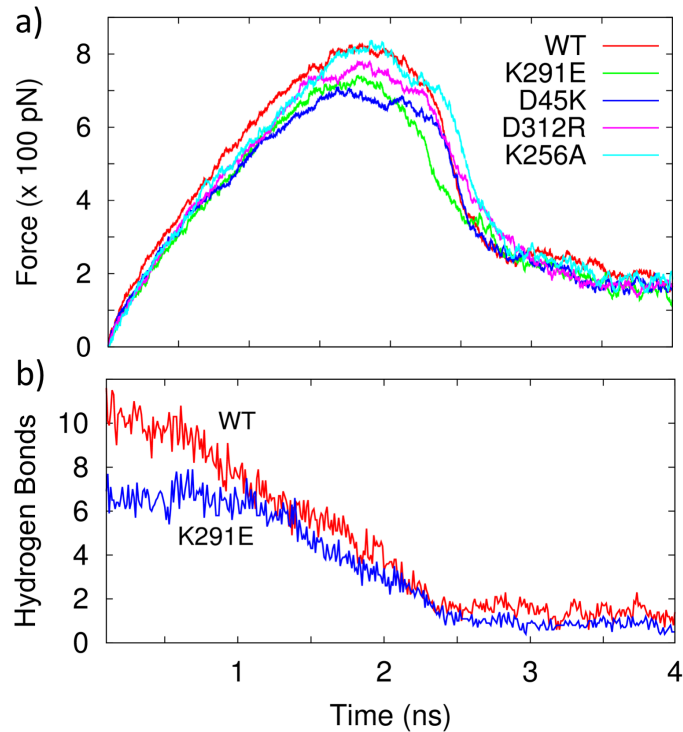
Figure 4.5 (a) Force required for separating the CTD away from the NTD at constant speed for the WT and different mutants. (b) The K291E mutation prevents formation of the 76-291 salt-bridge and also decreases the number of interdomain hydrogen bonds, further destabilizing the closed-form structure.

The D312R mutation removes the R148-D312 salt-bridge found in the WT. Removal of the 148-312 salt-bridge weakens the domain interface and reduces the force necessary to separate the domains (Figure 4.5a) compared to the WT. Another salt-bridge, D56-K256, is buried and is not a major contributor to interdomain stability (114). Upon alanine mutation K256A, which removes this salt-bridge, there is little change compared to the WT in the amount of force necessary to separate the domains, as shown in Figure 4.5a.

4.4 VP40 mutant cellular localization analysis

The effects of some mutations in the salt-bridge network were assessed in live HEK293 cells using an established EGFP-tagged VP40 system (115). The mutants assessed included D45K, R148A, E160A, and K291E, which should abrogate the aforementioned salt-bridges, as well as a control K90A. Lys90 was not shown to participate in interdomain salt-bridges and previously was shown to form virus like particles (VLPs) and localize to the plasma membrane inner leaflet, similar to WT VP40(116) in CHO-K1 cells. EGFP-VP40 localized to the PM with characteristic high intensity membrane protrusions (Figure 4.6a) (115), (116). Further quantification using a MATLAB script revealed that K291E increased the %PM localization of VP40 by ~20% (P<0.07) compared to the WT (Figure 4.6). These results suggest that removing the 76-291 salt-bridge may alter the time scale of VP40 structural transitions but without impairing the important interdomain associative forces likely provided by other salt-bridge and hydrophobic interactions. The D45K had a similar PM localization as K291E. It is important to note that the p-value for these measurements was >0.05. Because these experiments are done at high concentrations of VP40 at a significant timeframe post-transfection, time-dependent assays may be needed to detect the time dependent changes of the VP40 assembly to better assess the statistical significance of each mutation on plasma membrane localization and oligomerization.
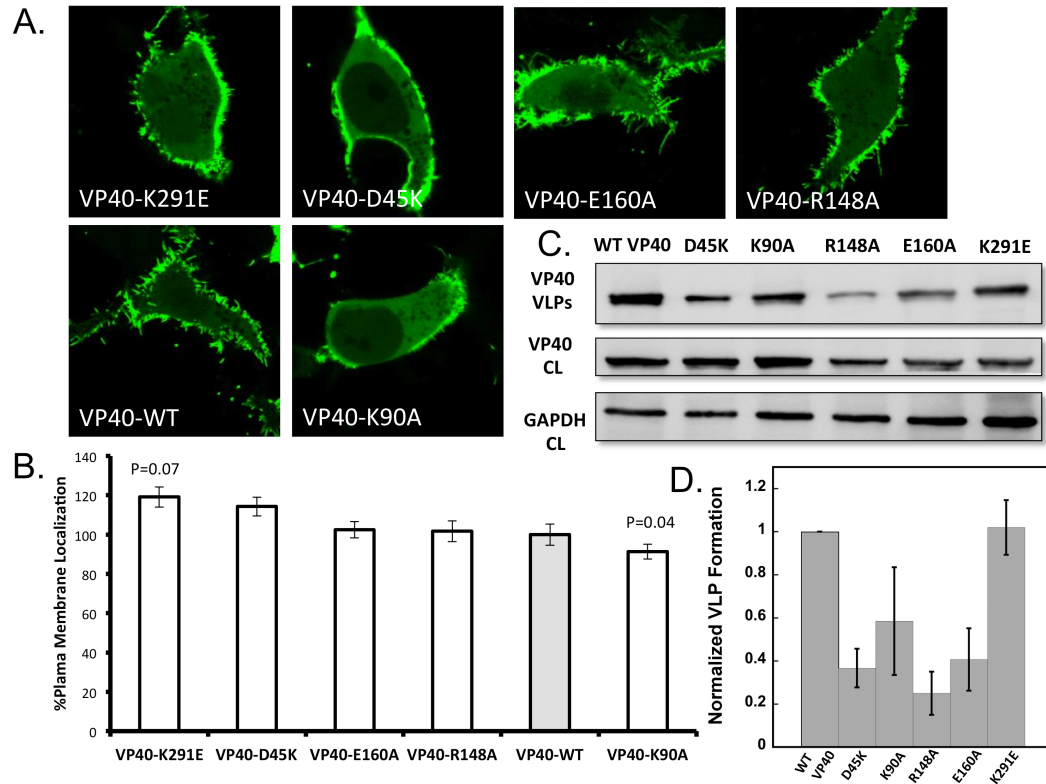
Figure 4.6 Cellular localization and VLP analysis of interdomain NTD-CTD salt-bridge mutants. (a) Representative images of VP40-EGFP constructs imaged 18-24 hours post transfection into HEK293 cells. All images were acquired with a zoom of 8.0, therefore the frame size is 20 μm and equivalent for all images. (b) Average plasma membrane localization for each construct. Image analysis was performed in MATLAB to determine the %PM localization for each. Error bars are ± the standard error of the mean. At least 28 cells were imaged per construct over three independent experiments on three different days. (c) Representative blots of VP40 and respective mutations in VLPs (top panel) or cell lysates (middle panel) or GADPH from cell lysates (bottom panel). (d) WT-VP40 VLP formation was set at 1 and each mutant was compared to WT-VP40 as fraction a of VLP formation. The bars represent the averages ± standard errors of the means for three independent experiments quantified using ImageJ. P-values for mutations were 0.002 for D45K, 0.001 for R148A, and 0.004 for E160A. K90A and K291E P-values were not statistically significant.

R148A or E160A did not contribute significantly to changes in PM localization. As R148-D312 was the first salt-bridge to break in the domain separation simulation, R148 likely plays a lesser role in mediating domain separation. As expected, K90A appeared similar to WT and exhibited plasma membrane localization albeit slightly less

69

than WT. As a whole, this live cell mutational analysis illustrates the differential functions of salt-bridge networks in VP40 conformational regulation and opens the door to large-scale cellular and in vitro analysis of VP40 conformational dynamics.

PM localization alone is not necessarily indicative of the oligomerization state of VP40 where hexamers and larger oligomers have been detected (117), (115), (118), (116). Thus, I monitored VP40 oligomerization state of EGFP-VP40 and respective mutations using N&B analysis as previously reported (117), (115), (118), (116). EGFP-VP40 and all mutations employed exhibited a similar extent of oligomerization (111) 18-24 hours after transfection suggesting mutations still permit assembly of VP40 in some fashion. However, the extent and type of oligomerization is difficult to discern with this technique, as N&B measures the size of oligomers based upon the assembly of each additional EGFP containing monomer. Thus, N&B is not able to decipher between a VP40 filament that provides the structure for viral egress and a filament that may be slightly twisted or less rigid based upon salt-bridge changes.

To determine if any of the VP40 salt-bridge mutations identified here influenced VLP formation, I measured VLP formation for WT and respective mutations using a VP40 antibody. VLPs were collected and measured 48 hours post-transfection using cell lysate expression of VP40 and GAPDH (total protein/cell content) as a control for each construct. Notably, several salt-bridge mutations had a statistically significant effect on VLP formation. D45K, R148A, and E160A all reduced VLP formation while K291E had a similar amount of VLPs as WT VP40. K90A showed a slight reduction in VLP formation but statistical analysis demonstrated this reduction was not statistically

significant. Thus, despite a similar level of PM localization and oligomerization, disruption of the D45-K326, R148-D312, or E160-K212 salt-bridges reduced VLP formation suggesting these interactions have an important role in formation of VP40 structural assemblies that mediate efficient VLP formation. Notably, previous mutation of K212 reduced VLP formation (119) by ~40%, further supporting the conclusions in this study.

4.5 The Role of Salt-Bridges in Domain association

The various functions of VP40 require that the NTD and CTD domains be able to exist in both a closed, associated form as well as an open, dissociated form. In both the closed and open forms, a flexible segment consisting of residues 187–201 links the domains (120). In general, multidomain proteins fold to the functional states by association of already folded domains (121). The rate-limiting step for the complete folding is often the correct matching of the domains at the interface, similar to protein-protein association of structured proteins that is generally "diffusion-limited" (122), (123), (124), (125). Dissociated domains that are connected only with a flexible linker will have both rotational and translational degrees of conformational freedom and the linker flexibility provides a large entropic barrier for correct interface matching. VP40 seems to have reduced this problem by introducing a second interdomain interaction to guide the domain association by using a handshake-like latch interaction of the CTD tail with a NTD flexible segment via the D45-K326 salt-bridge. The flexible nature of both the long CTD tail ending in K326 and the unstructured NTD segment containing the residue D45 increases the probability that these residues will interact in the open-form of

VP40 and form a salt-bridge. This interdomain salt-bridge handshake, along with the linker segment, reduces the number of possible relative orientations of the two domains as they approach and provides specificity for domain association. This increases the probability for a productive encounter that results in successful association of the CTD and the NTD. This is akin to the binding mechanism of flexible, structurally disordered proteins.[126],[127],[128] The contact of the CTD tail section with a loop in the NTD is stabilized by both the D45-K326 salt-bridge as well as hydrophobic interactions between the CTD-tail amino acids and amino acids in the NTD loop.

The initial structure for the association runs, displayed in Figure 4.7a was taken from a frame in the SMD dissociation simulation in which the domains were separated. This dissociated structure contained the latch-like CTD (tail)-NTD (loop) hydrophobic interactions and the D45−K326 salt-bridge as shown in Figure S2. Even with the intact latch-like interaction, the timescales for domain association may still be greater than microseconds. In order to achieve domain association in a reasonable computational time, the initial separated structure was carefully chosen so that the residues in the flexible linker segment were minimally perturbed. Furthermore, only the 15 residues in the linker segment (residues 187−201) and the last seven residues (320−326) in the CTD tail were allowed to be fully flexible and all other residues in both domains were harmonically restrained. Figure 4.7b displays the final frame of the association run at 100 ns, along with the superimposed native structure of the closed-form as a reference to display the good similarity. The associated structure was within 2.6Å rmsd compared to the native structure. The time course of the distance d_{cm} between the center-of-mass (cm) of the NTD and the cm of the CTD in Figure 4.7c gives the approximate time of association. At

approximately 12 ns, the distance is ~25 Å, which is the NTD(cm)–CTD(cm) distance in the WT VP40.
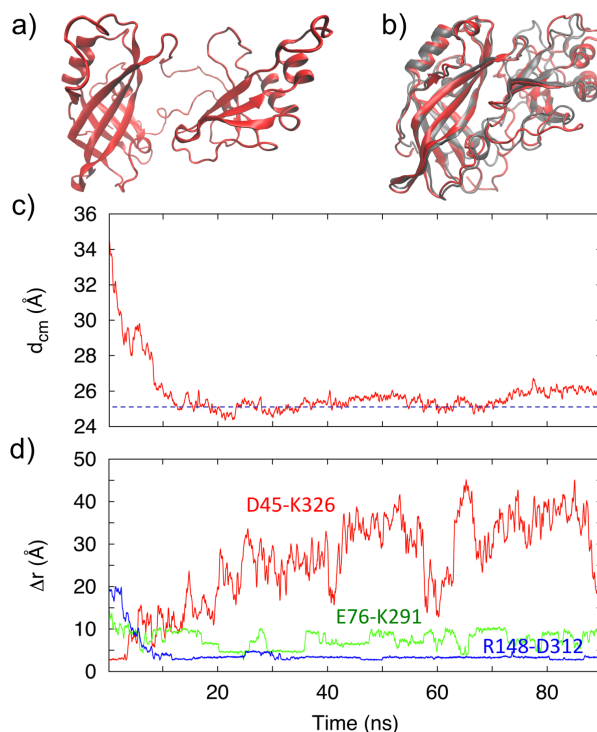
Figure 4.7 (a) Initial structure of the open-form of the VP40 used in the association MD simulations. (b) The final closed-formed structure at the end of the 100-ns association simulation, along with the PDB structure of the closed-form for reference. (c) Separation between the NTD center of mass (cm) and the CTD cm. The horizontal line at approximately 25Å represents the cm separation in the WT. (d) Time course of the separation between the amino acids that make each of the three important salt-bridges during domain-domain association.

I investigated the role played by the three salt-bridges that I found to be especially important in domain association or in stabilizing the closed-form: R148-D312, E76-K291 and D45-K326. Figure S4 (109) shows that initially in the dissociated form, only the D45-K326 salt-bridge is intact to guide the early stages of association. During the association, the D45–K326 salt-bridge is broken, but the E76-K291 and R148-D312 salt-bridges have formed to finalize association and stabilize the closed-form structure.

More detailed information of the time course of these three salt-bridges is given in Figure 4.7d. The D45-K326 salt-bridge provides initial specificity for association when the domains are fully separated and starting to approach each other. This D45-K326 salt-bridge breaks at 4 ns, which from Figure 4.7a is at a stage in which the domains are almost in contact, but still separated by 4Å compared to the fully associated conformation. At this stage, the domains are close enough that the other two salt-bridges, E76-K291 and R148–D312, have already started to form. These two salt-bridges are critical for guiding the domains during the middle stage of the association process after the D45-K326 salt-bridge has broken, but before short range interdomain interactions are active. For the CTD, I computed the average C_α rmsf when the latch with the D45-K326 salt-bridge was intact, and compared it to the rmsf of the CTD in when the latch is broken. As can be seen in Figure S5 (111), the rmsf values of the CTD residues that are at the domain interface are smaller and therefore the CTD is more constrained when the CTD tail makes the salt-bridge with the NTD loop.

The importance of other interactions in the association process was studied by comparing the SASA during dissociation and association as displayed in Figure S6 (ref. (111)). The SASA during the association begins to significantly drop around 8 ns, which is well after the D45-K326 guiding salt-bridge has broken. At 8 ns, the domains are very near each other but not yet fully associated. Therefore, as expected for the short-range hydrophobic interactions, they are helpful for guiding the final stages of association and are important for stabilizing the closed-form structure. The guidance from hydrogen bonds and hydrophobic interactions only at the latter stages of association highlights the importance of salt-bridges in the early and intermediate stages of the association process.

5. THE VP40 DIMER AND HEXAMER INTERACTIONS WITH THE PM

5.1 Association of VP40 Dimer with the PM

A recent study revealed that the binding of the human antibody HuScFv inhibited egress of Ebola VLPs from mammalian cells. One of the predicted interfaces between the antibodies and eVP40 was the shorter helix (H1) (residues 265-272) in the c-terminal domain (CTD) (129), suggesting that the shorter helix H1 plays a critical role in the Ebola virus life cycle. Investigations on other systems have shown that helices can mediate interactions between a protein and a membrane. Prakash et. al. (130) performed AAMD simulations that showed that the oncogenic mutant K-Ras interacts with anionic lipids with a preferred helix orientation that was parallel to the PM. Another study pointed out the important role of a CTD α-helix in the binding of the heterodimer actin-capping protein (CP) to a membrane (131). Combined molecular dynamics and experimental studies (132) on the peripheral membrane protein Osh4 in yeast *Saccharomyces cerevisiae* revealed a membrane binding region, an amphipathic lipid packing sensor (ALPS)-like motif consisting of a helix, with a major role in stabilizing the protein membrane interactions. In this work, I sought to understand the role of helix H1 and other residues in the binding of the eVP40 dimer with the PM using all-atom molecular dynamics (AAMD) simulations. To elucidate the atomic-level details of the association mechanism of eVP40 with the plasma membrane, I performed simulations with membranes consisting of different combinations of POPS, $PI(4,5)P_2$ (i.e. PIP_2), and POPI lipids. I found that lysine residues in loop regions of the dimer play an important role in the association process and interact with select lipids in the PM. The eVP40 dimer

easily associated with a PM consisting of anionic lipids POPS, POPI, and PIP_2. Our results also showed that after the initial association, the H1 α-helix that is in contact with the membrane can reorient itself so that it is aligned parallel to the PM. This parallel orientation can then facilitate oligomerization into hexamers that can change the shape of the membrane. In contrast, for a membrane without POPS, POPI, or PIP_2, the dimer is only transiently attached to the PM and in an orientation that does not encourage hexamerization.

Figure 5.1 below illustrates the eVP40 dimer. Residues that have been experimentally suggested as important for interactions with the PM (K224, K225, K270, K274, K275) are colored in blue. The CTD helix H1 on each monomer consisting of residues 265-272 is colored in red and I investigated its alignment with respect to the membrane. In addition to these residues, our MD study revealed that polar residues colored in green and lysine residues K221 and K236 are also important in the formation of the eVP40-PM complex. Chemical interventions that perturb the above mentioned orientation of the eVP40 dimer relative to the PM might disrupt VLP formation.
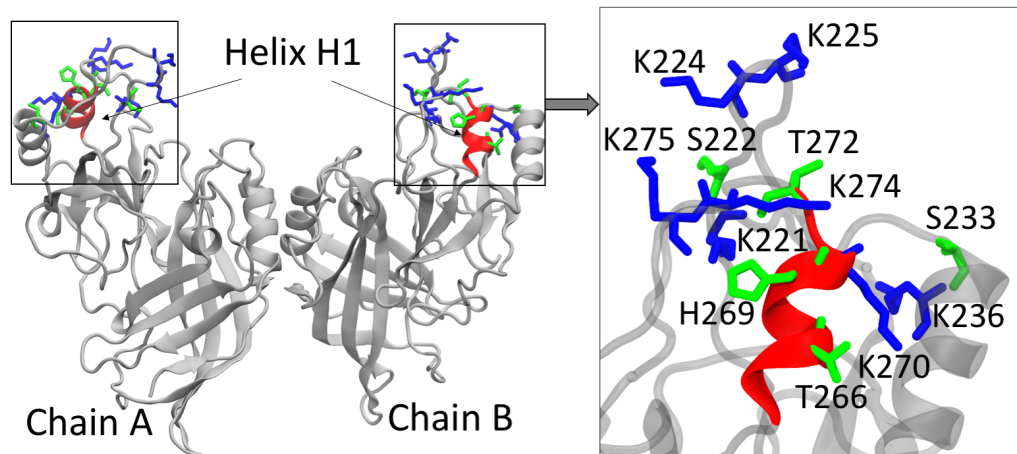
Figure 5.1 The eVP40 dimer with the H1 helices on each monomer highlighted in red. Lysine residues that are important in associating with the PM are colored in blue and important polar residues are colored in green.

For the POPS/POPI and the POPS/PIP$_2$ systems, the MD results agree with the previous experimental result (12) that suggested the strong electrostatic interactions between the positively charged lysine residues of the eVP40 dimer and negatively charged head groups of the PM are the major force in binding the eVP40 dimer to the PM. Figure 5.2 below displays snapshots of the time evolution of the systems. Figure 5.2 shows the initial and configurations of the dimer relative to the membrane for membrane systems a) POPS/POPI b) POPS/PIP$_2$ and c) no POPS/POPI/PIP$_2$. Initially, the eVP40 dimer was placed ~10Å below the lower leaflet of the PM so that at t=0 ns there are only weak interactions between the protein and the membrane. Figure 5.2a and 5.2b shows the orientation of the dimer when it has associated with the membranes containing POPS/POPI and POPS/PIP$_2$ respectively. Supplemental movie S1 (148) shows the dimer approaching and associating with the membrane. Important lysine residues on the dimer and important POPS and POPI lipid molecules in the membrane are highlighted, and

discussed below. In contrast, Figure 5.2c shows the orientation of the dimer when it has weakly associated with the membrane containing no POPS/POPI/PIP$_2$. Supplemental movie S2 (148) shows the dimer approaching the membrane with no POPS/POPI/PIP$_2$ and weakly associating at one end in the configuration depicted in Figure 5.2c.
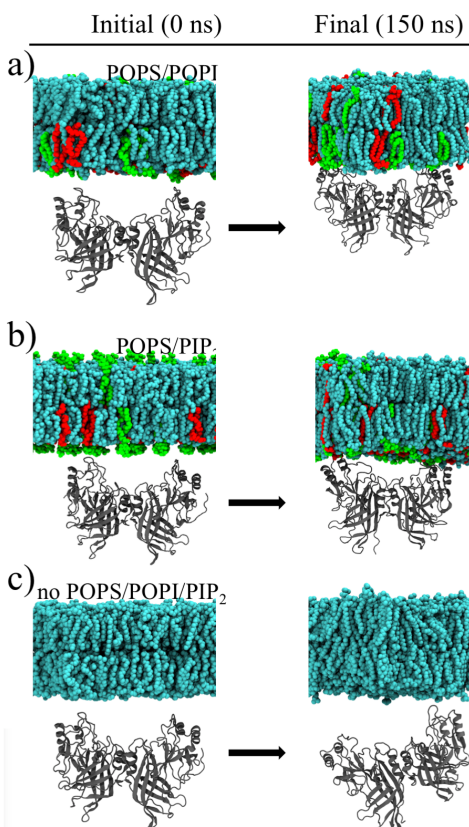
Figure 5.2 MD Snapshots of the eVP40 dimer and plasma membrane show the initial and final configurations of the dimer relative to the membrane for membrane systems a) POPS/POPI b) POPS/PIP2 and c) no POPS/POPI/PIP2

5.1.1 eVP40-PM association kinetics

In order to investigate the atomic-level details of the association process, I calculated the time dependence of two different parameters as displayed in Figure 5.3 that provide information on the approach of the dimer towards the membrane. The parameter d_{cm} represents the separation distance between the center-of-mass (cm) of the dimer and

the membrane. The C_α atoms in eVP40 were used to calculate the cm of the dimer, and the phosphorous (P) atoms in the lower leaflet of the PM were used to calculate the cm of the membrane. Initially, d_{cm} ~40Å, which is equivalent to a separation of ~10Å between the membrane and the nearest atoms of the dimer. For the systems with membranes containing POPS/POPI or POPS/PIP$_2$, d_{cm} decreased by approximately 10Å during the first 20 ns of the MD simulation as the dimer approached the membrane. Once the dimer associated with the membrane in the orientation shown in Figure 5.2a and 5.2b, d_{cm} remained stable at ~30Å. In contrast, the system containing the membrane without POPS/POPI/PIP$_2$ had a d_{cm} that never decreased. The orientation of the dimer shown in Figure 5.2c requires the dimer to rotate significantly compared to its original orientation. This rotation occurs over 60ns and moves the cm of the dimer farther from the membrane, which explains why d_{cm} increases for this system. Similar, timescales are displayed in Figure 5.3b, which shows small-scale rearrangement of atoms in the dimer due to interactions with the membrane. Figure 5.3b shows the root-mean-square-difference (rmsd) in position, compared to their initial position after minimization, averaged over all C_α in the dimer, as a function of time. The protein rmsd is calculated with rotations and translations subtracted out. For the systems with membranes containing POPS/POPI or POPS/PIP$_2$, the rmsd underwent major changes in the first 20 ns. For the system with the membrane without POPS/POPI/PIP$_2$, the rmsd changed more gradually and stabilized after 60 ns.
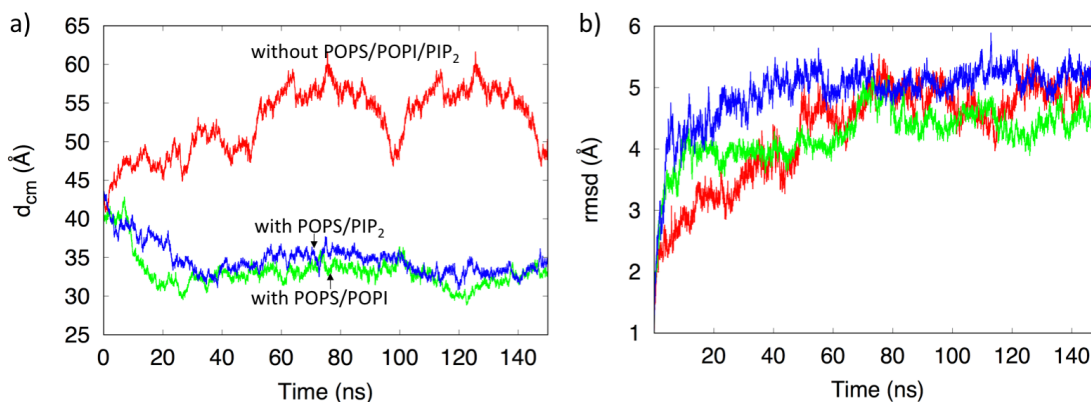
Figure 5.3. MD timescales for association of the eVP40 dimer with the plasma membrane. a) The parameter dcm represents the separation distance between the center-of-mass of the Cα atoms in eVP40 and the P atoms in the lower leaflet of the PM. b) For the dimer, the rmsd averaged over all C_α as a function of time. Translations and rotations of the dimer are subtracted out.

The association of the dimer with the membrane involves interactions of lysine residues in the loop segments with specific lipids in the PM. I discuss the details of these interactions in the next section. After that section, I discuss additional binding involving the H1 helix, which occurs on longer time-scales. The information on the H1 helix re-orientation discussed later is derived from an 850 ns simulation that I performed on the system with POPS/POPI.

5.1.2 eVP40-PM interactions

As hypothesized experimentally (12), I found that the association of the eVP40 dimer with the PM was dominated by electrostatic interactions. The initial docking was primarily due to the interaction of positively charged lysine residues in the CTD loop regions with the anionic heads of POPS, POPI, and PIP_2 lipids in the membrane. This is followed by the interaction of other residues in the CTD with the membrane. The number

of hydrogen bonds between eVP40 and the PM as a function of time are plotted in Figure 5.4. For the strongly associating POPS/POPI and POPS/PIP$_2$ systems, during the first 20 ns there was a sharp increase in the number of hydrogen bonds. The weakly associating no POPS/PIPI/PIP$_2$ system makes few hydrogen bonds. Our results on the protein-membrane bonding are consistent with the work on the peripheral membrane protein Osh4 in yeast *Saccharomyces cerevisiae* that showed hydrogen bonded interactions between protein and anionic lipids POPS/POPI stabilized the protein membrane complex.
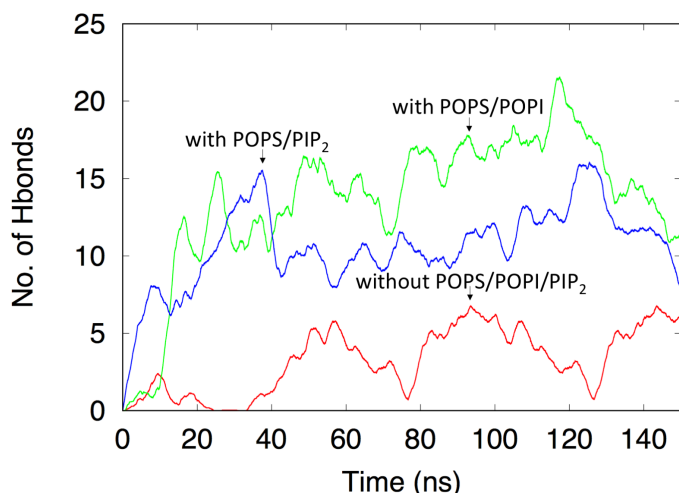
Figure 5.4. The time evolution of the number of hydrogen bonds between the eVP40 dimer and the PM.

There is strong electrostatic binding of the dimer with the membrane for the systems with POPS/POPI and POPS/PIP$_2$. The attractive interaction energy between the protein and PM is almost two-fold stronger when the dimer is associated with the membrane containing POPS/PIP$_2$ compared to when the dimer is associated with the POPS/POPI membrane. In the absence of these anionic lipids (system without POPS/POPI/PIP2), the system displays very little interaction energy, which implies that

the dimer is barely associated with the membrane. The selectivity of the dimer for interacting with POPS, POPI, or PIP_2 lipid molecules is explored in the next section.

5.1.3 Protein-lipid selectivity

Figure 5.5 displays the selectivity of the eVP40 dimer for different lipid types. Specifically, Figure 5.5 depicts the number of heavy atoms of lipids within 3.5 Å of any atom of the eVP40 dimer. The number of contact of POPS lipid atoms with the eVP40 dimer in Figure 5.5b increases as the eVP40 dimer approaches the PM in the system with POPS/POPI. The increase is sharp during the first 20 ns as the dimer approaches the membrane, and then increases more slowly until 100 ns as the dimer undergoes small-scale structural rearrangements in loop segments and helix H1. This agrees with the experimental result that eVP40 selectively binds with the POPS lipids in the PM (133) (109), (118), (115). In the absence of anionic lipids (no POPS/POPI/PIP_2), there was little selectivity of lipid molecules, as shown in Figure 5.5a. The eVP40 dimer displays the highest preference for the highly negatively-charged (-4) PIP_2 molecules, as shown in Figure 5.5c. The strong contacts between the dimer and PIP_2 molecules reduce the number of contact with other lipids such as POPS. This is consistent with the experimental work of Johnson et al. (118) which highlighted PIP_2 lipids that are crucial for eVP40 oligomerization, and also consistent with our coarse-grained MD study that observed enhanced selectivity of the dimer for PIP_2 lipids in the presence of a VP40 hexamer (134). Similarly, MD studies (135), (136) on membrane interactions of the auxilin-1 Phosphatase and Tensin Homolog (PTEN) have reported nanoclustering of PIP_2 lipids, which enhances the binding of auxilin to the lipid bilayer.
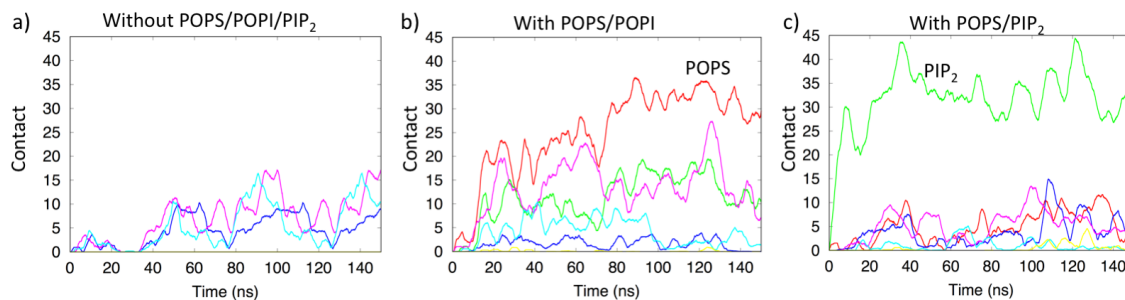
Figure 5.5 The number of heavy atoms in molecules of different lipid atoms within 3.5Å of the dimer for different membranes. a) Without POPS/POPI/PIP2, b) With POPS/POPI, c) with POPS/PIP2. Color scheme: blue-POPC, magenta-POPE, cyan-PSM, yellow-CHL.

I found that flexible loop regions of the dimer that contain positively charged lysine residues were especially important for interacting with the anionic POPS, POPI, or PIP$_2$ molecules. This is shown in Figure 5.6, which is a snapshot from the MD simulation with the POPS/PIP$_2$ membrane. Figure 5.6 shows that flexible loop regions in each monomer that contain positively charged lysine residues make strong contacts with regions of the membrane containing high concentrations of PIP$_2$ molecules. The dynamics of these flexible loop regions are investigated in the next section.
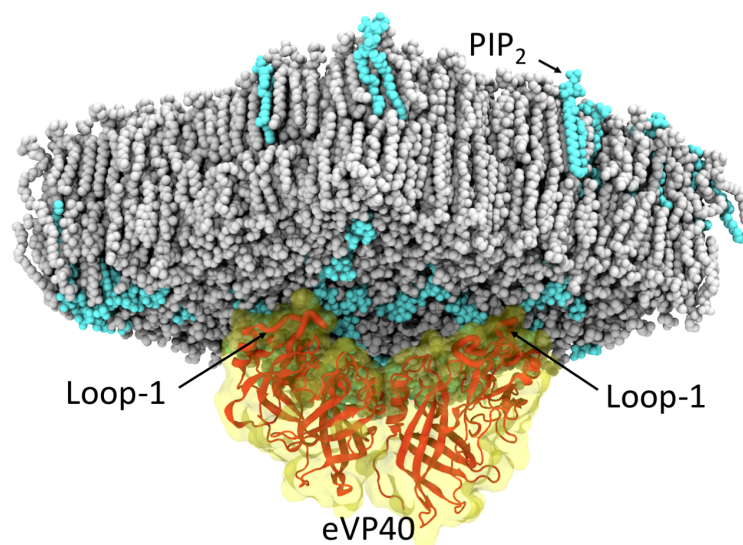
Figure 5.6 A snapshot from the MD simulation with the POPS/PIP2 membrane showing that loop regions (orange) in the dimer that contain positively charged lysine residues interact strongly with negatively charged PIP2 molecules (blue).

5.2 Conformational changes in the VP40 Dimer due to the PM Binding

To further understand the eVP40-PM association dynamics and complex stability, I continued our simulation system with POPS/POPI until 850ns. I explored the dynamics of the loop residues in the eVP40 dimer when interacting with the PM. The two loops are labelled as Loop-1 and Loop-2 as shown in Figure 5.7a. Loop-1 consists of residues 219-233 and includes two important lysine residues, K224 and K225. Loop-2 includes residues 274-283 and contains two important lysine residues, K274 and K275. Experimental studies found that deletion of these loop residues hindered VLP formation (12). Our MD investigations studied the difference in the flexibility of these loops during the dimer-PM association.

The positively charged lysine residues on Loop-1 and Loop-2 interact with negatively charged POPS and POPI molecules in the membrane. Despite the relatively

small fraction of POPS and POPI in the membrane (26%, Table S2), the eVP40 dimer manages to increase the probability of occurrence of favorable interactions between the protein and the membrane. During the initial association process in the first 50 ns, the rmsf is much larger than at later times. Therefore, as the dimer approaches the PM, the loops are flexible enough to allow lysine residues some freedom to be attracted to anionic lipid molecules. As shown in Figure 5.7b and 5.7c, the rmsf of the C_α atoms of the CTDs of chain A and chain B decrease after they interact with the PM. Thus, once the dimer is in contact with the PM, the loops regions become relatively fixed in place, which helps to maintain strong interactions of the dimer in the membrane.
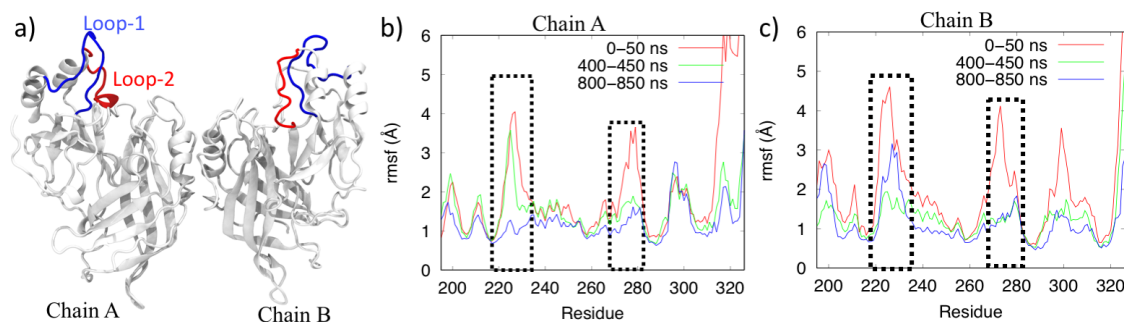
Figure 5.7 a) Two important loop regions in the eVP40 dimer that contain lysine residues that interact with the PM (Loop-1, residues 219-233 and Loop-2, residues 274-283) b) RMSF of the Cα atoms of the CTDs of the eVP40 dimer; the higher dashed bar represents Loop-1, and the shorter represents Loop-2.

5.3 Orientation of CTD helices in VP40 Dimer

The eVP40 dimer has an α-helix (H1) in each CTD (colored red in Figure 5.1) that can interact with the PM. Investigations on other systems have shown that helices can mediate interactions between a protein and a membrane (130), (131), (132). Here, I investigated the behavior of the eVP40 dimer's helices upon membrane association.

In the 850 ns simulation trajectory of the POPS/POPI system, I monitored the orientation of the H1 helices in each CTD of the dimer. The orientation of a helix with respect to the membrane is given by the tilt-angle ψ. The tilt-angle ψ is calculated as $\psi \equiv (90 - \theta)$, where the complimentary angle θ is the angle between the normal of the PM's lower leaflet and the vector along the helix. A parallel orientation of the helix with respect to the inner leaflet of the membrane would have $\psi = 0$. Figure 5.8a displays frames with the CTD helices (colored red) in various orientations. Figure 5.8b shows that over a period of 850 ns, the H1 helices in both Chain A and Chain B reorient themselves to approach a parallel orientation.
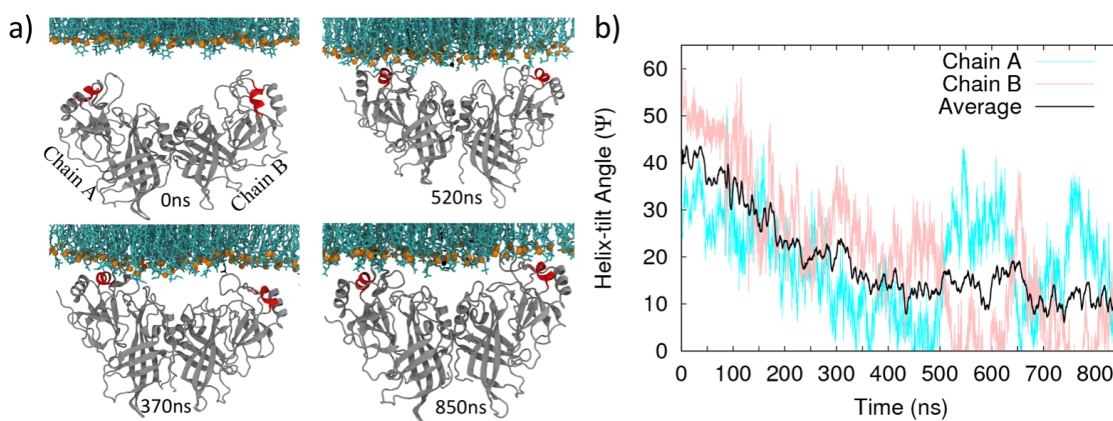
Figure 5.8 Orientation of the CTD helices (red) with respect to the membrane. a) MD frames showing the helices in various orientations at different times. b) Helix-tilt angle for the helix relative to the membrane. During the 850 ns MD simulation, helices in both Chain A and Chain B reorient themselves to approach an orientation parallel to the membrane.

The helix H1 consists of residues T266, L267, V268, H269, K270, L271 and T272. In examining later MD frames with $\psi \sim 10^{\circ}$, I found that an H1 parallel orientation allowed stronger interactions between POPS and POPI lipids in the membrane and polar

residues T266, H269, T272 and charged residue K270, as shown in Figure 5.9. As the

helix approaches a parallel orientation at later times, the number of bonds, especially for

K270 increases. This rearrangement of the helix is a conformational change of the eVP40

dimer that further stabilizes the eVP40-PM complex in an orientation that facilitates

subsequent hexamerization.
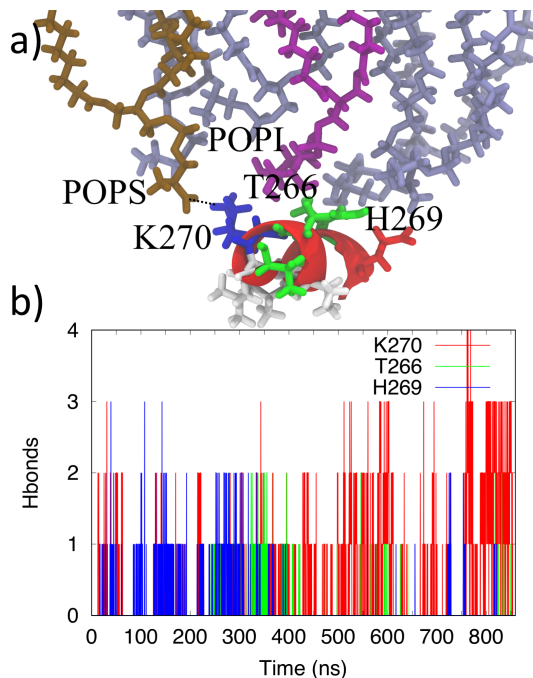
Figure 5.9 a) When helix H1 is aligned parallel to the membrane at later times, residues
T266, H269, K270, can interact strongly with anionic membrane lipids POPS and POPI.
b) The time evolution of the number of hydrogen bonds formed between T266, H269,
K270 and membrane lipids.

5.4 The hydrophobic residue V276 inserts into a membrane defect

Our MD studies also showed a possible mechanism for membrane penetration of

amino acids in the dimer. The electrostatic interactions between cationic protein residues

and lipid polar/charged groups can create membrane defects that allow insertion of

hydrophobic residues into the hydrophobic inner region of the membrane and increase the

binding. There is evidence for this in other systems. Lai et al. (137) proposed that electrostatic interactions between the GRP1 PH domain and lipids on the membrane were crucial to the stabilization of the equilibrium membrane binding structure of the PH domain. Their findings suggested that arginine and lysine residues (R277, K279) interacting with the charged phosphate groups on the membrane surface created transient lipid defects that allowed a hydrophobic residue (V278) to insert into the bilayer head group region and helped to bind the PH domain. The same effect was seen in combined MD and experimental studies of the binding of RecA protein at the surface of anionic membranes (138).

In our eVP40 dimer system, the hydrophobic residue V276 is close to the positively charged lysine residues K274 and K275 that strongly interact with the membrane. In our simulations, I observed that the electrostatic interactions between these lysine residues and the charged heads of the anionic POPS and POPI lipids created transient membrane defects which allowed the hydrophobic residue V276 to partially insert into the hydrophobic inner region of the membrane. These interactions might add extra stability to the protein membrane complex. Figure 5.10 illustrates a small membrane defect created by lysine residues and a partial penetration of V276 inside the membrane.
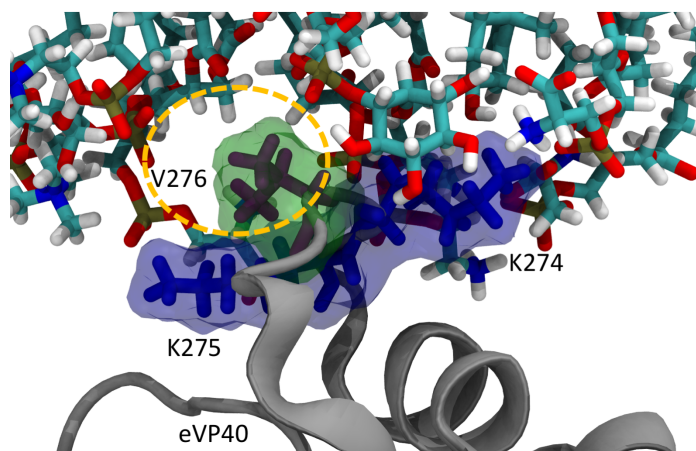
Figure 5.10 The hydrophobic V276 residue can partially penetrate the membrane at a small membrane defect created by lysine residues interacting with POPS and POPI molecules.

5.5 Clustering of PIP$_2$ Lipids in the Lower Leaflet of the PM

5.5.1 Hexamer and PM model

The initial x-ray crystal structure of the VP40 hexamer was taken from the protein data bank (PDB) [PDB entry 4ldd]. The hexamer is formed by three dimers that join via their N-terminal domain (NTD) interfaces. The C-terminal domains (CTD) of VP40 at the two ends are intact but the other four CTDs are missing in the crystal structure. Based on the orientations of the dimers in the hexameric structure[2], two of the middle CTDs point towards and interact with the membrane, whereas the other two CTDs point towards the cytoplasm. The missing CTDs that interact with the membrane were inserted using Modeller (139). These CTDs were oriented in such a way that the hydrophobic residues that are known to penetrate the membrane face towards the membrane (36), (117).

The membrane-VP40 hexamer system was generated using the Charmm-GUI (71) web server. An asymmetric lipid bilayer was generated in accordance with the high

complexity of the plasma membrane composition (140), (141). The percentage of various lipids (POPC:POPE:PSM:POPS:PIP$_2$:CHOL) in the modeled membrane were (41:8:23:4:4:20) in the outer (or upper) leaflet and (11: 37: 5: 16: 10: 21) in the inner (or lower) leaflet. The lipid-bilayer is generated large enough to accommodate the hexamer. The solvated membrane-protein system contained >600,000 atoms in the all-atom representations, but in the coarse-grained set up, the numbers were significantly reduced as described below. The lipid bilayer without protein was composed with the same ratios of the phospholipid mixture as described above for the bilayer with protein.

In order to investigate the VP40-membrane interactions, I prepared the hexamer-membrane system with the middle two CTDs facing the lower leaflet of the membrane so that the hydrophobic residues Leu213, Ile293, Leu295, and Val298 in a loop region were in direct contact with the membrane, as predicted by experiments (36), (117). The all-atom as well as coarse-grained representations of the system are displayed in Figure 5.11. The all-atom representation in Figure 5.11a includes the lower CTDs that interact with VP35 proteins that are involved in packaging and incorporating the viral RNA into VLP (142). As they do not influence the membrane binding, the lower CTDs were not included in the CGMD simulations (Figure 5.11b) to further reduce computational time.
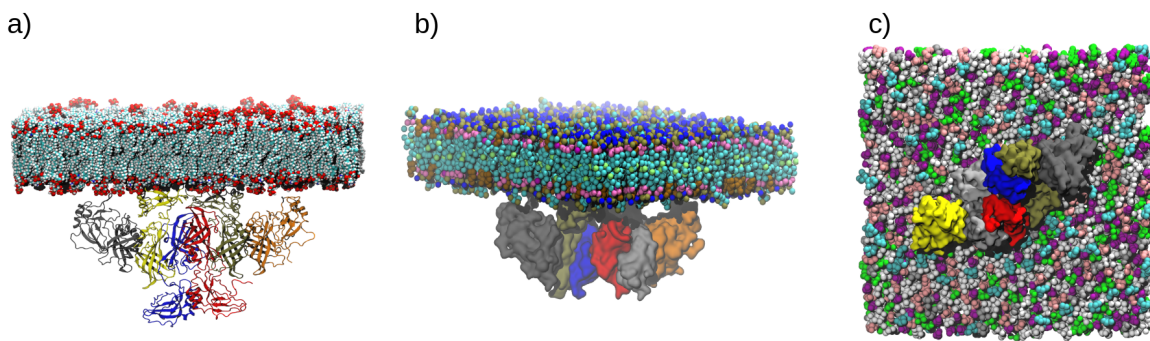
Figure 5.11 VP40 hexamer-membrane system. a) All-atom representation of VP40 hexamer at the PM lower-leaflet, side view. b) Coarse-grained representation, side view c) Distribution of various lipids in the initial setup, as viewed from the bottom. Lipids are colored as Light-Blue: PIP$_2$, Green: POPC, White: POPE, Gray: PSM, Pink: POPS, Purple: CHOL.

I performed a 10-μs CGMD simulation and monitored the dynamics of the VP40 hexamer-membrane interactions. The initial setup has several amino acids in the middle two CTDs interacting with the membrane. The protein center-of-mass was initially set approximately 54 Å below the membrane, measured from the z-coordinate of the center of mass of the lipid bilayer. During the MD simulations, the protein diffuses laterally around the lower leaflet and also increases the protein-membrane contacts. After ~ 3 μs, the CTDs at the two ends interact more strongly with the membrane. Figure 5.12a shows the relative distance between the lipid and protein center of masses (z_{cm}) as a function of time. As the two CTDs at the end start to interact strongly with the membrane, z_{cm} is significantly reduced.

In Figure 5.12b, I display the number of each type of lipid near the VP40 hexamer as a function of time. The normalized count for a lipid type is calculated as the number of that lipid type within 7Å of the protein divided by the total number of that lipid type in the lower leaflet of the PM. Interestingly, PIP$_2$ is the only lipid type that shows significant crowding around the protein. The normalized count for PIP$_2$ increases steadily until ~3 μs and plateaus to a much larger value (~25%) than the normalized counts of the other lipid types (<10%). It is worth noting that the reduction in z_{cm} in Figure 5.12a occurs after the occurrence in Figure 5.12b of significant PIP$_2$ crowding around the

protein. I further investigated the crowding of PIP$_2$ by exploring the lipid clustering in the membrane with or without protein, as discussed below.

Figure 5.12c shows a noticeable bulging or bending in the PM. This is because an increase in the number of contacts between the membrane and the end-CTDs cause the hydrophobic residues Leu213, Ile293, Leu295, and Val298 to penetrate further into the membrane.
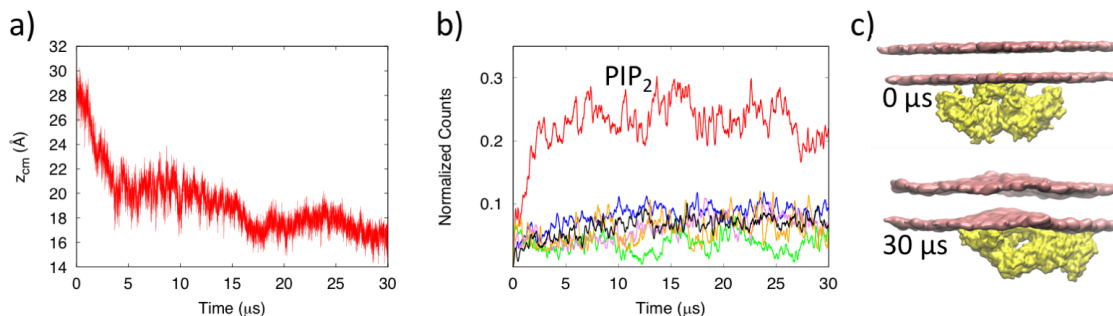
Figure 5.12 a) distance between the lipid and protein center of masses (z_{cm}), b) Normalized count of various lipids within 7Å of the VP40 hexamer as a function of time, c) Bending (or bulging) of the membrane due to enhanced lipid-protein interactions.

5.5.2 Lipid Clustering in the PM With and Without VP40 Hexamer

Different types of mammalian cell plasma membranes have different lipid compositions and the differences in lipid compositions can affect how proteins interact with the membranes (143). Through coarse-grained MD simulations, it has been shown that certain lipid types tend to cluster in the membrane (68). Especially important, the local lipid composition caused by clustering was found to directly affect the membrane curvature (68). Since membrane bending is important for VLP formation, I investigated

changes in the local lipid composition due to lipid clustering in the presence of the protein.

I investigated the distribution of PIP_2 in the lower leaflet of the plasma membrane in the presence of the VP40 hexamer and compared it with the distribution of PIP_2 without the protein. The top panel of Figure 5.13a, displays the VP40 hexamer-membrane systems as viewed from below (cytoplasmic side) at different times. The bottom panel of Figure 5.13a displays the same system as in the top panel but with the protein removed to better show the lipid distribution in the membrane. In the beginning (0 μs), the PIP_2 lipids are randomly distributed (as generated by Charmm-GUI). By 3 μs, a significant clustering of PIP_2 is observed, mostly in the periphery of the protein. Besides PIP_2, no other lipid type showed such clustering behavior. To compare the behavior of the lipids in the absence of protein, I performed a 10 μs simulation with only the membrane and display the results in Figure 5.13b. Initially (0 μs), the PIP_2 lipids are randomly distributed as in Figure 5.13a. By the end of the 10 μs simulation, PIP_2 shows some clustering but to a significantly lesser extent compared to the clustering in the presence of the protein (Figure 5.13a).
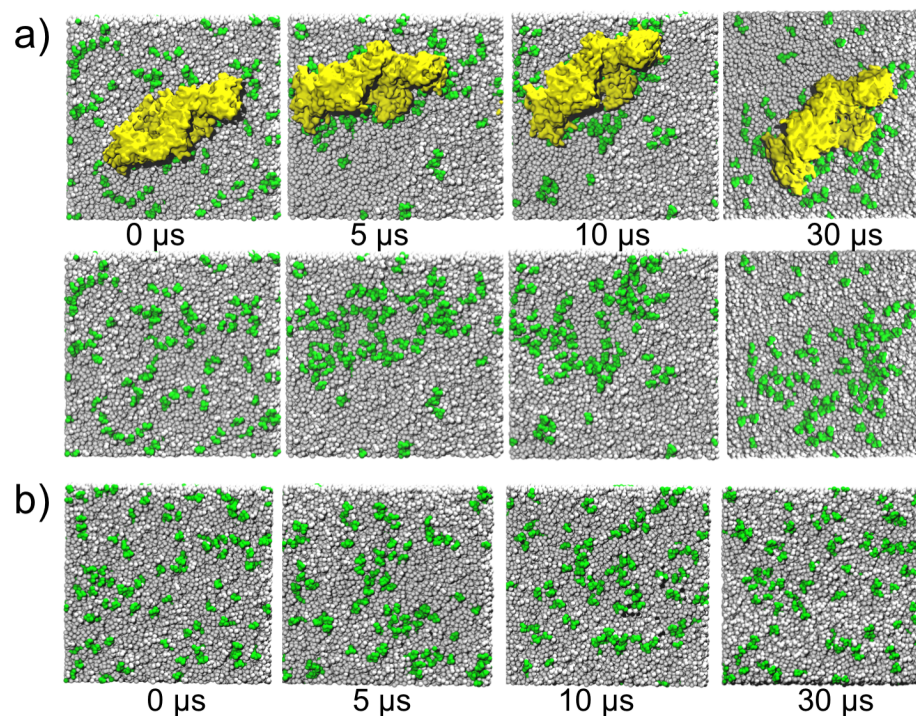
Figure 5.13 Distribution of lipids (PIP2: Green) in the lower leaflet of the PM. a) Results of the CGMD simulation with the presence of the VP40 hexamer. The top row shows the VP40 hexamer (Yellow), the row below are the same snapshots with the VP40 hexamer removed so that the lipids are visible. b) Results of the CGMD simulation in the absence of a protein.

Upon closer inspection of the PIP_2 clusters in the periphery of the protein, I observed that the PIP_2 head groups interact strongly with the basic and polar residues in VP40 as shown in Figure 5.14a. Since some of the CTD residues penetrate the membrane, the locations of the CTDs are noticeable in Figure 5.14b as gray areas surrounded by PIP_2 lipids (green). I calculated the radial pair distribution function g(r) for the various lipid types relative to the VP40 hexamer. The g(r) quantifies the relative abundance of a specific lipid type around the protein. As shown in Figure 5.14c, PIP_2 molecules significantly cluster around the protein, compared to other lipid types. Lipid

types POPE, POPS and CHOL show only a weak preference to be in contact with the

protein, and POPC seems to have the least preference.
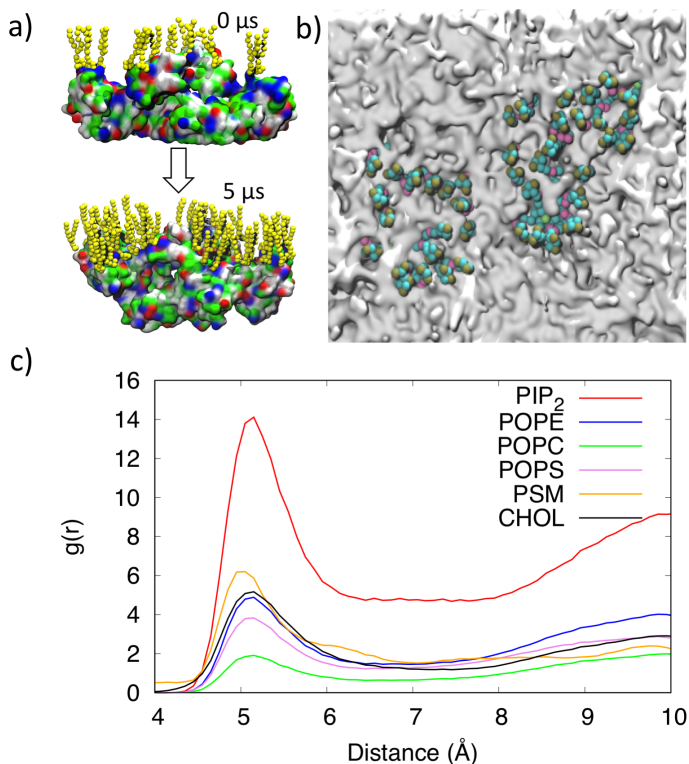
Figure 5.14 Distribution and clustering of PIP2 lipid$_s$ a) Distribution of the PIP2 lipids near the VP40 hexamer at 0 μs and 5 μs showing the negatively charged lipid head groups interacting with the positively charged or polar residues on the protein. b) Clustering of PIP2 lipids due to VP40. The VP40 was present in the simulation but is removed from the image to allow full view of the membrane. PIP$_2$ molecules cluster around gray regions that are the location of the VP40 CTDs. c) Radial pair distribution functions for various lipid types relative to the VP40 hexamer.

5.5.3 Quantification of Lipid Clustering

To quantitatively compare the distribution and clustering of the lipid molecules

with and without VP40, I calculated the fractional interaction matrix (68) for the lipids in

the lower leaflet of the PM. The fractional interaction matrix, also known as the

preferential partitioning of membrane components represents the normalized number of

contacts with a specific lipid type and has been used to study lipid phase separation into ordered and disordered domains in various bilayer systems containing peripheral and transmembrane proteins (144), as well as phase separations (69) and lipid clustering and membrane curvature in various bilayer systems (68). The fractional interaction matrix is calculated as

$$p_{ij} = \frac{c_{ij}/n_j}{\sum_k c_{ik}/n_k} \qquad\qquad (5.5.3.1)$$

Here, the preferential partitioning p_{ij} represents the number of contacts of the lipid type i relative to the lipid type j, normalized over the number of all lipid contacts with the lipid type i. The number of contacts c_{ij} is calculated based on the distance between the glycerol ester moiety (GL1/GL2 beads) of POPE, POPC, POPS, PIP$_2$ or amino alcohols (AM1/AM2 beads) of PSM or cholesterol ROH.

Figure 5.15a displays the fractional interaction matrix for the five different types of lipids in the lower leaflet of the bilayer at 10 µs in the presence of the VP40 hexamer. Two lipids are counted as being in contact if they are within 11 Å, which includes both the first and second solvation shells. The fractional interactions of each lipid type with all lipid types are given in the rows, and the values are normalized horizontally. If all five lipid types are randomly distributed, the entries would be 0.20. The fractional interaction for PIP$_2$ with itself (first entry, Row 5) is found to be 0.40, which means that about 40% of the PIP$_2$ contacts are with PIP$_2$ itself and represents clustering. The fractional interactions for any other lipid type (Rows 1-4) are approximately 0.20, indicating randomly distributed lipid contacts. This means that lipid types such as POPE, POPC and

PSM each have about the same probability to be in contact with themselves or with other lipid types. POPS is found to have the least preference to be surrounded by PIP_2.

To assess any enhancement in lipid clustering due to the VP40 hexamer, I compared the fractional interaction matrices for bilayers systems with and without the protein. Figure 5.15b displays the fractional interaction matrix for the lipids in the lower leaflet of the membrane-only bilayer system in the absence of VP40. In the absence of the protein, the PIP_2-PIP_2 fractional contact is 0.30, compared to 0.40 Figure 5.15a in the presence of protein. This suggests that the presence of the VP40 hexamer enhances the PIP_2 clustering by 33%. Such enhancement in PIP_2 clustering by VP40 can be functionally important in the viral life-cycle since local clustering of PIP_2 in the lower leaflet has been shown to result in curvature of the membrane (68).
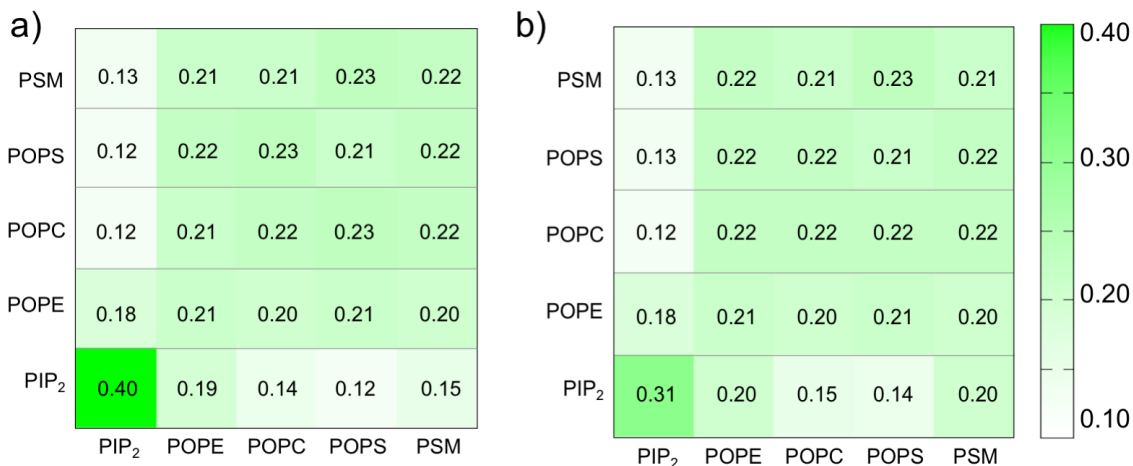
Figure 5.15 Fractional interaction matrix of lipids in the lower leaflet of the plasma membrane. a) With the VP40 hexamer. b) Without the VP40 hexamer.

5.5.4 Mechanism of Lipid Clustering

PIP_2-PIP_2 clustering can occur in the plasma membrane due to formation of hydrogen bond networks (145) with water molecules and/or Ca^{2+} that balance the

electrostatic repulsion between negatively charged head groups (146). I explored the molecular details of the PIP_2 interactions leading to PIP_2 clustering around VP40 and found that the repulsion between negatively charged phosphate head groups (P1 or P2 beads) can be balanced by the electrostatic attraction mediated by positively charged protein side chains (SC2 bead of Lys and Arg). Figure 5.16a shows a network of alternating Lys-PIP_2 interactions. When a cationic side chain mediates two PIP_2 head groups from either side as shown in Figure 5.16b, cholesterol molecules are found to fill the space thus created. This also explains why PIP_2 has higher self-clustering in the second solvation layer. I calculated the radial pair distribution of PIP_2 against all other lipid types and found that cholesterol is the most abundant lipid type around PIP_2, followed by POPE. This is reflected in Figure 5.16c by the high radial pair distribution for PIP_2-CHOL. In addition to cholesterol, PIP_2 is also significantly surrounded by POPE, due partly to the high percentage of POPE in the membrane, giving a high radial pair distribution for PIP_2-POPE (Figure 5.16c). This also helps in forming a higher PIP_2-PIP_2 distribution in the second solvation layer.
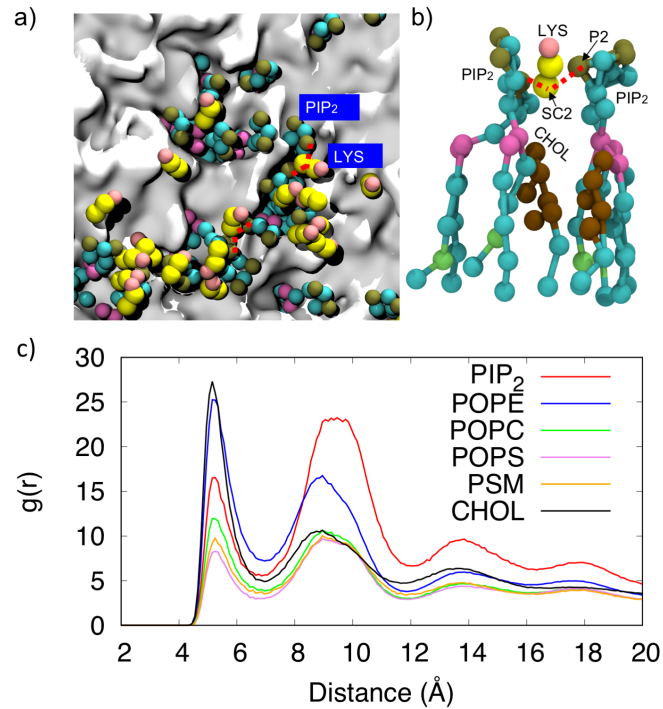
Figure 5.16 Illustration of the role of lysine residues in PIP_2 clustering. a) The network of interacting Lys and PIP2. b) Lysine side chains mediating PIP_2-PIP_2 interactions c) Radial pair distribution function of other lipids around PIP_2 lipids

6. CONCLUSIONS

I performed molecular dynamics simulations to investigate the molecular and atomic level dynamics of large-scale structural changes in transformer proteins. I focused my investigations on RfaH and Ebola VP40. I first studied the mechanisms of the all-α to all-β conformational changes in the RfaH-CTD. In order to facilitate a faster structural conversion, I used an implicit solvent EEF1 force-field with the replica exchange molecular dynamics (REMD) technique. I explored the free-energy landscape of the conformational space of the RfaH-CTD that spanned the α-helix and β-configurations. I used detailed all-atom simulations to refine the structure of the replica which is closest to the experimentally observed RfaH-CTD β-structure. I analyzed the time evolution during

99

the all-α to all-β transformation of various structural parameters such as rmsd, number of H-bonds, and secondary structure formation. The early onset of the loss of helicity of the CTD Helix I as compared to Helix II in the absence of the NTD sheds light on the importance of interfacial interactions between the two domains of RfaH. Results suggest that the folding into the β-scaffold proceeds with formation of an antiparallel β-hairpin structure. Through information transfer entropy calculations, I identified major driver-responder residues that are involved in helix unfolding as well as β-scaffold folding. Future investigations of the RfaH-CTD with detailed all-atom simulations for all steps in the α-helix to β-structural conversion process will allow a better understanding of the dynamics of this transformer protein, which may serve as a good model system for investigating protein structural transformations in general.

I continued my research to a full RfaH molecule and investigated the mechanisms of the structural transformation of the CTD in the full RfaH using various molecular dynamics techniques. Through dynamic network analysis, I determined the communities of amino acids that are involved in interdomain and intradomain communications. Examination of these communities showed that amino acid F126 in the CTD shows correlated motions with the NTD amino acids I93, S97, and L96. F126 also affects the RNAP binding site through interaction with I93, making it a possible functionally relevant residue.

Using targeted MD, I investigated the α→β as well as the β→α structural transformation of the CTD in the full RfaH. The observed folding and unfolding pathways are different for the forward and back conversions. The observed mechanism for the folding of the CTD into the β-barrel structure is consistent with those from

previous computational studies of the isolated CTD. In the presence of the NTD, however, the CTD helix unfolding process leading to the initiation of the β-structure is different. Instead of going through a completely unfolded random coil state, the α-helix to β-structure transition occurs through a collapsed globular state in which significant residual hydrophobic interactions are present. Also, the order of the unfolding of the CTD helices is reversed due to interactions with the NTD. I used the Adaptive Biasing Force method to calculate the free-energy profile for the structural transformation. I find that the CTD-NTD interdomain interactions constitute the main barrier in the CTD α-helix to β-barrel structural conversion. Once the interfacial interactions are broken, the structural conversion of the CTD is relatively easy. SMD calculations elucidated specific amino acids at the interdomain coupling between movement of the CTD helices and the NTD helix, and show how solvent accessible surface area is affected at the RNAP binding site.

Using various molecular dynamics simulations, I investigated the dynamics of the association and the dissociation processes of the NTD and CTD domains in the Ebola virus VP40 transformer protein. MD simulations reveal a novel interdomain salt-bridge interaction between D45-K326 as part of the NTD-CTD latch that was shown to be important for regulating VP40 oligomerization. The results suggest that the salt-bridge interaction between D45-K326 is helpful in guiding domain-domain association, whereas the E76-K291 interaction stabilizes the closed-form structure. VLP analysis demonstrated that some of these salt-bridges may be significant in forming the correct VP40 assembly structure for significant scission and VLP formation. For instance, mutation of the D45-K326, R148-D312, and E160-K212 salt-bridges led to a statistically significant reduction in VLPs. Live cell imaging using an EGFP-tagged VP40 system revealed that the

mutation K291E shows an enhanced PM localization. This may be because the K291E mutation that abrogates the salt-bridge allows easier domain dissociation, which is an important step in VP40 hexamerization at the plasma membrane leading to VLP formation. These computational and experimental live cell mutational analyses illustrate the differential functions of salt-bridge networks in VP40 conformational regulation and open the door to large-scale molecular, membrane, cellular and *in vitro* analysis, and computational studies of VP40 conformational dynamics. Further experiments are necessary to investigate the time-dependency of plasma membrane localization and the biophysics of assembly of VP40. For instance, VP40 localization, trafficking and oligomerization may need to be studied from initial stages of VP40 protein expression to scission of mature VLPs detectable at early stages of infection.

I also investigated the Ebola virus protein VP40 dimer interactions at the lower leaflet of the PM for different lipid compositions of the membrane. For the POPS/POPI and the POPS/PIP$_2$ systems, the strong electrostatic interactions between the positively-charged lysine residues of the eVP40 dimer and negatively-charged head groups of the PM are found to be the major driving force in binding the eVP40 dimer to the PM. During association with a membrane containing anionic POPS, POPI, or PIP$_2$, several different lysine residues in loop regions of the dimer interact with the anionic lipid molecules in the membrane. Subsequent to the initial association, CTD helices of the dimer reorient to add additional stabilizing interactions. The helix orientation that most enhanced the binding is when H1 helix is aligned parallel to the PM. This alignment of helix H1 allows electrostatic interactions between lysine K270 as well as polar residues H269, T272, T266 of the protein with the anionic lipids of the membrane.

Our MD simulations revealed an additional stabilizing effect due to penetration of the hydrophobic residues into the PM. The lysine residues can interact with anionic lipid molecules to create defects in the membrane and these defects allow hydrophobic residues such as V276 to partially insert into the membrane. The bound eVP40-membrane complex with partial insertion of hydrophobic residues into the membrane not only provides dimer stability but may also affect the lateral diffusion (137), (147) of the dimer at the PM. Our investigations provide a molecular level detail of the eVP40 dimer localizations at the lower leaflet of the PM and understanding these lipid-protein interactions can be helpful in designing chemical interventions that disrupt VLP formation.

To investigate the protein-lipid interactions for the Ebola virus protein VP40 hexamer at the lower leaflet of the plasma membrane, I used CGMD simulations. The PIP_2 lipid is found to have a tendency to cluster in general, but this clustering is significantly enhanced in the presence of the VP40 hexamer. PIP_2 was found to cluster around two CTDs of the VP40 hexamer. Each CTD has a hydrophobic loop that penetrates the membrane bilayer and lies adjacent to a cationic patch rich in Lys residues. It is found that the electrostatic interactions of negatively charged PIP_2 head groups with cationic Lys side chains mediate the PIP_2 clustering, with cholesterol filling the space between the PIP_2s. The overall mechanism is consistent with the hypothesis that PIP_2 stabilizes VP40 oligomers (118). PIP_2 is likely important for the initial association of VP40 to the plasma membrane bilayer, including formation of VP40 hexamers and larger oligomers from VP40 dimers (115). PIP_2 then becomes exposed on the plasma membrane outer leaflet as VP40 accumulates on the inner leaflet through a yet to be determined

mechanism (115). Clustering of PIP_2 would provide a mechanism to further concentrate VP40 oligomers and stabilize the underlying VP40 matrix layer, and to promote viral budding by causing membrane bending. Further biophysical investigation is warranted to determine if and how VP40 induces PIP_2 clustering *in vitro* and in cells, how PIP_2 contributes to membrane bending, and if cholesterol distribution in the plasma membrane affects VP40 assembly and budding. These studies may also prove useful in determining pharmacological strategies for inhibiting budding of the Ebola virus from the plasma membrane. For instance, therapeutic agents that can modulate membrane curvature or alter membrane fluidity in such a way to reduce plasma membrane bending may be sufficient to slow down the spread of the virus.

Overall Summary

Moonlighting proteins uses their same fold to perform two or more distinct biophysical or biochemical functions (Jefferry, 2015, J of Proteomics, 2009 , molecular biosystems). On the other hand, Metamorphic proteins possess structural plasticity from the same sequence of amino acid residues (Murzin, Biochemistry). Recently, a newly class of proteins is defined as "Transformer proteins" that carries the properties of moonlighting and metamorphic proteins and can undergo conformational change to perform different functions (Knauer, 2012, cell cycle). These proteins are rare in nature and some examples are transcription factor RfaH and Ebola virus protein VP40 (Bornholdt, 2013, GC 2015, 2016). The Carboxy-terminal domain of RfaH (RfaH-CTD) folds and refolds from all-α to all-β form to carryout different functions. However, the homolog NusG has structurally similar NTDs and possess 17% sequence identity on both domains has only

all-β form of CTD. The all-α form of RfaH-CTD masks the binding site of RNAP and inhibits the transcription process. The all-α form transform into all-β form after it is activated by the OPS element of ntDNA. The all-β form of RfaH-CTD enhances the translation by recruiting S10 ribosomes. Another example of a transformer protein is the Ebola virus protein VP40, which exists in different conformations depending on the required function: a butterfly shaped dimer is trafficked to the lower leaflet of the membrane, a hexamer to form the viral matrix, and an octamer ring structure to bind to RNA and regulate viral transcription.

The structural conversions in these transformer proteins are interesting from the protein folding point of view. Instead of a single native state and a unique function, these proteins have different native states with different functions which deviates from the common principle of protein folding. These proteins are also important in designing molecular switch that can be used as biosensors or smart Biomaterials used in biology, biomedicine and biotechnology (Ha and Loh , Chemistry, 2012).

I investigated the structural conversion in transformer proteins, transcription factor RfaH and Ebola virus protein VP40 using molecular dynamics simulations. Experiments (Burmann et. al) suggested that the expression of RfaH-CTD alone will spontaneously refold from all-α to all-β form. Using replica exchange MD (REMD), I investigated the all-α to all-β structural change in RfaH-CTD in isolated form. During REMD simulations of isolated RfaH-CTD, the structural integrity of shorter helix H1 was compromised and was unfolded faster compared to longer helix H2. The residues in the H1 were responders in the helix unfolding and were quantified using transfer entropy.

The formation of β-barrel structure were driven by residues in β3 and β4 hairpins that were formed first in the nucleation process. After investigating the molecular details of the structural conversion isolated RfaH-CTD, I performed simulations on full length RfaH where I used targeted MD simulations, Steered MD simulations (SMD) to understand the structural change of RfaH-CTD in the presence of tertiary contact with the NTD of RfaH. The NTD-CTD interdomain interactions was observed as a major barrier in the structural change of RfaH-CTD. Once the barrier is broken, the all-α to all-β of RfaH-CTD was found to be easier. Using SMD, I highlighted the interdomain saltbridge between E48-R138 as a major interaction. These results support the experiments (Burman et. al 2012) that the E48S mutant was readily detached from the NTD and was converted to β-barrel structure.

In the mean time when I was investigating a Transformer protein RfaH, there was an Ebola outbreak that caused many deaths and surprisingly one of the proteins in Ebola virus is a Transformer protein. The Ebola virus protein VP40 exists in different conformations depending on the required function: a butterfly shaped dimer is trafficked to the lower leaflet of the membrane, a hexamer to form the viral matrix, and an octamer ring structure to bind to RNA and regulate viral transcription. The common structural change in these pathways is the domain rearrangement which is VP40's transformer properties. I performed simulations to understand the domain rearrangement of Ebola VP40 monomer and highlighted important saltbridges in the interdomain region. The D45-K326 saltbridge was observed to be important in the domain dissociation and association. The force profiles showed mutants required less force to separate domains. Experiments performed showed a 20% increase in the plasma membrane localization of

mutant K291E which was due to easy detachment of the domain required for the formation of oligomers in the plasma membrane.

Ebola virus is a lipid-enveloped virus. As the virus enters the human cell, it hijacks the host cell machinery and replicate itself. Once the protein VP40 is encoded by its negative sense ssRNA, VP40 convert into dimer and is trafficked in the lower leaflet of the plasma membrane. Three dimers assemble in the membrane and undergo membrane rearrangement to form a hexamer. I performed simulation on the VP40 dimer and membrane systems and observed the electrostatic interaction as a major driving force in the association kinetics. The positive charged lysine residues in the VP40 dimer was found interacting with the negatively charged anionic lipids such as POPI, POPS and PIP_2. In the absence of these anionic lipids, VP40 dimer was associated weakly with the membrane. In addition to the electrostatic interaction, the additional stability of the protein-membrane complex was found to be due to the helix H1 alignment parallel to the lower leaflet for the membrane. This alignment allowed the efficient hydrogen bonded interactions between the protein and the membrane. I also observed a slight insertion of hydrophobic residues V276 inside the hydrophobic core of the membrane which might be important in the diffusion kinetics of VP40.

Finally, I investigated the VP40 hexamer and plasma membrane interactions using coarse grained MD (CGMD) simulations. Experiments performed by Johnson et al. highlighted the role of PIP_2 lipids in the oligomerization of VP40. Since POPS lipids was found exposed to the outer leaflet of the plasma membrane at the budding site, PIP_2 lipids are important at the interaction site of VP40 oligomers and the membrane. A computational study by Koldso et. al. on the unrelated membrane system showed a

correlation between the membrane curvature and the enrichment of PIP_2 lipids in the lower leaflet of the plasma membrane. This resulted in the membrane bending with concave upward. I performed CGMD on hexamer and membrane system and observed the enhanced clustering of PIP_2 lipids in the lower leaflet of the plasma membrane. This clustering helped in a slight bulging of the membrane at the interaction site. There is a general tendency of PIP_2 lipids to self-cluster, but I observed the 30% increase in the clustering of PIP_2 due to the VP40 hexamer. The radial pair distribution function of PIP_2 lipids is greater near the VP40 hexamer. I observed an extensive ionic interaction between the SC2 bead of lysine residues in the VP40 hexamer and P2 beads of PIP_2 lipids and the void between lipids is filled with the cholesterol.

REFERENCES

1.   Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(4096):223-230.

2.   Jeffery CJ (2009) Moonlighting proteins--an update. *Molecular bioSystems* 5(4):345-350.

3.   Jeffery CJ (2004) Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins. *Current opinion in structural biology* 14(6):663-668.

4.   Tokuriki N & Tawfik DS (2009) Protein dynamism and evolvability. *Science* 324(5924):203-207.

5.   Tompa P, Szasz C, & Buday L (2005) Structural disorder throws new light on moonlighting. *Trends in biochemical sciences* 30(9):484-489.

6.   Tuinstra RL*, et al.* (2008) Interconversion between two unrelated protein folds in the lymphotactin native state. *Proceedings of the National Academy of Sciences of the United States of America* 105(13):5057-5062.

7.   Mapelli M, Massimiliano L, Santaguida S, & Musacchio A (2007) The Mad2 conformational dimer: structure and implications for the spindle assembly checkpoint. *Cell* 131(4):730-743.

8.   Murzin AG (2008) Biochemistry. Metamorphic proteins. *Science* 320(5884):1725-1726.

9.   Knauer SH, Artsimovitch I, & Rosch P (2012) Transformer proteins. *Cell cycle* 11(23):4289-4290.

10.   Xiong L & Liu Z (2015) Molecular dynamics study on folding and allostery in RfaH. *Proteins*.

11.   Burmann BM*, et al.* (2012) An alpha helix to beta barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* 150(2):291-303.

12.   Bornholdt ZA*, et al.* (2013) Structural rearrangement of ebola virus VP40 begets multiple functions in the virus life cycle. *Cell* 154(4):763-774.

13.   Tomar SK, Knauer SH, NandyMazumdar M, Rösch P, & Artsimovitch I (2013) Interdomain contacts control folding of transcription factor RfaH. *Nucleic acids research* 41(22):10077-10085.

14.   Gc JB, Gerstman BS, Stahelin RV, & Chapagain PP (2016) The Ebola virus protein VP40 hexamer enhances the clustering of PI(4,5)P2 lipids in the plasma membrane. *Physical chemistry chemical physics* 18(41):28409-28417.

15.   Svetlov V & Nudler E (2012) Unfolding the bridge between transcription and translation. *Cell* 150(2):243-245.

16.   Mooney RA, Schweimer K, Rosch P, Gottesman M, & Landick R (2009) Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *Journal of molecular biology* 391(2):341-358.

17.   Belogurov GA, Sevostyanova A, Svetlov V, & Artsimovitch I (2010) Functional regions of the N-terminal domain of the antiterminator RfaH. *Molecular Microbiology* 76(2):286-301.

18.   Belogurov GA, Mooney RA, Svetlov V, Landick R, & Artsimovitch I (2008) Functional specialization of transcription elongation factors. *The EMBO Journal* 28(2):112-122.

19.   Knauer SH, Artsimovitch I, & Rösch P (2012) Transformer proteins. *Cell Cycle* 11(23):4289-4290.

20.   Zhou M, Ottenberg G, Sferrazza GF, & Lasmezas CI (2012) Highly neurotoxic monomeric alpha-helical prion protein. *Proceedings of the National Academy of Sciences of the United States of America* 109(8):3113-3118.

21.   Eisenberg D & Jucker M (2012) The amyloid state of proteins in human diseases. *Cell* 148(6):1188-1203.

22.   Straub JE & Thirumalai D (2011) Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annual Review Physics Chemistry* 62:437-463.

23.   Brundin P, Melki R, & Kopito R (2010) Prion-like transmission of protein aggregates in neurodegenerative diseases. *Nature Review Molecular Cell Biology* 11(4):301-307.

24.   Huang L*, et al.* (2010) Macromolecular crowding converts the human recombinant PrPC to the soluble neurotoxic β-oligomers. *The FASEB Journal* 24(9):3536-3543.

25.   Sang JC, *et al.* (2012) Slow spontaneous alpha-to-beta structural conversion in a non-denaturing neutral condition reveals the intrinsically disordered property of the disulfide-reduced recombinant mouse prion protein. *Prion* 6(5):489-497.

26.   Khandogin J & Brooks CL, 3rd (2007) Linking folding with aggregation in Alzheimer's beta-amyloid peptides. *Proceedings of the National Academy of Sciences of the United States of America* 104(43):16880-16885.

27.   Lashuel HA, Overk CR, Oueslati A, & Masliah E (2013) The many faces of alpha-synuclein: from structure and toxicity to therapeutic target. *Nature Review Neuroscience* 14(1):38-48.

28.   Steckmann T, Awan Z, Gerstman BS, & Chapagain PP (2012) Kinetics of peptide secondary structure conversion during amyloid beta-protein fibrillogenesis. *Journal of Theoritical Biology* 301:95-102.

29.   Kammerer RA, *et al.* (2004) Exploring amyloid formation by a de novo design. *Proceedings of the National Academy of Sciences of the United States of America* 101(13):4435-4440.

30.   Steinmetz MO, *et al.* (2008) Atomic models of de novo designed cc beta-Met amyloid-like fibrils. *Journal of molecular biology* 376(3):898-912.

31.   Lotz GP & Legleiter J (2013) The role of amyloidogenic protein oligomerization in neurodegenerative disease. *Journal of Molecular Medicine (Berl)* 91(6):653-664.

32.   Woolfson DN & Ryadnov MG (2006) Peptide-based fibrous biomaterials: Some things old, new and borrowed. *Current Opinion in Chemical Biology* 10(6):559-567.

33.   Radzimanowski J, Effantin G, & Weissenhorn W (2014) Conformational plasticity of the Ebola virus matrix protein. *Protein Science* 23(11):1519-1527.

34.   Gc JB, Gerstman BS, & Chapagain PP (2015) The Role of the Interdomain Interactions on RfaH Dynamics and Conformational Transformation. *The journal of physical chemistry. B* 119(40):12750-12759.

35.   Jasenosky LD, Neumann G, Lukashevich I, & Kawaoka Y (2001) Ebola virus VP40-induced particle formation and association with the lipid bilayer. *Journal of virology* 75(11):5205-5214.

36. Soni SP, Adu-Gyamfi E, Yong SS, Jee CS, & Stahelin RV (2013) The Ebola virus matrix protein deeply penetrates the plasma membrane: an important step in viral egress. *Biophysical journal* 104(9):1940-1949.

37. Stahelin RV (2014) Could the Ebola virus matrix protein VP40 be a drug target? *Expert opinion on therapeutic targets* 18(2):115-120.

38. De Vivo M, Masetti M, Bottegoni G, & Cavalli A (2016) Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of Medicinal Chemistry* 59(9):4035-4061.

39. Sugita T & Nakase T (1999) Non-universal usage of the leucine CUG codon and the molecular phylogeny of the genus Candida. *Syst Appl Microbiol* 22(1):79-86.

40. Feig M, Karanicolas J, & Brooks Iii CL (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling* 22(5):377-395.

41. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33-38, 27-38.

42. Phillips JC*, et al.* (2005) Scalable molecular dynamics with NAMD. *Journal of computational chemistry* 26(16):1781-1802.

43. MacKerell AD*, et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins *The journal of physical chemistry. B* 102(18):3586-3616.

44. Ulrich E*, et al.* (1995) A smooth particle mesh Ewald method. *The Journal of chemical physics* 103(19):8577-8593.

45. Isralewitz B, Gao M, & Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. *Current opinion in structural biology* 11(2):224-230.

46. Zheng W (2010) Anharmonic normal mode analysis of elastic network model improves the modeling of atomic fluctuations in protein crystal structures. *Biophysical journal* 98(12):3025-3034.

47. Liu Y, Hsin J, Kim H, Selvin PR, & Schulten K (2011) Extension of a three-helix bundle domain of myosin VI and key role of calmodulins. *Biophysical journal* 100(12):2964-2973.

48.   Park S & Schulten K (2004) Calculating potentials of mean force from steered molecular dynamics simulations. *The Journal of chemical physics* 120(13):5946-5961.

49.   Martin HS, Jha S, & Coveney PV (2014) Comparative analysis of nucleotide translocation through protein nanopores using steered molecular dynamics and an adaptive biasing force. *Journal of computational chemistry* 35(9):692-702.

50.   Ma J, Sigler PB, Xu Z, & Karplus M (2000) A dynamic model for the allosteric mechanism of GroEL. *Journal of molecular biology* 302(2):303-313.

51.   Cheng X, Wang H, Grant B, Sine SM, & McCammon JA (2006) Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors. *PLoS computational biology* 2(9):e134.

52.   Mashl RJ & Jakobsson E (2008) End-point targeted molecular dynamics: large-scale conformational changes in potassium channels. *Biophysical journal* 94(11):4307-4319.

53.   Compoint M, Picaud F, Ramseyer C, & Girardet C (2005) Targeted molecular dynamics of an open-state KcsA channel. *The Journal of chemical physics* 122(13):134707.

54.   Kong Y, Shen Y, Warth TE, & Ma J (2002) Conformational pathways in the gating of Escherichia coli mechanosensitive channel. *Proceedings of the National Academy of Sciences of the United States of America* 99(9):5999-6004.

55.   Mendieta J, Gago F, & Ramirez G (2005) Binding of 5'-GMP to the GluR2 AMPA receptor: insight from targeted molecular dynamics simulations. *Biochemistry* 44(44):14470-14476.

56.   Li C, Zhu Y, Wang Y, & Chen G (2014) Molecular dynamics simulation on the conformational transition of the mad2 protein from the open to the closed state. *International journal of molecular sciences* 15(4):5553-5569.

57.   Vanommeslaeghe K*, et al.* (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* 31(4):671-690.

58.   Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, & de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry. B* 111(27):7812-7824.

59.   Monticelli L*, et al.* (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of chemical theory and computation* 4(5):819-834.

60. Marrink SJ, de Vries AH, & Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. *Journal of Physical Chemistry B* 108(2):750-760.

61. Marrink SJ & Tieleman DP (2013) Perspective on the Martini model. *Chemical Society reviews* 42(16):6801-6822.

62. Goose JE & Sansom MS (2013) Reduced lateral mobility of lipids and proteins in crowded membranes. *PLoS computational biology* 9(4):e1003033.

63. Koldso H & Sansom MS (2012) Local Lipid Reorganization by a Transmembrane Protein Domain. *The journal of physical chemistry letters* 3(23):3498-3502.

64. Nawae W, Hannongbua S, & Ruengjitchatchawalya M (2014) Defining the membrane disruption mechanism of kalata B1 via coarse-grained molecular dynamics simulations. *Scientific reports* 4:3933.

65. Risselada HJ & Marrink SJ (2008) The molecular face of lipid rafts in model membranes. *Proceedings of the National Academy of Sciences of the United States of America* 105(45):17367-17372.

66. Mercredi PY*, et al.* (2016) Structural and Molecular Determinants of Membrane Binding by the HIV-1 Matrix Protein. *Journal of molecular biology* 428(8):1637-1655.

67. Charlier L*, et al.* (2014) Coarse-grained simulations of the HIV-1 matrix protein anchoring: revisiting its assembly on membrane domains. *Biophysical journal* 106(3):577-585.

68. Koldso H, Shorthouse D, Helie J, & Sansom MS (2014) Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers. *PLoS computational biology* 10(10):e1003911.

69. Basu I & Mukhopadhyay C (2015) In silico phase separation in the presence of GM1 in ternary and quaternary lipid bilayers. *Physical chemistry chemical physics : PCCP* 17(26):17130-17139.

70. Ingolfsson HI*, et al.* (2014) Lipid organization of the plasma membrane. *Journal of the American Chemical Society* 136(41):14554-14559.

71. Qi Y*, et al.* (2015) CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *Journal of chemical theory and computation* 11(9):4486-4494.

72.   Jo S, Kim T, Iyer VG, & Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* 29(11):1859-1865.

73.   Van Der Spoel D*, et al.* (2005) GROMACS: fast, flexible, and free. *Journal of computational chemistry* 26(16):1701-1718.

74.   Abraham MJ*, et al.* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19-25.

75.   de Jong DH, Baoukina S, Ingólfsson HI, & Marrink SJ (2016) Martini straight: Boosting performance using a shorter cutoff and GPUs. *Computer Physics Communications* 199:1-7.

76.   Hamelberg D, Mongan J, & McCammon JA (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics* 120(24):11919-11929.

77.   Hamelberg D, de Oliveira CA, & McCammon JA (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of chemical physics* 127(15):155102.

78.   Miao Y, Feixas F, Eun C, & McCammon JA (2015) Accelerated molecular dynamics simulations of protein folding. *Journal of computational chemistry* 36(20):1536-1549.

79.   Bai Q*, et al.* (2014) Computational study on the interaction between CCR5 and HIV-1 entry inhibitor maraviroc: insight from accelerated molecular dynamics simulation and free energy calculation. *Physical Chemistry Chemical Physics* 16(44):24332-24338.

80.   Gumbart JC, Roux B, & Chipot C (2013) Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *Journal of Chemical Theory and Computation* 9(1):794-802.

81.   Kamberaj H & van der Vaart A (2009) Extracting the causality of correlated motions from molecular dynamics simulations. *Biophysical journal* 97(6):1747-1755.

82.   Qi Y & Im W (2013) Quantification of Drive-Response Relationships Between Residues During Protein Folding. *Journal of Chemical Theory and Computation* 9(8).

83.   Hamed E & Keten S (2014) Hierarchical cascades of instability govern the mechanics of coiled coils: helix unfolding precedes coil unzipping. *Biophysical journal* 107(2):477-484.

84.   Kwansa AL, De Vita R, & Freeman JW (2014) Mechanical recruitment of N- and C-crosslinks in collagen type I. *Matrix biology : Journal of the International Society for Matrix Biology* 34:161-169.

85.   Sotomayor M & Schulten K (2007) Single-molecule experiments in vitro and in silico. *Science* 316(5828):1144-1148.

86.   Zheng W (2014) All-atom and coarse-grained simulations of the forced unfolding pathways of the SNARE complex. *Proteins* 82(7):1376-1386.

87.   He J, *et al.* (2013) Steered molecular dynamics simulation of the binding of the beta2 and beta3 regions in domain-swapped human cystatin C dimer. *Journal of molecular modeling* 19(2):825-832.

88.   Cuendet MA & Michielin O (2008) Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. *Biophysical journal* 95(8):3575-3590.

89.   Giorgino T & De Fabritiis G (2011) A High-Throughput Steered Molecular Dynamics Study on the Free Energy Profile of Ion Permeation through Gramicidin A. *Journal of Chemical Theory and Computation* 7(6):1943-1950.

90.   Cheung LS, *et al.* (2015) Characterization of monobody scaffold interactions with ligand via force spectroscopy and steered molecular dynamics. *Scientific reports* 5:8247.

91.   Xu L, *et al.* (2013) Using steered molecular dynamics to predict and assess Hsp70 substrate-binding domain mutants that alter prion propagation. *PLoS computational biology* 9(1):e1002896.

92.   Guzman DL, Randall A, Baldi P, & Guan Z (2010) Computational and single-molecule force studies of a macro domain protein reveal a key molecular determinant for mechanical stability. *Proceedings of the National Academy of Sciences of the United States of America* 107(5):1989-1994.

93.   Patel JS, Berteotti A, Ronsisvalle S, Rocchia W, & Cavalli A (2014) Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. *Journal of chemical information and modeling* 54(2):470-480.

94. Gendoo DMA & Harrison PM (2011) Discordant and chameleon sequences: Their distribution and implications for amyloidogenicity. *Protein Science* 20(3):567-579.

95. Gc JB, Bhandari YR, Gerstman BS, & Chapagain PP (2014) Molecular dynamics investigations of the alpha-helix to beta-barrel conformational transformation in the RfaH transcription factor. *The journal of physical chemistry. B* 118(19):5101-5108.

96. Singh S, Chiu CC, Reddy AS, & de Pablo JJ (2013) alpha-helix to beta-hairpin transition of human amylin monomer. *The Journal of chemical physics* 138(15):155101.

97. Qi R, Luo Y, Ma B, Nussinov R, & Wei G (2014) Conformational Distribution and alpha-Helix to beta-Sheet Transition of Human Amylin Fragment Dimer. *Biomacromolecules* 15(1):122-131.

98. Girvan M & Newman ME (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* 99(12):7821-7826.

99. Sethi A, Eargle J, Black AA, & Luthey-Schulten Z (2009) Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* 106(16):6620-6625.

100. McClendon CL, Kornev AP, Gilson MK, & Taylor SS (2014) Dynamic architecture of a protein kinase. *Proceedings of the National Academy of Sciences of the United States of America* 111(43):E4623-4631.

101. Schymkowitz J*, et al.* (2005) The FoldX web server: an online force field. *Nucleic acids research* 33(Web Server issue):W382-388.

102. Balasco N, Barone D, & Vitagliano L (2015) Structural conversion of the transformer protein RfaH: new insights derived from protein structure prediction and molecular dynamics simulations. *Journal of biomolecular structure & dynamics*:1-7.

103. Li S*, et al.* (2014) Mechanism of the All-α to All-β Conformational Transition of RfaH-CTD: Molecular Dynamics Simulation and Markov State Model. *Journal of Chemical Theory and Computation* 10(6):2255-2264.

104. Lemkul JA & Bevan DR (2008) A comparative molecular dynamics analysis of the amyloid beta-peptide in a lipid bilayer. *Archives of biochemistry and biophysics* 470(1):54-63.

105. Govindjee R, Ohno K, Chang CH, & Ebrey TG (1984) The C-terminal tail of bacteriorhodopsin--its conformation and role in proton pumping. *Prog Clinical Biology Research* 164:13-25.

106. Wang Q*, et al.* (2008) The extra C-terminal tail is involved in the conformation, stability changes and the N/C-domain interactions of the calmodulin-like protein from pearl oyster Pinctada fucata. *Biochimica Biophysica Acta* 1784(11):1514-1523.

107. Bandyopadhyay S, Banik U, Bhattacharyya B, Mandal NC, & Roy S (1995) Role of the C-terminal tail region in the self-assembly of lambda-repressor. *Biochemistry* 34(15):5090-5097.

108. Zhu Y*, et al.* (2015) Effects of the C-Terminal Tail on the Conformational Dynamics of Human Neuronal Calcium Sensor-1 Protein. *Journal of Physical Chemistry B* 119(44):14236-14244.

109. Scianimanico S*, et al.* (2000) Membrane association induces a conformational change in the Ebola virus matrix protein. *The EMBO journal* 19(24):6732-6741.

110. Silva LP, Vanzile M, Bavari S, Aman JM, & Schriemer DC (2012) Assembly of Ebola virus matrix protein VP40 is regulated by latch-like properties of N and C terminal tails. *PloS one* 7(7):e39978.

111. Gc JB*, et al.* (2016) Interdomain salt-bridges in the Ebola virus protein VP40 and their role in domain association and plasma membrane localization. *Protein Science* 25(9):1648-1658.

112. Glykos NM (2006) Software news and updates. Carma: a molecular dynamics analysis program. *Journal of computational chemistry* 27(14):1765-1768.

113. Bhaskara RM & Srinivasan N (2011) Stability of domain structures in multi-domain proteins. *Scientific reports* 1:40.

114. Waldburger CD, Schildbach JF, & Sauer RT (1995) Are buried salt bridges important for protein stability and conformational specificity? *Nature structural biology* 2(2):122-128.

115. Adu-Gyamfi E*, et al.* (2015) Host Cell Plasma Membrane Phosphatidylserine Regulates the Assembly and Budding of Ebola Virus. *Journal of virology* 89(18):9440-9453.

116. Adu-Gyamfi E*, et al.* (2014) A loop region in the N-terminal domain of Ebola virus VP40 is important in viral assembly, budding, and egress. *Viruses* 6(10):3837-3854.

117. Adu-Gyamfi E*, et al.* (2013) The Ebola virus matrix protein penetrates into the plasma membrane: a key step in viral protein 40 (VP40) oligomerization and viral egress. *The Journal of biological chemistry* 288(8):5779-5789.

118. Johnson KA, Taghon GJ, Scott JL, & Stahelin RV (2016) The Ebola Virus matrix protein, VP40, requires phosphatidylinositol 4,5-bisphosphate (PI(4,5)P2) for extensive oligomerization at the plasma membrane and viral egress. *Scientific Reports* 6:19125.

119. McCarthy SE, Johnson RF, Zhang YA, Sunyer JO, & Harty RN (2007) Role for amino acids 212KLR214 of Ebola virus VP40 in assembly and budding. *J Virol* 81(20):11452-11460.

120. Dessen A, Volchkov V, Dolnik O, Klenk HD, & Weissenhorn W (2000) Crystal structure of the matrix protein VP40 from Ebola virus. *The EMBO journal* 19(16):4228-4236.

121. Crieghton T (1992) *Proteins: Structures and Molecular Properties* (W. H. Freeman, New York, NY) Second Ed p 512.

122. Sugase K, Dyson HJ, & Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447(7147):1021-1025.

123. Schreiber G & Keating AE (2011) Protein binding specificity versus promiscuity. *Current opinion in structural biology* 21(1):50-61.

124. Rogers JM, Steward A, & Clarke J (2013) Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'. *Journal of the American Chemical Society* 135(4):1415-1422.

125. Arai M, Ferreon JC, & Wright PE (2012) Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *Journal of the American Chemical Society* 134(8):3792-3803.

126. Babu MM, van der Lee R, de Groot NS, & Gsponer J (2011) Intrinsically disordered proteins: regulation and disease. *Current opinion in structural biology* 21(3):432-440.

127. Shaw AS & Filbert EL (2009) Scaffold proteins and immune-cell signalling. *Nature reviews. Immunology* 9(1):47-56.

128. Munz M, Hein J, & Biggin PC (2012) The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. *PLoS computational biology* 8(11):e1002749.

129. Teimoori S*, et al.* (2016) Human transbodies to VP40 inhibit cellular egress of Ebola virus-like particles. *Biochem Biophys Res Commun* 479(2):245-252.

130. Prakash P, Zhou Y, Liang H, Hancock JF, & Gorfe AA (2016) Oncogenic K-Ras Binds to an Anionic Membrane in Two Distinct Orientations: A Molecular Dynamics Analysis. *Biophys J* 110(5):1125-1138.

131. Pleskot R, Pejchar P, Zarsky V, Staiger CJ, & Potocky M (2012) Structural insights into the inhibition of actin-capping protein by interactions with phosphatidic acid and phosphatidylinositol (4,5)-bisphosphate. *PLoS Comput Biol* 8(11):e1002765.

132. Monje-Galvan V & Klauda JB (2016) Peripheral membrane proteins: Tying the knot between experiment and computation. *Biochim Biophys Acta* 1858(7 Pt B):1584-1593.

133. Ruigrok RW*, et al.* (2000) Structural characterization and membrane binding properties of the matrix protein VP40 of Ebola virus. *J Mol Biol* 300(1):103-112.

134. Gc JB, Gerstman BS, Stahelin RV, & Chapagain PP (2016) The Ebola virus protein VP40 hexamer enhances the clustering of PI(4,5)P2 lipids in the plasma membrane. *Physical Chemistry Chemical Physics* 18(41):28409-28417.

135. Poyry S & Vattulainen I (2016) Role of charged lipids in membrane structures - Insight given by simulations. *Biochim Biophys Acta* 1858(10):2322-2333.

136. Kalli AC, Morgan G, & Sansom MS (2013) Interactions of the auxilin-1 PTEN-like domain with model membranes result in nanoclustering of phosphatidyl inositol phosphates. *Biophys J* 105(1):137-145.

137. Lai CL*, et al.* (2013) Molecular mechanism of membrane binding of the GRP1 PH domain. *J Mol Biol* 425(17):3073-3090.

138. Zhang L, Rajendram M, Weibel DB, Yethiraj A, & Cui Q (2016) Ionic Hydrogen Bonds and Lipid Packing Defects Determine the Binding Orientation and Insertion Depth of RecA on Multicomponent Lipid Bilayers. *J Phys Chem B* 120(33):8424-8437.

139. Eswar N*, et al.* (2006) Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* Chapter 5:Unit 5 6.

140. Zachowski A (1993) Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement. *The Biochemical journal* 294 (Pt 1):1-14.

141. van Meer G, Voelker DR, & Feigenson GW (2008) Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology* 9(2):112-124.

142. Johnson RF, McCarthy SE, Godlewski PJ, & Harty RN (2006) Ebola virus VP35-VP40 interaction is sufficient for packaging 3E-5E minigenome RNA into virus-like particles. *Journal of virology* 80(11):5135-5144.

143. Sharpe HJ, Stevens TJ, & Munro S (2010) A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* 142(1):158-169.

144. de Jong DH, Lopez CA, & Marrink SJ (2013) Molecular view on protein sorting into liquid-ordered membrane domains mediated by gangliosides and lipid anchors. *Faraday discussions* 161:347-363; discussion 419-359.

145. Levental I, Cebers A, & Janmey PA (2008) Combined electrostatics and hydrogen bonding determine intermolecular interactions between polyphosphoinositides. *J Am Chem Soc* 130(28):9025-9030.

146. Wang YH*, et al.* (2012) Divalent cation-induced cluster formation by polyphosphoinositides in model membranes. *Journal of American Chemical Soc* 134(7):3387-3395.

147. Ziemba BP & Falke JJ (2013) Lateral diffusion of peripheral membrane proteins on supported lipid bilayers is controlled by the additive frictional drags of (1) bound lipids and (2) protein domains penetrating into the bilayer hydrocarbon core. *Chemistry and Physics of Lipids* 172-173:67-77.

148. Jeevan B. GC, Bernard S. Gerstman, Prem P. Chapagain, Membrane Association and Localization dynamics of the Ebola Virus Matrix Protein VP40, BBA-Biomembranes, Under Revision

VITA

JEEVAN B. GC

EDUCATION

- Florida International University, Miami, FL, Ph.D. Physics, March 2017.
- Tribhuvan University, Kathmandu, Nepal, M. Sc., Physics, 2008.
- Tribhuvan University, Pokhara, Nepal, B. Sc., Physics, 2006.

PUBLICATIONS AND PRESENTATIONS

1. Jeevan B. GC, Bernard S. Gerstman, Robert V. Stahelin and Prem P. Chapagain, "The Ebola virus protein VP40 hexamer enhances the clustering of PIP2 lipids in the Plasma membrane". *Journal of Physical Chemistry Chemical Physics*, 2016, DOI: 10.1039/C6CP03776C. Cover Image Article

2. Jeevan B. GC, Kristen A. Johnson, Monica L. Husby, Cary T. Frick, Bernard S. Gerstman, Robert V. Stahelin, and Prem P. Chapagain, "Interdomain salt-bridges in Ebola virus protein VP40 and their role on domain association and plasma membrane localization". *Protein Science, 2016.* doi:10.1002/pro.2969.

3. Jeevan B. GC, Bernard S. Gerstman, and Prem P. Chapagain, "The Role of the Interdomain interactions on RfaH dynamics and conformational transformation". *Journal of Physical Chemistry* B, 119 (40), 12750-12759 (2015).

4. Jeevan B. GC, Yuba R. Bhandari, Bernard S. Gerstman, and Prem P. Chapagain, "Molecular dynamics investigations of the α-helix to β-barrel conformational transformation in the RfaH transcription factor". *Journal of Physical Chemistry* B, 118 (19), 5101-5108 (2014)

5. Jeevan B. GC, Bernard S. Gerstman, Prem P. Chapagain, "Membrane Associations and Localization Dynamics of The Ebola Virus Protein VP40". (2017, Submitted to BBA-Biomembranes (Under Revision)

6. Jeevan B. GC, Bernard S. Gerstman, Prem P. Chapagain, $PI(4,5)P_2$ clustering by the Ebola Virus Matrix Protein VP40, BPS March Meeting 2017, New Orleans, February 11-14.

7. J GC, B Gerstman, P Chapagain, The Molecular Dynamics Study of the Structural Conversions in the Transformer Protein RfaH, APS March Meeting, Baltimore, MD, March15-18, 2016.

8. J GC, B Gerstman, P Chapagain, Interdomain Interactions And The Mechanism Of Structural Transformation In RfaH, Biophysical Society Meeting, Los Angeles, CA, Feb27-March3, 2016.

9. J GC, B Gerstman, P Chapagain, Computational Study on the Dynamics and Structural Conversion of RfaH, American Physical Society, National Mentoring Conference, Miami, FL, 2015.

10. J GC, Y Bhandari, B Gerstman, P Chapagain, Molecular Dynamics Investigations of the alpha-helix to Beta-barrel Conformational Transformation in RfaH, APS March Meeting, San Antonio, Texas, 2015.

11. Jeevan B. GC, All-Alpha to All-Beta Structural Conversion in the Transcription Factor RfaH, Biophysical Society Meeting, Baltimore, Maryland, 2015 (Biophysical Journal, Volume 108, Issue 2).