7-8-2016

Computational Modeling of Allosteric Stimulation of Nipah Virus Host Binding Protein

Priyanka Dutta
University of South Florida, priyanka10.dutta@gmail.com

Computational Modeling of Allosteric Stimulation of Nipah Virus Host Binding Protein

by

Priyanka Dutta

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Cell Biology, Microbiology, and Molecular Biology
College of Arts and Sciences
University of South Florida

Major Professor: Sameer Varma, Ph.D.
Gary Daughdrill, Ph.D.
James Riordan, Ph.D.
Yu Chen, Ph.D.

Date of Approval:
June 23, 2016

Keywords: Water dynamics, Protein–protein interaction, Molecular dynamics, Structure prediction, Inverse machine learning, Paramyxovirus

To my best friend who is also my husband

&

To Baba and Ma

ACKNOWLEDGEMENT

I am grateful to my advisor Dr. Sameer Varma and committee members, Dr. Gary Daughdrill, Dr. James Riordan and Dr. Yu Chen who provided constant guidance and evaluated my progress. I specially thank the Department of Cell biology, Microbiology and Molecular Biology for supporting me financially as a graduate assistant. I also thank the staff members of the department for being patient and helpful at all times. I appreciate the companionship of the past and present lab members, especially Mr. Mohsen Botlani for being a friend and helping me maintain a positive attitude.

I express my gratitude to my family for their encouragement and support. I am grateful to my father Mr. Prodyot Dutta and mother Mrs. Tanima Dutta who taught me to dream and aspire. Special thanks to my father-in-law Mr. Sushanta Chandra and mother-in-law Mrs. Manju Chandra. Finally, this dissertation would not have been possible without the help and support of my husband Dr. Sayan Chandra who made me believe in myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Nipah belongs to the family of paramyxoviruses that cause numerous fatal diseases in humans and farm animals. There are no FDA approved drugs for Nipah or any of the paramyxoviruses. Designing antiviral therapies that are more resistant to viral mutations require understanding of molecular details underlying infection. This dissertation focuses on obtaining molecular insights into the very first step of infection by Nipah. Such details, in fact, remain unknown for all paramyxoviruses. Infection begins with the allosteric stimulation of Nipah virus host binding protein by host cell receptors. Understanding molecular details of this stimulation process have been challenging mainly because, just as in many eukaryotic proteins, including GPCRs, PDZ domains and T-cell receptors, host receptors induce only minor structural changes (< 2 Å) and, consequently, thermal fluctuations or dynamics play a key role. This work utilizes a powerful molecular dynamics based approach, which yields information on both structure and dynamics, laying the foundation for its future applications to other paramyxoviruses. It proposes a new model for the initial phase of stimulation of Nipah's host binding protein, and in general, highlights that (a) interfacial waters can play a crucial role in the inception and propagation of allosteric signals; (b) extensive inter-domain rearrangements can be triggered by minor changes in the structures of individual domains; and (c) mutations in dynamically stimulated proteins can induce non-local changes that spread across entire domains.

CHAPTER 1

INTRODUCTION

Viral infections have been one of the leading causes of health concern around the globe. In humans, the severity of viral infection can manifest into a mild or a morbid form, as seen in common cold (1, 2), or poliomyelitis (3, 4) respectively. In a report, Center for disease control and prevention (CDC) listed the 10 leading causes of death in the United States for the years 2013 and 2014, where death due to influenza and pneumonia were ranked 8th (5). Fatality due to viral infection is a common occurrence in other species as well, causing extensive damage to poultry, livestock, and domesticated animals. Furthermore, there have been reports of infections transmitted across species, which is of grave concern especially, in farms where livestock are raised together (6-14). Viral infections can be prevented with vaccinations or treated using antiviral therapies that generally interfere in the viral infection process (15-19).

The viral infection process begins with the attachment of the virus to the host cell, where entry proteins on the virus recognize and bind to specific receptors on the host. The virus entry is mediated either by membrane fusion of the host and viral membranes, as in the case of enveloped viruses, or the virus penetrates the membrane of the host cell as seen in non-enveloped viruses (15, 20-23). For enveloped viruses, the entry proteins are embedded on the viral membrane. This makes them an easy and attractive target for drug molecules to bind (17, 22, 24-26). Therefore, one of the promising

approaches for the treatment of viral infection is to design inhibitors acting on the entry protein (15, 18, 19, 27-29). Although designing entry inhibitors is a potentially compelling approach, there are challenges like, high mutation rates of a virus and low potency of inhibitors due to the mechanisms developed by a virus to evade these therapeutic agents. The various evasion strategies of the entry protein include, for example, oligomeric occlusion (30), glycosylation (31), conformational masking (32), multivalent interactions (33), etc. To overcome these challenges a detailed knowledge of the entry proteins and their mechanism of actions is necessary (15).

This dissertation deals with the study of the entry mechanism of the enveloped Nipah virus (NiV). It belongs to the family *Paramyxoviridae*, which are negative-sense, single-stranded RNA viruses, responsible for numerous diseases in humans and livestock. They are highly contagious and are transmitted through the respiratory route causing infections like croup, bronchiolitis, pneumonia, and also other systemic infections like measles, mumps and encephalitis (34-38). Paramyxovirus infection is reported to be the most common viral infection accounting for at least 5% of pediatric intensive care unit admissions with significant morbidity (39). Infections by respiratory syncytial virus (RSV), parainfluenza viruses (PIVs) and the human metapneumoviruses (hMPV) are known to outnumber the infections caused by influenza and other respiratory viruses (40-53). The past two decades have witnessed the emergence of paramyxoviruses via cross-species transmission, of which the genus of Henipavirus is known to be highly lethal. In countries like Bangladesh, Malaysia and India, NiV has caused encephalitis with ~70% mortality (54-56). Furthermore, reports of neurological problems and relapse of encephalitis years after initial infection, necessitates the

immediate attention to design antiviral drugs for treatment, not vaccination. The most frightening mode of transmission happens from human–to–human via physical contact, which makes highly dense populated regions vulnerable (57, 58). The lack of effective treatment, poor preventive measures and high pathogenicity makes NiV a biosafety level 4 virus. This exemplifies the gravity of the danger this virus poses in the event of an outbreak. Hence, it becomes essential to study the entry protein of NiV and its mechanism at the molecular level, which will provide fundamental information towards developing antiviral therapies.

The entry mechanism of NiV and most other paramyxoviruses (59-65) involve the concerted action of two membrane glycoproteins, the host binding protein and the fusion protein. The host binding protein recognizes and binds to its specific receptor, which stimulates it to activate (66-68). Upon activation, the host binding protein triggers the fusion protein, which facilitates the fusion of the host and viral membranes (Figure 1.1) (54, 65, 69-71).

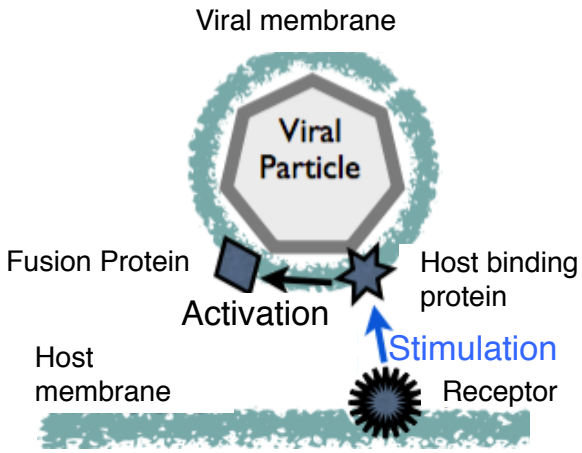


Figure 1.1 Schematic representation of the viral entry mechanism. The host binding protein on attaching to its receptor is stimulated to activate the fusion protein. The activated fusion protein in turn, fuses the virus and the host membranes.

This work focuses on the molecular mechanisms underlying the stimulation of the viral host binding protein. Over the past few decades, numerous different experimental strategies (60, 62, 63, 72-76) have been used to probe the stimulation mechanism of the paramyxovirus host binding proteins, and have yielded a wealth of information. X-ray crystallography and biochemical assays (59, 71, 77-86) show that there are two distinct domains in the host binding protein — one domain binds to host receptor and the other domain interacts with and activates the fusion protein. While the different family members of paramyxovirus bind to different host receptors, ranging from sugars to proteins (63, 68, 79, 86-89), the overall structure of the host binding protein is conserved across the paramyxovirus family. Additionally, while the mechanism of allosteric communication between the host binding and fusion activation domains is unknown, there is now growing consensus that this mechanism is conserved (62, 63, 74). The results from these experiments form the foundation of the different proposed models of fusion activation, which are discussed in Chapter 2. Although there is a cellular level understanding of the underlying mechanism of the stimulation of the host binding protein on attachment to its receptor, an insight at the molecular level is lacking.

The primary reason why the allosteric stimulation mechanism remains unknown, despite extensive structural and biochemical studies, is that the receptor binding domain undergoes little to no change in structure upon binding to the host receptor (Figure 1.2). The X-ray structures of apo and bound states of the receptor binding domain of the host binding protein have been analytically compared, and the calculated RMSD is found to be < 2 Å (72, 86, 90-92).

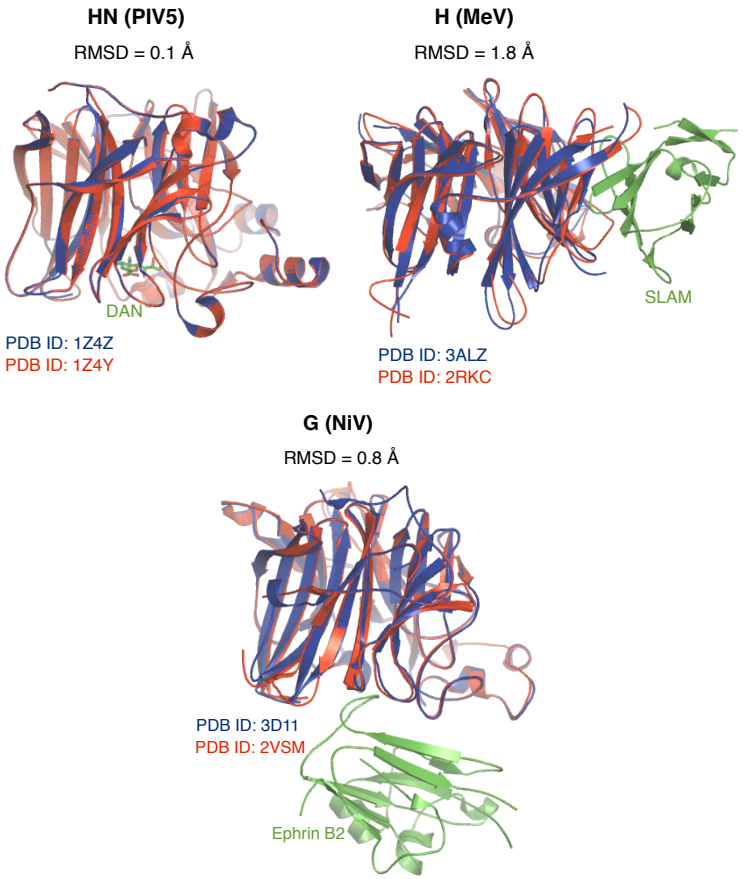


Figure 1.2 X-ray structures of representative receptor binding domains belonging to the three subfamilies of paramyxovirus host binding proteins (HN, H and G) (63). In each case, the X-ray structure of the receptor free state (blue) is superimposed over the structure of the receptor bound state (red), and the RMSD between the backbone atoms of these structures are indicated. The receptors are shown in green. The receptor binding induces only minor backbone changes in these receptor binding domains (93-95).

In fact, other well studied eukaryotic proteins, like G-protein couple receptors (GPCRs) and PDZ domains, have also exhibited < 2 Å displacement on ligand binding (96-98). Similar behavior is also reported in other eukaryotic proteins, for example, cyclic AMP dependent protein kinase (PKA) (99), nicotinic receptor (nAChR) (100), catabolite activator protein (CAP) (101), epidermal growth factor receptor (EGFR) (102, 103), and many more (104). Both experimental and computational studies corroborate

that the allosteric signaling depends not only on small structural changes, but also on changes in thermal fluctuations.

Molecular dynamics (MD) is a powerful technique that can provide direct atomistic level insights into the role of both the structure and dynamics in allosteric signaling. It is an established and reliable technique to study the relation between a protein's structure, its dynamics and function (105-107). Furthermore, one can obtain information about the time dependent motion of individual atoms, thus presenting the possibility to find answers to questions related to the properties of a model system with high molecular level details which can be used in synergism with experiments to obtain a better understanding of the system (106-112). MD simulations can be used to obtain an understanding and develop hypothesis about molecular behavior to connect the three-dimensional structure of the host binding protein to its dynamics. In Chapter 3, we provide a brief discussion on MD and other simulation techniques employed in this study.

Chapter 4 examines the details of signal inception at the receptor binding site of the host binding protein. The crystal structure of receptor-bound state of the receptor binding domain of the host binding protein shows that there are an exceptionally large number of water molecules at the protein-protein interface (Figure 1.3) (113). In fact, this extensive interface has the second highest number of water molecules. Do these water molecules play any role in signal inception? There are numerous molecular simulation studies of protein-protein interfaces where interstitial waters are not modeled explicitly (114-116), and can this interface be also modeled without discrete waters? Water is known to behave differently at interfaces. Do these interstitial waters behave

like interfacial waters, or because of their high numbers, do they exhibit bulk-like properties?

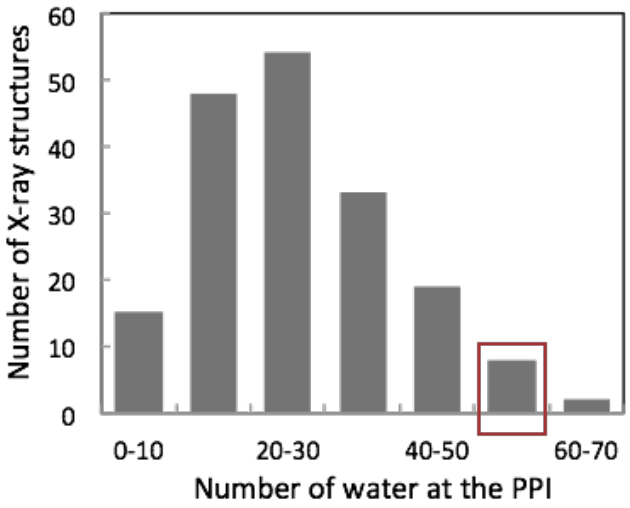


Figure 1.3 Histogram plot indicating average water molecules at protein–protein interfaces. The red box marks the bin corresponding to the number of water molecules at the interface formed between the NiV host binding protein and its receptor.

Chapter 5 examines the effect of receptor binding on the interface between two receptor binding domains. Interpretations from cellular assays have yielded contrasting models (60, 62, 63, 73). While there are no direct observations, one set of experiments suggests that the receptor-induced changes are large (69). However, only small receptor-induced changes have been reported in the cases of the PIV5 and NDV homologs of the host binding protein (59, 91, 117-119). But at the same time, it has been argued that the overall fusion stimulation mechanism is conserved across all subfamilies. How can such divergent explanations be interpreted? It is also known that receptor binding induces minor changes in the receptor binding domain, which leads to the question of how do the small changes transmit to the interface between two adjacent receptor binding domains?

Chapter 6 deals with the construction of the full length ectodomain, What is known from crystallography and the modeling work in Chapter 5 is the structure of the receptor binding domain and the interface of two receptor binding domains. In this regard the structure of domain that activates the fusion protein in not known. In addition the interface of the fusion activation domain and the receptor binding domain needs to be identified. It is important to construct the full length structure of the ectodomain to answer two main questions. Firstly, how does the signal get transmitted to the fusion protein activation domain? Secondly, what is the form of the signal at the fusion protein activation domain? Crystallographic studies on NDV and PIV5 homologs, and biochemical studies suggest that the overall architecture of the ectomains is conserved across all subfamilies, which provide the starting point for this study.

Finally, Chapter 7 summarizes the finding and outlines the future directions.

CHAPTER 2

NIPAH VIRUS

2.1 Family: *Paramyxoviridae*

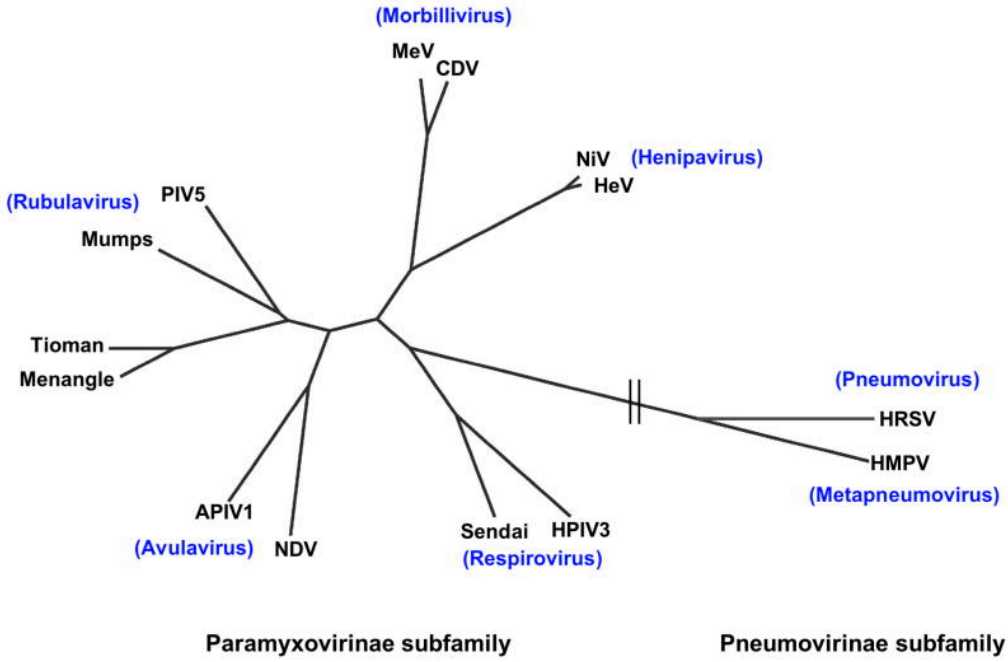


Figure 2.1 Phylogenetic tree of the *Paramyxoviridae* family. The tree was generated using Cobalt in NCBI (120). The fusion protein sequences were aligned by the fast minimum evolution method, and visualized using the Fig Tree program**.** Representative members of each genus of the *Paramyxovirinae* and *Pneumovirinae* subfamilies are shown (genera are shown in blue). Abbreviations: APIV-1, avian parainfluenza virus 1; CDV, canine distemper virus; HeV, Hendra virus; HMPV, human metapneumovirus; HPIV-3, human parainfluenza virus 3; HRSV, human respiratory syncytial virus; MeV, measles virus; NDV, Newcastle disease virus; NiV, Nipah virus; PIV-5, parainfluenza virus 5 (Reprinted with permission from (10). Copyright Cambridge University Press 2011).

Nipah virus belongs to the *Paramyxoviridae* family, which consists of non-segmented, negative-sense, single-stranded enveloped RNA viruses. They are classified into two subfamilies, *Paramyxovirinae* and *Pneumovirinae* (10, 54). The

subfamily *Paramyxovirinae* consists of five genera, namely Respirovirus, Morbillivirus, Rubulavirus, Avulavirus and Henipavirus, shown in Figure 2.1. Nipah virus belongs to the genus Henipavirus (121).

Paramyxoviruses are highly contagious and transmitted through respiratory route causing respiratory infections like croup, bronchiolitis and pneumonia, also systemic infections like measles, mumps and encephalitis. The lethality of these viruses is evident from the multiple accounts discussed below. Recent studies state that Paramyxovirus infections account for 5% of pediatric intensive care unit admissions with significant morbidity (39). Center for Disease Control and Prevention lists respiratory syncytial virus (RSV) as the most common cause of infection in young children worldwide (122). An extrapolation from an US based population analysis deduces 57,527 hospitalizations of children between 2-5 years annually (46). It is predicted that RSV infection will cause hospitalization of approximately 177,525 patients < 65 years with an annual death rate of 8% i.e. 14,000 adult patients (45). According to the World Health Organization review of March 2016, measles is the leading cause of death among young children accounting for 114900 deaths globally in 2014 (123). The Merck veterinary manual rates the avian paramyxovirus serotype 1 (PMV-1) or Newcastle disease virus (NDV) as the most virulent among all the known 11 PMV serotypes (124, 125), hence important as a pathogen for poultry. It is known to infect more than 236 species of poultry animals (126) and virtually all 8000 species of birds are susceptible to NDV infection (127). Paramyxoviruses, in general have a high rate of allopatric divergence (128) and hence exhibit one of the highest rates of cross-species transmission among RNA viruses (89).

2.2 Lethality and geographical distribution of Nipah

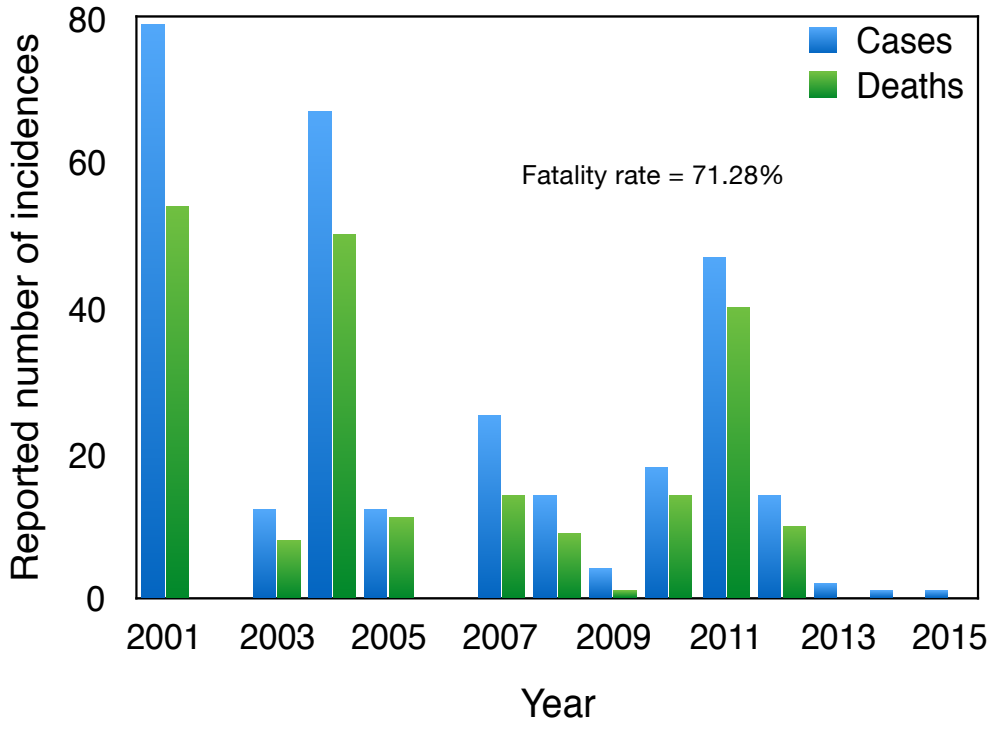


Figure 2.2 Total number of NiV infected people by year in Bangladesh and India.

The genus *Henipavirus* includes HeV and NiV. They have recently emerged as bat-borne viruses that are highly lethal in a broad range of mammalian hosts (7, 10, 12, 129-132) and NiV has also caused mortality in humans. Human infection by the HeV is very rare with only 7 cases being reported in the past 2 decades (133). On the other hand, the NiV presents itself as a more lethal member of the genus. This is evident from Figure 2.2, which shows the number of deaths in India and Bangladesh alone caused by the NiV infection in humans over the past 15 years (57, 58, 134-136). In addition, reports of infection in Malaysia since 1995 shows a human fatality rate of ~40% (55, 137). The major cause of death has been related to encephalitis on relapse.

Although, the first instance of human infection was reported in 1998, the Nipah virus was known to exist in its natural reservoir, the fruit bats of the genus *Pteropus*. It

is well known that for any species to be unharmed by a virus infection, it must contain the specific neutralizing antibodies, which are produced by way of a defense mechanism. Therefore, one can map out the spread of the virus both geographically and within various species by verifying its presence. In this case, the neutralizing antibodies have been found in bat species of Australia, Malaysia, Singapore, Cambodia, India, Bangladesh and Madagascar (138-142) across three continents. The transmission of NiV from the bats to humans has occurred via the following pathways as apparent from the epidemiological studies. The most frequently implicated and direct route is ingestion of fresh date palm sap, containing bat saliva. Though bats are the primary reservoir of the virus, it can be transferred to domesticated animals that can act as vectors and transmit the disease to humans in close contact. This commonly happens when the domesticated animal ingests fruits laden with bat saliva. Finally, the most frightening mode of transmission happens from person-to-person (57, 58, 143) via physical contact, which makes highly dense populated regions vulnerable. The lack of effective treatment, poor preventive measures and high pathogenicity makes NiV a biosafety level 4 virus. This exemplifies the gravity of the danger this virus poses in the event of an outbreak.

2.3 Proposed models of fusion activation

Over the years, different experimental techniques have been instrumental towards understanding and investigating the G protein. Here, we discuss the working mechanism of the receptor binding protein, which as mentioned in Chapter 1, is crucial to the infection mechanism. For the infection to initiate, the paramyxovirus entry requires the coordinated action of two viral membrane proteins: the receptor binding

protein and the fusion protein (F). The receptor binding protein on attaching to specific host receptor is stimulated to activate F. The activated F protein, in turn, facilitates the fusion of the virus and the host membranes. The receptor binding protein of paramyxoviruses can be designated as H, HN or G depending on its functionality (10, 61, 62, 74, 82, 88, 144). If the receptor binding protein has both hemaglutinin (sialic acid binding) and neuraminidase (sialic acid cleaving) functions, they are called HN proteins (Eg. PIV5) (91, 145, 146), while for those that lack the neuraminidase activity but can bind to sialic acid are called H proteins (Eg. MeV) (75, 81, 83, 84). In the case where they do not bind to the sialic acid, i.e. in Henipavirus, they are named the G protein (88).

The G protein of Nipah virus binds to cellular receptors ephrin B2/B3 of its hosts (66, 68, 79, 147). An approach to inhibit F-activation is to prevent the binding process, and this can be achieved by inactivating either of the proteins, G or ephrins, participating in binding. Experiments based on binding assays reveal that the G protein has extremely high (pico molar to subnano molar) binding affinity (66, 68) to ephrin, and owing to the fact that ephrin is found in neurons, smooth muscle, arterial endothelial cells and capillaries (148-151), the viral infection is found to be systemic in nature (66, 79, 152, 153). Additionally, ephrins are known to play a key role in cellular development, especially in the nervous and vascular systems (148, 154), which suggest that rendering the ephrins inactive in order to prevent the infection is not a viable option. The widespread occurrence of ephrins in vertebrates (148, 155-159) makes them potential targets across various species to the severe systemic infection caused by Nipah. This highlights the fact that the only approach to prevent its infection is by targeting the G protein. In order to do so, we need a fundamental understanding of the functionality of

the G protein at a molecular level so that the signal transduction can be intercepted to prevent the activation of G.

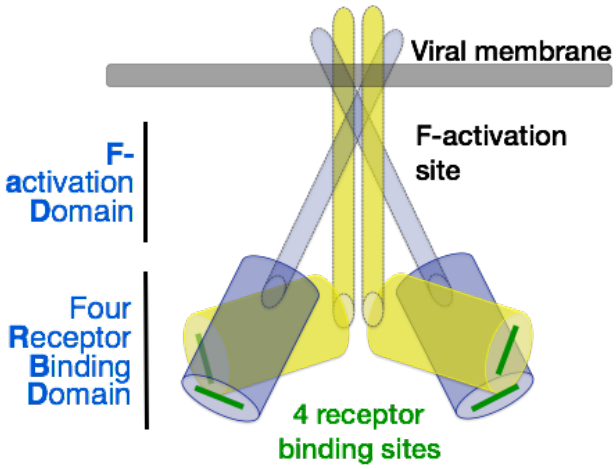


Figure 2.3 Schematic representation of the host binding protein highlighting the tetrameric architecture. The blue and yellow colors are used to highlight the two-fold symmetry of the dimer-of-dimers structure. The green bars indicate the receptor binding site.

Structurally the G protein is a type II membrane glycoprotein (82, 90) (schematic in Figure 2.3) that consists of a cytoplasmic tail, a transmembrane region, which anchors the protein to the viral envelope, and an ectodomain comprising of a N-terminal fusion protein (F) activation domain (FAD) and a C-terminal receptor binding domain (RBD). The receptor binding domain is composed of a propeller-shaped structure, where β-sheets are arranged cyclically around an axis through the center similar to the blades of a propeller. The center of the propeller forms the active site or the receptor binding site (shown as green bars in figure) and binds to its receptor ephrin, which gives it the name, the receptor binding domain (86, 90). Recently, from x-ray crystallography studies, Bowden *et. al.* were able to obtain partial information about the structure of the G protein. However, based on their experimental work and combining it with the results obtained by Yuan *et. al* (96-98)., and Welch *et. al*., it was possible to hypothesize the full length architecture of G as represented in Figure 2.3. The receptor binding sites

located on the RBD is the only component in this figure which was crystallographically resolved by Bowden *et. al*. Furthermore, experiments on the FAD evidence that it occurs as a tetramer with the presence of specific inter-monomer interactions, and shows that the tetrameric architecture is essential to retain the functional form of G (81, 117, 160).

Experiments including binding assays (66, 68), flow cytometry (70, 77), circular dichroism spectra (69, 70, 77), cell-cell fusion assays (69, 70), monoclonal antibody assays (69, 161), etc. help us to formulate an idea of the mechanism involved in fusion activation. Based on these experimental studies two models have been proposed to explain the fusion activation in the paramyxoviruses: (i) the dissociation or “clamp” model and (ii) the association or “provocateur” model (see Figure 2.4) (61, 63, 64, 73, 74, 162, 163). In case of the dissociation model, the receptor binding protein retains the F protein in its metastable, prefusion form. Upon receptor binding, the F is released, triggered to refold and drive merging of the viral and host membranes. In the association model the receptor binding protein actively triggers the metastable F protein by destabilizing it after binding to its receptor. A major difference between the two models is that, in the dissociation model the receptor binding protein retains the prefusion F conformation, whereas, in the association model the receptor binding protein destabilizes the prefusion F. The provocateur model suggests that F can be expressed in its prefusion form without the host binding protein (61). In case of Nipah virus, the experimental results support the dissociation model (61, 161, 164-167). There is now new evidence that the overall stimulation mechanism is partially conserved

across all members of the paramyxovirus family (62), but it still remains undetermined. In both the model, it is the G protein, which once stimulated triggers the F protein.

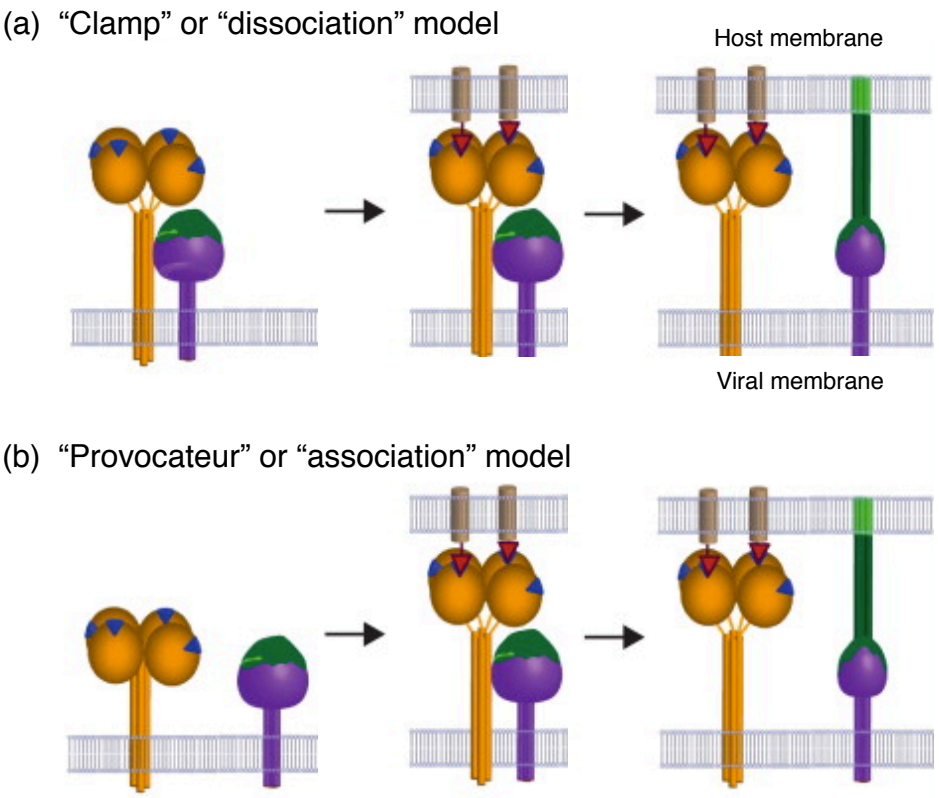


Figure 2.4 The protein-protein interaction between the host binding protein and F triggers the F protein. In case of **(a)** the “clamp” model, the host binding protein and F protein dissociate upon receptor engagement with the host binding protein, allowing the F protein to be triggered, while in **(b)** the “provocateur” model the F and the host binding protein associate on the surface of the viral membrane only to dissociate after receptor engagement. The host binding protein is illustrated in orange as a globular head (RBD) linked, via a flexible linker, to a four helix bundle stalk (FAD). The receptor binding site is shown as a blue triangle. The F protein is illustrated as a purple trimer with the domain that refolds in green. The receptor molecule is illustrated as a light brown cylinder with a red triangle as the attachment point (Reprinted with permission from (61). Copyright Elsevier B.V. 2014).

This hints to the fact by understanding the G stimulation process, one can gain a general insight into the activity of other receptor binding proteins in the paramyxovirus family. Currently, there are no molecular models to explain the stimulation of G by ephrin binding to activate F. We use a computational approach to observe the structural and dynamics changes induced by ephrin binding to G. We utilize this knowledge to

formulate a working hypothesis while performing our MD simulations and analyses (details in section 5.3 and 6.2.2). The information from these experiments often supplements our results, and helps create a platform for a comparative study.

Figure 2.5 Residues that affect fusion protein activation. The blue indicate residue when mutated resulted in hypo-fusogenecity and the yellow are the residues, which do not alter fusogenecity when mutated

Results from flow cytometry (70, 77), circular dichroism spectra (69, 70, 77), cell-cell fusion assays (69, 70) and monoclonal antibody assays (69, 161) indicate that triggering of the F protein involves the exposure of the FAD. It can be inferred from these experiments that the FAD is important for conformational integrity, G-F interactions and F triggering; hence this region is termed the F-activation domain (FAD). (71, 144, 146, 160-162, 166, 168-171). Additionally, biochemical and biophysical studies suggest that the ephrin induced conformational changes in G involves certain residues at the base of FAD critical for the activation of F constitute the F-activation site (69, 77). Figure 2.5 highlights the residues on RBD, which upon mutation affects the fusion protein activation and Table 2.1 lists the various experimentally observed

mutations done on FAD of G. Only those residues are listed whose contribution to fusogenecity have been successfully tested.

Table 1.1 List of experimental mutations done on Nipah virus. The * indicate fusogenicity similar to wild type, ↑ indicate hyper-fusogenicity and ↓ indicate hypo-fusogenicity and blank is non-fusogenic.

Mutation	Fusion	Year
T103A		2016 (165)
T117A	↑	
T119A		
S129A	↑	
N159A		2015 (70)
I160A	↑	
S161A		
P163A	*	
N164A	*	
P165A	*	
L166A	*	
P167A		
C146S		2012 (160)
C158S		
C162S		
N72Q	*	2012 (161)
N159Q		
I83A	*	2008 (166)
I94A	*	
I101A	*	
I105A	*	
I112A	*	
I120A	*	
I124A	*	
I131A	*	
I138A	*	
I155A	↓	
I160A	↓	
I170A	*	

The content of this table has been continually updated since the inception of the work presented in this dissertation, which began in 2012. The receptor binding protein activates the fusion protein only when stimulated. Thus, the mutations on G that affect the fusion positively (hyperfusogenic) or negatively (hypofusogenic) are the ones that are important to induce G stimulation. In all the cases, the cell surface expression and ephrin binding are determined and fusion is measured based on the syncytium formation (63, 66, 69, 71, 76, 77, 172). Taken together, it is known that the receptor binding sites are > 2 nm away from the F activation site (65, 71, 76, 80, 87, 144, 146, 160-162, 166, 168-171, 173). Furthermore, the RBDs are arranged as dimer-of-dimers, with the FAD serving as a two-fold axis of symmetry (65).

2.4 Dynamic allostery in fusion activation

The word “allosteric” was first coined in 1961 by Jacques Monod and François Jacob (174) from the Greek word *allo* meaning other or different and *steric* meaning solid, as used in steric hindrance. Over the past decades, the interpretation of allostery has evolved which we briefly discuss below.

The classical idea of allostery states that, if there are two topologically distinct sites, presumably distant from each other within one protein and each bind to different ligands, it is possible that they may interact despite being non-overlapping in their molecular structure (175, 176). This mechanism is illustrated in the top panel (allosteric protein) of Figure 2.6. Here, the protein contains two sites, namely, the allosteric and the active sites. The protein is considered functionally “inactive” in the absence of any ligand, however, when the ligand binds to the allosteric site, it induces a change in the functional/structural properties of the active site, thereby altering the protein to its

“active” state. The activation of the protein by this process completes the allosteric mechanism and allows the protein to interact with the substrate.

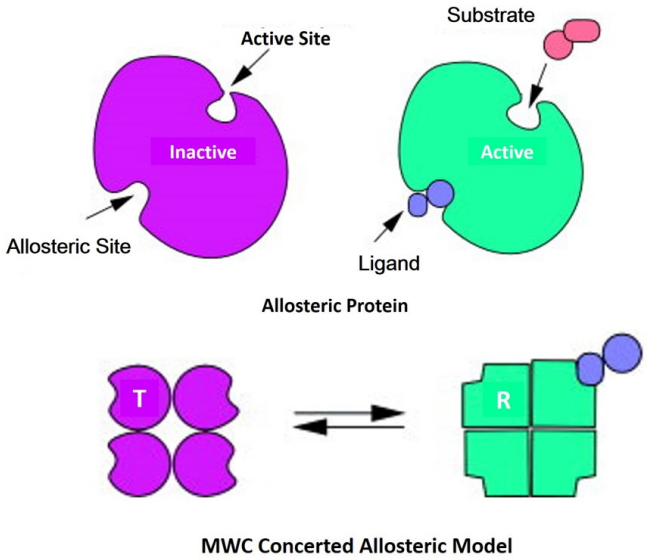


Figure 2.6 Schematic representation of the unified concept of allostery (Adapted with permission from (177). Copyright Wiley-Liss, Inc. 2004).

Later, in 1965,the Manod Wyman Changeux (MWC) model was proposed (178), which assumes that an allosteric enzyme comprised of multiple subunits can exist in two different conformations, i.e. active (R) or inactive (T). These subunits are expected to collectively maintain identical conformation as shown in the lower panel of Figure 2.6. According to the MWC model, allostery results in the inter-conversion of the protein from its R to T state or vice versa, in a concerted manner.

Two decades later, in 1984 researchers Cooper and Dryden, in their theoretical work introduced the concept of dynamically mediated allosteric interactions which revolutionized our understanding of allostery, and inspired more research in this direction (96). Their study demonstrated that these interactions essentially require long-range inter-site communication with the aid of atoms or structural groups dispersed throughout the protein, such that, it directly or indirectly experiences the presence of

ligands at each of the concerned sites (96). This was the first explicit articulation of dynamic allostery, which showed the theoretical relevance of entropic energetic contribution to biological functions. Here the authors demonstrated using statistical thermodynamic formalism that, changes in the frequency and amplitude of thermal fluctuations in a protein upon ligand binding could result in cooperative energies (order of a few kcal/mol) without perturbing the average structure. They showed that the ligand binding may have different effects on the protein: (a) the presence of a ligand can stabilize certain conformational sub-states over others and result in a shift in the mean of the probability distribution which is essentially the conformational change in the conventional sense; (b) the shape of the distribution might be affected where a narrower or broader distribution represents a conformational ensemble shift in the protein structure due to ligand binding. They explain allostery as a change in the conformational dynamics in the absence of a gross structural change in the protein, which is observable in the thermodynamics of the ligand attachment.

With the significant advancements in various experimental techniques and help from high computing power, researchers today are able to come up with a more refined understanding of the mechanics involved within a protein and its periphery. Based, on the outcome of both experimental and theoretical work, there is a growing consensus that all proteins are potentially allosteric (177, 179). A modern concept related to the population shift within conformations, also called the “new view” was introduced to describe allostery and accommodate the increasing evidence of the flexible nature of proteins. Several studies indicate that ligand binding at one site of the allosteric protein can effectively shift the population by redistributing the molecular ensembles, which,

results in conformational changes at some other sites. Thus, proteins should be treated as a dynamic ensemble of conformational states (177, 180). Based on this concept, experimentally, Otteman *et. al.* and Yu *et. al.* reported conformational changes as small as 1 Å which resulted in enormously amplified responses at sites 100 Å away (181, 182).

Moving forward with the established idea that proteins are flexible, whose conformations can be treated as statistical ensembles, and studying them from a thermodynamic point of view based on the foundation laid by Cooper and Dryden, the full dynamic spectrum of allosteric systems permits a classification of the regulatory strategies pertaining to its ensemble nature (176).

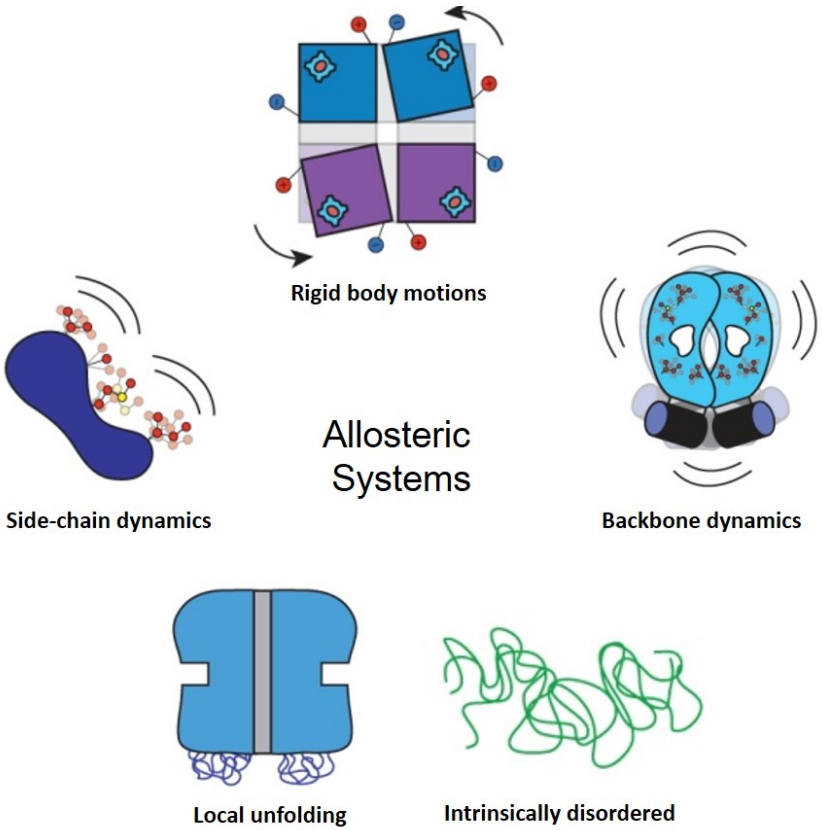


Figure 2.7 A schematic representation of the different allosteric systems that undergo change in the conformational ensemble upon activation (Adapted with permission from (176). Copyright Nature Publishing Group 2014).

Based on the above discussions, it is possible to categorize the different allosteric systems depending on the type of change in dynamics they exhibit upon activation by ligand binding as shown in Figure 2.7. Below, we present the 5 different dynamic allosteric processes known and cite relevant examples. (i) The rigid body motion, as shown in the figure is generally associated with structure based motion such as in the tetrameric form of haemoglobin (183, 184). (ii) The PDZ domain is known to exhibit modulation of ligand binding affinity via side chain dynamics (185). (iii) A unique example of backbone dynamics has been reported in CAP homodimer. The binding energetics of the ligand quenches its dynamics in the bound-state and induces a 90° change in the conformation of its DNA binding domain depicted using black cylinders in figure (186-188). (iv) An example where local unfolding plays a role in allostery is aminoglycoside *N*-(6′)-acetyltransferase II (AAC) — a homodimeric enzyme from *Enterococcus faecium* which has been reported to switch from positive cooperativity at low temperature to negative cooperativity at higher temperature (189). The local unfolding is illustrated using spaghetti-like lines in the figure. From a thermodynamic standpoint, such systems demonstrate a change in enthalpy of binding as the temperature increases, which is a signature of local binding (190). (v) The intrinsically disordered proteins (IDPs) are a challenge to the structure–function paradigm. They do not possess any stable tertiary structure under physiological conditions as shown in the figure (191-193), yet several IDPs have been reported to exhibit complex allosteric coupling behavior. For example, in α-synuclein, its N-terminal membrane-binding domain has been found to couple to its C-terminal disordered region via allostery (194).

The ensemble representation of allostery highlights the statistical nature of the allosteric coupling process unlike the classic notion of allosteric pathway that portrays a static picture of the protein. It conveys a deterministic picture of the signal propagation process, wherein each molecule at any instant is in the same conformation or at least can be equally well represented by the same average conformation. With the recent developments in computation it has been established that the ensemble model can be used as a potential framework for interpreting long-time scale MD simulations of allosteric proteins (195). The current speed and accuracy of all-atom MD simulations makes it a powerful tool to generate the energy landscape maps of the allosteric processes and also allows one to track the changes in dynamics of the conformational ensemble (176). We discuss the various computational methods implemented to study allosteric stimulation of Nipah virus host binding protein in the next chapter.

X-ray crystallography indicates that the receptor binding induces only a minor change in the structure o the receptor binding domain of G (Figure 1.2). Also other G analogues were included the HN and H proteins of PIV5 and MeV respectively. The results from the RMSD calculations indicate that receptor binding induces only small structural changes in their respective RBDs (63). Consistently in the Nipah RBD, the RMSD between the x-ray structures of its apo and ephrin bound states is < 2 Å (86, 90). Could this be an artifact of experimental technique due to the low temperatures? MD simulations at physiological temperature (93, 94) and were found to have similar low RMSD values. Under such conditions, receptor induced changes in thermal motions cannot be neglected. As we have learnt from many eukaryotic proteins like GPCR, PDZ domains, etc, where receptor binding also induces minor structural changes, signal

transduction occurs via a combination of receptor-induced changes in structure and dynamics. Such dynamic allostery is observed when the ligand binding at one site of the allosteric protein effectively shifts the population by redistributing the molecular ensembles, which, results in conformational changes at some other sites (96, 98, 177, 179, 180, 196). A host of experimental techniques including isothermal titration (197, 198), differential scanning calorimetry (199-201), fluorescence anisotropy detection (202-204) and circular dichroism (205, 206) are used to study allostery in proteins. However, nuclear magnetic resonance (NMR) stands out, as it allows the possibility to probe protein dynamics in a wide range of time scales: from picosecond to nanosecond dynamics of side chains and from microsecond to millisecond dynamics of the backbone (98, 188, 207-214) (215). Although, NMR provides an excellent resolution of less than 2 Å in identifying conformational changes, it is applicable to proteins with molecular masses less than 50 kDa (216, 217). In this regard, owing to the large size of the host binding protein (molecular mass of monomer ~65 kDa and tetramer ~ 280 kDa) (160), NMR is not a feasible technique to study its allostery. However, all-atom simulation techniques are a powerful alternative approach to investigate the proteins dynamics at the molecular scale (106, 109, 195, 218, 219).

CHAPTER 3

COMPUTATIONAL TECHNIQUES

All atom simulations can provide the intricate detail concerning individual particle motions as a function of time (106-108). A recent approach to understanding a system or to decipher its working mechanism has been to perform both experiments and simulations in tandem. For instance, simulations and NMR experiments (220, 221) help us obtain a better understanding about the structure, dynamics and thermodynamics of biological macromolecules. More often than not, simulations provide intricate details that may not be evident and obvious from experiments (222). Typically, based on the system and its property of interest, one can decide the simulation technique to adopt. Conducting simulations require an interdisciplinary approach, where concepts of physics, chemistry, biology, computer science and statistics are implemented in unison (223).

The study of proteins poses challenge owing to their complexities. Simulations allow us to gain information of such organic systems at a molecular level to study different properties of proteins such as, protein folding, protein dynamics, thermal stability, structure prediction etc. Any biological activity of a system, such as, protein – protein, protein – ligand or protein – water involves interactions that are mediated by multiple physical forces, which may have a static or dynamic origin. Depending on the user's focus of study, one can chose an appropriate technique to simulate such systems

over different time scales (195, 224, 225). For example, the study of protein dynamics requires one to perform a residue wise analysis of the translation, rotation and fluctuation in its conformity, which occurs over comparatively shorter time scales, and hence necessitates a simulation technique that samples the conformational space by providing a time resolution much shorter than the phenomena itself. In order to study any event occurring over longer time scales, such as protein folding, Monte Carlo (MC) simulations are performed. Although MC simulations allow us to study the global changes in protein dynamics, it does not provide any detail about the molecular motion of individual residues. Here, MD is found to be the best technique to simulate protein dynamics at longer time scales by tracking the time dependent evolution of individual molecular motions.

While performing MD simulations, the protein is placed in a solvent that is either implicit or explicit in description (114), and the choice of the solvent has critical impact on the analysis and outcome of the simulation. The solvent used to perform a simulation is primarily determined based on (i) size of the protein, and (ii) time scale of the biological event. In this regard, an optimal computation time can be achieved by choosing implicit solvent to study large proteins at longer time scales (~ μs), however, explicit solvent is usually considered to simulate biological events occurring at shorter time scales (~ ps to ns). For any macromolecule simulated using an implicit solvent, the motion of its atoms is governed by the Langevin equation, which implies that the dynamics of the macromolecule is stochastic in nature. However, when one implements explicit solvent, the time dependent atomic motions follow the classical Newtonian laws

of motion, essentially making the molecular dynamics of the system deterministic. We discuss this further detail in section 3.1.

In section 3.2, we discuss the details of structure prediction tools implemented to model the FAD, whose structure as discussed in Chapter 1 is unknown. The two tools used are (i) Homology modeling and (ii) *Ab initio* prediction. Homology modeling is a comparative technique, which takes advantage of the fact that evolutionarily related proteins have similar sequences, and therefore, by performing optimal structural superposition of the known and unknown structures, it is possible to determine the unknown structure. *Ab initio* prediction performs MC simulations that do not use forces, but compare energies of multiple conformations generated for the target sequence. One can identify the native structure corresponding to the lowest energy. Finally, in section 3.3, we discuss the technique based on inverse machine learning, which is used to quantify the changes in conformational ensembles arising from dynamic allostery.

3.1 Stochastic dynamics

While simulating proteins on longer time scales, it has to be taken into account that protein atoms are localized by their covalent interactions, while those in a fluid are not. One can use equations of motion to model the dynamics of individual fluid particles. If the masses (m_i) and initial position coordinates ($\boldsymbol{r_i}$) at $t = 0$, of N particles in a solvent are known, one can calculate velocity ($\boldsymbol{v_i} = \frac{d\boldsymbol{r_i}}{dt}$) and describe the dynamics of the system by Langevin equation (226, 227).

$$m_i \frac{d\boldsymbol{v_i}}{dt} = -m_i \gamma \boldsymbol{v}_i - \boldsymbol{\nabla} V(\boldsymbol{r}_i) + \boldsymbol{A}(t) \quad (3.1)$$

In Equation 3.1, γ is a friction coefficient, and $\boldsymbol{A}(t)$ is a fluctuating force that can be approximated by Gaussian white noise on the time scale considered. The protein and solvent environment exert the average effective potential $V(\boldsymbol{r}_i)$ that confines the motion (226). One can perform simulations by solving the Langevin equation, which forms the basis for stochastic dynamics. In order to simulate long-time phenomena, stochastic modeling of the internal motions is found to be a useful approach, in which only the relevant portion of the protein is explicitly included and the remainder of the molecule, as well as the solvent, serves to provide an effective potential, a frictional drag, and a heat bath (226, 228, 229). A significant amount of computational time can be saved by implementing this technique, where the solvent molecules are not explicitly present in the system, but are rather approximated by the frictional drag and the white noise associated with thermal motions of solvent molecules. In such a case, the solvent is referred to as being implicit. However, if one accounts for the actual presence of solvent molecules, i.e. explicit description of solvent, the simulations become time consuming. In such simulations, approximations involving frictional drag or consideration of the fluctuating force become unnecessary. So, for explicit solvents, by setting $\gamma = 0$ and removing the term $A(t)$, Equation 3.1 essentially takes the form of Newton's equation of motion (Equation 3.2), and the situation evolves to performing deterministic molecular dynamics (MD) simulations. Below, we present a brief introduction to the principals involved in MD. We have used the Groningen Machine for Chemical Simulations (GROMACS) version 4.3.5 application to perform all MD simulations (230).

3.1.1 Explicit solvent molecular dynamics simulation

The central idea of MD is based on classical mechanics, which iteratively solves the Newton's equations to study an N body system. Specifically, if the masses and initial position coordinates (at $t = 0$) of N particles are known, one can employ the MD approach (231). In MD, the force acting on an atom is computed by Newton's first law of motion (Equation 3.2).

$$\boldsymbol{F}_i = m_i \cdot \boldsymbol{a}_i = m_i \cdot \frac{d\boldsymbol{v_i}}{dt} = m_i \cdot \frac{d^2\boldsymbol{r_i}}{dt^2} \qquad (3.2)$$

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i} \qquad (3.3)$$

Here, m_i and r_i is the mass and position of the i th atom. It is to be noted that in order to compute the velocity ($\boldsymbol{v}_i$) and acceleration ($\boldsymbol{a}_i$), we must know the time interval (dt) over which the motion of the i th atom is considered to occur, in other words, the time interval between which the atom experiences a successive force ($\boldsymbol{F}_{i+1}$). The force, $\boldsymbol{F}_i$ is related to potential energy, V, as shown in Equation 3.3. Equations 3.2 and 3.3 are solved simultaneously over the time step (dt), which constitutes a single MD step. This process continues iteratively over a specified number of MD steps as per the user's discretion.

For a system in which the initial velocity of the N atoms is unknown, one must assign the respective velocities to the individual atoms. The initial velocities of the atoms are coupled to the temperature (T) of the system and it must be noted, that the velocities and their distribution are determined using the Maxwell-Boltzmann distribution function prior to the first MD step.

The potential energy, V, in Equation 3.3, is incorporated in MD as an input parameter, which is also known as the force field. These force fields have been parameterized to reproduce quantum mechanical calculations on small model systems, then adjusted to provide improved agreement with higher-quality *ab initio* data, crystallographic structures or experimental data (232). A force field consists of both bonded and non-bonded interactions. The bonded interactions can be classified as (i) two bodied, in which the distance between two atoms are considered, (ii) three-bodied, where the angle between atoms are taken into account and (iii) four-bodied in which the dihedral angles are maintained. The non-bonded interactions include the Lennard-Jones potential, electrostatic and the van der Waals interactions.

While running the MD simulations, one can set it up as an NVE (constant number, constant volume, and constant energy) ensemble or an NVT ensemble where the temperature is maintained constant. From the biological point of view, it is essential to maintain physiological conditions, which entails a temperature of 310 K and an atmospheric pressure of 1 bar, hence, we use the NPT ensemble unless otherwise mentioned. It is important to control the temperature of the system to prevent it from drifting during equilibration or as a result of force truncation or integration errors.

The MD simulations are performed inside a space-filling box, dimensions of which are carefully determined based on the size of the system and number of atoms. As the simulation progresses, and the positions of the atoms are updated, edge effects of finite size may develop, which can be eliminated by applying periodic boundary conditions. The effect of these boundary conditions can be visualized as a space-filling box surrounded by translated copies of itself. Hence, this creates an impression that the

system has no boundaries and is infinite. The artifact of periodic conditions replaces the artifact from unwanted boundaries in an isolated cluster.

Finally, the coordinates are written in an output file at regular intervals. A representation of these coordinates as a function of time is called a trajectory. It is important to monitor the temperature and energies at the required values throughout the simulation. Eventually, the system reaches an equilibrium state, which can be determined by observing the time evolution of parameters like temperature, energy, RMSD, or other distinct predefined factors. The equilibrated trajectory is then used for statistical averaging to calculate various macroscopic properties.

Implementation: All the MD simulations are carried out under isobaric-isothermal boundary conditions, and using Gromacs version 4.5.3 (230). Temperature is maintained at 310 K using an extended ensemble approach (233, 234) and with a coupling constant of 0.2 ps. An extended ensemble approach (235) is also used for maintaining pressure. Pressure is maintained at 1 bar using a coupling constant of 1 ps and a compressibility of 4.5×10^{-5} bar^{-1}. We also examine the effect of imposing alternative boundary conditions, including the canonical and micro-canonical boundary conditions, as discussed in the results section. The protein and ions are described using OPLS-AA parameters (236), and the water molecules are described using TIP4P parameters (237). While the TIP4P water model is known to over-estimate diffusion (238), it reproduces bulk water reorientational processes observed in the experimental spectrum (239). Note that while we do not model induced effects explicitly, which are important to both ionic interactions (240, 241) and hydrogen bonding (242-245), we discuss the implications of this approximation in the results section. Electrostatic

interactions are computed using the particle mesh Ewald scheme (246) with a Fourier grid spacing of 0.15 nm, a sixth-order interpolation, and a direct space cut-off of 10 Å. van der Waals interactions are computed explicitly for interatomic distance up to 10 Å. Charge neutrality of the three MD unit cells are maintained by selecting appropriate differences between the numbers of Na^+ and Cl^- ions. The bonds in proteins are constrained using the P-LINCS algorithm (247), and the geometries of the water molecules are constrained using SETTLE (248). These constraints permit use of an integration time step of 2 fs. The motion of the center of mass is reset every 100 ps. System specific MD parameters including system size and length of trajectory are provided in the respectively in Chapter 4 and Chapter 5.

3.1.2 Implicit solvent simulation

Implicit solvent is described by a continuum model, where the total free energy of the solvated molecule can be expressed as $E_{tot} = E_{vac} + \Delta G_{solv}$, (249) where E_{vac} is the molecule's energy in vacuum, and ΔG_{solv}, is the solvation free energy, in other words, the energy required to transfer the molecule from vacuum into the solvent. E_{vac} is a classical potential function that accounts for interactions and physical components, such as bond and angle stretching, torsional twist, and van der Waals and Coulomb interactions. The solvation free energy is composed of two components, an electrostatic and a non-electrostatic component, such that $\Delta G_{solv} = \Delta G_{el} + \Delta G_{sur}$. To account for the electrostatic component ΔG_{el}, generalized Born (GB) is implemented. It is a fast but approximate method to calculate molecular electrostatics in solvent described by the Poisson Boltzmann equation such that water is modeled as a dielectric continuum. The

total electrostatic force $\boldsymbol{F}_i$ on an atom i, is the net force from partial changes $\boldsymbol{F}_i^{charge}$ on the atom minus the GB force $\boldsymbol{F}_i^{GB}$ on the atom, hence

$$\boldsymbol{F}_i = \boldsymbol{F}_i^{charge} - \boldsymbol{F}_i^{GB} \qquad (3.4)$$

The $\boldsymbol{F}_i^{GB}$ on atoms is a derivative of the total GB energy, which is a function of the relative atom distance $\boldsymbol{r}_{ij}$, the effective distance f_{ij} and the dielectric between atoms i and j. The effective distance f_{ij} is determined by the GB function

$$f_{ij} = \sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{\frac{-r_{ij}^2}{4\alpha_i \alpha_j}}} \qquad (3.5)$$

where α is the Born radii of the atoms. The non-electrostatic component of free energy ΔG_{sur}, is found to be proportional to the total solvent-accessible surface area of the molecule, with a proportionality constant obtained from experimental solvation energies. This can be incorporated in the simulations by choosing a surface tension value, which accounts for the free energy associated with cavity formation, and is modeled as a linear function of the atomic surface area. The inclusion of the surface tension value is related to the frictional drag coefficient (γ) and fluctuating force ($\boldsymbol{A}(t)$), which as discussed above are essential approximations in describing implicit solvent. $\boldsymbol{A}(t)$ is a form of white noise that is represented by a Gaussian, and is related to γ by the fluctuation dissipation theorem (250). Therefore, we can perform stochastic dynamics to study protein in implicit solvent, by following a method similar to that discussed in section 3.1.1, but in this case, we replace Equation 3.2 by Equation 3.1.

Implementation: In the stochastic simulation the partial charges on the atoms and the connectivity between them are described using the OPLS-AA force field,(236) and their Born radii are computed at every time step using the Onufriev-Bashford-Case (OBC) scheme.(249) Both systems are simulated under isothermal conditions, where the temperature is maintained at 310 K using an extended ensemble approach(233, 234) and with a coupling constant of 0.2 ps. In accordance with the OBC scheme, the solvent is described using a dielectric constant of 80, and the electrostatics and van der Waals interactions are truncated beyond interatomic distances of 24 Å. Also in accordance with the OBC scheme, the free energy associated with cavity formation is modeled as a linear function of the atomic surface area using a surface tension of 0.005 kcal/mol/Å^2. The bonds in proteins are constrained using the P-LINCS algorithm.(247) The G protein and the G-B2 complex are simulated for 1.1 x 10^6 integration time steps, and the first 0.1 x 10^6 time steps are left out as equilibration.

3.1.3 Accelerated conformational sampling

As discussed in the previous section, MD simulations use atomistic empirical force fields, which require a time step of the order of femtoseconds for proper integration of the equations of motions. This permits an accurate resolution of the fastest motions (typically in the order of ps) in the system like bond stretching and bond bending. However, events like protein folding, domain re-orientation etc. often occur at longer timescales that may extend into microseconds or milliseconds. In this regard, simulating events spanning over such long time scales becomes a computing challenge as the total computing time can take several months. To overcome this problem a

variety of methods have been developed, commonly referred to as enhanced sampling techniques (231, 251).

Metadynamics is one of these methods, which uses the enhanced sampling technique to explore the infrequent rare events whose occurrence can extend into the millisecond time scale. We use metadynamics to probe different conformations and domain orientations in proteins. It is a method that facilitates sampling by the introduction of an additional bias potential, that acts on a set of selected number of degrees of freedom referred to as reaction coordinates or collective variables (CVs). At every MD step, a history dependent bias potential is deposited as a function of the CVs in the form of Gaussians hills as

$$V_G(S(x),t) = \sum_{t=t_G,2t_G,..nt_G} w\ exp\left(-\frac{(S(x)-s_t)^2}{2\delta s^2}\right) \qquad (3.6)$$

where for a new Gaussian added after a time interval t_G, the bias potential (V_G) at time t is related to the height (w) and width (δs) of the Gaussians as shown. Here, $s_t = S(x(t))$ is the value of the collective variable at time t. This potential fills the minima in the free energy surface as the MD progresses, and assists the system in escaping the energy minima to visit new regions in the configuration space that are practically inaccessible within reasonable computing time in the case of unbiased or conventional MD simulations (252-255).

Finally by adding up all the bias potential introduced at every time step, one can reconstructs the multidimensional free energy of the complex system, which is based on the artificial dynamics performed in the space defined by the CVs. The CVs basically provide a coarse-grained description of the system. It is important that the description of the CVs is such that it can distinguish between the initial and the finals states and

identify all the relevant intermediate states. To carry out metadynamics, we have used the Plumed patch with GROMACS (256).

Reports suggest that it is difficult to decide when to terminate a metadynamics run, and opinions may be divided based on the user. This is because the free energy does not converge to a definite value but fluctuates around the correct result in a single run. This uncertainty in the results is due to an average error, which is proportional to the square root of the bias potential deposition rate (257, 258). From a practical standpoint, continuing a run carries the risk that the system is irreversibly pushed in regions of configurational space that are physically irrelevant. To avoid this shortcoming in metadynamics, a slightly modified technique called 'well-tempered metadynamics' has been developed over the years, and implemented to great effect. This approach utilizes a self-healing umbrella sampling method and offers the possibility of controlling the regions of free energy surfaces that are physically meaningful to explore (259).

3.2 Protein structure prediction

As outlined in Chapter 1, our motivation is to generate the complete structure of the G protein, of which the FAD has no established structure. We take advantage of existing protein structure prediction tools to address this. Protein structure prediction is the process of developing the three dimensional structure of the protein from the knowledge of its amino acid sequences. They can be broadly categorized into three types based on their *modus operandi*: (i) prediction based on information available in sequence and structural databases, commonly known as homology modeling or comparative modeling, (ii) *ab initio* methods based entirely on physicochemical principles and (iii) threading, also known as the fold-recognition method is a method to

model proteins which have the same fold as proteins of known structures, but do not have any homologous proteins with known structure (260, 261). In this dissertation, we use homology modeling as well as *ab initio* methods to generate the structure of the FAD in the G protein, where the sequence of the G.

3.2.1 Homology modeling

Homology modeling is both a sequence and a template based technique. When this technique was first introduced in the 1960s, structures were constructed using the information obtained from structure alignment about the template core regions, its loops and side chains (260, 262). More recently, homology modeling encompasses the approach where one can approximate positions of conserved atoms in the templates to calculate coordinates of the final model (263, 264) or even use distance geometries and optimization techniques to satisfy spatial restraints obtained from the sequence-template alignments (265-267). In general, homology modeling involves four major steps: (i) finding known structures on which the sequence to be modeled is templated, (ii) aligning the sequence with the templates, (iii) building a model, and (iv) evaluating the final model (268, 269). The structure to be modeled is commonly referred to as the query or target sequence.

To address the first step, the template for modeling the sequence can be searched by using various sequence comparison tools such as BLAST, (270, 271) which matches the sequence similarities. Once the possible templates are identified, in order to perform step (ii), these sequences are aligned using programs for multiple sequence alignment like CLUSTAL OMEGA (272). For steps (iii) and (iv), we use MODELLER to build our three dimensional structures (273-275). Here, the modeling is

implemented by satisfying the spatial restraints in the template structure. In this real-space optimization method, the program starts by building the model using the distance and dihedral angle restraints of target sequence derived from the sequence alignment with 3D structures of the template. Certain force field terms are added to enforce proper stereochemistry (276). This together forms the objective function, which once optimized generates the model in Cartesian space. The end result is an all-atom model of a sequence, which is based on its alignment to one or more related protein structures (268). Also, based on the user's discretion, MODELLER can generate multiple possible structures, which are accompanied by their respective model evaluation scores, or a figure of merit. Therefore, by considering the scores, one can gauge the accuracy of the predictions and make an informed judgment of the final structure (277). It must be kept in mind that the accuracy of a model is directly proportional to the percentage sequence identity to its template (269, 277, 278).

We would like to add that MODELLER can also handle steps (i) and (ii) for generic cases, however, for our system, we did not opt these features. Instead, we manually used BLAST and CLUSTAL OMEGA as mentioned above. Later in Chapter 6, we discuss all the aspects mentioned above and explain our choice of this manual implementation.

3.2.2 *Ab initio* structure prediction

The basic assumption at the heart of this method is that a protein sequence tends to fold to a native conformation or ensemble of conformations near the global free-energy minimum. The most successful approach for *ab initio* structure prediction is achieved by utilizing the knowledge-based energy functions (279). Knowledge-based

potentials refer to the empirical energy terms derived from the statistics of the solved structures existing in the Protein Data Bank (PDB) (280). These terms include, (a) the generic and sequence independent terms like the hydrogen bonds and local backbone stiffness of a polypeptide chain (281) and, (b) the sequence dependent terms like pair wise residue contact potential (282), distance dependent atomic contact potential (283-285) and secondary structure propensities (286, 287). This method was first successfully demonstrated by Bowie and Eisenberg, who assembled small fragments of proteins taken from the PDB library to generate their protein models (288).

We use a tool developed on the knowledge-based potential named ROSETTA (289, 290) to perform *ab initio* structure prediction. Once the sequence is provided as input into ROSETTA, it serially fragments the complete sequence into sections, or local sequences, comprising of 3 residues. Following this, every local sequence is compared to the existing data in the PDB for structural similarity, and consequently a library is created depending on the information collected. Parallelly, another library is created for sections obtained by fragmenting the original sequence every ninth residue. These two libraries are then used to construct a crude low resolution structural model of the target sequence based on the conformations specified by the heavy backbone and the Cβ atoms. This low resolution structure is then subjected to an all-atom refinement procedure using an all-atom physics based energy function that include van der Waals interactions, pair wise solvation free energy, and an orientation-dependent hydrogen-bonding potential, thereby yielding a possible conformation. Then the conformation is refined where, multiple rounds of Monte Carlo sampling are carried out to minimize the positional uncertainty of the local sequences (291, 292), before the final stable structure

is determined. As suggested by the developers of ROSETTA, once must perform this process to generate a minimum of 10,000 structures, which provides a reasonably good statistical sampling. Finally, a score is associated with every structure and it is determined by considering a function, which consists of sequence-independent terms representing hard sphere packing, α-helix and β-strand packing, and the collection of β-strands in β-sheets along with the more important sequence-dependent terms representing hydrophobic burial and specific pair interactions such as electrostatics and disulfide bonds. One can identify the best structure by comparing the score of each conformation. We specifically use the Fold and Dock module of ROSETTA that allows the incorporation of symmetry and symmetric constraints (293), the details of which are provided in Chapter 6.

3.3 Analysis of conformational ensembles

To study the allostery involved in the stimulation of the host binding protein, one needs to analyze the ensemble of the three dimensional configuration to observe the intrinsic motions, which are known to be tightly related to the changes induced by ligand binding (182, 294-299). It is important to characterize the molecular motions quantitatively, because it provides a basis to directly correlate the changes in thermodynamic properties to its corresponding changes in molecular motion (107). This becomes challenging because it involves the comparison of two high dimensional data sets (300, 301).

The traditional approach to compare two conformational ensembles of proteins, $\mathbb{R} = \{r_1, r_2, \ldots, r_m\}$ and $\mathbb{R}' = \{r_1', r_2', \ldots, r_m'\}$, where r denotes a 3n-dimensional coordinate and m denote the number of conformations in the ensemble, is to compare

their respective summary statistics, like centers-of-masses (CoMs) and root mean square fluctuations (RMSFs). However, if a subset of the summary statistics of the two ensembles are found to be identical, it does not imply that the remaining 3n−6 summary statistics of two ensembles will also be identical (302). The general problem of finding and choosing a feature that appropriately distinguishes two ensembles can be overcome by comparing ensembles directly against each other, and prior to any dimensionality reduction. A further advantage of comparing ensembles directly against each other is that the resulting quantification naturally embodies differences in conformational fluctuations.

We use a method recently developed to compare the ensembles (93, 94). It quantifies the difference between two ensembles in terms of a metric, η, that satisfies two conditions: (i) $\eta(\mathbb{R} \to \mathbb{R}') = \eta(\mathbb{R}' \to \mathbb{R})$, and (ii) if $\eta(\mathbb{R} \to \mathbb{R}') = \eta(\mathbb{R}' \to \mathbb{R}'')$, then $\eta(\mathbb{R} \to \mathbb{R}') = \eta(\mathbb{R} \to \mathbb{R}'')$. This metric is also universal in that it is not bounded by system type or size, and can be used to examine differences in ensembles at any structural hierarchy for example the functional groups, amino acids, or the secondary structures.

Mathematically, η is a function of the geometrical overlap between conformational ensembles, $\mathbb{R}$ and $\mathbb{R}'$.

$$\eta = 1 - \|\mathbb{R} \cap \mathbb{R}'\| \qquad (2.6)$$

It is normalized, that is, $\eta \in [0, 1)$ and it takes up a value closer to unity as the difference between the ensembles increases. $\|\mathbb{R} \cap \mathbb{R}'\|$ is estimated by solving an inverse machine learning problem. In the traditional sense, machine learning is used for data classification (303-308) – the classification function, or machine $\left(\mathrm{F}\,(\mathrm{r})\right)$, is first

trained on a set of instances with known group identities, and then used for predicting the group identity of an unclassified instance. In principle, the conformational ensembles $\mathbb{R}$ and $\mathbb{R}'$ can also serve as training data to train a classification function, $\left(\mathrm{F}\left(\mathrm{r}\right)\right)$, which can, in turn, be used to predict whether an unseen conformation belongs to $\mathbb{R}$ or $\mathbb{R}'$. It is known that if $\left(\mathrm{F}\left(\mathrm{r}\right)\right)$ is constructed and trained appropriately, then the overlap between $\mathbb{R}$ and $\mathbb{R}'$ can be extracted from F (r) (94).

We calculate the η values RBD in its monomeric and dimeric forms and use it to our benefit to understand the effect of ephrin binding on the G protein.

CHAPTER 4

ROLE OF WATER IN SIGNAL INCEPTION

X-ray crystallography (86, 90) suggests that the interfaces of G with ephrins B2 and B3 are extensive, with interface areas greater than 2500 $Å^2$. In addition, the X-ray structure of the G-B2 complex contains a large number of interstitial water molecules, which is, in fact, three times greater than the average numbers of water molecules found in other protein-protein interfaces (shown in Figure 1.3) of comparable sizes (113, 309). In this regard, the following questions arise, (i) Does the resolution of water molecules in X-ray structures generally imply that their dynamics are sluggish, or is the interface between G and ephrins sufficiently wet for the dynamics of interstitial water molecules to be similar to the bulk? and (ii) In general, how do the dynamics of water at protein-protein interfaces compare against the dynamics of water at protein-water interfaces? Furthermore, water is known to govern the dynamics of the protein binding groves and often the protein response is dominated by the rearrangement of the water network on the protein surfaces (97, 310-312). Thus the question arises, does the interfacial water at the G-B2 interface contribute to the inception of the allosteric signal at the receptor binding site?

4.1 Interfacial water and protein function

The dynamical properties of water at protein surfaces have been studied extensively using both experimental and computational techniques (313-325).

Experimental techniques typically include nuclear magnetic resonance spectroscopy, magnetic resonance dispersion, dielectric relaxation, neutron scattering as well as time-resolved fluorescence. These techniques measure the time scales of rotational and translational dynamics of interfacial water molecules, but they detect different physical phenomena and, consequently, are subject to different interpretations. Computational techniques primarily constitute molecular dynamics simulations in which interatomic forces are obtained by treating polarization effects implicitly or explicitly, or at the quantum mechanical level, and the choice of the specific treatment is determined by the system size, observation time and spatial resolution. Together, these techniques allow observation of a wide range of time and length scales. While a complete picture of the complex interactions of hydration water molecules that accounts for all the data is still lacking, there is little doubt that the dynamics of interfacial water molecules are statistically different from those in the bulk. In fact, there is also now growing consensus that water molecules at the protein-water interface translate and rotate slower compared to the bulk (320, 326-358). Additionally, the extent of the shift depends on the complex interplay of protein surface chemistry and topology, that is, whether water molecules are present in cavities or near hydrophobic/hydrophilic moieties, or near backbone/side chain groups, or near concave/convex surfaces.

In this chapter, we extend such studies to systems where water molecules at the interface of the G protein and ephrin to decipher the role of interfacial water in allosteric signal inception. As mentioned earlier in Chapter 1, in one of the two protein-protein complex, the Nipah virus G protein binds to cellular ephrin B2, and in the other, the same G protein binds to ephrin B3. These two ephrins share only a modest sequence

identity of 50% (Figure 4.1), even at the protein-protein interface. The interactions of G with these ephrins constitute the first step in Nipah infection (66, 68, 79).

Figure 4.1 Seqeunce alignment of ephrin B2 (UniProt ID: P52799) and ephrin B3 (UniProt ID: Q15768) showing the modest identity.

4.2 Properties of interstitial water at Nipah virus-host protein-protein interface

To characterize the properties of water molecules present in the interfaces of the G-B2 and G-B3 protein complexes, we generated separate $\sim {}^{1}/_{2}\,\mu s$ long MD trajectories of the G-B2 and G-B3 complexes, respectively, in explicit solvent. We begin the analysis by defining a scheme to distinguish between interstitial and bulk water molecules. The MD unit cells corresponding to the apo, B2 bound and B3 bound states of G contained a total of 30367, 40900, and 40666 water molecules, respectively.

4.2.1 Statistical definition of interstitial water

The interfaces formed between G and ephrins are known to be uneven. As a result, this rules out a definitive scheme for discriminating between the interstitial and bulk water molecules. Hence a scheme based on density distribution was adopted.

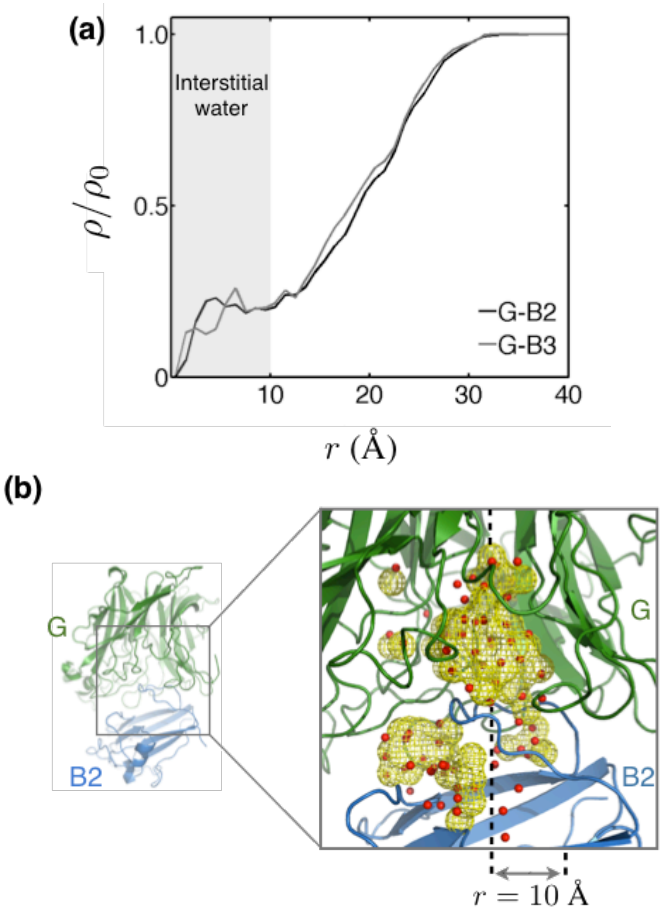


Figure 4.2 Water distribution in the interstitial regions of the G-ephrin complexes. **(a)** Normalized density of water (ρ/ρ_0) as a function of the perpendicular distance from the axes joining the geometric centers of G and ephrins **(b)** Partial view of the G-B2 complex showing a superimposition of the 69 interstitial water molecules resolved in the x-ray structure (red spheres), and the 65 highest occupancy regions observed in the MD simulation (yellow mesh). The axis of the right circular cylindrical region that connects the geometric centers of the G and B2 is drawn as a dashed black line (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

Figure 4.2a shows the water density variation as a function of the perpendicular distance from the axes joining the geometric centers of the G and the ephrins. In this

calculation only those water molecules that lie within the right circular cylindrical regions bounded by disks containing the geometric centers of G and ephrins have been considered. The heights of these cylinders are the distances between the geometric centers of the G and the ephrins. The average distances between the geometric centers of the G and the ephrins are 32.8 ± 0.2 Å and 34.7 ± 0.2 Å, respectively, for the G-B2 and G-B3 complexes. As shown in Figure 4.2a, the density distribution function exhibits an inflection point at a radial distances of 10 Å. Hence, the interstitial regions is defined as a 20 Å wide cylinder which thereby includes the region around the inflection point. The average numbers of water molecules in the interstitial regions of the G-B2 and G-B3 complexes are 65.3 ± 4.0 and 67.6 ± 3.9, respectively. These averages are comparable to the 69 water molecules resolved in this interstitial region in the X-ray structure of the G-B2 complex. Furthermore, the 65 highest occupancy regions in the MD simulation of the G-B2 complex coincide well with the positions of the water molecules resolved in in the X-ray structure (Figure 43b).

4.2.2 Diffusion Coefficients

Figure 4.3 shows the integrated form of the velocity autocorrelation of interstitial water molecules:

$$D(\tau) = {}^{1}/_{3} \int_{0}^{\tau} \langle \langle \boldsymbol{v}(0) \cdot \boldsymbol{v}(\tau) \rangle \rangle \, dt \qquad (4.1)$$

The double angular brackets in Equation 4.1 denote averages computed over the ensemble as well as the number of water molecules in the interstitial region. The diffusion coefficient of the interstitial water molecules is obtained from the Green-Kubo relationship (359) as a limiting case $D = D(\tau \to \infty)$. The diffusion coefficient obtained from periodic systems, however, needs to be corrected for finite size effects.

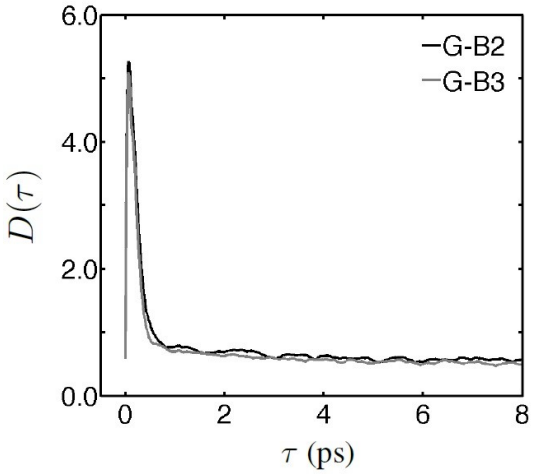


Figure 4.3 Integrated form of the velocity autocorrelation of water molecules, D(τ), occupying the interstitial regions in the G-B2 and G-B3 protein complexes (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

According to the hydrodynamic theory of diffusion,(360, 361) the leading order correction to the diffusion coefficient obtained from a cubic periodic cell is inversely proportional to the length of the cell,

$$D_{self} = D + \frac{k_b T \xi}{6\pi\eta L} = D + \frac{a}{L} \qquad (4.2)$$

In the equation above, η is the viscosity, ξ = 2.837 is a constant, and L is the length of the cubic cell. The higher order corrections to Equation 4.2 are relatively small. Instead of computing the viscosity needed for estimating the correction, we estimate the correction factor α empirically (360, 361). To accomplish this, we computed the diffusion coefficient of bulk water D from three different cubic cells of lengths 24.83, 31.85 and 40.70 Å (Figure 4.4).

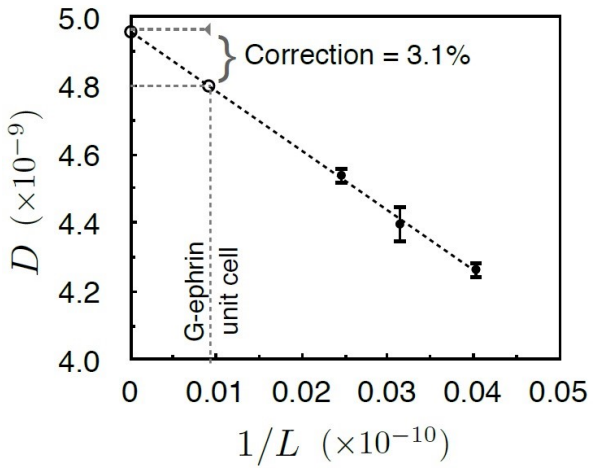


Figure 4.4 Diffusion coefficient D of water estimated from cubic periodic cells with different lengths L. D has units of m^2/s and 1/L has units of m^{-1} (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

In all the three simulations, water densities were maintained at 0.987 kg/dm^3. D is estimated from the velocity autocorrelations. While the slope of the line fitted to these data points yields *α*, the ordinate intercept yields the diffusion coefficient of bulk water corrected for finite size effects. The value obtained for D_{self} is 4.96 10^{-9} m^2/s. Different system sizes were used to simulate the G-ephrin complexes, and the finite size correction was calculated to be 3.1% of D. After accounting for this correction, we find that the diffusion coefficients of interstitial water molecules in the G-B2 and G-B3 complexes are 0.62 ± 0.07 x 10^{-9} and 0.57 ± 0.16 x 10^{-9} m^2/s, respectively. The standard deviations were obtained by dividing the trajectory into four separate blocks. The estimated values of the diffusion coefficients of interstitial water molecules are found to be almost an order in magnitude smaller than the computed bulk water diffusion coefficient. This difference is significantly larger than the mismatch between the computed and experimental estimates for the diffusion coefficient of bulk water. The experimental estimate for the diffusion coefficient of bulk water at 310 K is reported to be 3.1 x 10^{-9} m^2/s (362, 363). Our results indicate that despite the exceptionally high

wetness of the interfaces,(309) the interstitial water molecules diffuse fairly sluggishly. In addition, the effect of the difference in the chemical environment of the two ephrins on water diffusion is negligible.

4.2.3 Residence times

In this section we investigate if the water molecules in the interstitial regions exchange with the bulk? Figure 4.5 shows the residence time correlation,

$$R(\tau) = \frac{\langle s(0)s(\tau)\rangle}{\langle s\rangle} \qquad (4.3)$$

of the water molecules present in the interstitial regions of the G-B2 and G-B3 complexes. In Equation 4.3, the product $s(0)s(\tau)$ takes up a value of unity if a water molecule occupies the interstitial region continuously over a time interval τ. Note that in the estimation of averages, we do not include the water molecules present continuously from the beginning to the end of the analysis portion of the trajectory.

Interestingly, only one such “trapped” water in the G-B2 complex and four such water molecules in the G-B3 complex are seen. These trapped water molecules amount to less than 6% of the total water molecules in the interstitial regions. Modeling the residence time correlation as a sum of two exponential functions,(321)

$$R(\tau) = Ae^{-\tau/\tau_1} + (1-A)e^{-\tau/\tau_2} \qquad (3.4)$$

reveals two distinct subpopulations, *A* and (1 - *A*), of fast and slow exchanging water molecules. We use unweighted least square fitting to determine the values of *A*, τ_1 and τ_2. In the interstitial region of the G-B2 complex, we find that 93% of the water molecules have residence times of τ_1 = 1.4 ps, and the remaining fraction of water molecules have longer residence times of τ_2 = 64.7 ps. Consistently, a similar

distribution of the fast and slow exchanging water molecules in the interstitial region of the G-B3 complex, with *A* = 91%, τ_1 = 1.3 ps and τ_2 = 44.7 ps is calculated.

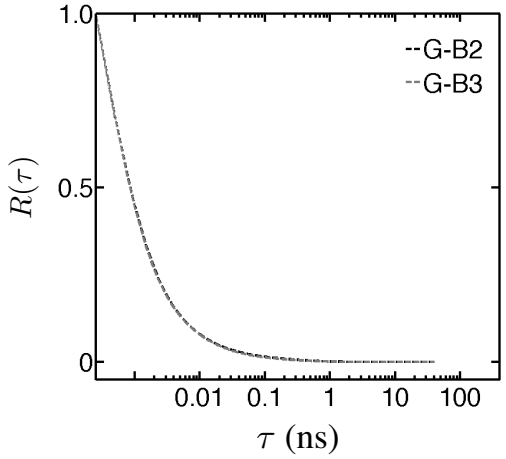


Figure 4.5 Residence time correlation of the water molecules, R(τ), occupying the interstitial regions in the G-B2 and G-B3 protein complexes (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

In these calculations the denotation of a continuous occupancy over time is subject to the choice of the observation time interval. Nevertheless, it is expected that the observation time intervals shorter than the ones used ($\Delta\tau$ < 0.25 ps) will reveal additional discontinuities, resulting in smaller computed residence times(364). Therefore, it can be safely concluded from this analysis are that more than 94% of the water molecules in the interstitial regions exchange with the bulk and that they have residence times less than 100 ps. More importantly, the interstitial water molecules tend to occupy crystallographic positions (Figure 4.3b), and exchange with bulk solvent. It is to be stressed that these residence times should not be compared directly to the reported residence times of water molecules at protein-water interfaces because R(τ) is also a function of the shape and size of the observation volume.

4.2.4 Hydrogen bond dynamics

To evaluate quantitatively the dynamics of hydrogen bonds made by interstitial water molecules a method based on autocorrelation functions of hydrogen bond populations (365-367) is implemented. In this method the hydrogen bond correlation function is defined as

$$c(\tau) = \frac{\langle h(0)h(\tau)\rangle}{\langle h\rangle} \qquad (4.5)$$

where, $h(\tau)$ is a hydrogen bond indicator function. $h(\tau)$ assumes a value of unity if a tagged hydrogen bond at τ= 0 also exists at time τ. Otherwise, $h(\tau)$ = 0. Therefore, $c(\tau)$ describes the probability that tagged hydrogen bond at τ = 0 is also bonded at a time τ, regardless of whether it breaks intermittently during the time interval τ. To account for the intermittent breaking of hydrogen bonds, a second correlation function is defined,

$$n(\tau) = \frac{\langle h(0)[1-h(\tau)]H(\tau)\rangle}{\langle h\rangle} \qquad (4.6)$$

In this correlation function, $H(\tau)$ takes up a value of unity if a tagged hydrogen bonded pair at τ = 0 is not hydrogen-bonded at time τ, but the donor and acceptor atoms of the tagged pair are within some predefined distance. Otherwise, $H(\tau)$ = 0. Essentially, $n(\tau)$ describes the probability that the donor and acceptor atoms of tagged hydrogen bond at τ = 0 are not hydrogen bonded at τ, but are within some distance that allows the pair to re-engage in hydrogen bonding. In the original Luzar-Chandler model,(366-368) this cut-off distance was chosen as 3.5 Å, which corresponds roughly to the first minimum in the oxygen-oxygen radial distribution function of bulk water (369, 370). Consequently, this choice of cut-off distance implies that donor-acceptor pairs that are within the first coordination shell are considered to have the ability to re-engage in

hydrogen bonding. Associating $c(\tau)$ and $n(\tau)$ with populations of unbroken and broken hydrogen bonds at time τ, respectively, the kinetics of hydrogen bond formation can be modeled as

$$\frac{dc(\tau)}{dt} = -kc(\tau) + k'n(\tau) \qquad (4.7)$$

where k and k' are rate constants. The inverse of the rate constant k is considered to be the average hydrogen bond life-time, that is, $\tau_{HB} = 1/k$. The time derivative of $c(\tau)$ is essentially a reactive flux correlation function,(366, 367)

$$\frac{dc(\tau)}{dt} = \frac{\langle \dot{h}(0)[1-h(\tau)] \rangle}{\langle h \rangle} \qquad (4.8)$$

where $\dot{h}(0) = dh/d\tau|_{\tau=0}$.

To determine the hydrogen bond life times of interstitial water molecules from Equation 4.7, we first determine the correlation functions $c(\tau)$, $n(\tau)$ and $dc(\tau)/dt$, and then use least square fitting to solve for the rate constants k and k' (371). A geometric definition of hydrogen bond used in the Luzar-Chandler model is adopted in this study. Accordingly to their definition, a pair of donor (D) and acceptor (A) atoms, with at least one of them being an interstitial water oxygen, are considered hydrogen bonded if they are separated by less than 3.5 Å, and simultaneously the A-D-H angle is less than 30º. The angle of 30º corresponds roughly to the amplitude of librations that break hydrogen bonds, as estimated from Debye-Waller factors. Various other definitions for hydrogen bonds have been proposed in the literature, based on energetics or electronic structure, and they yield similar results for hydrogen bonding kinetics in bulk water (372). In our calculations, we consider all the nitrogen and oxygen atoms within the proteins to be potential acceptors and the subset of acceptors that are bonded to the hydrogen atoms

as donors. Also, we do not include the weak hydrogen bonds involving protein carbon atoms (373-376) in the analysis.

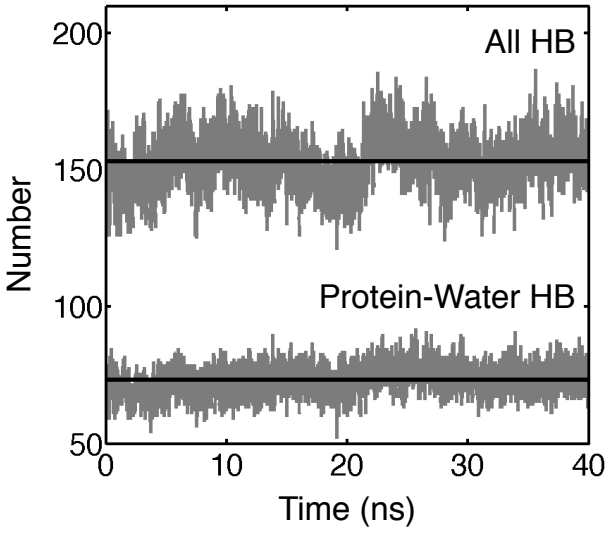


Figure 4.6 Time evolution of the number of hydrogen bonds (HB) made by interstitial water molecules in the G-B2 complex. The solid lines overlaying the number evolutions correspond to average values. The total number of hydrogen bonds is a sum of the numbers of protein-water and water-water hydrogen bonds (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

Our rationale for not considering weak hydrogen bonds is that their thermodynamic stability emerges primarily from induced effects, which are not modeled explicitly in non-polarizable force field simulations (241, 375, 377).

Figure 4.6 shows the time evolution of the number of hydrogen bonds made by interstitial water molecules in the G-B2 complex. As mentioned above, there are on an average 65 water molecules present in the interstitial region of the G-B2 complex. The interstitial water molecules make about 153 ± 10 distinct hydrogen bonds, out of which 48% involve protein functional groups.

An analysis of the protein-water hydrogen bonds shows that in the majority of cases water serves as a donor. It can be seen that on an average, each interstitial water molecule is engaged in forming 3.3 hydrogen bonds, which is comparable to the

number of hydrogen bonds made by a water molecule in the bulk phase. Separate simulations of bulk water indicate that a water molecule makes on an average 3.4 hydrogen bonds in the bulk phase, which is consistent with previous studies of the TIP4P water model (242). In addition, about 10% of the interstitial water molecules (6 ± 2) form hydrogen bonds simultaneously with both proteins in the complex, essentially bridging the interaction between them. A visual inspection of these bridging water molecules indicates that they are not clustered at any specific site of the protein-protein interface. The interstitial water molecules in the G-B3 complex exhibit similar overall statistics, and make 165 ± 10 distinct hydrogen bonds, out of which about 50% are with protein functional groups. In this case, 10 ± 3 water molecules serve to bridge the two proteins.

Figure 4.7 shows the autocorrelation functions of hydrogen bonds made by interstitial water molecules in the G-B2 complex. We have calculated correlation functions for all hydrogen bonds made by the interstitial water molecules and also separately for the water-water and water-protein hydrogen bonds. These correlations are estimated using observation time intervals of 2 fs, which is necessary for computing $dc(\tau)/dt$ (368). The estimated values of $dc(\tau)/dt$ obtained from the numerical derivative of $c(\tau)$ do not match those obtained from Equation 3.8 for observation time intervals greater than 2 fs. The correlation profiles for interstitial water molecules in the G-B3 complex are similar to those in the G-B2 complex.

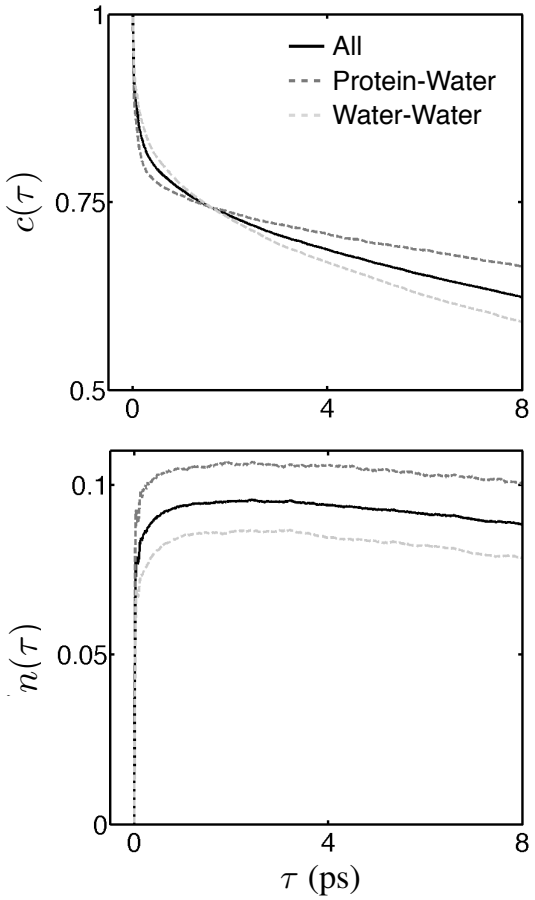


Figure 4.7 Autocorrelation functions of hydrogen bond populations, $c(\tau)$ and $n(\tau)$, of the interstitial water molecules in the G-B2 complex (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

The rate constants obtained from these correlation functions are listed in Table 4.1. The data in Table 4.1 also examines the effect of external boundary conditions on hydrogen bond dynamics. We find that for bulk water, canonical boundary conditions produce a hydrogen bond lifetime similar to that obtained from a Gibbs ensemble.

In addition, choosing a relaxation time constant for the Nose–Hoover chain that weakens temperature coupling by a factor of 5 has a negligible effect on τ_{HB}. Simulating bulk water under microcanonical boundary conditions, however, produces a τ_{HB} longer than that obtained from a canonical ensemble. Nevertheless, this difference is not related *per se* to the altered thermodynamic boundary conditions.

Table 4.1 Hydrogen bond lifetimes of water molecules τ_{HB} in the bulk phase and in the interstitial regions of the G-B2 and G-B3 complexes. τ_T is the relaxation time constant of the Nose – Hoover chain (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

System		Ensemble	k	k'	τ_{HB} (ps)
Bulk	Experiment				2.6
		NPT	0.56	1.17	1.79
		NVT$_{(\tau_{HB}=1)}$	0.55	1.11	1.82
		NVT$_{(\tau_{HB}=5)}$	0.56	1.17	1.79
		*NVT**$_{(\tau_{HB}=1)}$	0.33	1.01	3.03
		*NVE**$_{(T=313\ \pm 5K)}$	0.32	0.95	3.13
G–B2	all	*NPT*	0.23	1.44	4.44
	Wat–Wat		0.24	1.58	4.22
	Wat–Pro		0.18	1.10	5.53
	all	*NVE**$_{(T=310\ \pm 1K)}$	0.12	0.87	8.13
	Wat–Wat		0.14	0.92	7.35
	Wat–Pro		0.09	0.67	10.99
G–B3	All	*NPT*	0.28	1.71	3.55
	Wat–Wat		0.34	2.01	2.91
	Wat–Pro		0.16	0.96	6.21

It arises primarily from the changes introduced (shown by asterisk in Table 4.1) into the Hamiltonian and integration algorithm for obtaining energy conservation and temperature stability in the microcanonical ensembles (230). In the case of the two protein – protein complexes, the life times are estimated for all the hydrogen bonds made by the interstitial water molecules and also separately for the hydrogen bonds made between the interstitial water molecules (Wat – Wat) and between the interstitial water molecules and the protein (Wat – Pro). The experimental data shown in Table 3.1 is taken from elsewhere (378).

The following changes were introduced – (i) the restraints on all bonds, including those of water molecules, are released and, consequently, a smaller integration time step of 0.5 fs is used; (ii) while long range electrostatics are computed using the particle

mesh Ewald scheme, short range electrostatic interactions are truncated using a switch function in which they are reduced linearly between 9 and 10 Å; and finally, (iii) the van der Waals interactions are computed by re-normalizing them such that they decay to zero smoothly over the 10 Å of direct space. Implementing these changes into the canonical ensemble produces a τ_{HB} quantitatively comparable to that obtained from the microcanonical ensemble. We therefore attribute the differences in the interstitial water dynamics between the Gibbs and the microcanonical ensembles to the altered simulation protocol and not to the altered boundary conditions.

We find that irrespective of the external boundary conditions the hydrogen bond lifetimes (τ_{HB}) of interstitial water molecules are 2-3 times longer than those in the bulk. This increase in τ_{HB} is not entirely due to the longer lifetimes of hydrogen bonds involving protein side chains. The water-water hydrogen bond lifetimes in the G-B2 interface are two times longer than those in bulk water. Nevertheless, these shifts are within the range expected at protein-water interfaces (320, 321, 323, 325) and so these results suggest that the hydrogen bond dynamics at the protein-protein interfaces are statistically similar to those at the protein-water interfaces. Explicit polarization effects have not been accounted for in these calculations. However, it is known from literature that the inclusion of polarization, results in longer hydrogen bond lifetimes (242, 244, 245) perhaps arising from damped librations. In this regard, the hydrogen bond lifetimes computed for the interstitial water molecules in G-B2/B3 complexes can be considered to be on the lower limit.

4.2.5 Dipole correlations

At this point, our results suggest that the hydrogen bond lifetimes of interstitial water molecules are longer compared to those in the bulk water. So, the question that arises is, how do the overall rotational dynamics of the water molecules in the interstitial regions differ from those in the bulk water? To examine this we compute the dipole correlation function,

$$M_{\parallel}(\tau) = \frac{\langle \sum \mu_{\parallel}(0) \sum \mu_{\parallel}(\tau)/n_0 n_\tau \rangle}{\langle (\sum \mu_{\parallel}(0)/n_0)^2 \rangle} \qquad (4.9)$$

where $\mu_{\parallel}$ is the component of the water dipole moment parallel to the axis joining the geometric centers of G and ephrins. n_0 and n_τ refer to the number of interstitial water molecules in frames separated by a time interval τ. Under the isotropic conditions maintained in bulk water, for all $n_\tau = n_0$, $M_{\parallel}(\tau) = M(\tau)$. Figure 4.8a shows $M_{\parallel}(\tau)$ estimated for interstitial water molecules in the G-B2 and G-B3 complexes. Due to the observational time interval being 0.25 ps, this analysis does not capture the sub-picosecond (ultrafast) rotational relaxation of water molecules (379). For both the complexes, the correlation function decays asymptotically to a nonzero value, which indicates an orientational preference for the interstitial water molecules. This is confirmed by evaluating the radial distribution of water dipoles about the axes joining the geometric centers of the G and the ephrins (Figure 4.8b). We attribute this orientational preference to the asymmetric distribution of charged residues at the G-ephrin interface. Additionally, we observe that there are no Na^+ or Cl^- ions present in these interstitial regions, as determined from their respective radial distribution functions (shown in Figure 4.9). Hence, we can safely eliminate any influence on $\mu_{\parallel}$ arising from the presence of ions at the protein-protein interfaces.

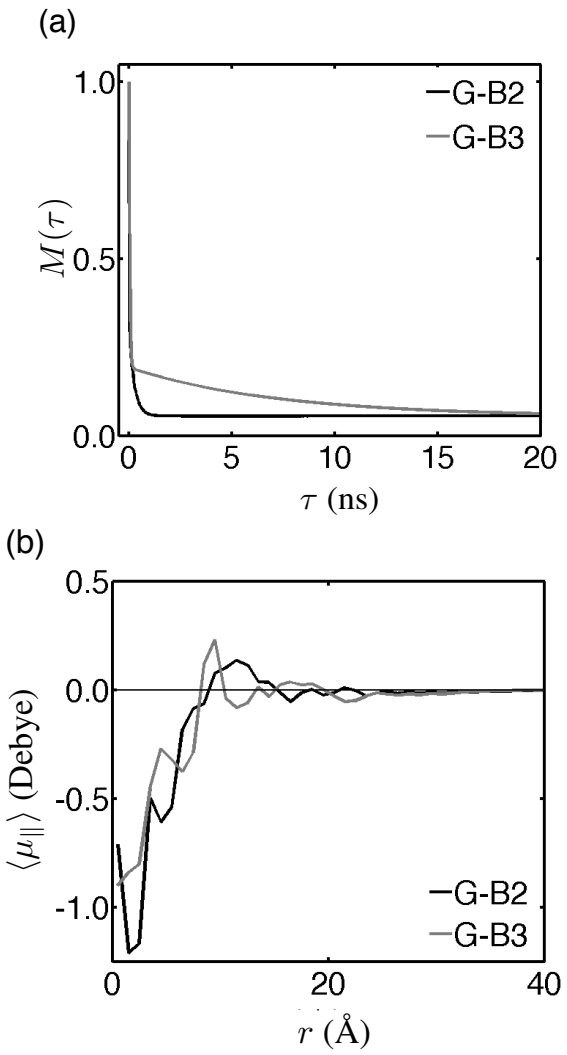


Figure 4.8 (a) Dipole correlations, $M_{\|}(\tau)$, of the interstitial water molecules in the G-B2 and G-B3 complexes. **(b)** Distribution of water dipoles, $\mu_{\|}$, estimated as a function of the radial distance, r, from the axis joining the geometric centers of G and B2/B3 (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

Modeling the dipole correlation function as a sum of two exponential functions,

$$M_{\|}(\tau) = (1 - B)\left[Ae^{-\tau/\tau_1} + (1 - A)e^{-\tau/\tau_2}\right] + B \qquad (4.10)$$

reveals two distinct subpopulations, *A* and (1-*A*), corresponding to the fast and slow rotating water molecules. The constant *B* in Equation 4.10 accounts for the limiting non-zero value of M($\tau \rightarrow \infty$) and refers to the fraction of water molecules that exhibit a permanent orientational preference along the axes joining the geometric centres of the

G and the ephrins. A weighted least square fitting was done to determine the values of *A*, *B*, τ_1 and τ_2. In the interstitial region of the G-B2 complex, about 5.6% of the water molecules exhibit a permanent orientational preference along the geometric centres of G and B2. The results indicate that 69% of the water molecules have re-orientational times of τ_1 = 6.0 ps, about 25% have longer re-orientational times of τ_2 = 248.8 ps. While the fast relaxation time is of the same order as that of the relaxation time in bulk phase, more than a quarter of interstitial water molecules relax at rates 2 orders in magnitude slower than bulk water, which is reminiscent of an anisotropic polar environment.

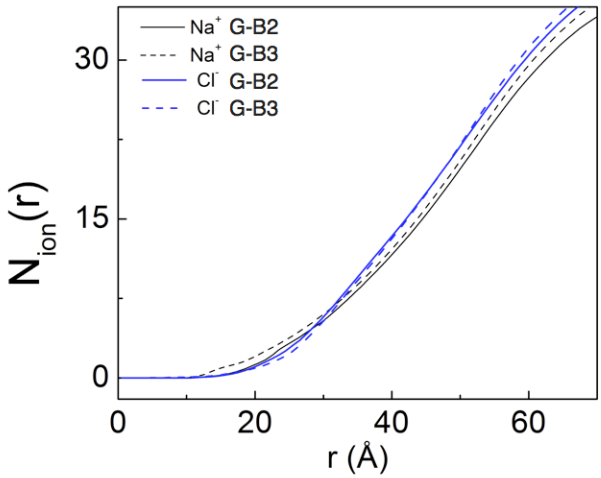


Figure 4.9 The number of ions estimated as a function of the radial distance, r, from the axis joining the geometric centers of G and B2/B3.

An alternative interpretation of this result is that the set of interstitial water molecules have two different relaxations timescales, one corresponding to that of bulk water, and another that is a hundred-fold slower. In the interstitial region of the G-B3 complex, the rotational relaxation of water is much slower, with 81% of the water molecules relaxing with a τ_1 = 38 ps, only 14% relaxing with a τ_2 = 7317 ps. A similar residual fraction of water molecules permanently orient between the sequences of ephrins B2 and B3. The

difference in relaxation rates of interstitial water molecules in the G-B2 and G-B3 complexes emerges primarily from differences between the sequences of ephrins B2 and B3. This is because the G-B2 and G-B3 complexes are structurally similar (RMSD < 2 Å). In general, these results show that chemistry can affect dipole relaxation rates significantly, and highlight that the dielectric response of interstitial solvent is unlike that of bulk water.

4.3 Explicit solvent vs. implicit solvent simulations

Consistent with the X-ray structure of the G-B2 complex,(90) our explicit solvent MD simulations indicate that the G-ephrin complexes accommodate an exceptionally high number of water molecules in their interstitial regions. Additionally, while the interstitial water molecules tend to occupy crystallographic sites, most interstitial water molecules exchange with the bulk solvent every hundred picoseconds. But what specific physiological role do these interstitial water molecules serve?

The binding of ephrin B2/B3 to G causes G to activate another viral protein, F (66, 68, 79, 90). Upon activation, F mediates virus-host membrane fusion. Since ephrins and the F protein bind to G at mutually exclusive sites, the effect of ephrin binding must transduce to the F-binding site of G to activate F (66, 68, 79, 90, 94). This F-activating allosteric signal, is contained within the changes in the conformational density of G brought about by ephrin binding (94). If $\mathbb{R}_{apo}$ and $\mathbb{R}_{bound}$ represent, respectively, the conformational densities sampled by the G protein in its apo and ephrin-bound states, then the F-activating signal is contained within the ephrin-induced conformational density shift $\Delta\mathbb{R} := \mathbb{R}_{apo} \rightarrow \mathbb{R}_{bound}$. Now if the water molecules at the G-ephrin interface were to contribute to the allosteric activation of F, then they must contribute to $\Delta\mathbb{R}$.

The difference between two three dimensional (3D) conformational ensembles was quantified using a method based on support vector machines (93, 94) as described in section 3.3. For the analyses shown in this chapter, the optimized values for the Lagrange multiplier (*C*) and the width of the transformational Kernel (γ) chosen were 10^2 and 10^{-1} respectively. The mean absolute error between the computed and analytical values of the discriminability index (η) was found to be 3.2%.

To examine whether the interstitial water molecules contribute to $\Delta\mathbb{R}$, we determine $\Delta\mathbb{R}$ from explicit solvent simulations and compare them quantitatively to the $\Delta\mathbb{R}$ obtained from a separate set of implicit solvent simulations. We expect explicit and implicit solvent models to yield different $\Delta\mathbb{R}$ because (a) the two models describe bulk solvent differently, and (b) implicit solvent models do not account for the discrete nature of water molecules at the G-ephrin interface. Given the high numbers of water molecules at the G-ephrin interface, the lack of their specific volumes in the implicit solvent model can alter the G-ephrin interface. In addition, since about 10% of the interstitial water molecules hydrogen bond simultaneously with both proteins, bridging the interaction between the two proteins, their absence could directly alter the G-B2 binding modes. The absence of discrete water molecules at the G-B2 interface could, therefore, lead to altered $\Delta\mathbb{R}$.

Figure 4.10 shows two primary differences in the G-B2 interface predicted from explicit and implicit solvent simulations. Compared to the explicit solvent model, (a) the implicit solvent model predicts a smaller distance between G and B2, which presumably results from the absence of discrete water molecules at the interface, and (b) the implicit solvent model predicts a smaller width of the G cavity into which the GH loop of B2

inserts, which can also be attributed to the absence of discrete water molecules at the interface. These two structural differences signify that in the absence of explicit water molecules in the implicit solvent model, the G-B2 interface is more compact.

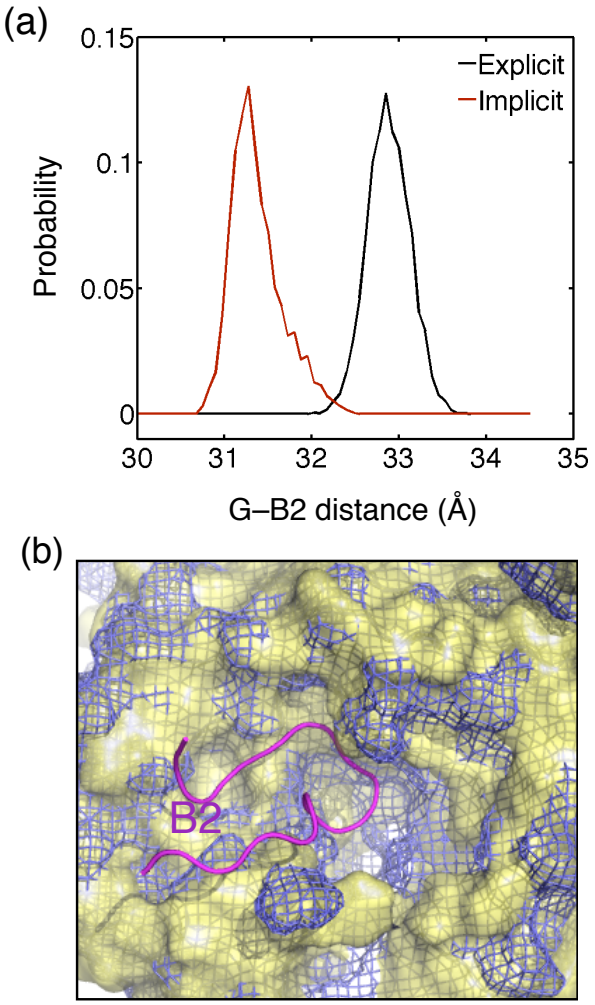


Figure 4.10 Effect of the treatment of the solvent on the structural properties of the G-B2 interface **(a)** Probability distribution of distances between the geometric centers of G and B2 proteins obtained from implicit and explicit solvent simulations of the G-B2 complex. **(b)** Partial view of the G protein showing its B2 binding site. The structure of the G protein taken from the explicit solvent simulation of the G-B2 complex is rendered as a yellow van der Waals surface, and the structure of the G protein taken from the implicit solvent simulation of the G-B2 complex is rendered as a blue van der Waals mesh. The GH loop of B2 that inserts into the G cavity is shown as a magenta ribbon (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

Figure 4.11a compares the $\Delta\mathbb{R}$ obtained from explicit and implicit solvent simulations. The $\Delta\mathbb{R}$ are estimated separately for each amino acid of G in terms of a

quantity that we refer to as discriminability (η) (93, 94). This quantity is normalized and bounded, that is, $\eta \in [0,1)$ and it takes up a value closer to unity as the difference between the conformational densities increases.

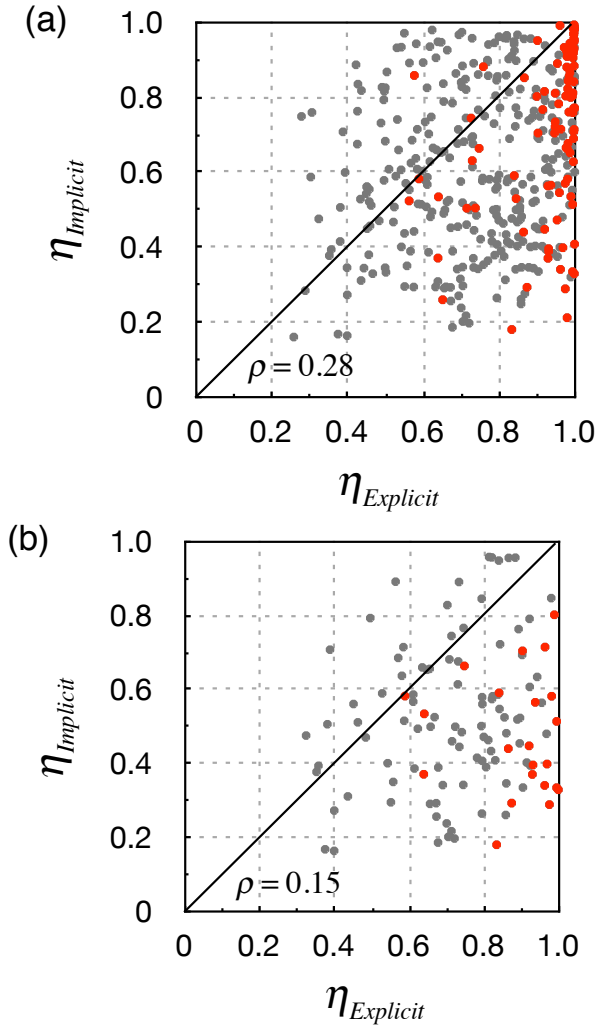


Figure 4.11 Comparison of B2-induced conformational density shifts in G, as estimated from implicit and explicit solvent simulations. **(a)** The 416 dots represent the conformational density shifts for the 416 residues in G. The dots colored red correspond to residues that are part of the allosteric signalling pathway. **(b)** The 114 dots represent the conformational density shifts of a subset of the residues of G that satisfy the condition given by Equation 4.11 (Reprinted with permission from (95). Copyright American Chemical Soceity 2014).

We determine separately for each residue in the G head domain between its representative ensembles in the apo and the B2 bound states. Since we are simulating

416-residue segment of the G head domain, a comparison between the two G ensembles yields 416 η values, one η for each residue. Each of the two ensembles is represented by 2001 conformations extracted at regular intervals from their respective simulations. Doubling or reducing the ensemble size by a factor of 2 affects the quantification minimally (93).

Prior to extracting the ensemble of a residue from the G head domain, all the simulated configurations of G head domain are least square fitted on to the X-ray coordinates of the G head domain. Structure fitting is necessary to remove the bias of η against whole molecule rotation and translation, as that is not the goal of this comparison. We expect the least-square algorithm to be adequate for structure fitting because the structural differences between the apo and bound states are small (380, 381). Also, during ensemble comparison we consider only the heavy atoms. The $\Delta\mathbb{R}$ values estimated from the explicit and implicit solvent simulations are statistically different with a Pearson correlation of 0.28 (Figure 4.11a). This difference is even more pronounced for residues that are known to participate in allosteric signaling (94). This difference reflects the overall effect of treating the solvent using a mean field approximation.

Although the absence of discrete interstitial water molecules in the implicit solvent simulation contribute to this difference, this analysis does not delineate their specific role. To gain further insight into the specific contribution of the interstitial water molecules to $\Delta\mathbb{R}$, further analysis was done. We identify the subset of residues in the G protein whose conformational densities in the apo state are unaffected by the treatment of the bulk solvent. This is done by estimating the difference between the

conformational density values obtained from both the explicit and implicit solvent simulations for each residue in the apo state of the G protein. This difference is denoted as $\eta_{imp\leftrightarrow exp}$. This is followed by filtering out the residues whose $\eta_{imp\leftrightarrow exp}$ are smaller than a certain tolerance. We choose $d^2 = BT/8\pi^2 T_{xray}$ as the tolerance, which is the mean square deviation of a residue obtained from crystallographic B factors (86). The ratio T/T_{xray} rescales the B factors from X-ray diffraction temperature T_{xray} = 100 K to physiological temperature T = 310 K (380, 381). Consequently, if a given residue meets the condition in Equation 4.11,

$$\eta_{imp\leftrightarrow exp} < \mathrm{erf}\,(d/\sqrt{2}) \qquad (4.11)$$

then the difference between its conformational density obtained from the explicit and implicit solvent simulations is smaller than the spread in the residue's electron density observed in the X-ray diffraction data. Note that the error function in Equation 4.11 represents the transformation of the tolerance to the appropriate Hilbert space where η is estimated (94). Out of 416 about 114 residues of G meet this criterion, and even for these residues the explicit and implicit solvent simulations produce statistically different B2-induced shifts (with a Pearson correlation of 0.15) in the conformational density (Figure 4.11b). Since the conformational densities of these residues in the apo state are not affected by the treatment of solvent, this difference reflects the specific effect of treating the G-B2 interaction using a mean field approximation.

Taken together with the results from Figure 4.11, this analysis suggests that the differences observed in $\Delta\mathbb{R}$ estimated from the implicit and explicit solvent models are partially due to the absence of explicit interstitial water molecules in the implicit solvent model, thereby supporting the hypothesis that the interstitial water molecules also

contribute to $\Delta\mathbb{R}$. This finding, however, does not generalize the idea that implicit solvent models should not be used for investigating protein-protein complexes. In fact, there are numerous examples in literature where protein-protein interactions have been modeled successfully using implicit solvent models (115, 116, 249, 382). The G-ephrin complex is unique in the sense that it sandwiches an exceptionally large amount of water at its interface, and our results indicate that the treatment of explicit solvent becomes critical for such cases.

4.4 Summary

In this chapter, we study systems where water molecules occupy interstitial regions between two proteins. Specifically, we consider two natural protein-protein complexes, both of which are formed independently during the fusion of Nipah viruses with host cells. In one complex, the Nipah virus G protein binds to cellular the ephrin B2, and in the other the same G protein binds to the ephrin B3. While the two complexes are structurally similar, the two ephrins share only a modest sequence identity of ~50%, even in the portions that form the interface.

The atomistic MD simulations reveal that while the interstitial water molecules tend to occupy crystallographic sites, most water molecules exhibit residence times of less than 100 ps in the interstitial region. Therefore, we argue that the crystallographic sites for water should not be viewed as sites for “bound” water molecules, but rather as preferred sites for water occupancy. The dynamical properties of the interstitial water molecules in the two complexes are quantitatively different, the trend in their shifts with respect to bulk values are similar. Since the two ephrins are topologically similar, the quantitative differences in water dynamics emerge primarily from the differences in the

sequences of the two ephrins. The effect of chemical difference is seen predominantly in dipole relaxation rates, and not as much in diffusion rates or residence times or hydrogen-bond lifetimes.

In addition, despite the exceptional wetness of the protein-protein interfaces, the dynamics of interstitial water molecules are considerably slower compared to the bulk. In the G complexes the interstitial water molecules diffuse at rates 10 times slower compared to bulk water. Additionally, the interstitial water molecules exhibit hydrogen bond lifetimes 2-3 times longer than bulk water. This increase is not entirely due to the presence of protein-water molecules hydrogen bonds. The water-water hydrogen bond lifetimes also increase in the interstitial regions, indicating that the rattling events that break hydrogen bonds are slower. Nevertheless, these shifts in hydrogen bond lifetimes are within the range expected at protein-water interfaces, and so these results suggest that the hydrogen bond dynamics at the protein-protein interfaces are statistically similar to those at the protein-water interfaces. Our results further indicate that the majority of interstitial water molecules exhibit dipole relaxation times similar to those in the bulk, however, there is an appreciable fraction whose relaxation times are 100-1000 fold longer than bulk water.

To gauge the functional relevance of the interstitial water molecules, we have performed a quantitative study on how the implicit solvent models compare against the explicit solvent models in producing ephrin-induced shifts in the G conformational density. The ephrin-induced shifts in the G conformational density are critical to the allosteric activation of the viral fusion protein, F. The implicit solvent model predicts a more compact G-B2 interface compared to the explicit solvent model, with G and B2

being physically closer to each other presumably because of the absence of the discrete interstitial water molecules at the G-B2 interface. The two models yield strikingly different induced changes in the G conformational density, even for those amino acids whose conformational densities in the apo state are unaffected by the treatment of the bulk solvent. Together these results suggest that the interstitial water molecules contribute to the allosteric activation of F, and therefore, are functionally important for a proper description of allosteric transitions. Hence, from here on, we always use explicit solvent while performing MD simulations.

CHAPTER 5

HOST RECEPTOR INDUCED CHANGES IN RECEPTOR BINDING DOMAIN

5.1 Background

Experiments show that the G protein interacts with the ephrin and F through separate sites located on two different domains, the RBD and the FAD. No model explaining this allosteric coupling has been proposed yet. In fact, the analogous mechanisms in other paramyxoviruses also remain undetermined. The structural organization of G is such that allosteric coupling must involve at least one of the two interfaces – the RBD-FAD interface and/or the RBD-RBD interface shown in Figure 1.2. Here we study the RBD-RBD dimer of the Nipah G protein in its ephrin free and ephrin bound states by performing molecular dynamics simulations.

The primary goal in this chapter is to understand the specific effects of ephrin on the RBD-RBD interface, which remains unknown, but has been implicated to play a vital role in the allosteric stimulation of G (93, 94, 147). Additionally, we have carried out MD simulations of the RBD-RBD dimer in the event of a triple mutation, $V_{209}V_{210}G_{211}\rightarrow AAA$. The residues VVG are part of the RBD, and distant from both the ephrin-RDB and the RBD-FAD interfaces. It is known that their mutation to the alanines affects neither the expression of G nor its binding to ephrin (77). Yet, the triple mutation abolishes the ability of G to activate F. If we find that the stimulation-deficient mutant does not modify

the response of ephrin-binding on the RBD-RBD rearrangement, then we will conclude that the ephrin-induced rearrangement in the RBD-RBD interface is not a sufficient condition for the G stimulation. Together, this MD study will inform us of the effect of ephrin binding on the RBD-RBD interface and will also provide insight into the role of RBD-RBD interfacial rearrangements in triggering G stimulation.

The details of the MD parameters are provided in Chapter 3. The systems discussed in this chapter, the ephrin-free state comprises of 356,770 particles, and the ephrin-bound state comprises of 435,254 particles. We also estimate the free energy of the RBD-RBD interface in both its ephrin-free and ephrin- bound states by carrying out accelerated conformational sampling (252, 383, 384).

5.2 Wild type dimer

5.2.1 Construction of model

Although there are no experimentally established structures of the RBD-RBD dimer of NiV G, we were able to construct the initial dimer model for carrying out the MD simulations. This was done by incorporating and integrating the following information from multiple experimental reports. Firstly, the X-ray structures for the isolated NiV RBD as well as its complex with ephrin (86, 90) were identified. Secondly, both the ephrin free and ephrin bound structures of the NiV RBD, which have been subjected to MD at physiological temperature, and have been found to be stable (93, 94). Thirdly, Bowden *et* al. (80) had proposed a RBD-RBD interface for the G protein of the HeV (PDB ID: 2X9M), which served as a suitable template to construct the initial model of the RBD-RBD interface of Nipah G.

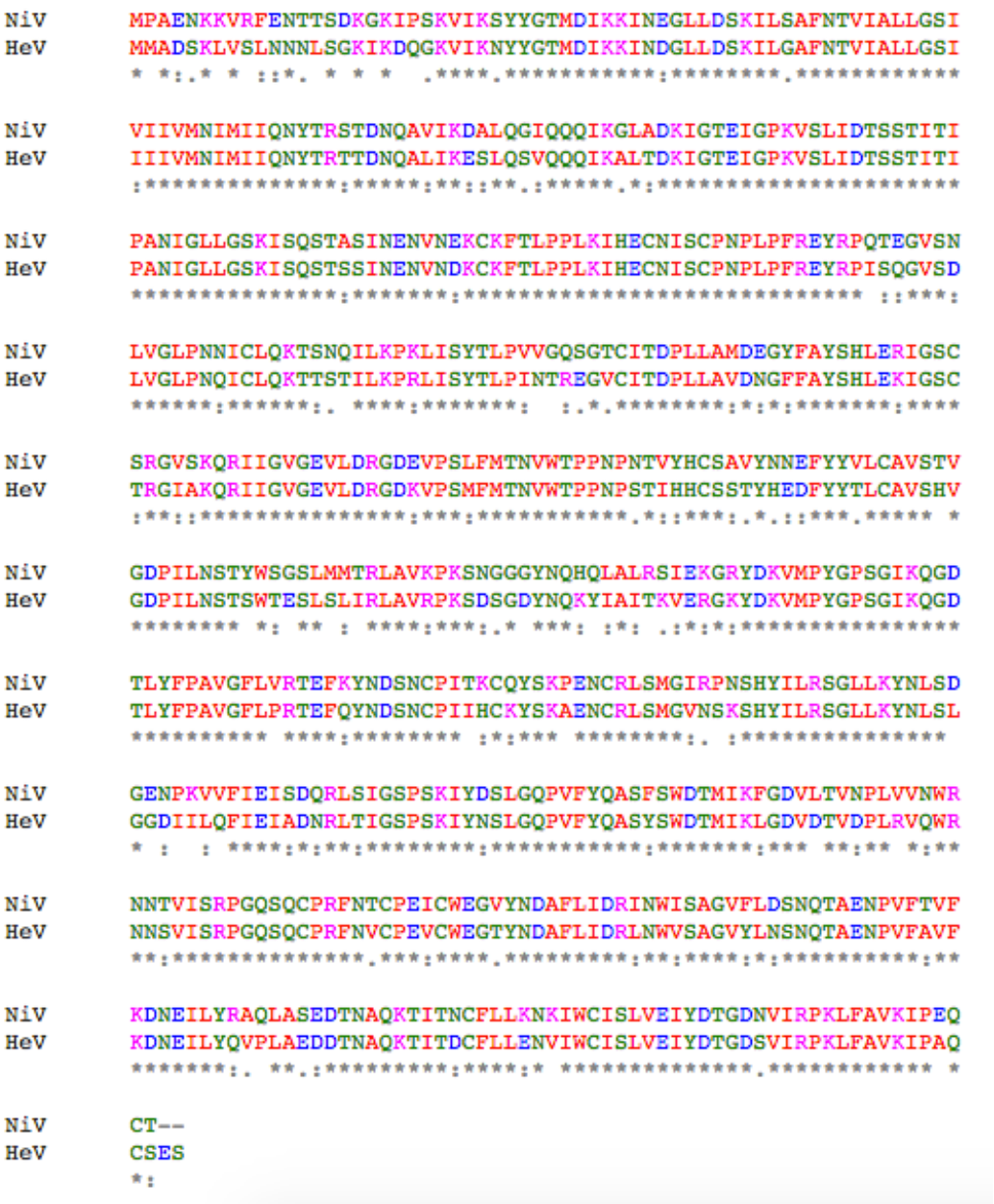


Figure 5.1 Aligned sequences of the G proteins of NiV (UniProt Q9IH62) and HeV (UniProt O89343) viruses. Sequence identity is 78.5 % and sequence similarity of 89.2 %.

This approach was implemented because (a) the G protein of HeV is a closely related homolog of the NiV G protein (89% sequence similarity, Figure 5.1), and (b) X-ray structures of the ephrin-free and the ephrin-bound states of the HeV virus RBD closely match the respective X-ray structures of the NiV RBD (Figure 5.2) (80, 86, 90). The sequence alignment shown in Figure 5.1 was done using Clustal Omega (272).

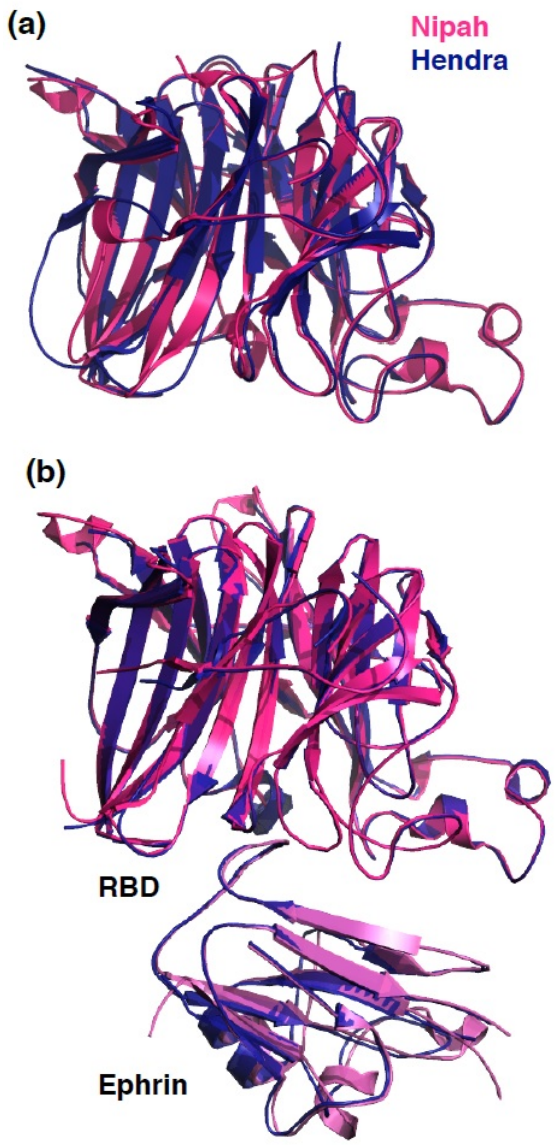


Figure 5.2 Superimposed X-ray structures of NiV and HeV RBDs in their **(a)** Apo states, and **(b)** ephrin bound states.

Firstly, the X-ray structures for the isolated RBD as well as its complex with ephrin (86, 90) were identified. Secondly, both the ephrin free and the ephrin bound structures of the RBD have been found to be stable (93, 94). Thirdly, Bowden *et* al. (80) had proposed a RBD-RBD interface for the G protein of the HeV (PDB ID: 2X9M) which served as a suitable template to construct the initial model of the RBD-RBD interface of the G protein of NiV. This approach was implemented because (a) the G protein of HeV

is a closely related homolog of the NiV G protein (89% sequence similarity, Figure 5.1), and (b) X-ray structures of the ephrin free and the ephrin bound states of the HeV RBD closely match the respective X-ray structures of the NiV RBD (Figure 4.2) (80, 86, 90). The sequence alignment shown in Figure 5.1 was done using Clustal Omega (272).

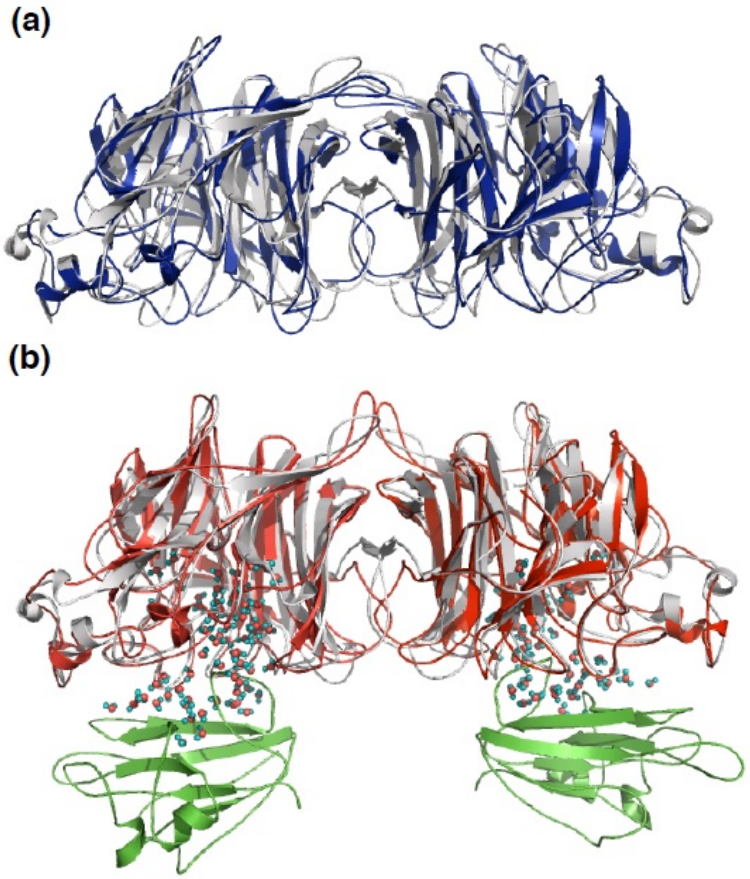


Figure 5.3 Initial models of the RBD-RBD interface in the **(a)** ephrin free state, and **(b)** ephrin bound state. The structure in grey is the RBD-RBD dimer (PDB ID: 2XM9) of HeV, which is used for templating the initial models of the NiV RBD-RBD dimer.

The RBD-RBD interface of HeV G protein was proposed (80) by consolidating data concerning the (a) packing interactions within the crystals, (b) conservation patterns within the RBD-RBD interfaces of analogous receptor binding proteins of other paramyxoviruses, and (c) the distribution of the N-linked glycosylation sites on the RBD. In particular, the distribution of the glycosylation sites on the RBDs of NiV and HeV are

such that they permit only one specific face of the RBD to dimerize with an adjacent RBD – the remaining faces of the RBDs contain protruding glycosyl chains that result in steric clashes. Therefore, there remains no ambiguity concerning the dimerization face of the RBD, although Bowden *et* al. (50) had pointed out that the relative orientation between the two RBDs in the ephrin free and bound states couldn't be conclusively assured. Nevertheless, the NiV RBD-RBD model constructed using the HeV template serve as an excellent starting point for MD simulations, which we utilize to determine the relative orientations between the RBDs.

In order to construct the initial model of the RBD-RBD interface in the ephrin free state, we took the monomeric form of Nipah's RBD (94) from our earlier simulations (Chapter 3). The thermalized conformation of the RBD (640 ns snapshot) was taken and two of its copies were geometrically fitted individually onto the two RBDs of HeV's RBD-RBD dimer. The two geometric fits yielded identical least squared fit values as expected because the RBD-RBD interface is known to be symmetric. The fits showed a very low RMSD of < 2 Å thereby substantiating our supposition. The templated model is shown in Figure 5.3.

We used the same protocol to construct the initial model of the RBD-RBD dimer in the ephrin bound state, but in this case we took a thermalized conformation at 460 ns from our simulation of the NiV's ephrin-bound RBD monomer (94). Even in this case, the geometric fits were excellent (RMSD < 2 Å). The reason that the structures of both the ephrin free and ephrin bound RBDs fit excellently on to the RBD of the HeV (80) is because, (Figure 5.2) the difference between the ephrin free and ephrin bound structures of the RBD is small (86, 90, 93, 94), as discussed in section 1.2. Note that

after fitting the RBD of the ephrin-RBD complex to the RBD of the HeV RBD-RBD template, we applied the resulting rotational matrix to ephrin. The water molecules sandwiched between ephrin and the RBD were retained and subjected to the rotational matrix. As mentioned in the previous chapter, these interstitial water molecules are critical not only to the structural integrity of the RBD-ephrin interface, but also to the inception of the ephrin binding signal at the RBD-ephrin interface (95). The two constructed RBD-RBD dimers were then energy-minimized, solvated separately in salt solutions and then subjected to MD. The ephrin free state and the ephrin bound state comprised of 356,770 and 435,254 particles respectively.

5.2.2 Effect of receptor binding on dimer

We examine how the small changes induced by ephrin in individual RBDs affect the interface between two RBDs. Since a single RBD-RBD template was used for constructing the initial models of both the ephrin free and ephrin bound dimers, the orientations between the two RBDs in these initial models are identical.

The two templated dimer models were subjected to separate MD simulations. Figure 5.4 shows the molecular definitions of the collective variables, d_{CoM}, θ_{tilt}, and θ_{roll}, used to describe the interface between the two RBDs. Figure 5.5.a tracks the time evolution of the three collective variables that describe the interface between the two RBDs in a dimer: d_{CoM}, θ_{tilt}, and θ_{roll}. The RBD-RBD interface of the ephrin bound state is strikingly different from that of the ephrin free state. Repeating these simulations by assigning different initial velocity distributions in the simulations yields the same result confirming that the RBD-RBD interface of the ephrin bound state is markedly different from that of the ephrin free state. Figure 5.5a further shows that the two

simulations conducted of the ephrin bound state yield identical RBD-RBD orientations; however, the two simulations of the ephrin free state yield slightly different RBD-RBD orientations. To understand the latter, we visualize the RBD-RBD interfaces obtained from these simulations with respect to the position of the FAD as shown in Figure 5.5b.

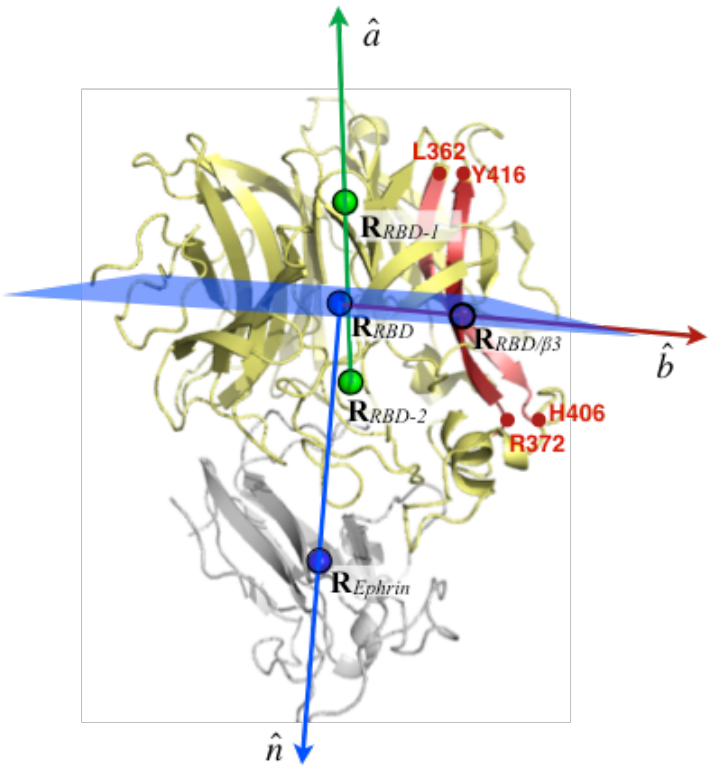


Figure 5.4 Illustration of the molecular definitions of collective variables. A RBD is drawn as a yellow cartoon and ephrin is drawn as a grey cartoon. d_{CoM} is the distance between the center of masses ($\mathbb{R}_{RBD}$) of the backbone atoms of two RBDs. θ_{tilt} is the angle between the central axes $\hat{\boldsymbol{a}}$ of the two RBDs. We construct $\hat{\boldsymbol{a}} = (\mathbb{R}_{RBD-1} - \mathbb{R}_{RBD-2})/|\mathbb{R}_{RBD-1}||\mathbb{R}_{RBD-2}|$ by defining two points $\mathbb{R}_{RBD-1}$and $\mathbb{R}_{RBD-2}$.These two points are the centers of masses of the backbone atoms of the two halves of RBD divided by the plane shown in blue. This plane is defined by the normal $\hat{\boldsymbol{n}} = (\mathbb{R}_{Ephrin} - \mathbb{R}_{RBD})/|\mathbb{R}_{Ephrin}||\mathbb{R}_{RBD}|$ and the point $\mathbb{R}_{RBD}$. θ_{roll} is the angle of rotation of the RBD about its central axis. Geometrically it is the angle between the axes $\hat{\boldsymbol{b}}$ of the two RBDs. This axis is defined as $\hat{\boldsymbol{b}} = (\mathbb{R}_{RBD/\beta3} - \mathbb{R}_{RBD})/|\mathbb{R}_{RBD/\beta3}||\mathbb{R}_{RBD}|$, where $\mathbb{R}_{RBD/\beta3}$is the center of mass of the backbone atoms of $\beta3$ blade of RBD. The $\beta3$ blade is highlighted in red and its terminal residues are indicated. It is chosen over the other β-blades because its structure and dynamics undergo the least change upon RBD-ephrin complexation (93, 94).

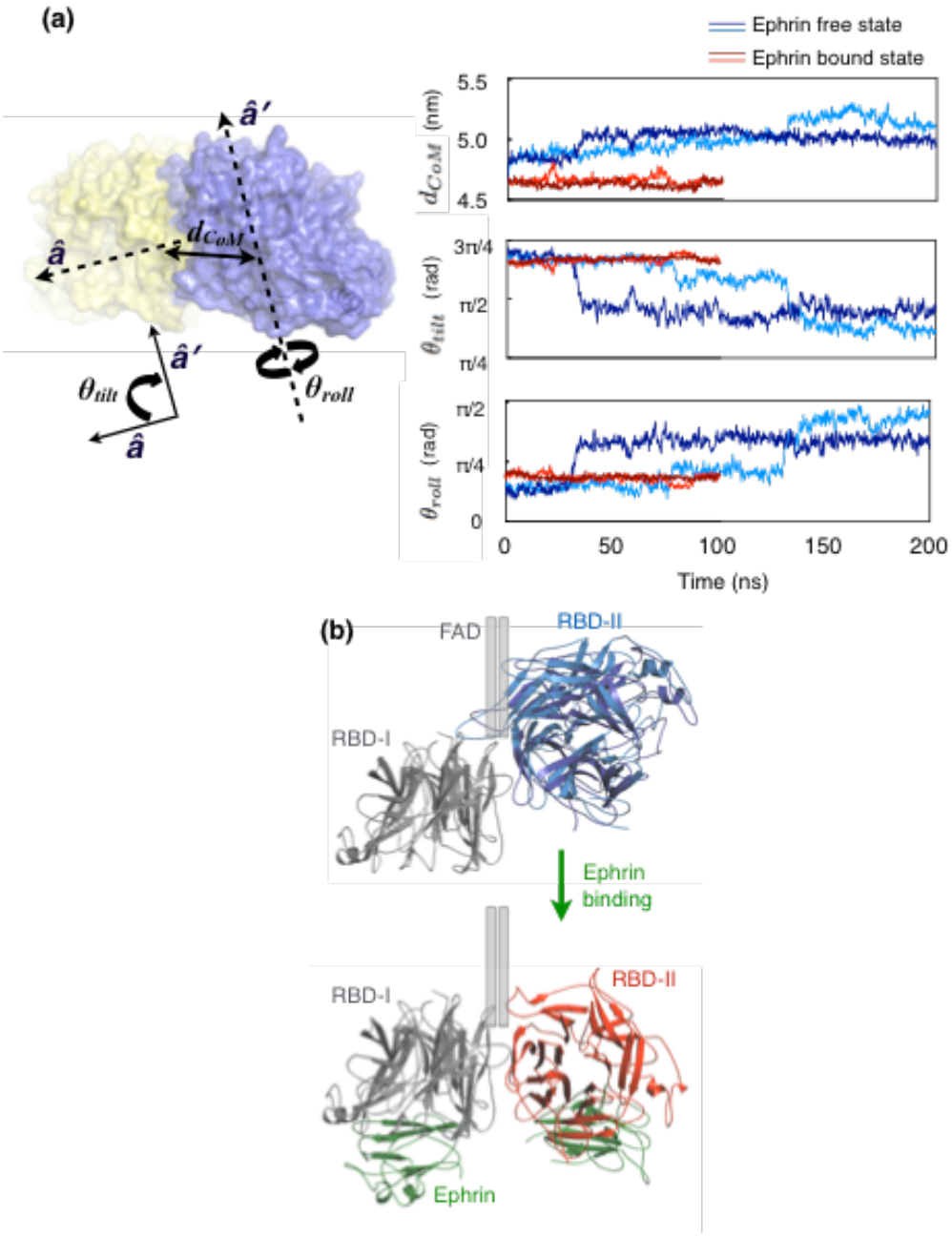


Figure 5.5 (a) Time evolutions of collective variables that describe the interface between the two RBDs of a dimer. The two lines for each of the ephrin free and ephrin bound states indicate two separate MD simulations. d_{CoM}, θ_{tilt}, and θ_{roll} are discussed in Figure 5.4. **(b)** Final snapshots of the RBD-RBD interface in MD simulations.

Note that two superimposed structures are shown for the ephrin free state, to highlight the slightly different RBD-RBD geometries. The location of the FAD relative to the RBD-RBD dimer is depicted according to the structure of the full length ectodomain

proposed by Broder and coworkers (65), which was homology modeled on the X-ray structures of the G analogs in the Newcastle Disease Virus and the parainfluenza virus (80, 87). We emphasize that the FAD tends to interact more extensively with the RBDs in the ephrin free state, as compared to the ephrin bound state. Therefore, the reason the two simulations of the ephrin free state produce slightly different RBD-RBD interfaces could be due to the absence of the RBD-FAD interface in our simulations. Based on the outcome of these simulations, we can conclude that the ephrin binding induces a significant change in the RBD-RBD orientation.

Time scales that can be simulated for classical MD are in the range of hundreds of nanoseconds, however *rare events* in biological systems are known to occur at order of magnitude larger (252, 253). MD simulations to attain such large timescales is computationally expensive, hence to observe the biological events one can utilize methodologies aimed at accelerating rare events using the available computer time with improved efficiency. Metadynamics is one of many techniques available that allows enhanced sampling in MD simulations and reconstructing the free-energy surface as a function of few selected degrees of freedom, the collective variables (CVs) (described in Figure 5.4).

We simulate the ephrin-free state and ephrin-bound state. We note that while free energy profiles obtained from a single metadynamics simulation converge theoretically under the long time scale limit (385-388), achieving convergence for the systems studied here is challenging from a practical standpoint, as they consist of hundreds of degrees of freedom that can potentially contribute to free energies, and our system sizes exceed 350K particles. For these systems we, therefore, generate five

separate (75 ns long) metadynamics trajectories for each state, and then use the resulting standard deviations as our estimate for convergence.

From the conformational sampling using metadynamics we find over hundreds of local minima on the averaged hypersurfaces $U(d_{CoM}, \theta_{tilt}, \theta_{roll})$ of both the ephrin free and the ephrin bound states. Nevertheless, grouping those that lie within each other's standard deviations leads to the identification of only seven statistically discernible clusters in the ephrin free state and eleven clusters in the ephrin bound state. We represent the clusters using their respective deepest energy wells, $\langle U\left(d_{CoM}^{min}, \theta_{tilt}^{min}, \theta_{roll}^{min}\right)\rangle$. Figure 4.7 shows the four deepest energy wells for each state, as well as the RBD-RBD orientations corresponding to them. $\langle U(X)\rangle|_{Y^{min},Z^{min}}$ is the free energy along the variable X, with variables $Y = Y^{min}$ and $Z = Z^{min}$.

Table 5.1 Collective variables of the four deepest statistically discernible minima in the receptor free state and the receptor bound states.

s_{min}	Receptor free state				Receptor bound state			
	Min1	Min2	Min3	Min4	Min1	Min2	Min3	Min4
d_{CoM} (nm)	5.10	5.10	5.00	5.35	4.65	4.70	4.75	4.90
θ_{tilt} (rad)	1.35	1.20	1.80	1.50	2.05	2.00	1.95	1.85
θ_{roll} (rad)	1.10	1.25	0.70	1.05	0.60	0.65	0.70	0.75

The coordinates of these minima are provided in Table 5.1. Figure 5.6c illustrates the orientation of the G protein in its ephrin free and bound states. The RBD-RBD conformations corresponding to these energy minima are also shown, along with their $U(d_{CoM}, \theta_{tilt}, \theta_{roll})$ values scaled with respect to the deepest minimum. We depict only the asymmetric case of the relative orientation between the stalk and the RBD dimer however, it is possible that both the RBDs rotate in relation to the stalk. These two sets of preferred conformations are clearly different from each other.

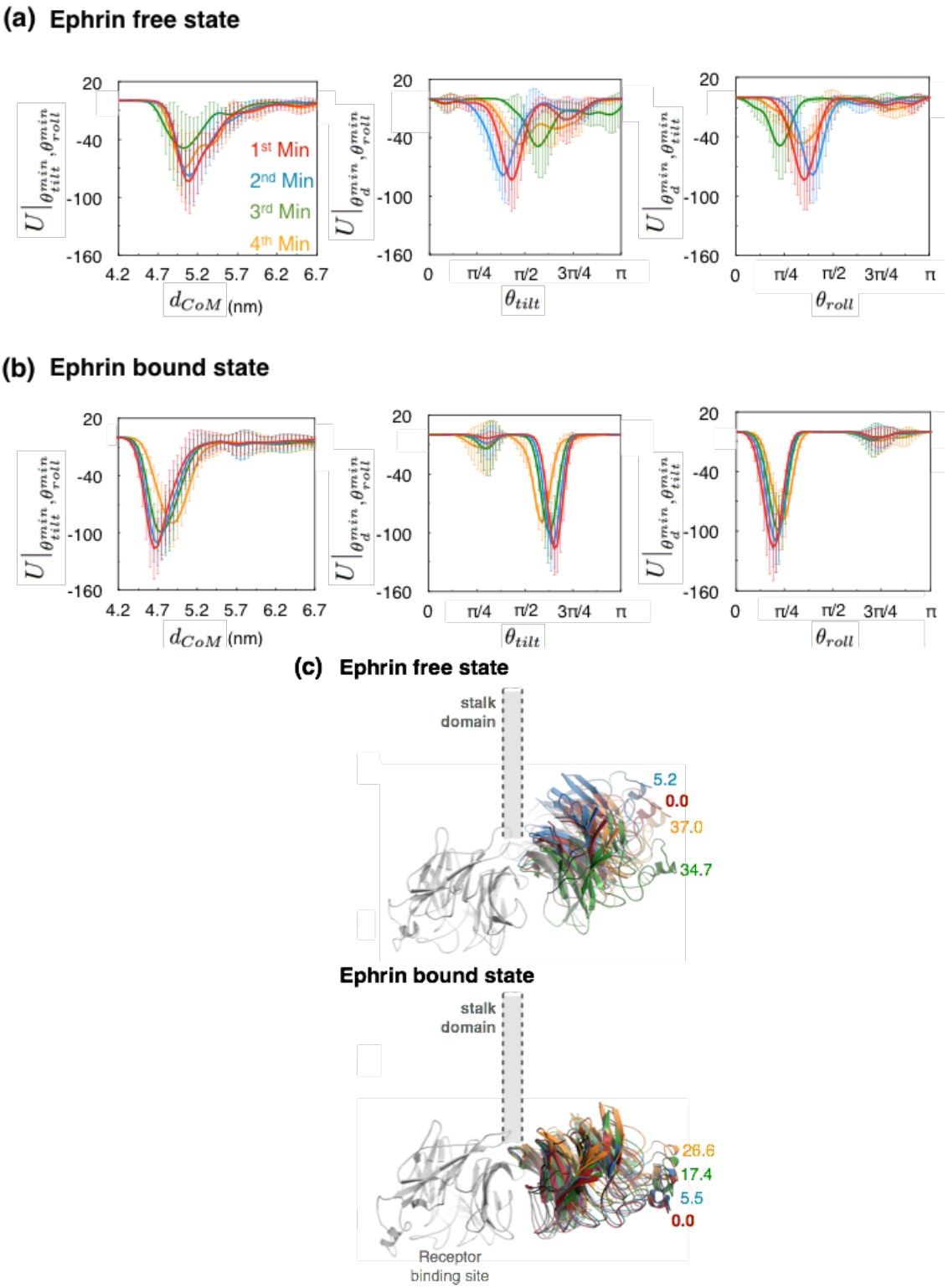


Figure 5.6 The four deepest free energy minima on the RBD-RBD interfacial free energy hypersurfaces $U(d_{CoM}, \theta_{tilt}, \theta_{roll})$ determined for the **(a)** ephrin-free and **(b)** ephrin-bound states and **(c)** shows illustration of the same. All energies are in the units of kJ/mol.

The RBD-RBD conformations corresponding to the deepest wells in the ephrin bound state are explored during sampling of the ephrin free state, but they do not emerge as the energetically preferred conformations in the ephrin free state.

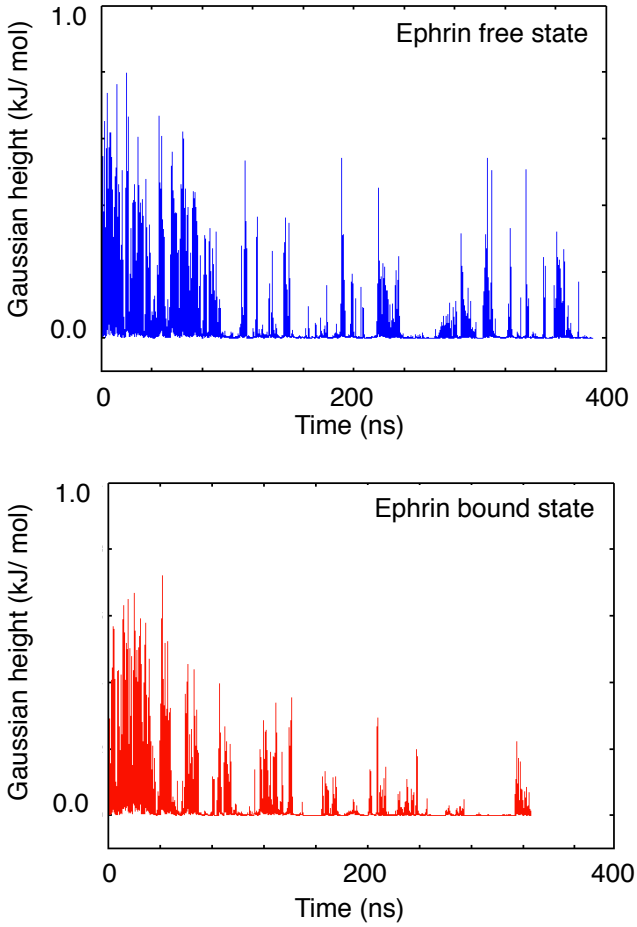


Figure 5.7 Time evolutions of Gaussian heights obtained from well–tempered accelerated conformational sampling of ephrin free and bound states.

The results obtained from the metadynamics simulations are found to be in agreement with our understanding of the RBD-RBD dimer system as explored by the MD studies. However the large error bars from the five separate simulations indicate that the system is not converged. We therefore simulate the ephrin free state and ephrin bound state using well-tempered accelerated conformational sampling. As shown in Figure 5.7, the non-diminishing nature of the Gaussian heights indicate that the

simulations for both the ephrin free and ephrin bound states need to be further continued until convergence is achieved. The coordinates of d_{CoM}, θ_{tilt} and θ_{roll} at the deepest minima of the ephrin bound states are 4.88 nm, 1.20 rad and 0.80 rad while that of ephrin free state are 4.48 nm, 2.00 rad and 0.30 rad respectively. Hence, irrespective of the simulation technique adopted, we arrive at the same conclusion that the RBD-RBD orientation in the ephrin free and the ephrin bound states are dissimilar. Hence, from here on we have used MD as the simulation technique to further probe the RBD-RBD dimer system.

Next, we test the reversibility of the structural transition, that is, whether the RBD-RBD interface in the ephrin free state changes to that of the ephrin bound interface if the ephrins are attached to the RBDs when the RBDs in the dimer are oriented about each in the ephrin free state. Four separate MD simulations were preformed to verify this.

We initiated all these simulations using an RBD-RBD orientation representative of the ephrin free state ($d_{CoM} = 5.0$ nm, $\theta_{tilt} = 1.3$ rad, and $\theta_{roll} = 1.1$ rad). Two of these MD simulations were started after re-solvating the RBD-RBD dimer in salt solution and energy minimizing. These two simulations served as the controls (blue lines in Figure 5.8) and we expect that the RBD-RBD orientation was maintained throughout the simulation. The remaining two simulations were initiated following an additional preparatory step where we substituted the two ephrin free RBDs with pre-equilibrated conformations of the ephrin bound RBDs. The latter two simulations shown by red lines (Figure 5.8) are, therefore, of the ephrin-bound state, but initiated from an RBD-RBD interface preferred in the absence of ephrin. If the ephrin induced structural transition

were indeed reversible, then this substitution would result in the RBD-RBD interface to return to the orientation we observed when the RBDs were bound to ephrin. This is exactly what we observe. The RBDs in these simulations reorient and the interfaces return to the orientation as shown in Figure 5.5b.

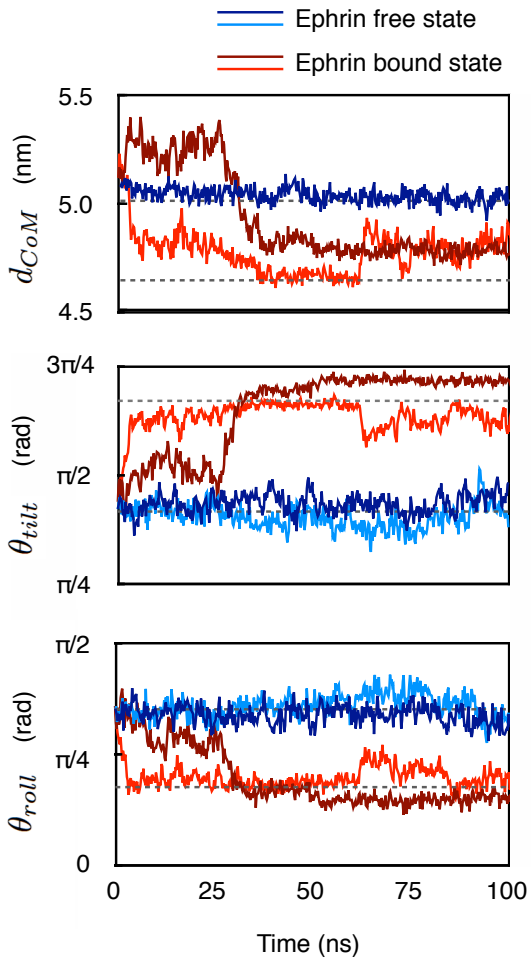


Figure 5.8 Time evolutions of interfacial collective variables, d_{CoM}, θ_{tilt}, and θ_{roll} in four separate MD simulations.

Next, we attempt to understand how the ephrin binding induces a structural transition in the RBD-RBD interface? First we examine the ephrin-RBD interface. Since the interface consists of four salt bridges (90), it is plausible that the ephrin influences the RBD-RBD interface via through-space electrostatics. However, it is known from experiments that alanine-substitution of the non-polar residues in the ephrin binding

site, such as W504, impact the G-stimulation negatively without affecting the ephrin binding (172). This suggests that ephrin's influence on the RBD-RBD interface is not entirely due to through-space electrostatics which hints to the fact that the inception of the signal at the ephrin-RBD interface must constitute the changes in the conformational ensembles of the residues in RBD's ephrin binding site. We quantify the ephrin induced shifts in the conformational ensembles of RBD's residues in terms of a normalized metric η, which is a function of the geometrical overlap between the two conformational ensembles (see section 3.3). For computing η, we constructed the conformational ensembles of the ephrin free and ephrin bound states, that is, $\mathbb{R}_{apo}$ and $\mathbb{R}_{bnd}$, by extracting snapshots at regular intervals from the equilibrated section of their respective trajectories (Figure 4.6a and Figure 4.8). We generated two separate trajectories for each state, which provided enhanced sampling for the extracted conformations, similar to our previous study on the PHPT1 protein (389). Prior to extracting the coordinates of a residue, the entire conformation of the RBD was least-square fitted onto the X-ray structure of the RBD, which was necessary to remove any bias in η against whole molecule rotation and translation (93). Figure 5.9a shows the η values calculated for all the residues in the RBD, highlighting those that are interacting with ephrin. Residues belonging to the RBD's ephrin binding site, that is, residues that are within 5Å from the ephrin in the X-ray structure of the ephrin-RBD complex (90), are highlighted in darker lines. The horizontal dashed line indicates $\eta = erf(1/\sqrt{2})$, which is equal to a CoM shift of 1Å where there is no change in fluctuation (93). As we expected, the ephrin binding indeed alters the conformational ensembles of all the residues in RBD's ephrin binding site.

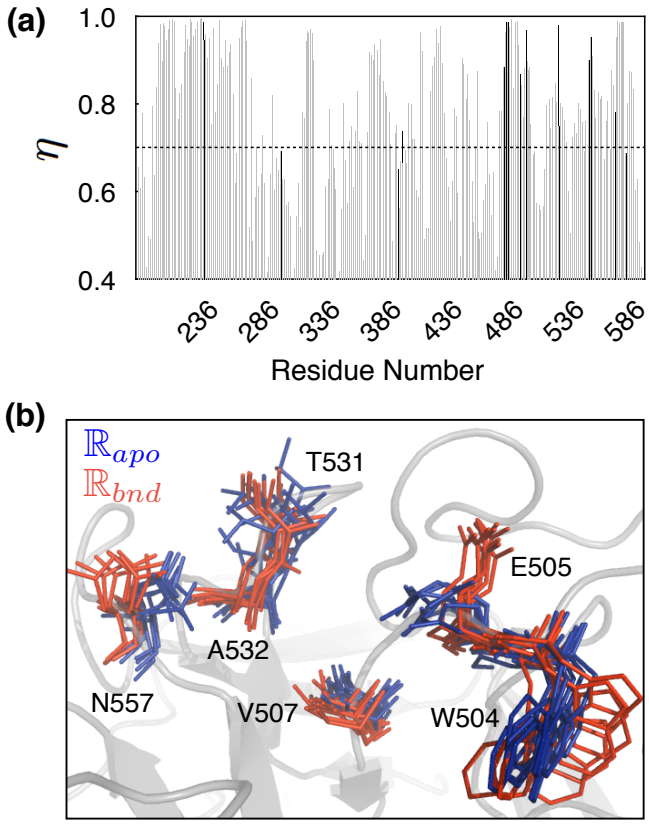


Figure 5.9 (a) Ephrin induced conformational ensemble shifts (η) of the residues in the RBD. **(b)** Conformational ensembles, $\mathbb{R}_{apo}$ and $\mathbb{R}_{bnd}$, of selected residues belonging to RBDs ephrin-binding site.

The comparison of the conformational ensembles in the ephrin free and ephrin bound states also informs us that the ephrin induced changes in RBD's ephrin binding site are not restricted to changes in the CoMs of the amino acid backbones. The effect of ephrin binding results in the reorientation of the amino acid side chains, dampening of the fluctuations of the residues and, surprisingly, the enhancement of the fluctuations of certain other residues (Figure 5.9b).

Furthermore, the extent of the conformational ensemble change of a residue is not correlated with the residue's contribution to the G-stimulation (Table 5.2, shown later). A residue's influence on the G stimulation is the effect its mutation has on fusion, as ephrin binding is known to be unaffected by these mutations. This is evident from the

site-directed mutagenesis experiments reported earlier (67, 172). These changes in the RBD's ephrin binding site combined with the through-space electrostatics seem to trigger the RBD-RBD interfacial reorientation.

Table 5.2 List of η of residues constituting the ephrin binding site in RBD.

Residue	Mutation	η	Fusogenicity
C216		0.9930	
G238		0.9931	
S239		0.9930	
C240		0.9957	
S241		0.9872	
R242		0.9453	
L305		0.3013	
Y389		0.8514	
S390		0.9114	
N394		0.7808	
I401		0.6504	
R402		0.5632	
N404		0.7377	
F458		0.8680	
P488		0.8839	
G489		0.9649	
Q490		0.9872	
S491		0.9860	
Q492		0.9767	
P500		0.9341	
E501		0.8690	
W504	A	0.7698	0-20% of wt
E505	A	0.8831	20-40% of wt
G506		0.9694	
V507	S	0.8971	
Q530	A	0.8470	40-50% of wt
T531	A	0.8046	40-50% of wt
A532	K	0.9808	20-40% of wt
E533	Q	0.7490	0-20% of wt
D555		0.7616	
N557	A	0.9010	20-40% of wt
A558		0.8891	
Q559		0.9527	
E579		0.7815	
I580		0.9538	
Y581	A	0.9902	No expression
I588		0.6876	

Experimental results show that mutations of these residues impact the G-stimulation negatively, some to a greater extent than others (67, 172), but with minimal effect on ephrin-binding. It is to be noted that the ephrin binding alters the backbone CoMs of Q530, A532, E533 and N557. The residue E505 undergoes a side chain reorientation. Ephrin binding dampens the fluctuations of T531 and Q530, but enhances the fluctuations of W504. The structure of the RBD shown (Figure 5.9b) in the background in grey is representative of the ephrin bound state, and is included solely for visualization.

Next, we examine the RBD-RBD interfaces in the ephrin free and the ephrin bound states. As shown in Figure 4.6b, the RBD-RBD interface in the ephrin bound state is more extensive compared to the ephrin free state. This however does not imply that ephrin binding leads only to the formation of new RBD-RBD contacts. In fact, ephrin binding disrupts 8 out of 20 inter-RBD residue-residue contacts and creates 15 new inter-RBD residue-residue contacts (Figure 5.10).

Additionally, since the RBD-RBD conformation in one state is not preferred to the other state, we conclude that the ephrin induces inter-RBD rotation by disfavoring certain contacts and preferentially favoring other inter-RBD contacts. Furthermore, as shown in Figure 5.10 the residues constituting the RBD-RBD interface do not exhibit any specific preference toward residue chemistry or polarity, in fact, our results evidence that about one half of the residues constituting the RBD-RBD interface are non-polar.

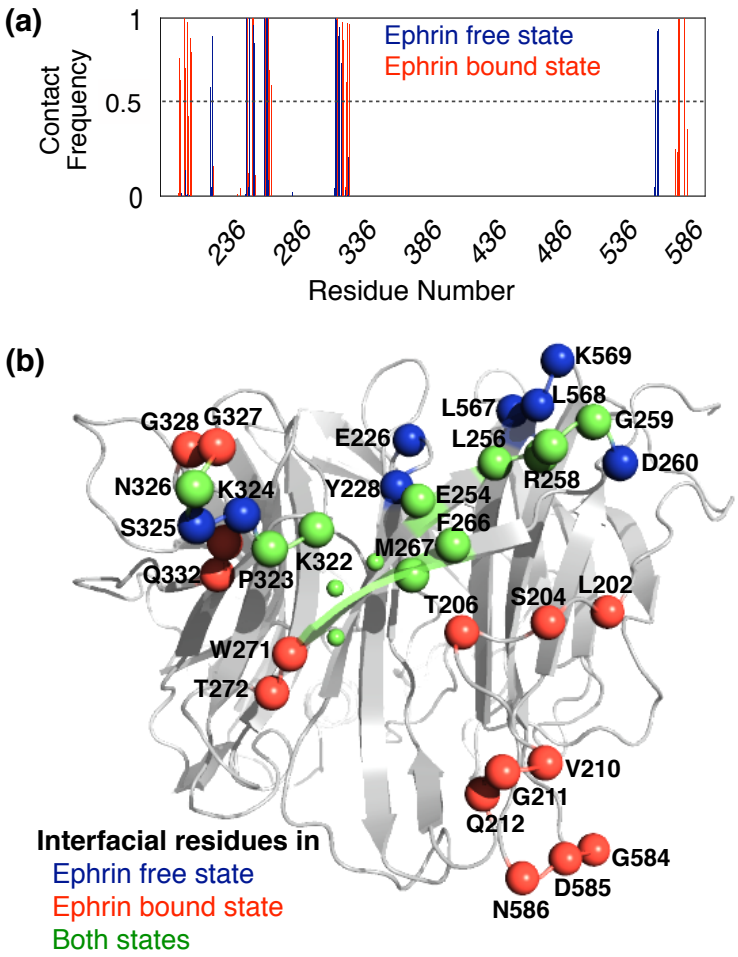


Figure 5.10 (a) Frequency of the inter-RBD contacts. A residue is considered to be in contact with adjacent RBD is it is within 5 Å of the adjacent RBD. **(b)** Residues constituting the RBD-RBD interface in the ephrin free and ephrin bound states. A residue is considered to constitute the RBD-RBD interface if its contact frequency with the adjacent RBD is greater than 50%.

Interestingly, while none of the residues constituting the RBD-RBD interface undergo any large changes in backbone CoMs, almost all of them undergo some form of conformational ensemble shifts as shown in Figure 5.11. The conformational ensembles are depicted using 8 representative snapshots taken from two independent sets of MD runs. The yellow color is used to denote the adjacent RBD in the RBD-RBD dimer. The intrinsic conformational ensembles do not exhibit any ephrin induced changes. Also, we do not observe any systematic trend in the type of the conformational

ensemble shifts. This is apparent from the difference in the type of changes seen in the residues, for example, some residues (D585) undergo changes primarily in their backbone CoMs, while other residues (T206) show shifts in their side chain orientations. In some cases, the residues (G328) exhibit only fluctuation changes. Consequently, a direct relationship between the ephrin-induced ensemble shifts and their contributions to RBD-RBD interfacial reorientations cannot be drawn. There are specific ephrin-induced changes, whose contributions to inter-RBD reorientation appear rationalizable (for example, residues D585 and R589). In the ephrin free state, the RBD's are distant from each other, however, the ephrin bound state behaves differently where they form an intra-RBD salt bridge, and stay close to their respective counterparts in the adjacent RBD. It is therefore plausible that the ephrin binding brings these two residues in proximity to each RBD, negating the electrostatic repulsion between them and their respective counterparts. This allows the two RBDs to form a compact dimeric structure. It is expected that an alanine-substitution of one of these residues will hinder the formation of the RBD-RBD orientation that is seen in the ephrin bound state, and thereby, impact the G stimulation negatively. On the other hand, a double analine-substitution should impact the G stimulation positively. However, one can expect, moderate salt concentrations to counter the effects of such substitutions.

Combining the results so far, we infer that ephrin binding alters the conformational ensembles of several residues in both, the ephrin binding site as well as the RBD-RBD interface. While these changes in the conformational ensemble are small, they trigger a large reorientation of the RBD-RBD interface. The nature of the interfacial rearrangement is such, that it enhances the solvent-exposure of the FAD. This finding is

in good agreement with a proposed model on the fusion regulation of the NiV, essentially stating, that the stimulation of G by ephrin exposes the FAD, which, in turn allows G to activate F (61).

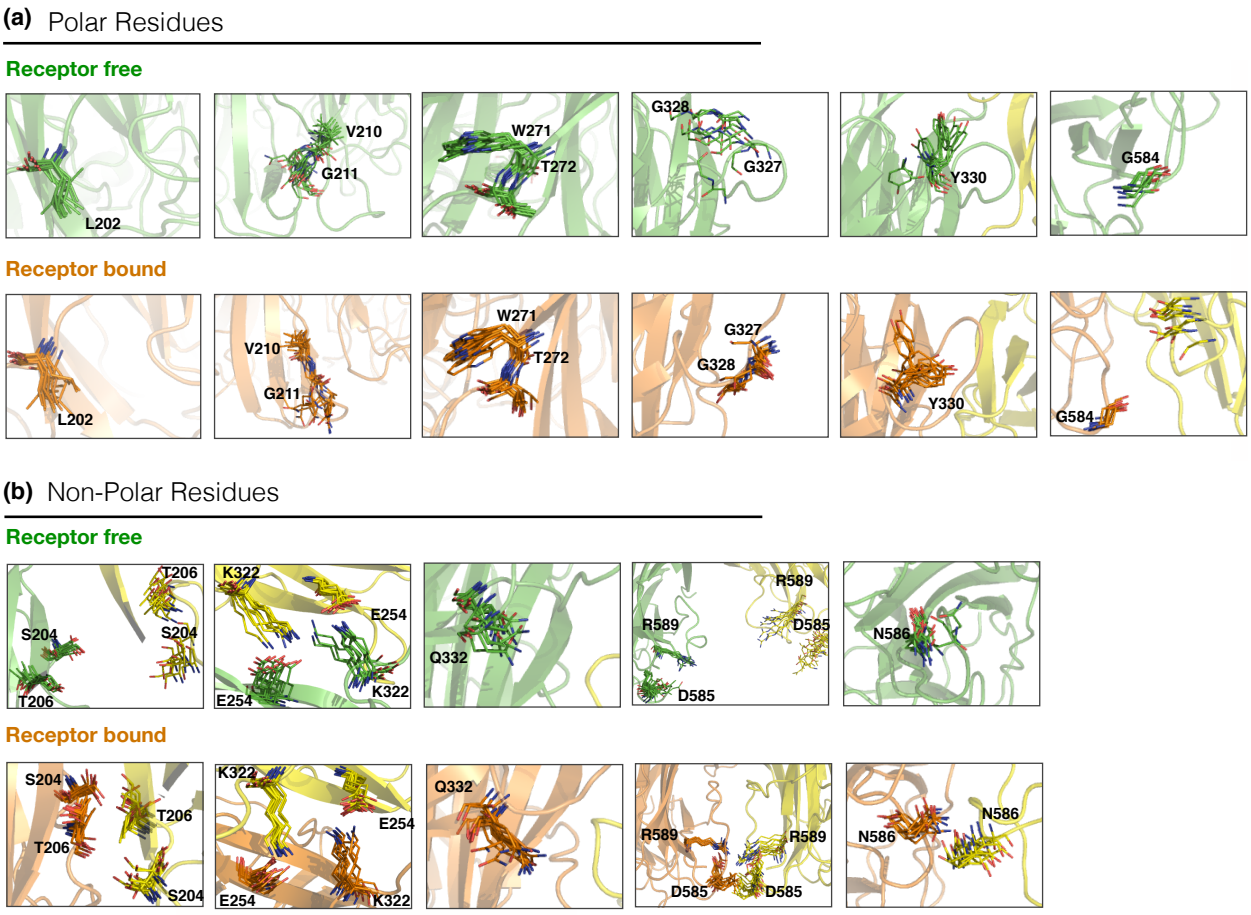


Figure 5.11 Comparison of conformational ensembles of RBD-RBD interfacial residues in their ephrin-free and ephrin-bound state.

5.3 Molecular dynamics simulation of stimulation-deficient mutant

Five of the residues that constitute the RBD-RBD interface in the ephrin bound state are L202, S204, T206, V210 and G211. These residues have been mutated to alanine in experiments, and all of these mutations are known to have a negative impact on G-stimulation (77). In particular, the triple mutation, $V_{209}V_{210}V_{211}$→AAA leads to the complete loss in fusion without affecting the ephrin binding. In the discussion below, we

find answers to the following questions, (i) do the residues contribute to the stability of the RBD-RBD interface in the ephrin bound state? And, (ii) does the alanine substitution prevent the G stimulation by disfavoring ephrin induced interfacial reorientation?

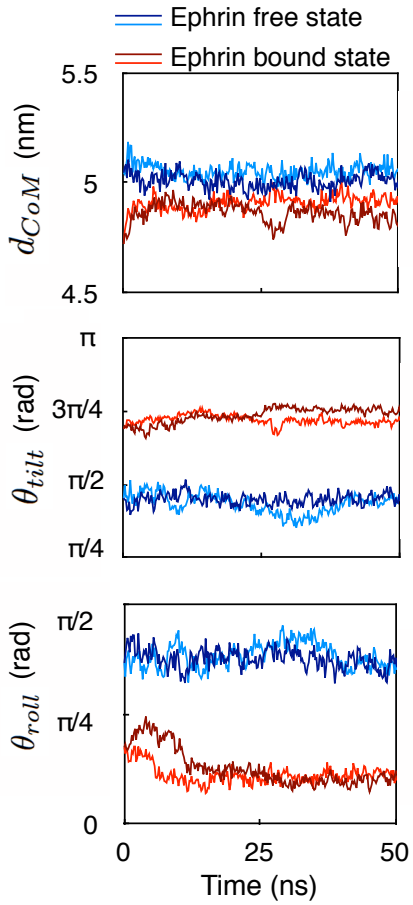


Figure 5.12 Time evolutions of interfacial collective variables, d_{CoM}, θ_{tilt}, and θ_{roll}, in four separate MD simulations of the triple-mutant, VVG→AAA.

To address this, we performed MD simulations of the triple mutant, VVG→AAA. In the first step, we introduced the triple mutation in the monomeric forms of the RBD. The 640 ns and 460 ns snapshots of the MD trajectories simulated for the analysis of the wild-type form in Chapter 3 (94, 95) were used to introduce the triple mutation in the monomeric forms of, the ephrin free and the ephrin bound RBDs.

These mutated structures were energy-optimized, re-solvated in salt solutions, and then equilibrated for 200 ns, at which point, the conformational RMSDs and potential energies were thermalized. The starting conformations of the ephrin free and ephrin bound dimers were constructed from the final snapshots of these simulations by least-square fitting two copies of the RBDs on their respective RBD-RBD orientations observed in the absence and presence of ephrin. These constructed dimers were again energy optimized, re-solvated in the salt solutions and then subjected to MD. We conducted two simulations for each state, and tracked the time evolution of the collective variables that describe the interface between the two RBDs in the dimers as shown in Figure 5.12.

Figure 5.13 compares the collective variables and the representative RBD-RBD conformations obtained from these simulations to that of the wild type dimer. The alanine-substitution of VVG affects the RBD-RBD interface of the ephrin bound state, suggesting that these residues contribute to the stability of the RBD-RBD interface in the ephrin bound state.

Surprisingly, the effect of ephrin binding to the mutated RBD still induces a large interfacial reorientation that brings the two RBDs closer to each other, and away from the C-terminal region of the stalk domain. Effectively, the nature of the ephrin induced interfacial rearrangement preferentially enhances the solvent-exposure of the FAD. Based on experimental results, it is known that the triple mutation abrogates G-stimulation. The outcome of the MD simulations suggest that the ephrin-induced solvent exposure of the stalk may be important, however, it is solely not the sufficient condition for G stimulation.

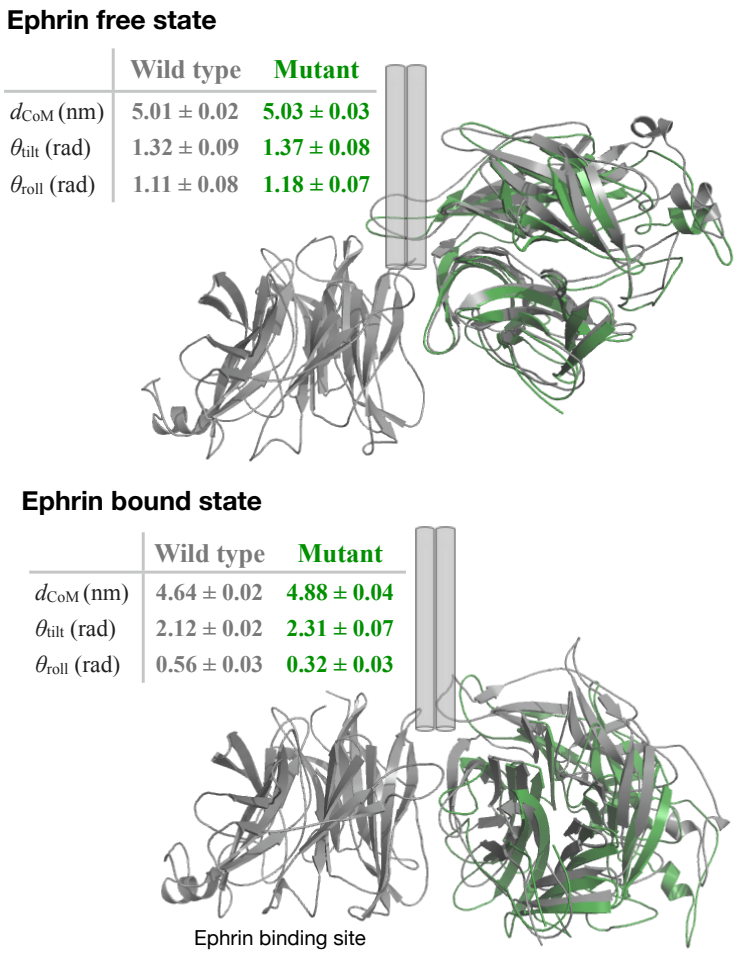


Figure 5.13 Effect of the triple mutation, $V_{209}V_{210}G_{211}$→AAA, on the RBD-RBD interfaces in the ephrin free and ephrin bound states. The standard deviations are estimated from block averaging.

So *how does the triple mutation abrogate the G-stimulation?* The overall ephrin-binding signal that transduces to the FAD is essentially part of the conformational ensemble shifts in the RBD induced by ephrin binding (94, 96-98). The triple mutation must therefore be altering the original signal in the wild type RBD, $\Delta\mathbb{R} := \mathbb{R}_{apo} \rightarrow \mathbb{R}_{bnd}$ (quantified in Figure 4.9a), to a different signal $\Delta\mathbb{R}^m := \mathbb{R}^m_{apo} \rightarrow \mathbb{R}^m_{bnd}$.

To gain insight into the mutation induced shifts in the signal, we determine the subset of residues whose $\Delta\mathbb{R} \neq \Delta\mathbb{R}^m$ based on the technique introduced in section

3.3. This was done by constructing the ensembles $\mathbb{R}^m_{apo}$ and $\mathbb{R}^m_{bnd}$ from the MD trajectories of the mutated dimers and then estimating the difference η^m between them, similar to how η was determined as shown in Figure 5.9a from the ensembles $\mathbb{R}_{apo}$ and $\mathbb{R}_{bnd}$. Then, we estimate η_{apo}, and η_{bnd}, which are the differences between the conformational ensembles $\mathbb{R}_{apo}$ and $\mathbb{R}_{bnd}$, and $\mathbb{R}_{bnd}$ and $\mathbb{R}^m_{bnd}$ respectively. The set of residues, which satisfy the inequality, $\Delta\mathbb{R} \neq \Delta\mathbb{R}^m$ are those that satisfy at least one of the following three conditions:

$$|\eta - \eta^m| > 2{\times}MAE,$$

$$\eta_{apo} > \mathrm{erf}\left(1/\sqrt{2}\right) and$$

$$\eta_{bnd} > \mathrm{erf}\,(1/\sqrt{2}) \qquad (5.1)$$

where MAE is the mean absolute error of our method for quantifying differences between the Guassian ensembles. The first inequality in Equation 5.1, ensures that the difference between the magnitudes of the ephrin induced ensemble shift in the wild-type and the mutant RBDs is greater than the error of our method. In the latter two inequalities, the upper limit $\mathrm{erf}\left(1/\sqrt{2}\right)$ corresponds to a shift in the CoM of 1 Å in the Hilbert space where η values are estimated when there is no change in the thermal fluctuations (94). The latter two inequalities therefore place a tolerance on the mutation induced ensemble shift in the ephrin-free and the ephrin-bound states.

Applying the conditions stated in Equation 5.1, and choosing a MAE = 3.2%, our analysis indicated that about 75% residues exhibit $\Delta\mathbb{R} \neq \Delta\mathbb{R}^m$. Choosing a larger MAE = 5.8%, we find that about 60% of the residues exhibit $\Delta\mathbb{R} \neq \Delta\mathbb{R}^m$. Figure 5.13a

compares η against η_m, and highlights the residues that satisfy the inequality $\Delta\mathbb{R} \neq \Delta\mathbb{R}^m$.

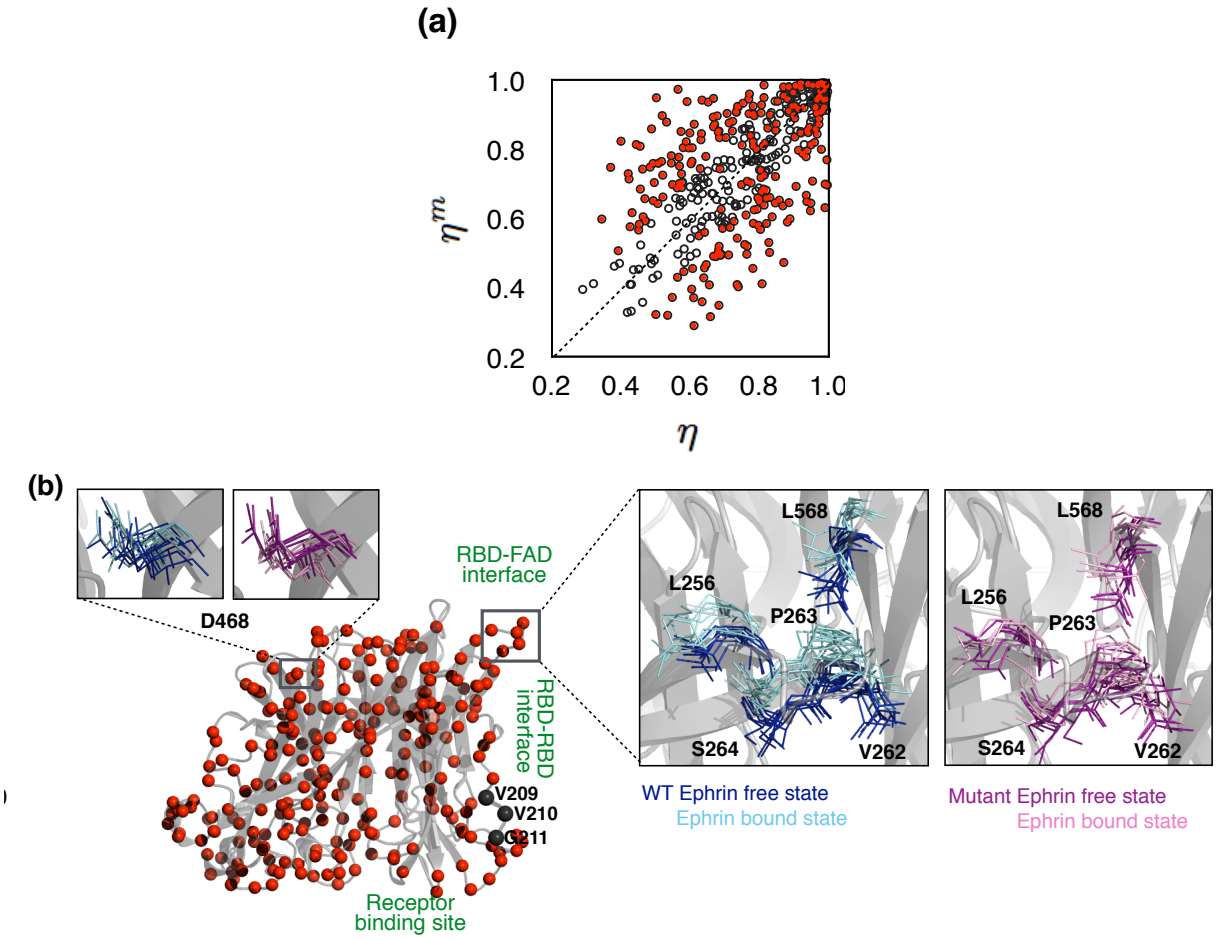


Figure 5.14 (a) Comparison of the ephrin-induced conformational ensemble shifts between the wild type RBD (η) and the mutated RBD (η^m). Residues identified to exhibit $\Delta\mathbb{R} \neq \Delta\mathbb{R}^m$ using MAE= 5.8% in Equation 5.1 are highlighted in red. **(b)** X-ray structure of RBD (center) highlighting the subset of residues (red spheres) that exhibit $\Delta\mathbb{R} \neq \Delta\mathbb{R}^m$. Also shown are conformational ensembles of selected residues, including those proximal to the RBD-FAD interface. Note that the RBD structures shown in the backgrounds in grey are representative structures, and are included solely for visual orientation.

Visualizing the identities of these residues on the X-ray structure of the RBD, it is evident that these residues are located near the mutation site and appear to spread across the entire RBD (Figure 5.14b). In general, one would expect that the extent of

the shift may be inversely related to the distance from the mutation site; however, no such relationship can be inferred from the data shown in Figure 5.12.

Figure 5.14b shows the conformational ensembles of selected residues, including those near the RBD-FAD interface. Experiments show that the alanine-substitution of D468 impacts the G-stimulation negatively, suggesting that it is part of the signal transduction pathway (147). In the wild type form, ephrin induces a distinct shift in D468's conformational ensemble, however, this shift is missing in the event of the triple mutation. Figure 5.14b shows the conformational ensembles of a cluster of five other residues proximal to the RBD-FAD interface. Similar to the D468 case, these residues are not perturbed by ephrin binding in the event of the mutation. Hence, our MD simulations suggest that the VVG mutation abrogates the G stimulation by suppressing the propagation of the ephrin binding signal via the RBD-FAD interface.

5.4 Summary

The results shown in this chapter investigates how the RBD-RBD interface of NiV's G protein is affected by ephrin binding. For the MD simulations, the initial model of the RBD-RBD interface was templated based on the RBD-RBD interface of the homologous G protein of the HeV virus (80). We show evidence that the ambiguity in the templated initial model lies not in the identity of RBD's dimerization face, but in the relative orientation between the two RBDs, which we explore using MD, for both the ephrin free and the ephrin bound cases.

The ephrin binding is found to induce a large change in the RBD-RBD interfacial orientation, which is also reversible. Ephrin induces this reorientation by disfavoring certain contacts and also preferentially favoring other inter-RBD contacts. The residues

constituting the RBD-RBD interface do not exhibit any specific preference toward residue chemistry or polarity, and almost all of these residues undergo some form of conformational ensemble shift, associated either with a change in side-chain orientation or change in fluctuation or change in backbone CoM. None of the residues undergo any large change in its intrinsic structure. Essentially, ephrin induces large inter-RBD reorientations mediated by minor changes in individual RBDs.

Visualizing the ephrin-induced inter-RBD rearrangement in the context of the position of FAD, our results demonstrate that the interfacial rearrangement favours the enhancement of solvent-exposure of the FAD. To gain further insight, we also simulated the effect of ephrin binding on the RBD-RBD interface of a stimulation deficient mutant. The mutation affects the interfacial arrangement in the ephrin bound state, and also enhances the solvent-exposure of the FAD. We therefore conclude that the ephrin induced solvent exposure of the stalk may be important to G stimulation, but is not a sufficient condition. However, there is no experimental structure of the full length ectodomain of the G protein, and this inference is derived purely on the basis of a model of the ectodomain proposed by Broder and coworkers (65), which was homology modeled using the X-ray structures of the full length ectodomains of the receptor binding proteins of other paramyxoviruses. Our simulations clearly evidences that ephrin induces equivalent RBD-RBD interfacial rearrangements in both wild type and stimulation-deficient RBD dimers.

A statistical analysis of ephrin induced conformational ensemble shifts in the wild type and stimulation-deficient mutant dimer shows that the mutation has a global effect on the conformational ensemble of the RBD. Additionally, we show that the mutation

suppresses ephrin induced shifts in residues located near the RBD-FAD interface, despite the fact that the mutation is at the RBD-RBD interface. This indicates that the mutation abrogates G stimulation by suppressing the signal that is mediated to the FAD via the RBD-FAD interface.

CHAPTER 6

TOWARDS CONSTRUCTION OF FULL LENGTH ECTODOMAIN MODEL

6.1 Background

Presently, we are able to obtain a molecular level understanding of signal transduction in the RBD-RBD dimer, but in order to understand what is the form of the signal and how it is transmitted through the RBD-FAD interface, one needs to identify the interface which requisites knowledge of the structure of the FAD. The FAD contains the F-activation site, which is the destination of the allosteric signal that is initiated at the receptor binding site leading to the stimulation of the G protein makes it intriguing, since one can use the start and end points to map the allosteric signaling pathways. An insight into this allosteric pathway of G-stimulation is crucial for the integrated understanding of the viral entry process that involves G-stimulation followed by F-activation that initiates the fusion of the host and viral membranes.

Experiments suggest constitutively active nature of the FAD, where a truncated G protein containing only a segment of the FAD without the RBD, was found to activate the F protein leading to syncytia. However, it is not confirmed if the structure of the FAD remains unchanged when the RBD is cleaved. It was inferred from these experiments that the RBD is not only responsible for G-stimulation upon binding to the appropriate receptor, but it is also important to prevent premature F-activation, which is primarily achieved by concealing the F-activation site on the FAD (69). This implies that on ephrin

binding there must be an inter-RBD rearrangement resulting in the solvent exposure of the FAD. Consequently, one can infer that in the case of a G-stimulation deficient mutant, the F-activation site should remain concealed, suggesting that there must not be any inter-RBD rearrangements. In contrast, MD simulation in Chapter 5, indicate that both the wild type and the stimulation-deficient G undergo similarly large RBD-RBD rearrangements, suggesting that the solvent exposure of FAD is not the only criteria for G-stimulation. This leads to one of the following scenarios, (i) the signal gets progressively suppressed at the FAD, or (ii) it is not transmitted altogether beyond the RBD-FAD interface to the F-activation site in the mutant.

Nevertheless, currently there does not exist any experimentally resolved structure of the FAD. Hence, it becomes essential to model the FAD. As mentioned in section 2.1 disordered regions can be involved in allosteric signaling, hence it becomes imperative to verify the presence or lack of such disordered regions in the FAD. We therefore use the protein disorder prediction tool PrDOS (390), which is a widely used tool among many others (391-393), to evaluate the possibility of structural disorder. Every residue in the sequence is assigned a binary disordered/ordered tag, in the range [0,1], which corresponds to probability values above/below 0.5, respectively (394). A probability above 0.5 is reserved for a disordered residue. The disorder probability for each residue of the G protein (1 – 602 amio acids) is plotted in Figure 6.1. Although residues in the regions 194 – 198 and 392 – 398, in the RBD (residue 177 – 602) are predicted as disordered, we know from x-ray crystallography that it is the six-bladed β-propeller C-terminal region (90). The residues in the range 1 – 100 form the N-terminal cytoplasmic tail and the transmembrane helix. Since we aim to model the FAD, we

focus on the disorder probability of the region between residues 101 – 176, corresponding to the FAD.

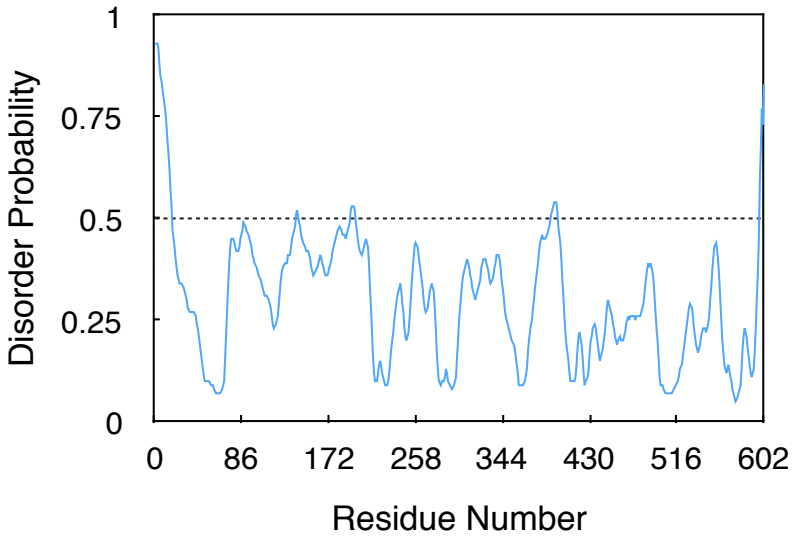


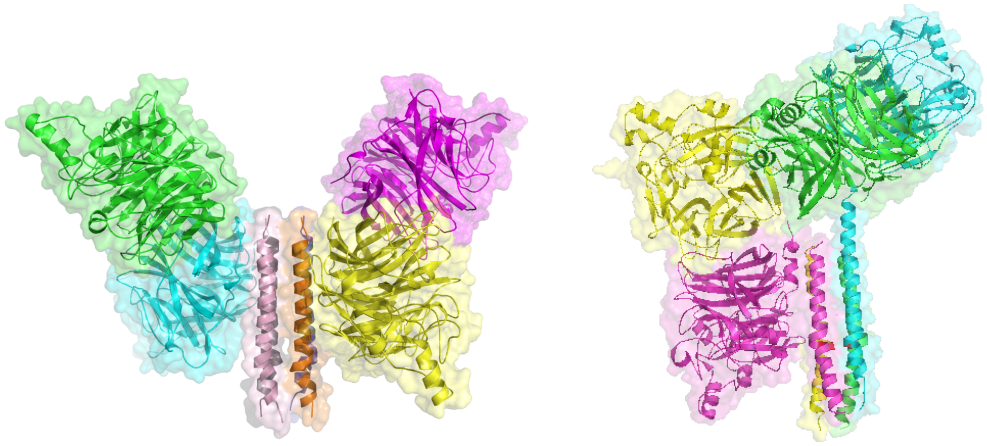
Figure 6.1 The disorder probability prediction for each residue. The residues with probability higher than 0.5, marked by the black dashed line are predicted as disordered.

The peak seen between residues 139 – 144 (NENVND), can be attributed to the presence of a cluster of polar residues, which is known to result in disorder (298, 395, 396). However six residues are too few to establish disorder. Reports on sequence alignment of the G to other paramyxoviruses show the presence of a similar cluster of polar residues, which is part of the protein sequence that connect the RBD and FAD (160, 397). Any specific contribution of this region towards the functionality of G has not been reported (398-400).

6.2 Overall strategy

FAD of two other paramyxoviruses, namely the HN proteins of NDV and PIV5 have been crystallographically resolved (91, 92). Both of them consist of a parallel tetrameric α-helical coiled coil structure commonly referred to as four-helix bundle (4HB) shown in Figure 6.2 (91, 92, 146). As mentioned earlier in Chapters 1 and 2, there is no

structural information on the FAD of the G protein, however the secondary structure prediction and sequence alignments to existing proteins indicate that the FAD of G has similar α-helical structure (74, 166).



(a) NDV (b) PIV5

Figure 6.2 Tetrameric architecture of the viruses belonging to the *Paramyxoviridea* family **(a)** NDV and **(b)** PIV5. Each monomeric unit is shown in a different color. Due to the high symmetry in the structure only two helices of the 4HB are distinctly visible.

Although the sequence identity of the G protein with the HN of NDV and PIV5 is very low, precisely 17.78% and 20.30% respectively (Figure 6.3) (272), we know from several experimental reports that the G is a tetramer with the FAD forming the 4HB similar to the structure of the HN of NDV and PIV5 (shown in Figure 6.2). It is to be noted that the length of FAD in the G protein is significantly longer, approximately 40 amino acids, than the other HN proteins. The difference in length especially in the regions between residues 100-183 are visible as gaps, in the aligned sequence shown in Figure 6.3. These extra regions lead to certain unique structural features in the G protein, which are unseen in other paramyxoviruses, which is discussed in details in the later sections.

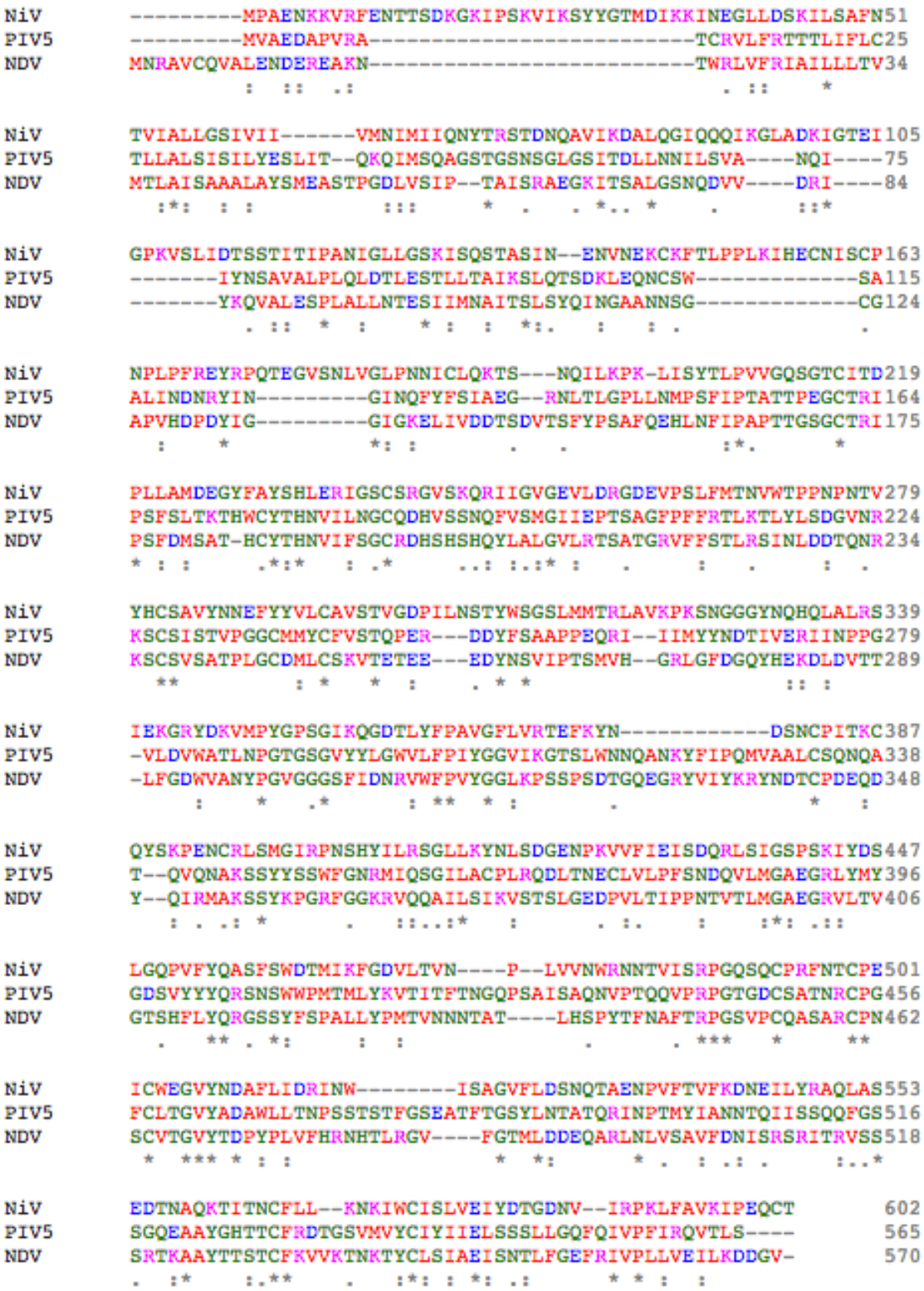


Figure 6.3 Aligned sequences of the G proteins of NiV (UniProt Q9IH62) and HN proteins of PIV5 (UniProt P04850) and NDV (UniProt P12554).

Our overall strategy involves constructing the constituent regions of FAD and using the information on RBD-RBD interfaces in bound/apo states from Chapter 5 we will compile the full-length ectodomain by mapping onto a low resolution cryo-EM surface density. This is an established technique used to construct a full length protein by compiling the existing three dimensional structural information of its different constituting domains (401-405), discussed later.

6.3 Molecular modeling approach

Based on the sequence, structural similarity to other paramyxovirus host binding proteins and data available from experiments done on the FAD of G protein, one can identify three distinct regions and model their structures. Our aim is construct the separate regions and then piece them together to obtain the final full length FAD. We will assemble these regions to obtain a final model of the FAD, and eventually construct the ectodomain (using a method discussed in section 6.3). Reports of sequence alignment of the region between residue 101 – 145 with other paramyxoviruses, most importantly PIV5 and NDV indicate the presence of 4 helical bundle (160). There are three cysteine residues in FAD between the region 146-162, which are known from mutagenesis experiments to be crucial for the maintenance of the dimer-of-dimer architecture of the G protein by forming disulphide bonds (160). The last 14 amino acids, residues 163-176 are known to be unique to the *Henipavirus* (160) and are rich in prolines, which forms the linker between the FAD and RBD. Based on the above information, we divide the FAD into three regions, (i) residues 101-145 constitute the helical bundle region, (ii) the disulphide-bridged region consist of residues 146-162 and (iii) the proline-rich linker region defined by residues 163-176. We present an elaborate

discussion in the following sections on how the experimental results help us to obtain an understanding of the possible structures of these individual regions, thereby, aiding in making an appropriate choice of regions specific modeling tools.

Figure 6.4 Sequence of the FAD highlighting the residues forming the basis for dividing the FAD into three separate regions. The hydrophobic repeat in the helical bundle region is highlighted in blue, the cysteine residues are shown in red in the disulphide–bridged region and the proline residues are indicated in green in the proline–rich linker region.

Modeling a protein is possible by using a myriad of tools available for academic use. As discussed in Chapter 3, the use of the tool depends upon the user's requirements, desired end product, amount of experimental information available on the query protein, and also on the opportunity to refine/validate the model by predicting experimental mutations. In this section we specifically focus on modeling the three regions mention above and discuss our approach and the rationale behind our decisions to model each region.

6.3.1 Homology modeling of helical bundle region

In the early 1950s it was postulated that there is a range of possible sequence periodicities in the regular packing of the helices that can favor the α-helical configurations with the number of residues per turn, such as $4/1$ (4 residues over a single turn), $7/2$ (7 residues over 2 turn), $11/3$, $15/4$, and $18/5$, which are often referred to as tetrad, heptad, hendecad and so on (406, 407). These repeats play a crucial role in the determination of the packing interactions in the helices to form the coiled coil structures. Several of these coiled coil structures are known to contain

discontinuities in a periodically recurring pattern (408-412), suggestive of a structural model with local changes that can possibly impact the protein functionalities. The discontinuity can be of different forms, for example a four heptads plus a skipped residue making a total of 29 residues of a local segment that is repeated throughout the protein (408, 409), or four heptads followed by a skipped residues that shifts the heptad repeat pattern in the ensuing three heptads, which is seen in the hemagglutinin glycoprotein of the influenza virus or the HA protein (411). This disruption in the repeat pattern is known to affect the pitch profile, which is the rate of winding of the α-helices around each other in the coiled coil structure and is also a key determinant of both intramolecular and intermolecular interactions (413). Such interruptions in the repeat pattern can initiate or terminate the super-helical twists, thereby altering the pitch of the coiled coil structure.

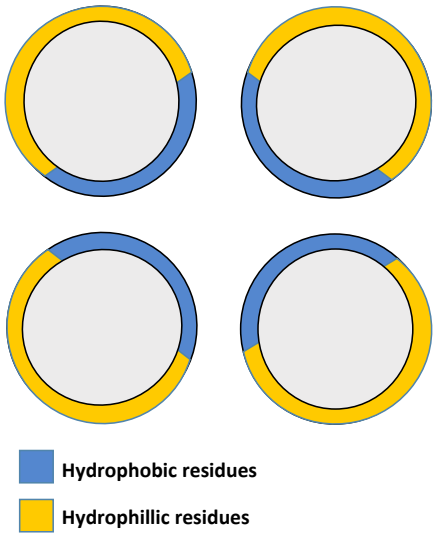


Figure 6.5 The general arrangement of hydrophobic residues in 4 helical bundle conformation shown using a schematic of the top view. The hydrophobic residues, shown in blue reside on the inside, away from the solvent while the hydrophilic residues (yellow) prefer to interact with the solvent.

A four-helix-bundle or the 4HB topological arrangement is attained by proteins when the interfaces between the helices consist mostly of hydrophobic residues such

that the polar side chains are exposed on the surface interacting with the solvent environment, as shown in the schematic in Figure 6.5. Often the 4HB form coiled coils, which are bundles of α-helices that are wound into super-helical structures. (406) They can run in the same directions called parallel or in opposite directions termed antiparallel (414). It is the α-helices packing interactions (Figure 6.6) of the residues at the hydrophobic core that contribute to the nature of the final coiled coil structure, which can also be interpreted from the sequence of the hydrophobic residue repeats within each individual helix (406, 415). Depending on the repeat motif mentioned earlier, the hydrophobic residues can be arranged in the sequence such that every first and fourth residue is hydrophobic in a heptad, or every first, fourth and eighth residue of the hendecad forms the core and so on (414). The possible arrangements that can lead to specific interactions are shown in Figure 6.6. As shown, these residues at the core of the helical bundle can in general be classified into parallel, perpendicular or acute interactions depending on the orientation of the side chain of the residues in an individual helix with respect to the adjacent helices (415). These interactions are primarily dominated by the first and fourth residues of the repeating motifs, which form the *a*-layer and the *d*-layer respectively in the hydrophobic core of the helix. The specific arrangements and orientations of the side chains result from the various non-covalent interactions at the core that give the structure its stability. The nature of the *a*-layer and *d*-layer, whether continuous or discontinuous can manifest into columnar helical bundles or twisted super helices respectively.

The sequences of the 4HB regions of the FAD of the paramyxoviruses show a predominant existence of the heptad or hendecad repeat pattern (146, 160). In the

heptad repeat every first and fourth residue is hydrophobic, thereby allowing the two turns over seven residues to form the helix, while in the case of the hendecad, every first, fourth and eighth residue is hydrophobic leading to the three turns of a helix over eleven residues (406). Thus identifying the appropriate motif occurring in the FAD of G is essential in selecting the template to be used during homology modeling and is crucial for the proper prediction of the FAD region of the G protein.

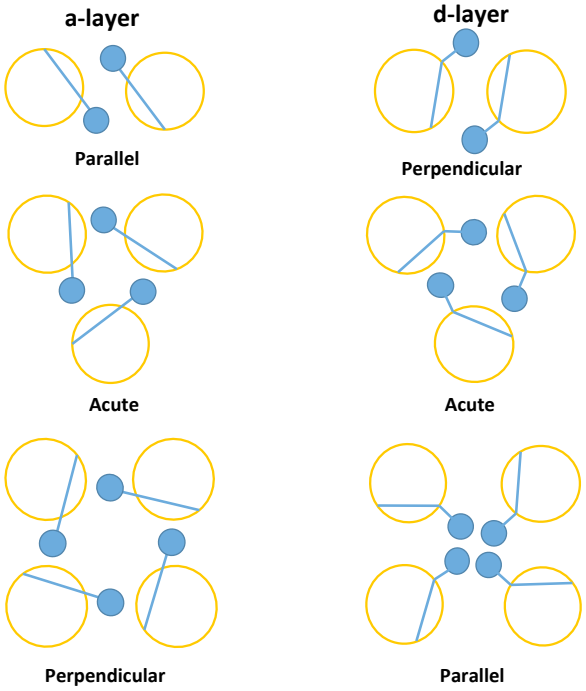


Figure 6.6 The packing interactions at the core of the helical bundles between the hydrophobic residues highlighting the different interactions of the side chains indicated as blue ball and sticks.

The template for modeling the sequence can be searched by using the various sequence comparison tools such as BLAST (270, 271) based solely on sequence similarity. A BLAST analysis of the sequence of the FAD returned a total of eighteen proteins with sequence identity more than 30%, of which sixteen belonged to the same genus (*Henipavirus*), however none of these proteins have an experimentally resolved three dimensional structure in the database (416). Although, there are no three

dimensional structures with high sequence identity, we can utilize the FAD structures of other viruses belonging to the *Paramyxoviridae* family as the basis to template the structure of FAD of G. Once we identify the appropriate template, we can perform homology modeling, which is a sequence and three dimensional structured template based technique (273).

Here we discuss our basis for choosing an appropriate template prior to homology modelling the FAD of G, by performing a comparative study of sequences (Figure 6.3) and structures (Figure 6.4) of HN of PIV5 and NDV. Although the sequence identity of the G protein of NiV as a whole to the HN of NDV and PIV5 is as low as 17.78% and 20.30% respectively, it belongs to the same family of the virus; moreover multiple experiments indicate similar conformations of the proteins and also the conserved core mechanism of their functionalities (10, 59, 61-64, 76, 88).

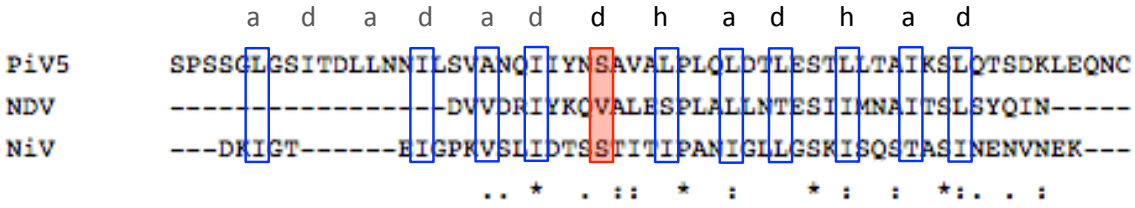


Figure 6.7 Sequence of FAD of NiV G aligned to HN of NDV and PIV5. The 4HB heptad repeats of PIV5 are labeled *a* and *d* in grey, followed by the transition point highlighted in red and the hendecad repeat labeled *a*, *d* and *h* in black. Only the repeats common in NiV are highlighted in blue.

The isoleucine, proline, and two serines are conserved in the sequences of NiV G, NDV HN and PIV5 HN, (indicated by stars at the bottom row in the sequence alignment profile in Figure 6.7) however, a difference emerges when the hydrophobic repeats are compared (417). The hydrophobic repeats for PIV5 are labeled in the top row, while only those common to the NiV are highlighted with blue boxes in the figure. The repeats of several paramyxoviruses are known to be shifted – the motif changes

from a heptad repeat to a hendecad. The transition point between heptad and hendecad is highlighted in red in Figure 6.7 (160). The first and fourth residue of the heptad repeat are labeled on the top row as *a* and *d* respectively in grey, signifying its contribution in forming the *a*-layer and the *d*-layer. After the transition point the first, fourth and eighth residues are depicted as *a*, *d* and *h* respectively. The PIV5 HN has a serine at this transition point which gives the 4HB a kink to initiate a slight super-helical twist (146), while in case of the NDV HN, this transition point is occupied by a valine, which being hydrophobic in nature renders the 4HB conformation to form a continuous and uninterrupted coiled coil structure (92). To visualize the impact this subtle detail in the sequence can have on the tertiary structure, we show the 4HB conformation of PIV5 in Figure 6.8.

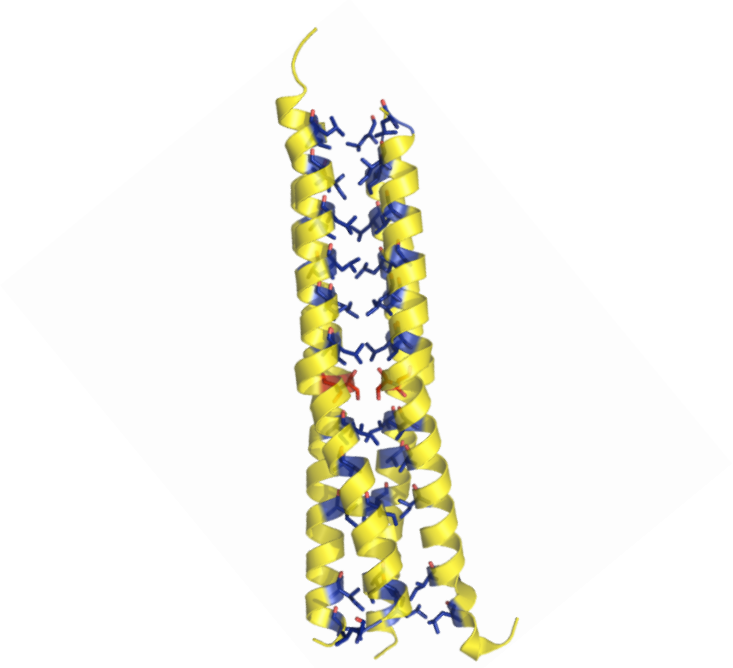


Figure 6.8 The x-ray crystallographic structure of host binding protein of PIV5 4 helical bundle FAD. The hydrophobic residues forming the a-layer and the d-layers of the motifs at the core are shown in blue. The serine at the transition point between heptad and hendecad repeat is highlighted in red.

Unlike the NDV HN, the NiV G protein has a serine at the transition point. Moreover the repeat pattern of heptad disrupted by the serine and followed by hendecad repeat is similar to that of PIV5. Hence, we decide to use PIV5 as our template for homology modeling.

The four helices of the PIV5 structure are not identical (91, 92, 146, 160); they have minute differences between each other as illustrated in Figure 6.9. The RMSDs are calculated (shown in figure) for all three helices with respect to one helix. To retain the uniqueness of each of the helices one should model them separately.

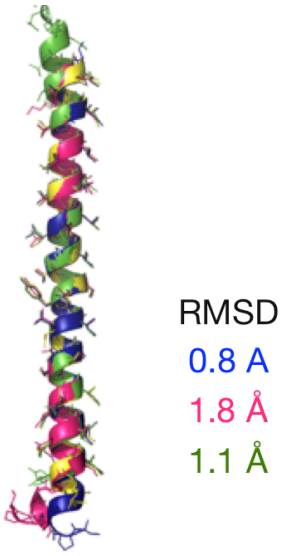


Figure 6.9 Root mean squared deviation (RMSD) of the three helices shown in colors blue, pink and green with respect to the yellow helix.

Based on the method discussed in section 3.2.1, we first aligned the sequences of PIV5 HN and NiV G as shown in Figure 6.7 using CLUSTAL OMEGA (272). We homology modeled the four helices by generating several structures using MODELLER version 9.15 (273, 274). The obtained structures must be assessed to rank them and identify the native structure. Since the native structure of a protein is generally the lowest free energy of all states, one can argue that the assessment of the final structure can be done by evaluating the free energy surface of a protein derived by thorough

sampling of the potential energy surface. However, since the potential energy surface is defined by molecular mechanics force fields, certain errors may originate from the approximations in the force fields (105, 418). To minimize such errors, an alternate method can be used which involves a scoring function whose global minimum corresponds to the native structure from multiple sampled structures of different sequences available in the database (419-425). The DOPE score, is the outcome of the default scoring function in MODELLER (273-275). It is a statistical potential specifically optimized to evaluate the homology modeled structure and is an acronym for Discrete Optimized Protein Energy. Here, the reference state explicitly depends on the sizes of the native structures, which are used to derive the statistical potential and this method claims to improve the results leading to increased accuracy of protein structure assessment (418). The score is so defined that the structure with the lowest value is the one closest to the native form. Hence, we identify the DOPE scores of the helices, and then construct the final model based on the structural alignment of the helix with the lowest DOPE score. This is done by aligning the Cα atoms of the hydrophobic core highlighted in Figure 6.7 of the NiV G onto the Cα atoms of the PIV5 HN 4HB structure.

We proceed by identifying a suitable approach to obtain the final model. In the first approach, which we call the multiple sequence alignment approach, the structural information from all four helices are utilized to generate 2000 structures. From these modeled structures, one helix is identified corresponding to the lowest DOPE score. Then, we superimpose it onto the four helices of the template structure. This results in a symmetrical 4HB region as shown in Figure 6.10. However, the model obtained by this

approach is not ideal since it does not preserve the naturally occurring difference between the helices noticed in the 4HB of PIV5 (depicted in Figure 6.9).

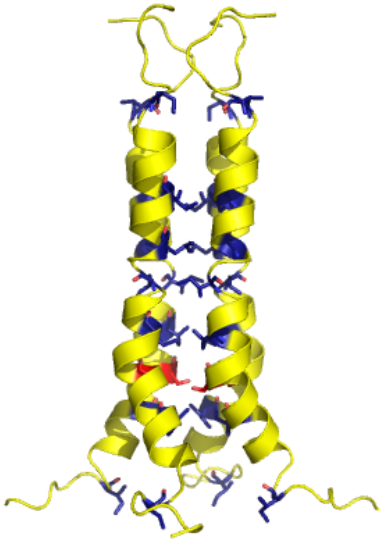


Figure 6.10 The homology model obtained by the multiple sequence alignment approach, such that a single helix is generated from the structural information of all four helices and superimposed, which generates this perfectly symmetrical structure. The hydrophobic core residues are colored in blue and the serine in red.

In our second approach, called individual alignment approach, each sequence is aligned to the 4HB region of G so that the structural alignment of each helix is individually retained. We generate 500 structures of each helix totaling to 2000 structures. The lowest score structure is then selected from the four clusters containing the 500 structures of separate helices and templated to construct the final 4HB of G as shown in Figure 6.11. The residues forming the hydrophobic core are highlighted to emphasize the retention of the innate packing information of PIV5 HN (shown in Figure 6.8). Therefore, this approach provides us a reliable structure of the final model of the 4HB region of FAD.

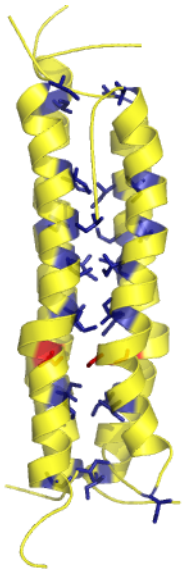


Figure 6.11 The structure obtained from homology modeling using structural information of each of the helices separately. The hydrophobic residues forming the core are colored in blue and the serine is depicted in red.

6.3.2. Ab initio structure prediction of disulphide-bridged region

As shown in the sequence alignment in Figure 6.3 there are a total of seventeen cysteine residues of which fourteen are in the RBD and are essential for the six bladed β-propeller forming seven disulphide bonds (86). The remaining three cysteine residues C146, C158, and C162 lie in the FAD as shown in the Figure 6.12.

To study the functional role of these cysteine residues, their involvement in intermolecular disulfide bond formation and their contribution to the oligomeric structure of G, the individual cysteine residues were mutated to serine experimentally (160). Results from Western blotting and SDS-PAGE indicate that cysteine residues 158 and/or 162 maybe involved in intermolecular disulfide bond formation. It was undetermined whether both of these cysteine residues formed disulfide bonds or if due to the proximity of these residues to each other, a mutation in one residue affected the ability of the other to form a disulfide bond (160). These mutants were reported as fusion defective after studying its effect on F activation, that was verified by an analysis

of syncytia or multinucleated giant cell formations (160). Thus, the effect of the mutations on fusion provides conclusive evidence that these disulphide bonds are essential to maintain the oligomeric structure of G. Based on the above results; a possible arrangement of the disulphide bonds was estimated as illustrated in Figure 6.12. Since, we do not have any known structure that can be used as a template to model this region, we cannot use homology modeling as a tool here. In this regard, we resort to *ab initio* methods and apply the experimentally obtained constraints in the disulphide bond region to perform predictive modeling.

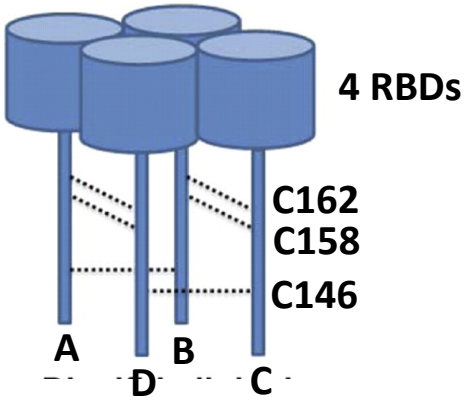


Figure 6.12 C158 and C162 form double inter-subunit disulfide bonds (dotted lines) and the dimer-of-dimer structure is formed by the inter-subunit disulfide bond formation through C146 (dotted line). The cylindrical barrels depict the RBD. The monomers are labeled as A, B, C and D, to indicate the constraints implemented in the simulation (Adapted with permission from (160). Copyright 2012 American Society for Microbiology).

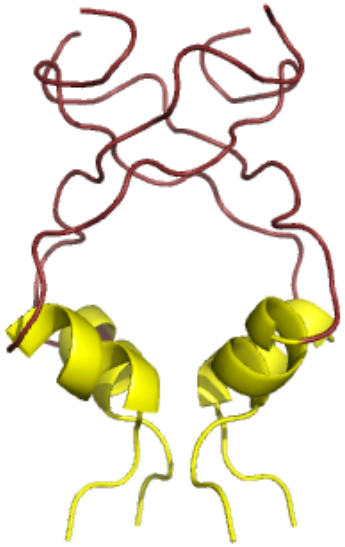
Based on the method discussed in section 3.2.2 we used the fold and dock module of ROSETTA, that allows *ab initio* prediction method, symmetric protein assembly and inclusion of constraints (293). The fold and dock module is a combination of the symmetric assembly protocol and the ROSETTA *ab initio* structure prediction protocol, where the internal degrees of freedom of a monomer, and rigid body degrees of freedom between monomers that are symmetric are simultaneously sampled (290,

292, 293, 426). It works efficiently for the prediction of structures of intertwined symmetric assemblies (427-429). The implementation of the symmetric assembly protocol allows predicting the structure of a symmetrical protein assembly based on the structure of a single subunit, and the constraints incorporated between the subunits. Hence, the applicability of this module suits the scenario in this region, where the single subunits are the monomers (A, B, C, or D), and the inter-monomer constraint is determined by the nature of disulphide bonds (see Figure 6.12).

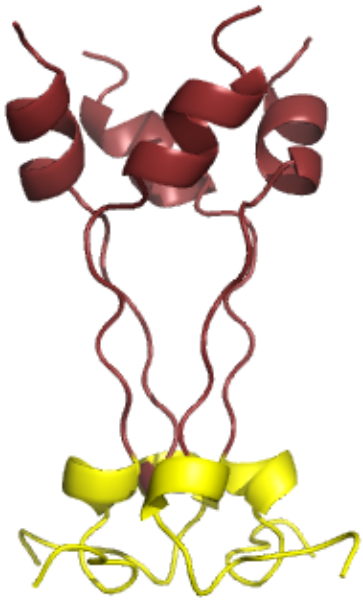
First, fragment libraries of three and nine residue lengths are generated using the target sequence, by matching them with various structures available in the PDB database. We use the Robetta server that provides an automated tool to create PDB based fragment libraries for all the overlapping fragments of three and nine residues (426). As seen in Figure 6.4, there are seventeen residues (C146 –C162) in the disulphide-bridged region. Prior to starting the *ab initio* calculations, we select the target sequence such that it contains the entire disulphide bridge region and ten residues (A136 – K145) from the 4HB region. By including the residues from the 4HB region, and then performing the *ab initio* calculations, we essentially add another constraint that allows us to filter out and eliminate the structures (among thousands of possible structures generated by using ROSETTA) in which the helicity within the residues A136 – K145 are not maintained. Inclusion of these residues also allows us to connect the two constituting FAD fragments together.

We generated 10000 structures, which were then grouped following an energy clustering procedure using the cluster application in ROSETTA, (described in chapter 3) yielding 15 distinct clusters. We find, that more than 50% of the total number of

structures falls into two clusters, indicating that the final structure may belong to one of the two. In Figure 6.13, we show the specific structures corresponding to the minimum energy in these two clusters.



Cluster 1



Cluster 2

Figure 6.13 The two clusters with the lowest energies obtained after clustering the 10000 structures. The N-terminal of cluster 2 retains the helices in the 4HB region, highlighted in yellow.

Visualization of the two structures from the respective clusters show that while cluster 1 appears to develop a bulge in the middle where the individual monomers bend outwards, in the case of cluster 2, it maintains a more columnar, cylindrical conformity. In cluster 1, we find that the individual helices which are a continuation of the 4HB region seem to point outward, hinting an inconsistency in the structure in comparison with the final model of 4HB obtained from homology modeling (Figure 6.11). Therefore, it can be safely concluded that the conformation represented in cluster 2 is better related to the final model of the 4HB region, i.e. the structure of the residues of the 4HB region shown (yellow) in Figure 6.13 conforms in a manner to maintain continuity with the 4HB region (Figure 6.11), and hence is more favorable.

6.3.3 Structure prediction of proline-rich linker region using accelerated conformation sampling

In general, the proline rich region is found in certain proteins and is known to play a crucial role in assembly and regulation of intracellular signaling (430-433). This region is not present in the paramyxovirus family except in the *Henipavirus* genus, and is known to connect the RBD and the FAD (160). As shown in Figure 6.14, there exists a gap in the alignment profile when the sequences of G are aligned to *Morbillivirus* H proteins using the conserved PP-XX-I/V motif, confirming that the proline-rich region is unique to G. In this region, out of the fourteen residues (P163 – E176), four of them are prolines.

The structure of a proline residue contains a nitrogen atom, which is covalently bound within a five-membered ring, an unusual structure where the side chain is cyclized back onto the backbone amide position (434, 435). This restricts the residue's

backbone conformational motion. In addition, the presence of N-CH_2 group restricts the motions of the residue preceding the proline. This -CH_2 group prevents it from being a hydrogen bond donor. Together these restrictions make proline a 'helix and β-sheet breaker', giving the simplified view that proline disrupts secondary structure by inhibiting the backbone to conform to α-helix or β-sheet conformation. Another alternate interpretation is, proline overrides other forms of secondary structure to conform to its own kind of secondary structure (433-435).

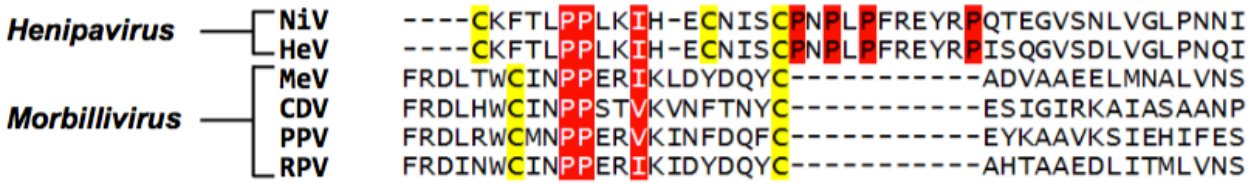


Figure 6.14 Alignment of G proteins of *Henipavirus* to H proteins of *Morbillivirus* to emphasize the uniqueness of the proline-rich region. The gap shown using dashed line when the sequences of G are aligned to *Morbillivirus* H proteins using the conserved PP-XX-I/V motif highlighted in red with white lettering (Adapted with permission from (160). Copyright 2012 American Society for Microbiology).

In the case of proteins containing the proline-rich region, the sequences are known to have specific motifs corresponding to their functional roles (434, 435). In this regard, there exists no sequence match between the motif of G and any other protein containing a proline-rich region. Due to this, we cannot template the proline-rich region of G to that of another protein, thereby preventing us from the use of any template based structure prediction technique. Therefore, for the purpose of our work, we are interested to identify if there is a preferred distance between the RBD and FAD. This can be done by estimating the end-to-end distance of the proline rich region in its preferred conformation, which can be obtained my monitoring its evolution in the free energy space.

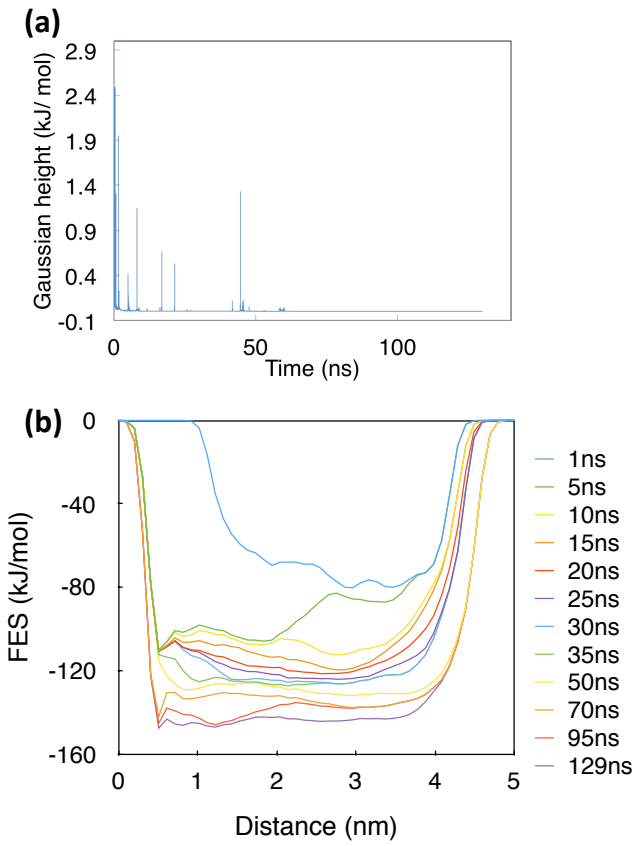


Figure 6.15 (a) Time evolution of Gaussian heights, and **(b)** shows the corresponding free energy surface as a function of CV (distance), calculated at different intervals of the simulations.

We apply well-tempered accelerated conformational sampling technique (section 3.1.3) to explore the various preferred energy orientations of the proline rich linker region (436). It is a method that facilitates sampling by the introduction of an additional bias potential or force that acts on a set of selected number of degrees of freedom referred to as CVs. A history dependent bias potential is deposited as a function of the CVs in the form of Gaussians. Our CV is the distance between centers–of–masses of the C–terminal and N–terminal of this fragment. The structure corresponding to the minimum energy can be identified as that with the preferred conformation.

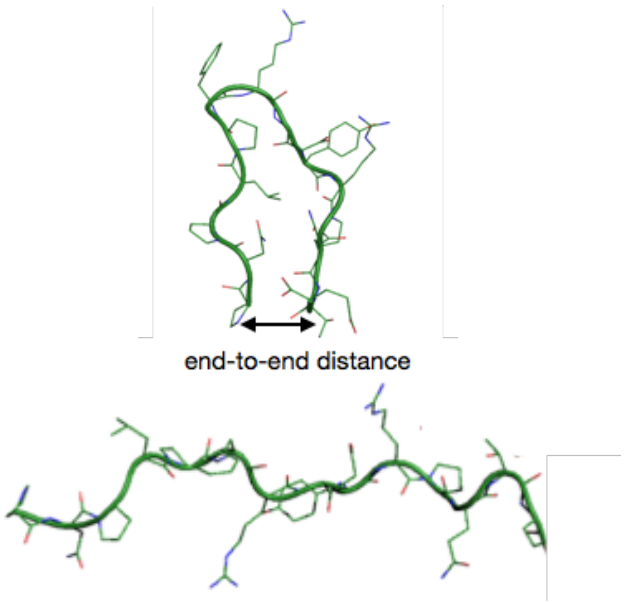


Figure 6.16 Representative preferred conformations obtained from accelerated MD simulations of sequence containing 14 residues.

The 14 residues in the proline-rich linker region of G are such that the end residues constitute of a non-polar proline at the N-termnal and a polar glutamic acid at the C-terminal. We simulate the sequence containing the 14 residues by setting the lower bound for the CV at 0.5 nm. The end-to-end distance of the residues of the N–termini and C–termini provides the preferred conformation of the linker to connect the RBD to FAD. Hence, an end-to-end distance of < 0.5 nm is not feasible to accommodate the RBD and FAD. To identify the preferred conformation we calculate the free energy at different intervals as the simulation progresses. The simulation was continued until convergence was achieved which is evident from the time evolution of Gaussian heights (Figure 6.15a). The convergence is indicated by the diminishing nature of Gaussian heights. Figure 6.16b shows the free energy surface as a function of the end–to–end distance of the proline rich linker region. In Figure 6.15b, the free energy surface does not show a distinct minimum value for distance between 0.5 nm – 4 nm. In Figure 6.16, we show two representative conformations of the proline rich

linker region when the end–to–end distance is 0.5 nm and 4.5 nm. Thus there is no preferred end-to-end distance, which means that the proline rich region serves simply as a flexible linker. Or perhaps its structure depends on interactions with the FAD and RBD, which will be examined when combining these structures using cryo-EM.

6.4 Mapping molecular model onto cryo-electron microscopy protein surface

Cryo-EM is a form of transmission electron microscopy, where biological samples are cooled down to cryogenic temperatures (Liq. N_2), in order to minimize thermal fluctuations. The cryogenic environment is essential to obtain an electron density map of the molecules with reduced uncertainty in their positions. Additionally, one can achieve atomic-resolution of complex, dynamic molecular assemblies, by integrating cryo-EM density maps to the different modalities for structure determination, such as X-ray crystallography and nuclear magnetic resonance spectroscopy.

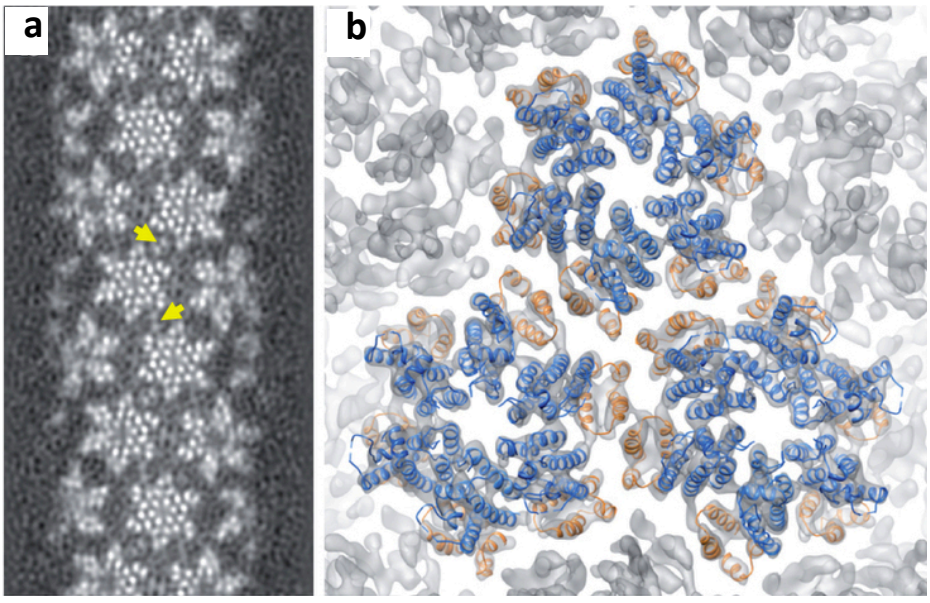


Figure 6.17 a. Electron density map of the CA tube of HIV-1 capsid. b. MDFF model of the HIV-1 capsid assembly, superimposed with the electron density map. Three CA hexamers, with N-terminal domains shown in blue and C-terminal domain in orange (Adapted with permission from (405). Copyright 2013 Nature Publishing Group).

Figure 6.17a is shown as an example of a typical cryo-EM density map obtained for CA tube of HIV-1 capsid (437). Such cryo-EM maps can be subjected to methods such as rigid-fragment fitting (438-440), as implemented in DireX (441), Flex-EM (442) applications, or flexible fitting as done in Rosetta (443) and FRODA (444) to refine the structural information. These methods use low-frequency normal modes (445) deformable elastic networks (441) and cross correlation (446) or least squares difference between experimental and simulated maps (447) to drive the structure into the cryo-EM density. Other methods use Molecular Dynamics Flexible Fitting (401, 402, 448) (MDFF) where molecular dynamics (MD) is performed to match structures obtained from crystallography and electron microscopy data. Numerous structures have been successfully resolved using MDFF, including ribosomes some(403, 449-451), photosynthetic proteins (452, 453), chaperonins (454), myosin (455), etc. Figure 6.17b shows an atomic model for the entire assembly of CA hexameric protein of HIV-1 capsid (456), obtained using MDFF. In this case, Zhao *et. al.* built the linker and missing loops using homology modeling, and then the atomic structures of the N-terminal domain and the C-terminal domain were docked into the electron density, which was followed by applying MDFF to get the final structure (405).

The discussion above highlights the reliability, utility and advantage of generating the structure of a macromolecule by mapping its molecular model onto the protein surface obtained from cryo-EM microscopy. Therefore, we identify this technique as a prospective route to obtain the complete structure of the G protein at a molecular level, which can be conveniently achieved by acquiring the cryo-EM density map from our

collaborators. This presents the opportunity for future research to gain a comprehensive understanding of the complete structure and allosteric signaling in G.

6.5 Summary

The FAD contains the F-activation site, which is the destination of the allosteric signal that is initiated at the receptor binding site leading to the stimulation of the G protein which makes it intriguing, since one can use the start and end points to map the allosteric signaling pathways. Therefore, in order to understand how the signal from the RBD is transmitted through the RBD-FAD interface, one needs to identify the interface which requisites knowledge of the structure of the FAD.

We model the FAD by breaking it down into three regions. We use the protein disorder prediction tool, to identify any residues forming structural disorder. We note that G is structured, which allows us to use structure prediction tools to model the regions in FAD.

The 4HB region is modeled using homology modeling, which required us to identify an appropriate template. For this, we compare the 4HB regions of the HN of PIV5 and NDV paramyxoviruses, and infer that HN of PIV5 is a suitable template owing to its similarity in the hydrophobic repeat to that of the G protein. We performed homology modeling using two approaches, namely, the multiple sequence alignment approach and the individual alignment approach. We find, in case of the individual alignment approach, the innate packing information of HN of PIV5 is retained. Therefore, this approach provides us a reliable structure of the final model of the 4HB region of FAD.

Experiments show, that the disulphide bonds in the FAD are essential to maintain the oligomeric structure of G. There exists no known structure that can be used as a template to model the disulphide-bridged region, hence, we cannot use homology modelling as a tool. In this regard, we resort to ab initio methods and apply the experimentally obtained constraints in the disulphide bond region to perform predictive modelling. Prior to starting the ab initio calculations, we select the target sequence such that it contains the entire disulphide bridge region and ten residues (A136 – K145) from the 4HB region. By including the residues from the 4HB region, and then performing the ab initio calculations, we essentially add another constraint that allows us to filter out and eliminate the structures in which the helicity within the residues A136 – K145 are not maintained. The predicted structures are grouped following an energy clustering procedure, which resulted in 15 distinct clusters, of which 2 contained more than 50% of the generated structures. We then identify one as the most favorable based on the criteria of the constraints applied.

In the proline rich linker region, out of the 14 residues (P163 – E176), four of them are prolines. There exists no sequence match between the motif of G and any other protein containing a proline rich linker region. Hence, we cannot template the proline rich linker region of G to that of another protein, which prevents us from the use of any template based structure prediction technique. Therefore, we focus on identifying the length of the proline rich linker region, to estimate the distance between the RBD and FAD. We proceed by estimating the end-to-end distance of the proline rich linker region in its preferred conformation, which can be obtained my monitoring its evolution in the free energy space. We apply well-tempered accelerated conformational sampling

technique to explore the various preferred energy orientations of the proline rich linker region. We find that the location of the minimum energy is not distinct, meaning the distance between the two ends of the proline rich linker does not have any preferred conformation.

CHAPTER 7

CONCLUSION AND FUTURE DIRECTIONS

This work demonstrates the applicability of molecular dynamics to identify and characterize the details underlying dynamic allosteric stimulation of the host binding protein of NiV, paving the way for its application to the study of other paramyxoviruses. The research presented in this dissertation provides a new approach to model a protein structure, and study the allosteric mechanism involved.

Figure 7.1 summarizes the findings obtained while studying the allosteric stimulation of the host binding protein of NiV. We find, that interstitial water behave a lot like interfacial waters and its explicit description is important for modeling signal inception. Our study of two receptors (ephrin B2/ B3) with dissimilar sequences (< 50% sequence similarity), indicate a quantitative difference in the dynamics of interstitial water, but the trends in the shifts with respect to the values of bulk water are found to be similar. Despite the high wetness of the protein-protein interfaces, the dynamics of interstitial water is considerably slower compared to the bulk.

We find that ephrin binding induces a large change in RBD-RBD interfacial orientation, which is reversible. Ephrin induces this reorientation by disfavoring certain contacts and also preferentially favoring other inter-RBD contacts. The residues constituting the RBD-RBD interface do not exhibit any specific preference toward residue chemistry or polarity, and almost all of these residues undergo some form of

conformational ensemble shift, whether it is a change in side-chain orientation or change in fluctuation or change in backbone center-of-mass. None of the residues, however, undergo any large change in intrinsic structure.

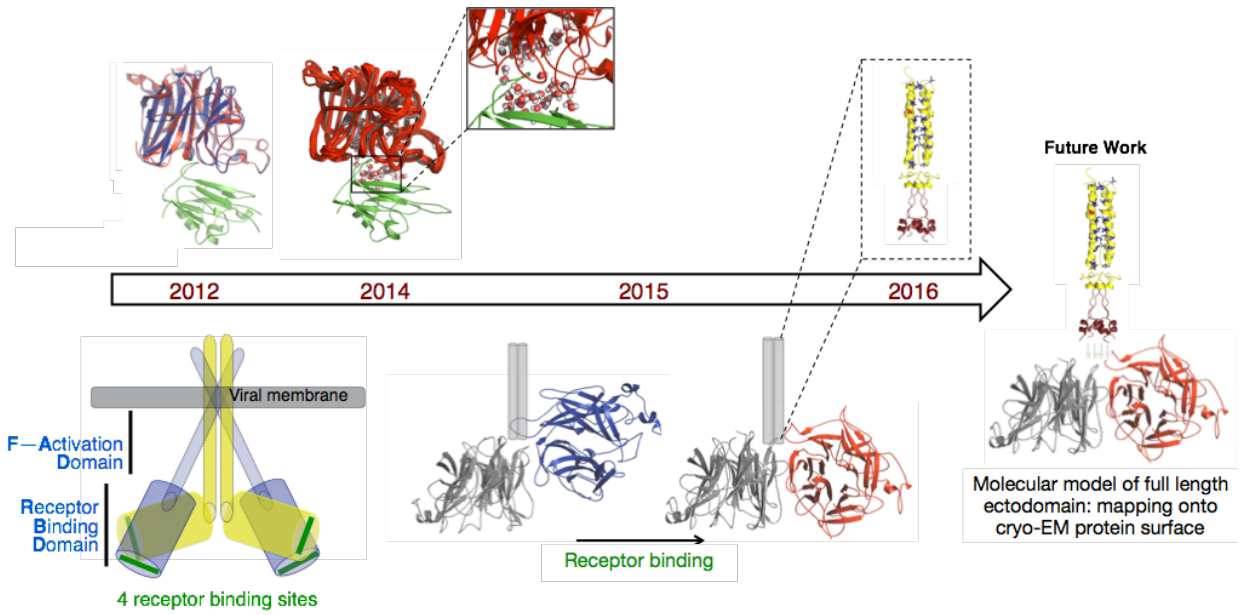


Figure 7.1 Timeline representation of progress of research. The x-ray crystallographically resolved receptor binding domain in the apo (blue) and bound (red) states together with the proposed dimer-of-dimer architecture (schematic shown) formed the starting point. In 2014, the crucial role of interfacial water in the inception and propagation of allosteric signal was established. In 2015, a new model was proposed wherein extensive inter-domain rearrangements triggered by minor structural changes in the individual domains. In 2016, the constituent regions of the F-activation domain were modeled.

Essentially, ephrin induces large inter-RBD reorientations through only minor changes in individual RBDs. Visualizing the ephrin-induced inter-RBD rearrangement in the context of the position of FAD, we find that the interfacial rearrangement is such that it will enhance the solvent-exposure of the FAD. This finding essentially supports a proposed model of fusion regulation of the NiV where stimulation of G by ephrin exposes FAD, which in turn, allows G to activate F. To gain further insight, we also simulate the effect of ephrin binding on the RBD-RBD interface of a stimulation-deficient

mutant. We find that while the mutation does affect interfacial arrangement in the ephrin bound state; the ephrin induced interfacial rearrangement is still such that it will enhance the solvent-exposure of the FAD. We therefore, conclude that while ephrin induced solvent exposure of FAD may be important to G stimulation; it by itself is not a sufficient condition.

We model the constituent regions of the fusion protein activation domain involved allosteric pathway in the full length ectodomain of the host binding protein, using structural prediction tools by implementing a bottom–up approach. We use a combination of homology modeling, *ab initio* structure prediction and accelerated conformational sampling techniques to model the three regions. The full length G protein structure can be constructed in order to extensively study the allosteric mechanism induced by receptor binding. This can be done by mapping the developed molecular model onto cryo-EM data of the protein surface as discussed in Section 6.3. The outcome of this research prompts intriguing questions that need to be addressed: (i) What is the molecular detail of the allosteric pathway in the full length ectodomain? (ii) Can we identify this allosteric communication pathway to identify specific residues as targets to design allosteric drugs? The answers to these questions hold promise towards advancement in the treatment of infections by NiV and other paramyxoviruses.

In general, from our analysis of interstitial water at the G-B2 and G-B3 interface, we see that the interstitial water molecules exchange with the bulk water, and also have lower diffusion coefficients suggesting their sluggish behavior due to the formation of hydrogen bonds with the proteins. But, is this dynamic behavior of interstitial water independent of the protein system to which they belong? To verify this, we simulated

and carried out similar analyses of the interstitial water molecules in two different complexes, namely, the BLIPII–TEM1 and BLIPII–TEM22. We justify the choice of these two complexes and provide preliminary data in Appendix A. A more detailed analytical approach is necessary, which can be addressed by performing simulations on other complexes. How do the following influence the dynamics of interstitial water: (i) topology of the protein–protein interface, (ii) volume of the protein–protein interface, (iii) physicochemical nature of the protein–protein interface, and (iv) size of constituting proteins? Future work in this direction will help us to gain knowledge about the global behavior of interstitial water, if any, which may be instrumental in developing specific solvent models to define interstitial water.

Do the water molecules on the protein-water interface at the protein surface play any role in allosteric signaling? We establish that interfacial water at the protein-protein interface plays a crucial role in the inception and propagation of the allosteric signal. Explicit description of solvent at the interstitial region is essential for the proper inception of the allosteric signal at the receptor binding site of the host binding protein. Building on this, we use explicit solvent and identify the receptor induced allosteric stimulation of the dimeric receptor binding domain. We find that extensive inter-domain rearrangements are triggered by minor structural changes in individual domains. Based on our analysis, we make testable forward predictions concerning the specific mechanism of allosteric signaling. We also find that mutations in dynamically stimulated proteins can induce non-local changes that extend to the entire protein.

In addition, what remain undetermined are the specific relationships between the conformational ensemble shifts of topologically distant allosteric sites. Identifying such

causal links between distant sites remains an area of active research in the field of protein allostery, and such relationships are relatively unknown even for several well-studied proteins like GPCRs and PDZ domains. This is essentially a N–body correlation problem, but the correlation that needs to be studied is not in conformational fluctuations, but in their ensemble shifts. It is to be noted that analytical tools to perform such calculations are being developed in our lab, which will help to identify such causal links between distant sites.

REFERENCES

1. Plattet, P., and R. K. Plemper. 2013. Envelope Protein Dynamics in Paramyxovirus Entry. Mbio 4.
2. Allan, G. M., and B. Arroll. 2014. Prevention and treatment of the common cold: making sense of the evidence. Canadian Medical Association Journal 186:190-199.
3. Punpanich, W., and T. Chotpitayasunondh. 2012. A review on the clinical spectrum and natural history of human influenza. International Journal of Infectious Diseases 16:E714-E723.
4. Musso, D., and D. J. Gubler. 2016. Zika Virus. Clinical microbiology reviews 29:487-524.
5. Hogle, J. M. 2002. Poliovirus cell entry: Common structural themes in viral cell entry pathways. Annual Review of Microbiology 56:677-702.
6. Murphy, S. L. B. S., K. D. M. A. Kochanek, J. M. D. Xu, and E. P. D. Arias. 2015. Mortality in the United States, 2014. In NCHS Data brief, U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
   Centers for Disease Control and Prevention.
7. Howard, C. R., and N. F. Fletcher. 2012. Emerging virus diseases: can we ever expect the unexpected? Emerging Microbes & Infections 1.
8. Field, H. E., J. S. Mackenzie, and P. Daszak. 2007. Henipaviruses: Emerging Paramyxoviruses associated with fruit bats. Curr Top Microbiol 315:133-159.
9. Parrish, C. R., E. C. Holmes, D. M. Morens, E.-C. Park, D. S. Burke, C. H. Calisher, C. A. Laughlin, L. J. Saif, and P. Daszak. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology Reviews 72:457-+.
10. Domingo, E. 2010. Mechanisms of viral emergence. Veterinary Research 41.
11. Aguilar, H. C., and B. Lee. 2011. Emerging paramyxoviruses: molecular mechanisms and antiviral strategies. Expert Rev Mol Med 13.
12. Locatelli, S., and M. Peeters. 2012. Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. Aids 26:659-673.
13. Aljofan, M. 2013. Hendra and Nipah infection: Emerging paramyxoviruses. Virus Research 177:119-126.
14. Mackenzie, J. S. 2005. Emerging zoonotic encephalitis viruses: Lessons from Southeast Asia and Oceania. Journal of Neurovirology 11:434-440.
15. Wong, K. T. 2010. Emerging epidemic viral encephalitides with a special focus on henipaviruses. Acta Neuropathologica 120:317-325.
16. Dimitrov, D. S. 2004. Virus entry: Molecular mechanisms and biomedical applications. Nature Reviews Microbiology 2:109-122.

17. Kliger, Y., S. A. Gallo, S. G. Peisajovich, I. Munoz-Barroso, S. Avkin, R. Blumenthal, and Y. Shai. 2001. Mode of action of an antiviral peptide from HIV-1. Inhibition at a post-lipid mixing stage. J Biol Chem 276:1391-1397.
18. Leonard, J. T., and K. Roy. 2006. The HIV entry inhibitors revisited. Current Medicinal Chemistry 13:911-934.
19. Badani, H., R. F. Garry, and W. C. Wimley. 2014. Peptide entry inhibitors of enveloped viruses: The importance of interfacial hydrophobicity. Biochimica Et Biophysica Acta-Biomembranes 1838:2180-2197.
20. Fofana, I., N. Jilg, R. T. Chung, and T. F. Baumert. 2014. Entry inhibitors and future treatment of hepatitis C. Antiviral Research 104:136-142.
21. Tsai, B. 2007. Penetration of nonenveloped viruses into the cytoplasm. Annual Review of Cell and Developmental Biology 23:23-43.
22. Firquet, S., S. Beaujard, P.-E. Lobert, F. Sane, D. Caloone, D. Izard, and D. Hober. 2015. Survival of Enveloped and Non-Enveloped Viruses on Inanimate Surfaces. Microbes and Environments 30:140-144.
23. Dimitrov, D. S. 2000. Cell biology of virus entry. Cell 101:697-702.
24. Chandran, K., D. L. Farsetta, and M. L. Nibert. 2002. Strategy for nonenveloped virus entry: a hydrophobic conformer of the reovirus membrane penetration protein mu 1 mediates membrane disruption. Journal of Virology 76:9920-9933.
25. Flores, A., and E. Quesada. 2013. Entry Inhibitors Directed Towards Glycoprotein gp120: An Overview on a Promising Target for HIV-1 Therapy. Current Medicinal Chemistry 20:751-771.
26. Lempp, F. A., and S. Urban. 2014. Inhibitors of Hepatitis B Virus Attachment and Entry. Intervirology 57:151-157.
27. Zhou, Y., J. Agudelo, K. Lu, D. H. Goetz, E. Hansell, Y. T. Chen, W. R. Roush, J. McKerrow, C. S. Craik, S. M. Amberg, and G. Simmons. 2011. Inhibitors of SARS-CoV entry - Identification using an internally-controlled dual envelope pseudovirion assay. Antiviral Research 92:187-194.
28. Zhou, Y., and G. Simmons. 2012. Development of novel entry inhibitors targeting emerging viruses. Expert Review of Anti-Infective Therapy 10:1129-1138.
29. Qian, X.-J., Y.-Z. Zhu, P. Zhao, and Z.-T. Qi. 2016. Entry inhibitors: New advances in HCV treatment. Emerging Microbes & Infections 5.
30. Baldick, C. J., M. J. Wichroski, A. Pendri, A. W. Walsh, J. Fang, C. E. Mazzucco, K. A. Pokornowski, R. E. Rose, B. J. Eggers, M. Hsu, W. Zhai, G. Zhai, S. W. Gerritz, M. A. Poss, N. A. Meanwell, M. I. Cockett, and D. J. Tenney. 2010. A Novel Small Molecule Inhibitor of Hepatitis C Virus Entry. Plos Pathogens 6.
31. Bogan, A. A., and K. S. Thorn. 1998. Anatomy of hot spots in protein interfaces. Journal of molecular biology 280:1-9.
32. Wei, X. P., J. M. Decker, S. Y. Wang, H. X. Hui, J. C. Kappes, X. Y. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong, and G. M. Shaw. 2003. Antibody neutralization and escape by HIV-1. Nature 422:307-312.

33. Kwong, P. D., M. L. Doyle, D. J. Casper, C. Cicala, S. A. Leavitt, S. Majeed, T. D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, H. Katinger, P. Parren, J. Robinson, D. Van Ryk, L. P. Wang, D. R. Burton, E. Freire, R. Wyatt, J. Sodroski, W. A. Hendrickson, and J. Arthos. 2002. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature 420:678-682.
34. Moulaei, T., S. R. Shenoy, B. Giomarelli, C. Thomas, J. B. McMahon, Z. Dauter, B. R. O'Keefe, and A. Wlodawer. 2010. Monomerization of Viral Entry Inhibitor Griffithsin Elucidates the Relationship between Multivalent Binding to Carbohydrates and anti-HIV Activity. Structure 18:1104-1115.
35. Virtue, E. R., G. A. Marsh, and L.-F. Wang. 2009. Paramyxoviruses infecting humans: the old, the new and the unknown. Future Microbiology 4:537-554.
36. Nagai, Y., and T. Yoshida. 1984. VIRAL PATHOGENESIS MECHANISM OF ACUTE AND PERSISTENT INFECTIONS WITH PARAMYXOVIRUSES. Nagoya Journal of Medical Science 46:1-18.
37. Gogoi, P., K. Ganar, and S. Kumar. Avian Paramyxovirus: A Brief Review.
38. Chant, K., R. Chan, M. Smith, D. E. Dwyer, P. Kirkland, and N. S. W. E. Grp. 1998. Probable human infection with a newly described virus in the family Paramyxoviridae. Emerging Infectious Diseases 4:273-275.
39. Mahy, B. W. J. 2011. The Biology of Paramyxoviruses Introduction.
40. Tong, A. S. W., K. L. Hon, Y. C. K. Tsang, R. W. Y. Chan, C. C. Chan, T. F. Leung, and P. K. S. Chan. 2016. Paramyxovirus Infection: Mortality and Morbidity in a Pediatric Intensive Care Unit. J Trop Pediatr.
41. Law, B. J., V. Decarvalho, D. Scheifele, J. Langley, N. Macdonald, J. Simons, J. McDonald, E. Wang, I. Mitchell, and O. Hammerberg. 1993. RESPIRATORY SYNCYTIAL VIRUS-INFECTIONS IN HOSPITALIZED CANADIAN CHILDREN - REGIONAL DIFFERENCES IN PATIENT POPULATIONS AND MANAGEMENT-PRACTICES. Pediatric Infectious Disease Journal 12:659-663.

42. Null, D., C. Bimle, L. Weisman, K. Johnson, J. Steichen, S. Singh, E. Wang, E. Asztalos, A. M. Loeffler, P. H. Azimi, J. M. Lieberman, N. E. O'Donnell, R. J. Cooke, K. McCormick, W. Koo, M. Hammami, A. D. Milner, P. Gaon, S. Nachman, K. P. Tarpey, P. J. Sanchez, R. S. Broyles, D. Bratcher, M. V. Ball, F. J. Duda, P. M. DeCuir, B. Pollara, L. S. Nelson, M. Balbus, M. J. Schultz, B. E. Chipps, L. B. Givner, M. O'Shea, M. Everard, K. Pfeffer, A. J. Page, P. H. Dennehy, J. Modlin, T. Rhodes, N. DeVincenzo, B. Nickerson, A. Arrieta, F. D. Boucher, R. E. Keeney, T. E. Young, J. C. Stevens, R. Ariagno, M. Adams, M. J. Polak, S. K. Lynch, J. S. Gerdes, M. Kuba, M. Aouthmany, K. LaMar, G. Y. W. Chang, M. M. Shelton, S. K. W. Hadeed, U. Vasan, A. Hennessy, R. Yogev, R. C. Welliver, D. Tristram, C. Albin, T. T. Jefferson, G. D. Purdy, C. M. Buckner, J. Schlessel, C. Sia, B. Tan, K. Sankaran, C. J. Morley, D. K. White, H. C. Meissner, I. Frantz, S. A. Desai, C. W. Stanley, R. Inwood, L. E. Solecki, E. Wald, K. Smail, R. E. Fox, V. Taciak, C. L. Park, D. Vidyasagar, G. Redding, D. E. Mayock, P. D. Reuman, E. M. Bifano, B. Estrada, M. Y. Mancao, B. Hook, G. McDavid, J. L. Rowen, J. A. Patel, J. Robinson, B. Lee, W. Rodriguez, J. Arrobio, J. R. Hocker, C. McConnell, G. Piedimonte, I. Sosenko, B. Patel, S. Shervinski, P. E. Stobie, K. Perea, S. A. Chartrand, M. C. Wilson, R. de Lemos, R. Ramanathan, B. A. Barnett, S. R. Luber, W. V. Raszka, D. S. Holsclaw, D. L. Klein, B. J. Law, M. J. Balsan, B. H. Douglass, B. P. O'Sullivan, R. Spaulding, R. B. Van Dyke, A. Merza, J. O. Hendley, R. J. Boyle, P. A. Hughes, M. J. Horgan, R. C. Maynard, K. Teufert, M. Majure, J. A. George, D. R. Kuerschner, V. Ghai, D. Thomas, N. MacDonald, T. Kovesi, M. Blayney, C. S. Mani, V. H. San Joaquin, L. Roberts, A. Wiznia, M. Rosenberg, P. Karna, D. L. Murray, W. Lenney, S. Clayton, D. G. Oelberg, M. M. Reininger, S. C. Eppes, J. A. Childs, W. C. Gruber, T. R. Hazinski, E. A. Steinberg, L. C. Lopez, G. R. Elliott, J. R. Groothuis, E. Simoes, A. Hakim, F. Mimouni, L. Rubin, S. K. Sood, H. F. Sadiq, T. G. Marshall, D. Miller, M. R. Drayton, S. O'Neill, P. Chetcuti, D. L. Trupp, J. Heart, E. R. Cooper, E. R. Brown, A. Chetty, T. B. Rice, D. Rupar, C. T. Cho, R. D. Leff, S. D. Levine, J. K. Kolls, M. Hall, S. L. Smith, L. Schwartz, R. J. Lemen, C. B. Hall, C. E. Long, H. Panitch, S. M. Kolb, J. L. Colombo, C. G. Judy, B. L. Golembe, J. D. Anderson, J. McDonald, D. McCormack, D. P. Ruggerie, C. Triplett, M. W. Odom, G. Lopez-Cox, M. H. Sawyer, J. D. Connor, J. E. Fergie, K. Purcell, A. D. Kantak, D. M. Fihe, H. D. Davies, L. Mitchell, K. N. S. Subramanian, Y. Smith, E. B. St John, J. W. Stolz, C. Sheils, F. Cox, W. Foshee, R. Diaz, S. Coonce, H. L. Keyserling, C. B. Padrick, C. Lamprecht, F. R. Livingston, J. M. Langley, P. Weller, G. J. Cropp, A. Sola, M. L. Kumar, C. D. Lapin, P. Carlisle, R. Martz, M. Radetsky, S. Midani, M. H. Rathore, E. J. Riff, G. F. Shay, E. Hogvall, D. W. Russell, A. H. Thomson, S. M. Lawrey, K. Hardy, K. Harvey, R. Turner, E. O. Cox, A. Dana, A. Madan, J. Wallace, D. C. Stevens, B. Asmar, N. A. Selewski, J. Kelly, T. Klerr, S. Spruill, E. M. Conner, D. Carlin, F. H. Top, and I.-R. S. V. S. G. The. 1998. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. Pediatrics 102:531-537.

43. Meissner, H. C., R. C. Welliver, S. A. Chartrand, B. J. Law, L. E. Weisman, H. L. Dorkin, and W. J. Rodriguez. 1999. Immunoprophylaxis with palivizumab, a humanized respiratory syncytial virus monoclonal antibody, for prevention of respiratory syncytial virus infection in high risk infants: a consensus opinion. Pediatric Infectious Disease Journal 18:223-231.
44. Law, B. J., X. Carbonell-Estrany, and E. A. F. Simoes. 2002. An update on respiratory syncytial virus epidemiology: a developed country perspective. Respiratory Medicine 96:S1-S7.
45. Feltes, T. F., A. K. Cabalka, C. Meissner, F. M. Piazza, D. A. Carlin, F. H. Top, E. M. Connor, H. M. Sondheimer, and G. Cardiac Synagis Study. 2003. Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease. Journal of Pediatrics 143:532-540.
46. Falsey, A. R., P. A. Hennessey, M. A. Formica, C. Cox, and E. E. Walsh. 2005. Respiratory syncytial virus infection in elderly and high-risk adults. New England Journal of Medicine 352:1749-1759.
47. Hall, C. B., G. A. Weinberg, M. K. Iwane, A. K. Blumkin, K. M. Edwards, M. A. Staat, P. Auinger, M. R. Griffin, K. A. Poehling, D. Erdman, C. G. Grijalva, Y. W. Zhu, and P. Szilagyi. 2009. The Burden of Respiratory Syncytial Virus Infection in Young Children. New England Journal of Medicine 360:588-598.
48. Hon, K.-L., and A. K. C. Leung. 2009. Severe childhood respiratory viral infections. Advances in pediatrics 56:47-73.
49. Paget, S. P., D. N. Andresen, A. M. Kesson, and J. R. Egan. 2011. Comparison of human metapneumovirus and respiratory syncytial virus in children admitted to a paediatric intensive care unit. Journal of Paediatrics and Child Health 47:737-741.
50. Hon, K. L., T. F. Leung, W. Y. Cheng, N. M. W. Ko, W. K. Tang, W. W. Wong, W. H. P. Yeung, and P. K. S. Chan. 2012. Respiratory syncytial virus morbidity, premorbid factors, seasonality, and implications for prophylaxis. Journal of Critical Care 27:464-468.
51. Leung, T. F., D. S. Y. Lam, T. Y. Miu, K. L. Hon, C. S. K. Chau, S. W. Ku, R. S. Y. Lee, P. Y. Chow, W. K. Chiu, D. K. K. Ng, and R. Hong Kong Soc Paediat. 2014. Epidemiology and risk factors for severe respiratory syncytial virus infections requiring pediatric intensive care admission in Hong Kong children. Infection 42:343-350.
52. Harnden, A. 2005. Human metapneumovirus. British Journal of General Practice 55:84-85.
53. Martin, E. T., J. Kuypers, J. Heugel, and J. A. Englund. 2008. Clinical disease and viral load in children infected with respiratory syncytial virus or human metapneumovirus. Diagnostic Microbiology and Infectious Disease 62:382-388.
54. Principi, N., and S. Esposito. 2014. Paediatric human metapneumovirus infection: Epidemiology, prevention and therapy. J Clin Virol 59:141-147.
55. Eaton, B. T., C. C. Broder, D. Middleton, and L. F. Wang. 2006. Hendra and Nipah viruses: different and dangerous. Nature Reviews Microbiology 4:23-35.
56. Chua, K. B. 2003. Nipah virus outbreak in Malaysia. J Clin Virol 26:265-275.

57. Wong, S. C., M. H. Ooi, A. R. Abdullah, S. Y. Wong, S. Krishnan, P. H. Tio, P. C. Pek, B. F. Lai, A. Mohan, J. Muhi, A. Kiyu, M. T. Arif, and M. J. Cardosa. 2008. A decade of Japanese encephalities surveillance in Sarawak, Malaysia: 1997-2006. Tropical Medicine & International Health 13:52-55.
58. Luby, S. P., M. Rahman, M. J. Hossain, L. S. Blum, M. M. Husain, E. Gurley, R. Khan, B.-N. Ahmed, S. Rahman, N. Nahar, E. Kenah, J. A. Comer, and T. G. Ksiazek. 2006. Foodborne transmission of Nipah virus, Bangladesh. Emerging Infectious Diseases 12:1888-1894.
59. Luby, S. P., M. J. Hossain, E. S. Gurley, B.-N. Ahmed, S. Banu, S. U. Khan, N. Homaira, P. A. Rota, P. E. Rollin, J. A. Comer, E. Kenah, T. G. Ksiazek, and M. Rahman. 2009. Recurrent Zoonotic Transmission of Nipah Virus into Humans, Bangladesh, 2001-2007. Emerging Infectious Diseases 15:1229-1235.
60. Bose, S., A. S. Song, T. S. Jardetzky, and R. A. Lamb. 2014. Fusion Activation through Attachment Protein Stalk Domains Indicates a Conserved Core Mechanism of Paramyxovirus Entry into Cells. Journal of Virology 88:3925-3941.
61. Iorio, R. M., V. R. Melanson, and P. J. Mahon. 2009. Glycoprotein interactions in paramyxovirus fusion. Future Virology 4:335-351.
62. Jardetzky, T. S., and R. A. Lamb. 2014. Activation of paramyxovirus membrane fusion and virus entry. Current Opinion in Virology 5:24-33.
63. Palgen, J.-L., E. M. Jurgens, A. Moscona, M. Porotto, and L. M. Palermo. 2015. Unity in diversity: shared mechanism of entry among paramyxoviruses. Progress in molecular biology and translational science 129:1-32.
64. Plemper, R. K., M. A. Brindley, and R. M. Iorio. 2011. Structural and Mechanistic Studies of Measles Virus Illuminate Paramyxovirus Entry. Plos Pathogens 7.
65. Steffen, D. L., K. Xu, D. B. Nikolov, and C. C. Broder. 2012. Henipavirus Mediated Membrane Fusion, Virus Entry and Targeted Therapeutics. Viruses-Basel 4:280-308.
66. Negrete, O. A., E. L. Levroney, H. C. Aguilar, A. Bertolotti-Ciarlet, R. Nazarian, S. Tajyar, and B. Lee. 2005. EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. Nature 436:401-405.
67. Negrete, O. A., D. Chu, H. C. Aguilar, and B. Lee. 2007. Single amino acid changes in the Nipah and Hendra virus attachment glycoproteins distinguish EphrinB2 from EphrinB3 usage. Journal of Virology 81:10804-10814.
68. Negrete, O. A., M. C. Wolf, H. C. Aguilar, S. Enterlein, W. Wang, E. Muehlberger, S. V. Su, A. Bertolotti-Ciarlet, R. Flick, and B. Lee. 2006. Two key residues in ephrinB3 are critical for its use as an alternative receptor for Nipah virus. Plos Pathogens 2:78-86.
69. Liu Q, Stone J A, Bradel-Tretheway B, Dabundo J, Benavides Montano J A, Santos-Montanez J, Biering S B, Nicola A V, Iorio R M, Lu X, and A. H. C. 2013. Unraveling a Three-Step Spatiotemporal Mechanism of Triggering of Receptor-Induced Nipah Virus Fusion and Cell Entry. PLoS Pathogen 9.
70. Liu, Q., B. Bradel-Tretheway, A. I. Monreal, J. P. Saludes, X. Lu, A. V. Nicola, and H. C. Aguilar. 2015. Nipah Virus Attachment Glycoprotein Stalk C-Terminal Region Links Receptor Binding to Fusion Triggering. Journal of Virology 89:1838-1850.

71. Talekar, A., I. DeVito, Z. Salah, S. G. Palmer, A. Chattopadhyay, J. K. Rose, R. Xu, I. A. Wilson, A. Moscona, and M. Porotto. 2013. Identification of a Region in the Stalk Domain of the Nipah Virus Receptor Binding Protein That Is Critical for Fusion Activation. Journal of Virology 87:10980-10996.
72. Brindley, M. A., M. Takeda, P. Plattet, and R. K. Plemper. 2012. Triggering the measles virus membrane fusion machinery. Proceedings of the National Academy of Sciences of the United States of America 109:E3018-E3027.
73. Ader, N., M. Brindley, M. Avila, C. Orvell, B. Horvat, G. Hiltensperger, J. Schneider-Schaulies, M. Vandevelde, A. Zurbriggen, R. K. Plemper, and P. Plattet. 2013. Mechanism for Active Membrane Fusion Triggering by Morbillivirus Attachment Protein. Journal of Virology 87:314-326.
74. Lee, J. K., A. Prussi, T. Paal, L. K. White, J. P. Snyder, and R. K. Plemper. 2008. Functional interaction between paramyxovirus fusion and attachment proteins. Journal of Biological Chemistry 283:16561-16572.
75. Samuel, O., and Y. Shai. 2001. Participation of two fusion peptides in measles virus-induced membrane fusion: Emerging similarity with other paramyxoviruses. Biochemistry 40:1340-1349.
76. Lee, B., and Z. A. Ataman. 2011. Modes of paramyxovirus fusion: a Henipavirus perspective. Trends in Microbiology 19:389-399.
77. Aguilar, H. C., Z. A. Ataman, V. Aspericueta, A. Q. Fang, M. Stroud, O. A. Negrete, R. A. Kammerer, and B. Lee. 2009. A Novel Receptor-induced Activation Site in the Nipah Virus Attachment Glycoprotein (G) Involved in Triggering the Fusion Glycoprotein (F). Journal of Biological Chemistry 284:1628-1635.
78. Ayllon, J., E. Villar, and I. Munoz-Barroso. 2010. Mutations in the ectodomain of newcastle disease virus fusion protein confer a hemagglutinin-neuraminidase-independent phenotype. J Virol 84:1066-1075.
79. Bonaparte, M. I., A. S. Dimitrov, K. N. Bossart, G. Crameri, B. A. Mungal, K. A. Bishop, V. Choudhry, D. S. Dimitrov, L. F. Wang, B. T. Eaton, and C. C. Broder. 2005. Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. Proceedings of the National Academy of Sciences of the United States of America 102:10652-10657.
80. Bowden, T. A., M. Crispin, D. J. Harvey, E. Y. Jones, and D. I. Stuart. 2010. Dimeric Architecture of the Hendra Virus Attachment Glycoprotein: Evidence for a Conserved Mode of Assembly. Journal of Virology 84:6208-6217.
81. Brindley, M. A., R. Suter, I. Schestak, G. Kiss, E. R. Wright, and R. K. Plemper. 2013. A Stabilized Headless Measles Virus Attachment Protein Stalk Efficiently Triggers Membrane Fusion. Journal of Virology 87:11693-11703.
82. Langedijk, J. P. M., F. J. Daus, and J. T. VanOirschot. 1997. Sequence and structure alignment of Paramyxoviridae attachment proteins and discovery of enzymatic activity for a morbillivirus hemagglutinin. Journal of Virology 71:6155-6167.
83. Navaratnarajah, C. K., V. H. J. Leonard, and R. Cattaneo. 2009. Measles Virus Glycoprotein Complex Assembly, Receptor Attachment, and Cell Entry. In Measles: History and Basic Biology. D. E. Griffin, and M. B. A. Oldstone, editors. 59-76.

84. Navaratnarajah, C. K., N. Oezguen, L. Rupp, L. Kay, V. H. J. Leonard, W. Braun, and R. Cattaneo. 2011. The heads of the measles virus attachment protein move to transmit the fusion-triggering signal. Nature Structural & Molecular Biology 18:128-U183.
85. Xu, K., Y.-P. Chan, K. R. Rajashankar, D. Khetawat, L. Yan, M. V. Kolev, C. C. Broder, and D. B. Nikolov. 2012. New Insights into the Hendra Virus Attachment and Entry Process from Structures of the Virus G Glycoprotein and Its Complex with Ephrin-B2. Plos One 7.
86. Xu, K., K. R. Rajashankar, Y. P. Chan, J. P. Himanen, C. C. Broder, and D. B. Nikolov. 2008. Host cell recognition by the henipaviruses: Crystal structures of the Nipah G attachment glycoprotein and its complex with ephrin-B3. Proceedings of the National Academy of Sciences of the United States of America 105:9953-9958.
87. Bossart, K. N., G. Crameri, A. S. Dimitrov, B. A. Mungall, Y. R. Feng, J. R. Patch, A. Choudhary, L. F. Wang, B. T. Eaton, and C. C. Broder. 2005. Receptor binding, fusion inhibition, and induction of cross-reactive neutralizing antibodies by a soluble G glycoprotein of Hendra virus. Journal of Virology 79:6690-6702.
88. Chang, A., and R. E. Dutch. 2012. Paramyxovirus Fusion and Entry: Multiple Paths to a Common End. Viruses-Basel 4:613-636.
89. Zeltina, A., T. A. Bowden, and B. Lee. 2016. Emerging Paramyxoviruses: Receptor Tropism and Zoonotic Potential. PLoS pathogens 12:e1005390-e1005390.
90. Bowden, T. A., A. R. Aricescu, R. J. C. Gilbert, J. M. Grimes, E. Y. Jones, and D. I. Stuart. 2008. Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. Nature Structural & Molecular Biology 15:567-572.
91. Welch, B. D., P. Yuan, S. Bose, C. A. Kors, R. A. Lamb, and T. S. Jardetzky. 2013. Structure of the Parainfluenza Virus 5 (PIV5) Hemagglutinin-Neuraminidase (HN) Ectodomain. Plos Pathogens 9.
92. Yuan, P., K. A. Swanson, G. P. Leser, R. G. Paterson, R. A. Lamb, and T. S. Jardetzky. 2011. Structure of the Newcastle disease virus hemagglutinin-neuraminidase (HN) ectodomain reveals a four-helix bundle stalk. Proceedings of the National Academy of Sciences of the United States of America 108:14920-14925.
93. Leighty, R. E., and S. Varma. 2013. Quantifying Changes in Intrinsic Molecular Motion Using Support Vector Machines. Journal of Chemical Theory and Computation 9:868-875.
94. Varma, S., M. Botlani, and R. E. Leighty. 2014. Discerning intersecting fusion-activation pathways in the Nipah virus using machine learning. Proteins-Structure Function and Bioinformatics.
95. Dutta, P., M. Botlani, and S. Varma. 2014. Water dynamics at protein-protein interfaces: A molecular dynamics study of virus-host receptor complexes. Submitted.
96. Cooper, A., and D. T. F. Dryden. 1984. ALLOSTERY WITHOUT CONFORMATIONAL CHANGE - A PLAUSIBLE MODEL. European Biophysics Journal with Biophysics Letters 11:103-109.

97. Cui, Q., and M. Karplus. 2008. Allostery and cooperativity revisited. Protein Science 17:1295-1307.
98. Nussinov, R., and C.-J. Tsai. 2013. Allostery in Disease and in Drug Discovery. Cell 153:293-305.
99. Taylor, S. S., R. Ilouz, P. Zhang, and A. P. Kornev. 2012. Assembly of allosteric macromolecular switches: lessons from PKA. Nature Reviews Molecular Cell Biology 13:646-658.
100. Changeux, J.-P., P.-J. Corringer, and U. Maskos. 2015. The nicotinic acetylcholine receptor: From molecular biology to cognition. Neuropharmacology 96:135-136.
101. Kalodimos, C. G., and Nyas. 2012. Protein function and allostery: a dynamic relationship. Blavatnik Awards for Young Scientists 2011 1260:81-86.
102. Arkhipov, A., Y. Shan, R. Das, N. F. Endres, M. P. Eastwood, D. E. Wemmer, J. Kuriyan, and D. E. Shaw. 2013. Architecture and Membrane Interactions of the EGF Receptor. Cell 152:557-569.
103. Endres, N. F., R. Das, A. W. Smith, A. Arkhipov, E. Kovacs, Y. Huang, J. G. Pelton, Y. Shan, D. E. Shaw, D. E. Wemmer, J. T. Groves, and J. Kuriyan. 2013. Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. Cell 152:543-556.
104. Changeux, J.-P. 2013. 50 years of allosteric interactions: the twists and turns of the models. Nature Reviews Molecular Cell Biology 14:819-829.
105. Brooks, C. L., M. Karplus, and B. M. Pettitt. 1988. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. Wiley, New York.
106. Karplus, M., and G. A. Petsko. 1990. MOLECULAR-DYNAMICS SIMULATIONS IN BIOLOGY. Nature 347:631-639.
107. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. Nature Structural Biology 9:646-652.
108. Karplus, M., and J. Kuriyan. 2005. Molecular dynamics and protein function. Proceedings of the National Academy of Sciences of the United States of America 102:6679-6685.
109. Wang, W., O. Donini, C. M. Reyes, and P. A. Kollman. 2001. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. Annual Review of Biophysics and Biomolecular Structure 30:211-243.
110. Hansson, T., C. Oostenbrink, and W. F. van Gunsteren. 2002. Molecular dynamics simulations. Current Opinion in Structural Biology 12:190-196.
111. van Gunsteren, W., D. Bakowies, R. Burgi, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, A. Glattli, T. Hansson, C. Oostenbrink, C. Peter, J. Pitera, L. Schuler, T. Soares, and H. B. Yu. 2001. Molecular dynamics simulation of biomolecular systems. Chimia 55:856-860.
112. Van Gunsteren, W. F. 1988. THE ROLE OF COMPUTER SIMULATION TECHNIQUES IN PROTEIN ENGINEERING. Protein Engineering 2:5-14.

113. Ahmed, M. H., F. Spyrakis, P. Cozzini, P. K. Tripathi, A. Mozzarelli, J. N. Scarsdale, M. A. Safo, and G. E. Kellogg. 2011. Bound Water at Protein-Protein Interfaces: Partners, Roles and Hydrophobic Bubbles as a Conserved Motif. Plos One 6.
114. Anandakrishnan, R., A. Drozdetski, R. C. Walker, and A. V. Onufriev. 2015. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. Biophysical journal 108:1153-1164.
115. Chen, J., C. L. Brooks, III, and J. Khandogin. 2008. Recent advances in implicit solvent-based methods for biomolecular simulations. Current Opinion in Structural Biology 18:140-148.
116. Wang, T., and R. C. Wade. 2003. Implicit solvent models for flexible protein-protein docking by molecular dynamics simulation. Proteins-Structure Function and Genetics 50:158-169.
117. Yuan, P., G. P. Leser, B. Demeler, R. A. Lamb, and T. S. Jardetzky. 2008. Domain architecture and oligomerization properties of the paramyxovirus PIV 5 hemagglutinin-neuraminidase (HN) protein. Virology 378:282-291.
118. Ferreira, L., I. Munoz-Barroso, F. Marcos, V. L. Shnyrov, and E. Villar. 2004. Sialidase, receptor-binding and fusion-promotion activities of Newcastle disease virus haemagglutinin-neuraminidase glycoprotein: a mutational and kinetic study. J Gen Virol 85:1981-1988.
119. Munoz-Barroso, I., C. Cobaleda, G. Zhadan, V. Shnyrov, and E. Villar. 1997. Dynamic properties of Newcastle Disease Virus envelope and their relations with viral hemagglutinin-neuraminidase membrane glycoprotein. Biochim Biophys Acta 1327:17-31.
120. Papadopoulos, J. S., and R. Agarwala. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 23:1073-1079.
121. Mayo, M. A. 2002. Virus taxonomy - Houston 2002. Archives of virology 147:1071-1076.
122. 2016. Respiratory Syncytial Virus Infection (RSV): Trends and Surveillance.
123. Organization, W. H. 2016. Media Center Measles Factsheet.
124. Patti J. Miller, D., PhD. 2014. Merck Veterinary Manual: Newcastle disease in poultry (Avian pneumoencephalitis, exotic or velogenic Newcastle disease).
125. Hines, N. L., and C. L. Miller. 2012. Avian paramyxovirus serotype-1: a review of disease distribution, clinical symptoms, and laboratory diagnostics. Veterinary medicine international 2012:708216-708216.
126. Kapczynski, D. R., C. L. Afonso, and P. J. Miller. 2013. Immune responses of poultry to Newcastle disease virus. Developmental and Comparative Immunology 41:447-453.
127. Kaleta, E. F., and C. Baldauf. 1988. Newcastle disease in free-living and pet birds. Developments in Veterinary Virology 8:197-246.
128. Kitchen, A., L. A. Shackelton, and E. C. Holmes. 2011. Family level phylogenies reveal modes of macroevolution in RNA viruses. Proceedings of the National Academy of Sciences of the United States of America 108:238-243.
129. Marsh, G. A., and L.-F. Wang. 2012. Hendra and Nipah viruses: why are they so deadly? Current Opinion in Virology 2:242-247.

130. Bellini, W. J., B. H. Harcourt, N. Bowden, and P. A. Rota. 2005. Nipah virus: An emergent paramyxovirus causing severe encephalitis in humans. Journal of Neurovirology 11:481-487.
131. Chua, K. B., W. J. Bellini, P. A. Rota, B. H. Harcourt, A. Tamin, S. K. Lam, T. G. Ksiazek, P. E. Rollin, S. R. Zaki, W. J. Shieh, C. S. Goldsmith, D. J. Gubler, J. T. Roehrig, B. Eaton, A. R. Gould, J. Olson, H. Field, P. Daniels, A. E. Ling, C. J. Peters, L. J. Anderson, and B. W. J. Mahy. 2000. Nipah virus: A recently emergent deadly paramyxovirus. Science 288:1432-1435.
132. Wild, T. F. 2009. Henipaviruses: A new family of emerging Paramyxoviruses. Pathol Biol 57:188-196.
133. 2014. Centers for Disease Control and Prevention : Hendra virus disease.
134. Nipah virus outbreaks in the WHO South-East Asia Region.
135. Nipah virus infection.
136. 2012. Institute of Epidemiology, Disease Control and Research : Outbreak.
137. Lam, S. K. 2003. Nipah virus - a potential agent of bioterrorism? Antiviral Research 57:113-119.
138. Epstein, J. H., V. Prakash, C. S. Smith, P. Daszak, A. B. McLaughlin, G. Meehan, H. E. Field, and A. A. Cunningham. 2008. Henipavirus infection in fruit bats (Pteropus giganteus), India. Emerging Infectious Diseases 14:1309-1311.
139. Hsu, V. P., M. J. Hossain, U. D. Parashar, M. M. Ali, T. G. Ksiazek, I. Kuzmin, M. Niezgoda, C. Rupprecht, J. Bresee, and R. F. Breiman. 2004. Nipath virus encephalitis reemergence, Bangladesh. Emerging Infectious Diseases 10:2082-2087.
140. Iehle, C., G. Razafitrimo, J. Razainirina, N. Andriaholinirina, S. M. Goodman, C. Faure, M.-C. Georges-Courbot, D. Rousset, and J.-M. Reynes. 2007. Henipavirus and Tioman virus antibodies in pteropodid bats, Madagascar. Emerging Infectious Diseases 13:159-161.
141. Reynes, J. M., D. Counor, S. Ong, C. Faure, V. Seng, S. Molia, J. Walston, M. C. Georges-Courbot, V. Deubel, and J. L. Sarthou. 2005. Nipah virus in lyle's flying foxes, Cambodia. Emerging Infectious Diseases 11:1042-1047.
142. Wacharapluesadee, S., B. Lumlertdacha, K. Boongird, S. Wanghongsa, L. Chanhome, P. Rollin, P. Stockton, C. E. Rupprecht, T. G. Ksiazek, and T. Hemachudha. 2005. Bat Nipah virus, Thailand. Emerging Infectious Diseases 11:1949-1951.
143. Luby, S. P., E. S. Gurley, and M. J. Hossain. 2009. Transmission of Human Infection with Nipah Virus. Clinical Infectious Diseases 49:1743-1748.
144. Tanabayashi, K., and R. W. Compans. 1996. Functional interaction of paramyxovirus glycoproteins: Identification of a domain in sendai virus HN which promotes cell fusion. Journal of Virology 70:6112-6118.
145. Rodriguez, J. J., and C. M. Horvath. 2004. Host evasion by emerging paramyxoviruses: Hendra Virus and Nipah Virus V proteins inhibit interferon signaling. Viral Immunol 17:210-219.
146. Bose, S., B. D. Welch, C. A. Kors, P. Yuan, T. S. Jardetzky, and R. A. Lamb. 2011. Structure and Mutagenesis of the Parainfluenza Virus 5 Hemagglutinin-Neuraminidase Stalk Domain Reveals a Four-Helix Bundle and the Role of the Stalk in Fusion Promotion. Journal of Virology 85:12855-12866.

147. Bishop, K. A., T. S. Stantchev, A. C. Hickey, D. Khetawat, K. N. Bossart, V. Krasnoperov, P. Gill, Y. R. Feng, L. Wang, B. T. Eaton, L.-F. Wang, and C. C. Broder. 2007. Identification of hendra virus g glycoprotein residues that are critical for receptor binding. Journal of Virology 81:5893-5901.
148. Frisen, J., J. Holmberg, and M. Barbacid. 1999. Ephrins and their Eph receptors: multitalented directors of embryonic development. Embo Journal 18:5159-5165.
149. Wang, H. U., Z. F. Chen, and D. J. Anderson. 1998. Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4. Cell 93:741-753.
150. Shin, D., G. Garcia-Cardena, S. I. Hayashi, S. Gerety, T. Asahara, G. Stavrakis, J. Isner, J. Folkman, M. A. Gimbrone, and D. J. Anderson. 2001. Expression of ephrinB2 identifies a stable genetic difference between arterial and venous vascular smooth muscle as well as endothelial cells, and marks subsets of microvessels at sites of adult neovascularization. Developmental Biology 230:139-150.
151. Gale, N. W., P. Baluk, L. Pan, M. Kwan, J. Holash, T. M. DeChiara, D. M. McDonald, and G. D. Yancopoulos. 2001. Ephrin-B2 selectively marks arterial vessels and neovascularization sites in the adult, with expression in both endothelial and smooth-muscle cells. Developmental Biology 230:151-160.
152. Ksiazek, T. G., P. A. Rota, and P. E. Rollin. 2011. A review of Nipah and Hendra viruses with an historical aside. Virus Research 162:173-183.
153. Escaffre, O., V. Borisevich, and B. Rockx. 2013. Pathogenesis of Hendra and Nipah virus infection in humans. Journal of Infection in Developing Countries 7:308-311.
154. Flanagan, J. G., and P. Vanderhaeghen. 1998. The ephrins and Eph receptors in neural development. Annual Review of Neuroscience 21:309-345.
155. Gale, N. W., and G. D. Yancopoulos. 1999. Growth factors acting via endothelial cell-specific receptor tyrosine kinases: VEGFs, angiopoietins, and ephrins in vascular development. Genes & Development 13:1055-1066.
156. Himanen, J. P., and D. B. Nikolov. 2003. Eph receptors and ephrins. International Journal of Biochemistry & Cell Biology 35:130-134.
157. Himanen, J. P., and D. B. Nikolov. 2003. Eph signaling: a structural view. Trends in Neurosciences 26:46-51.
158. Lee, B., Z. A. Ataman, and L. Jin. 2008. Evil versus 'eph-ective' use of ephrin-B2. Nature Structural & Molecular Biology 15:540-542.
159. Nikolov, D. B., K. Xu, and J. P. Himanen. 2013. Eph/ephrin recognition and the role of Eph/ephrin clusters in signaling initiation. Biochimica et biophysica acta 1834:2160-2165.
160. Maar, D., B. Harmon, D. Chu, B. Schulz, H. C. Aguilar, B. Lee, and O. A. Negrete. 2012. Cysteines in the Stalk of the Nipah Virus G Glycoprotein Are Located in a Distinct Subdomain Critical for Fusion Activation. Journal of Virology 86:6632-6642.
161. Biering, S. B., A. Huang, A. T. Vu, L. R. Robinson, B. Bradel-Tretheway, E. Choi, B. Lee, and H. C. Aguilar. 2012. N-Glycans on the Nipah Virus Attachment Glycoprotein Modulate Fusion and Viral Entry as They Protect against Antibody Neutralization. Journal of Virology 86:11991-12002.

162. Paal, T., M. A. Brindley, C. St Clair, A. Prussia, D. Gaus, S. A. Krumm, J. P. Snyder, and R. K. Plemper. 2009. Probing the Spatial Organization of Measles Virus Fusion Complexes. Journal of Virology 83:10480-10493.
163. Smith, E. C., A. Popa, A. Chang, C. Masante, and R. E. Dutch. 2009. Viral entry mechanisms: the increasing diversity of paramyxovirus entry. Febs Journal 276:7217-7227.
164. Aguilar, H. C., K. A. Matreyek, C. M. Filone, S. T. Hashimi, E. L. Levroney, O. A. Negrete, A. Bertolotti-Ciarlet, D. Y. Choi, I. McHardy, J. A. Fulcher, S. V. Su, M. C. Wolf, L. Kohatsu, L. G. Baum, and B. Lee. 2006. N-glycans on Nipah virus fusion protein protect against neutralization but reduce membrane fusion and viral entry. Journal of Virology 80:4878-4889.
165. Stone, J. A., A. V. Nicola, L. G. Baum, and H. C. Aguilar. 2016. Multiple Novel Functions of Henipavirus O-glycans: The First O-glycan Functions Identified in the Paramyxovirus Family. PLoS pathogens 12:e1005445-e1005445.
166. Bishop, K. A., A. C. Hickey, D. Khetawat, J. R. Patch, K. N. Bossart, Z. Zhu, L.-F. Wang, D. S. Dimitrov, and C. C. Broder. 2008. Residues in the Stalk Domain of the Hendra Virus G Glycoprotein Modulate Conformational Changes Associated with Receptor Binding. Journal of Virology 82:11398-11409.
167. Plemper, R. K., A. L. Hammond, D. Gerlier, A. K. Fielding, and R. Cattaneo. 2002. Strength of envelope protein interaction modulates cytopathicity of measles virus. Journal of Virology 76:5051-5061.
168. Deng, R. T., Z. Y. Wang, A. M. Mirza, and R. M. Iorio. 1995. LOCALIZATION OF A DOMAIN ON THE PARAMYXOVIRUS ATTACHMENT PROTEIN REQUIRED FOR THE PROMOTION OF CELLULAR FUSION BY ITS HOMOLOGOUS FUSION PROTEIN SPIKE. Virology 209:457-469.
169. StoneHulslander, J., and T. G. Morrison. 1997. Detection of an interaction between the HN and F proteins in Newcastle disease virus-infected cells. Journal of Virology 71:6287-6295.
170. Stone-Hulslander, J., and T. G. Morrison. 1999. Mutational analysis of heptad repeats in the membrane-proximal region of Newcastle disease virus HN protein. Journal of Virology 73:3630-3637.
171. Porotto, M., Z. Salah, I. DeVito, A. Talekar, S. G. Palmer, R. Xu, I. A. Wilson, and A. Moscona. 2012. The Second Receptor Binding Site of the Globular Head of the Newcastle Disease Virus Hemagglutinin-Neuraminidase Activates the Stalk of Multiple Paramyxovirus Receptor Binding Proteins To Trigger Fusion. Journal of Virology 86:5730-5741.
172. Guillaume, V., H. Aslan, M. Ainouze, M. Guerbois, T. F. Wild, R. Buckland, and J. P. M. Langedijk. 2006. Evidence of a potential receptor-binding site on the Nipah virus G protein (NiV-G): Identification of globular head residues with a role in fusion promotion and their localization on an NiV-G structural model. Journal of Virology 80:7546-7554.
173. Melanson, V. R., and R. M. Iorio. 2006. Addition of N-glycans in the stalk of the newcastle disease virus HN protein blocks its interaction with the F protein and prevents fusion. Journal of Virology 80:623-633.

174. Monod, J., and F. Jacob. 1961. GENERAL CONCLUSIONS - TELEONOMIC MECHANISMS IN CELLULAR METABOLISM, GROWTH, AND DIFFERENTIATION. Cold Spring Harbor Symposia on Quantitative Biology 26:389-&.
175. Gunasekaran, K., B. Y. Ma, and R. Nussinov. 2004. Is allostery an intrinsic property of all dynamic proteins? Proteins-Structure Function and Bioinformatics 57:433-443.
176. Monod, J., J. P. Changeux, and F. Jacob. 1963. ALLOSTERIC PROTEINS AND CELLULAR CONTROL SYSTEMS. Journal of molecular biology 6:306-&.
177. Motlagh, H. N., J. O. Wrabl, J. Li, and V. J. Hilser. 2014. The ensemble nature of allostery. Nature 508:331-339.
178. Monod, J., J. Wyman, and J. P. Changeux. 1965. ON NATURE OF ALLOSTERIC TRANSITIONS - A PLAUSIBLE MODEL. Journal of molecular biology 12:88-&.
179. Ma, B. Y., T. Elkayam, H. Wolfson, and R. Nussinov. 2003. Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proceedings of the National Academy of Sciences of the United States of America 100:5772-5777.
180. Kern, D., and E. R. P. Zuiderweg. 2003. The role of dynamics in allosteric regulation. Current Opinion in Structural Biology 13:748-757.
181. Ottemann, K. M., W. Z. Xiao, Y. K. Shin, and D. E. Koshland. 1999. A piston model for transmembrane signaling of the aspartate receptor. Science 285:1751-1754.
182. Yu, E. W., and D. E. Koshland. 2001. Propagating conformational changes over long (and short) distances in proteins. Proceedings of the National Academy of Sciences of the United States of America 98:9517-9520.
183. Perutz, M. F. 1970. STEREOCHEMISTRY OF COOPERATIVE EFFECTS IN HAEMOGLOBIN. Nature 228:726-&.
184. Perutz, M. F., A. J. Wilkinson, M. Paoli, and G. G. Dodson. 1998. The stereochemical mechanism of the cooperative effects in hemoglobin revisited. Annual Review of Biophysics and Biomolecular Structure 27:1-34.
185. Petit, C. M., J. Zhang, P. J. Sapienza, E. J. Fuentes, and A. L. Lee. 2009. Hidden dynamic allostery in a PDZ domain. Proceedings of the National Academy of Sciences of the United States of America 106:18249-18254.
186. Popovych, N., S. Sun, R. H. Ebright, and C. G. Kalodimos. 2006. Dynamically driven protein allostery. Nature Structural & Molecular Biology 13:831-838.
187. Tzeng, S.-R., and C. G. Kalodimos. 2012. Protein activity regulation by conformational entropy. Nature 488:236-240.
188. Tzeng, S.-R., and C. G. Kalodimos. 2009. Dynamic activation of an allosteric regulatory protein. Nature 462:368-U139.
189. Freiburger, L. A., O. M. Baettig, T. Sprules, A. M. Berghuis, K. Auclair, and A. K. Mittermaier. 2011. Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. Nature Structural & Molecular Biology 18:288-U270.

190. Schrank, T. P., D. W. Bolen, and V. J. Hilser. 2009. Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. Proceedings of the National Academy of Sciences of the United States of America 106:16984-16989.
191. Uversky, V. N., C. J. Oldfield, and A. K. Dunker. 2005. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. Journal of Molecular Recognition 18:343-384.
192. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. Journal of molecular biology 293:321-331.
193. Tompa, P. 2011. Unstructural biology coming of age. Current Opinion in Structural Biology 21:419-425.
194. Sevcsik, E., A. J. Trexler, J. M. Dunn, and E. Rhoades. 2011. Allostery in a Disordered Protein: Oxidative Modifications to alpha-Synuclein Act Distally To Regulate Membrane Binding. Journal of the American Chemical Society 133:7152-7158.
195. Dror, R. O., R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw. 2012. Biomolecular Simulation: A Computational Microscope for Molecular Biology. Annual Review of Biophysics, Vol 41 41:429-452.
196. Xu, D., C. J. Tsai, and R. Nussinov. 1997. Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Engineering 10:999-1012.
197. Martinez-Julvez, M., O. Abian, S. Vega, M. Medina, and A. Velazquez-Campoy. 2012. Studying the Allosteric Energy Cycle by Isothermal Titration Calorimetry. Allostery: Methods and Protocols 796:53-70.
198. Strawn, R., T. Stockner, M. Melichercik, L. Jin, W.-F. Xue, J. Carey, and R. Ettrich. 2011. SYNERGY OF MOLECULAR DYNAMICS AND ISOTHERMAL TITRATION CALORIMETRY IN STUDIES OF ALLOSTERY. Methods in Enzymology: Biothermodynamics, Vol 492, Pt D 492:151-188.
199. Sturtevant, J. M. 1987. BIOCHEMICAL APPLICATIONS OF DIFFERENTIAL SCANNING CALORIMETRY. Annual Review of Physical Chemistry 38:463-488.
200. Shnyrov, V. L., G. G. Zhadan, C. Cobaleda, A. Sagrera, I. Munoz-Barroso, and E. Villar. 1997. A differential scanning calorimetric study of Newcastle disease virus: identification of proteins involved in thermal transitions. Arch Biochem Biophys 341:89-97.
201. Bruylants, G., J. Wouters, and C. Michaux. 2005. Differential scanning calorimetry in life science: Thermodynamics, stability, molecular recognition and application in drug design. Current Medicinal Chemistry 12:2011-2020.
202. Zhao, Q., Q. Lv, and H. Wang. 2014. Identification of Allosteric Nucleotide Sites of Tetramethylrhodamine-Labeled Aptamer for Noncompetitive Aptamer-Based Fluorescence Anisotropy Detection of a Small Molecule, Ochratoxin A. Analytical Chemistry 86:1238-1245.
203. Foda, Z. H., Y. Shan, E. T. Kim, D. E. Shaw, and M. A. Seeliger. 2015. A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. Nature Communications 6.

204. Zhang, D., Q. Zhao, B. Zhao, and H. Wang. 2012. Fluorescence Anisotropy Reduction of Allosteric Aptamer for Sensitive and Specific Protein Signaling. Analytical Chemistry 84:3070-3074.
205. Kelly, S. M., and N. C. Price. 2000. The Use of Circular Dichroism in the Investigation of Protein Structure and Function. Current Protein & Peptide Science 1:349-384.
206. Zayner, J. P., C. Antoniou, A. R. French, R. J. Hause, Jr., and T. R. Sosnick. 2013. Investigating Models of Protein Function and Allostery With a Widespread Mutational Analysis of a Light-Activated Protein. Biophysical journal 105:1027-1036.
207. Tzeng, S.-R., and C. G. Kalodimos. 2011. Protein dynamics and allostery: an NMR view. Current Opinion in Structural Biology 21:62-67.
208. Mittermaier, A., and L. E. Kay. 2006. Review - New tools provide new insights in NMR studies of protein dynamics. Science 312:224-228.
209. Smock, R. G., and L. M. Gierasch. 2009. Sending Signals Dynamically. Science 324:198-203.
210. Das, R., V. Esposito, M. Abu-Abed, G. S. Anand, S. S. Taylor, and G. Melacini. 2007. cAMP activation of PKA defines an ancient signaling mechanism. Proceedings of the National Academy of Sciences of the United States of America 104:93-98.
211. Masterson, L. R., A. Mascioni, N. J. Traaseth, S. S. Taylor, and G. Veglia. 2008. Allosteric cooperativity in protein kinase A. Proceedings of the National Academy of Sciences of the United States of America 105:506-511.
212. Manley, G., and J. P. Loria. 2012. NMR insights into protein allostery. Archives of Biochemistry and Biophysics 519:223-231.
213. Tzeng, S.-R., and C. G. Kalodimos. 2013. Allosteric inhibition through suppression of transient conformational states. Nature Chemical Biology 9:462-+.
214. Shi, L., and L. E. Kay. 2014. Tracing an allosteric pathway regulating the activity of the HslV protease. Proceedings of the National Academy of Sciences of the United States of America 111:2140-2145.
215. Dokholyan, N. V. 2016. Controlling Allosteric Networks in Proteins. Chem Rev.
216. Carlomagno, T. 2014. Present and future of NMR for RNA-protein complexes: A perspective of integrated structural biology. Journal of Magnetic Resonance 241:126-136.
217. Yu, H. T. 1999. Extending the size limit of protein nuclear magnetic resonance. Proceedings of the National Academy of Sciences of the United States of America 96:332-334.
218. Beveridge, D. L., and F. M. Dicapua. 1989. FREE-ENERGY VIA MOLECULAR SIMULATION - APPLICATIONS TO CHEMICAL AND BIOMOLECULAR SYSTEMS. Annual Review of Biophysics and Biophysical Chemistry 18:431-492.
219. Jorgensen, W. L. 2004. The many roles of computation in drug discovery. Science 303:1813-1818.
220. Wrabl, J. O., D. Shortle, and T. B. Woolf. 2000. Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: Molecular dynamics simulations of native and denatured staphylococcal nuclease. Proteins-Structure Function and Genetics 38:123-133.

221. Dobson, C. M., and M. Karplus. 1986. Internal motion of proteins: nuclear magnetic resonance measurements and dynamic simulations. Methods in enzymology 131:362-389.
222. Henzler-Wildman, K., and D. Kern. 2007. Dynamic personalities of proteins. Nature 450:964-972.
223. Steinhauser, M. O., and S. Hiermaier. 2009. A Review of Computational Methods in Materials Science: Examples from Shock-Wave and Polymer Physics. International Journal of Molecular Sciences 10:5135-5216.
224. Scheraga, H. A., M. Khalili, and A. Liwo. 2007. Protein-folding dynamics: Overview of molecular simulation techniques. Annual Review of Physical Chemistry 58:57-83.
225. Klepeis, J. L., K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw. 2009. Long-timescale molecular dynamics simulations of protein structure and function. Current Opinion in Structural Biology 19:120-127.
226. Nadler, W., A. T. Brunger, K. Schulten, and M. Karplus. 1987. MOLECULAR AND STOCHASTIC DYNAMICS OF PROTEINS. Proceedings of the National Academy of Sciences of the United States of America 84:7933-7937.
227. Bou-Rabee, N. 2014. Time Integrators for Molecular Dynamics. Entropy 16:138-162.
228. Attard, P. 2002. Stochastic molecular dynamics: A combined Monte Carlo and molecular dynamics technique for isothermal simulations. Journal of Chemical Physics 116:9616-9619.
229. Chandrasekhar, S. 1943. Stochastic problems in physics and astronomy. Reviews of Modern Physics 15:0001-0089.
230. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation 4:435-447.
231. Frenkel, D., and B. Smit. 2001. Understanding Molecular Simulation: From Algorithm to Application. Academic Press.
232. Beauchamp, K. A., Y.-S. Lin, R. Das, and V. S. Pande. 2012. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. Journal of Chemical Theory and Computation 8:1409-1414.
233. Nose, S. 1984. A MOLECULAR-DYNAMICS METHOD FOR SIMULATIONS IN THE CANONICAL ENSEMBLE. Molecular Physics 52:255-268.
234. Hoover, W. G. 1985. CANONICAL DYNAMICS - EQUILIBRIUM PHASE-SPACE DISTRIBUTIONS. Physical Review A 31:1695-1697.
235. Parrinello, M., and A. Rahman. 1981. POLYMORPHIC TRANSITIONS IN SINGLE-CRYSTALS - A NEW MOLECULAR-DYNAMICS METHOD. Journal of Applied Physics 52:7182-7190.
236. Kaminski, G. A., R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. Journal of Physical Chemistry B 105:6474-6487.
237. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER. Journal of Chemical Physics 79:926-935.

238. Guevara-Carrion, G., J. Vrabec, and H. Hasse. 2011. Prediction of self-diffusion coefficient and shear viscosity of water and its binary mixtures with methanol and ethanol by molecular simulation. Journal of Chemical Physics 134.
239. Neumann, M. 1986. Dielectric relaxation in water. Computer simulations with the TIP4P potential. The journal of Chemical Physics. 1567.
240. Varma, S., and S. B. Rempe. 2010. Multibody Effects in Ion Binding and Selectivity. Biophysical journal 99:3394-3401.
241. Rossi, M., A. Tkatchenko, S. B. Rempe, and S. Varma. 2013. Role of methyl-induced polarization in ion binding. Proceedings of the National Academy of Sciences of the United States of America 110:12978-12983.
242. Xu, H. F., H. A. Stern, and B. J. Berne. 2002. Can water polarizability be ignored in hydrogen bond kinetics? Journal of Physical Chemistry B 106:2054-2060.
243. Kim, B. C., T. Young, E. Harder, R. A. Friesner, and B. J. Berne. 2005. Structure and dynamics of the solvation of bovine pancreatic trypsin inhibitor in explicit water: A comparative study of the effects of solvent and protein polarizability. Journal of Physical Chemistry B 109:16529-16538.
244. Lee, H.-S., and M. E. Tuckerman. 2007. Dynamical properties of liquid water from ab initio molecular dynamics performed in the complete basis set limit. Journal of Chemical Physics 126.
245. Johnson, M. E., C. Malardier-Jugroot, R. K. Murarka, and T. Head-Gordon. 2009. Hydration Water Dynamics Near Biological Interfaces. Journal of Physical Chemistry B 113:4082-4092.
246. Darden, T., D. York, and L. Pedersen. 1993. PARTICLE MESH EWALD - AN N.LOG(N) METHOD FOR EWALD SUMS IN LARGE SYSTEMS. Journal of Chemical Physics 98:10089-10092.
247. Hess, B. 2008. P-LINCS: A parallel linear constraint solver for molecular simulation. Journal of Chemical Theory and Computation 4:116-122.
248. Miyamoto, S., and P. A. Kollman. 1992. SETTLE - AN ANALYTICAL VERSION OF THE SHAKE AND RATTLE ALGORITHM FOR RIGID WATER MODELS. Journal of Computational Chemistry 13:952-962.
249. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins-Structure Function and Bioinformatics 55:383-394.
250. Kubo, R. 1966. FLUCTUATION-DISSIPATION THEOREM. Reports on Progress in Physics 29:255-&.
251. Dellago, C., and P. G. Bolhuis. 2009. Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events. Advanced Computer Simulation Approaches for Soft Matter Sciences Iii 221:167-233.
252. Laio, A., and F. L. Gervasio. 2008. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Reports on Progress in Physics 71.
253. Barducci, A., M. Bonomi, and M. Parrinello. 2011. Metadynamics. Wiley Interdisciplinary Reviews-Computational Molecular Science 1:826-843.
254. Tiwary, P., and M. Parrinello. 2013. From Metadynamics to Dynamics. Physical Review Letters 111.

255. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. Proceedings of the National Academy of Sciences of the United States of America 99:12562-12566.
256. Bonomi, M., D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello. 2009. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. Computer Physics Communications 180:1961-1972.
257. Laio, A., A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello. 2005. Assessing the accuracy of metadynamics. Journal of Physical Chemistry B 109:6714-6721.
258. Bussi, G., A. Laio, and M. Parrinello. 2006. Equilibrium free energies from nonequilibrium metadynamics. Physical Review Letters 96.
259. Barducci, A., G. Bussi, and M. Parrinello. 2008. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. Physical Review Letters 100.
260. Westhead, D. R., and J. M. Thornton. 1998. Protein structure prediction. Current Opinion in Biotechnology 9:383-389.
261. Al-Lazikani, B., J. Jung, Z. X. Xiang, and B. Honig. 2001. Protein structure prediction. Current Opinion in Chemical Biology 5:51-56.
262. Browne, W. J., A. C. T. North, and D. C. Phillips. 1969. A POSSIBLE 3-DIMENSIONAL STRUCTURE OF BOVINE ALPHA-LACTALBUMIN BASED ON THAT OF HENS EGG-WHITE LYSOZYME. Journal of molecular biology 42:65-&.
263. Levitt, M., and R. Sharon. 1988. ACCURATE SIMULATION OF PROTEIN DYNAMICS IN SOLUTION. Proceedings of the National Academy of Sciences of the United States of America 85:7557-7561.
264. Levitt, M. 1992. ACCURATE MODELING OF PROTEIN CONFORMATION BY AUTOMATIC SEGMENT MATCHING. Journal of molecular biology 226:507-533.
265. Sali, A., and T. L. Blundell. 1993. COMPARATIVE PROTEIN MODELING BY SATISFACTION OF SPATIAL RESTRAINTS. Journal of molecular biology 234:779-815.
266. Kolinski, A., M. R. Betancourt, D. Kihara, P. Rotkiewicz, and J. Skolnick. 2001. Generalized comparative modeling (GENECOMP): A combination of sequence comparison, threading, and lattice modeling for protein structure prediction and refinement. Proteins-Structure Function and Genetics 44:133-149.
267. Havel, T. F., and M. E. Snow. 1991. A NEW METHOD FOR BUILDING PROTEIN CONFORMATIONS FROM SEQUENCE ALIGNMENTS WITH HOMOLOGS OF KNOWN STRUCTURE. Journal of molecular biology 217:1-7.
268. Baker, D., and A. Sali. 2001. Protein structure prediction and structural genomics. Science 294:93-96.
269. Marti-Renom, M. A., A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. 2000. Comparative protein structure modeling of genes and genomes. Annual Review of Biophysics and Biomolecular Structure 29:291-325.
270. Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden. 2008. NCBIBLAST: a better web interface. Nucleic Acids Research 36:W5-W9.

271. Boratyn, G. M., C. Camacho, P. S. Cooper, G. Coulouris, A. Fong, N. Ma, T. L. Madden, W. T. Matten, S. D. McGinnis, Y. Merezhuk, Y. Raytselis, E. W. Sayers, T. Tao, J. Ye, and I. Zaretskaya. 2013. BLAST: a more efficient report with usability improvements. Nucleic Acids Research 41:W29-W33.
272. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soeding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7.
273. Eswar, N., D. Eramian, B. Webb, M.-Y. Shen, and A. Sali. 2008. Protein structure modeling with MODELLER. In Methods in Molecular Biology. B. Kobe, M. Guss, and T. Huber, editors. 145-159.
274. Webb, B., and A. Sali. 2014. Comparative Protein Structure Modeling Using MODELLER. Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.] 47:5.6.1-32.
275. Webb, B., and A. Sali. 2014. Protein Structure Modeling with MODELLER. In Protein Structure Prediction, 3rd Edition. D. Kihara, editor. 1-15.
276. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. Journal of Physical Chemistry B 102:3586-3616.
277. Sanchez, R., and A. Sali. 1998. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. Proceedings of the National Academy of Sciences of the United States of America 95:13597-13602.
278. Koehl, P., and M. Levitt. 1999. A brighter future for protein structure prediction (Third Community Wide Experimental on the Critical Assessment of Techniques for Protein Structure Prediction) (California, USA, December 13017, 1998). Nature Structural Biology 6:108-111.
279. Bonneau, R., and D. Baker. 2001. Ab initio protein structure prediction: Progress and prospects. Annual Review of Biophysics and Biomolecular Structure 30:173-189.
280. Skolnick, J. 2006. In quest of an empirical potential for protein structure prediction. Current Opinion in Structural Biology 16:166-171.
281. Zhang, Y., A. Kolinski, and J. Skolnick. 2003. TOUCHSTONE II: A new approach to ab initio protein structure prediction. Biophysical journal 85:1145-1164.
282. Skolnick, J., L. Jaroszewski, A. Kolinski, and A. Godzik. 1997. Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? Protein Science 6:676-688.
283. Samudrala, R., and J. Moult. 1998. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. Journal of molecular biology 275:895-916.
284. Lu, H., and J. Skolnick. 2001. A distance-dependent atomic knowledge-based potential for improved protein structure selection. Proteins-Structure Function and Genetics 44:223-232.

285. Zhou, H. Y., and Y. Q. Zhou. 2002. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Science 11:2714-2726.
286. Zhang, Y., and J. Skolnick. 2005. The protein structure prediction problem could be solved using the current PDB library. Proceedings of the National Academy of Sciences of the United States of America 102:1029-1034.
287. Zhang, Y., I. A. Hubner, A. K. Arakaki, E. Shakhnovich, and J. Skolnick. 2006. On the origin and highly likely completeness of single-domain protein structures. Proceedings of the National Academy of Sciences of the United States of America 103:2605-2610.
288. Bowie, J. U., and D. Eisenberg. 1994. AN EVOLUTIONARY APPROACH TO FOLDING SMALL ALPHA-HELICAL PROTEINS THAT USES SEQUENCE INFORMATION AND AN EMPIRICAL GUIDING FITNESS FUNCTION. Proceedings of the National Academy of Sciences of the United States of America 91:4436-4440.
289. Simons, K. T., C. Kooperberg, E. Huang, and D. Baker. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. Journal of molecular biology 268:209-225.
290. Simons, K. T., R. Bonneau, I. Ruczinski, and D. Baker. 1999. Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins Suppl 3:171-176.
291. Li, Z. Q., and H. A. Scheraga. 1987. MONTE-CARLO-MINIMIZATION APPROACH TO THE MULTIPLE-MINIMA PROBLEM IN PROTEIN FOLDING. Proceedings of the National Academy of Sciences of the United States of America 84:6611-6615.
292. Kaufmann, K. W., G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, and J. Meiler. 2010. Practically Useful: What the ROSETTA Protein Modeling Suite Can Do for You. Biochemistry 49:2987-2998.
293. DiMaio, F., A. Leaver-Fay, P. Bradley, D. Baker, and I. Andre. 2011. Modeling Symmetric Macromolecular Structures in Rosetta3. Plos One 6.
294. Koshland, D. E. 1958. APPLICATION OF A THEORY OF ENZYME SPECIFICITY TO PROTEIN SYNTHESIS. Proceedings of the National Academy of Sciences of the United States of America 44:98-104.
295. Huber, R., and W. S. Bennett. 1983. FUNCTIONAL-SIGNIFICANCE OF FLEXIBILITY IN PROTEINS. Biopolymers 22:261-279.
296. Kumar, S., B. Y. Ma, C. J. Tsai, N. Sinha, and R. Nussinov. 2000. Folding and binding cascades: Dynamic landscapes and population shifts. Protein Science 9:10-19.
297. Gutteridge, A., and J. Thornton. 2005. Conformational changes observed in enzyme crystal structures upon substrate binding. Journal of molecular biology 346:21-28.
298. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology 6:197-208.

299. Tobi, D., and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. Proceedings of the National Academy of Sciences of the United States of America 102:18908-18913.
300. Damm, K. L., and H. A. Carlson. 2006. Gaussian-weighted RMSD superposition of proteins: A structural comparison for flexible proteins and predicted protein structures. Biophysical journal 90:4558-4573.
301. Wolfe, K. C., and G. S. Chirikjian. 2012. Quantitative Comparison of Conformational Ensembles. Entropy 14:213-232.
302. McClendon, C. L., L. Hua, G. Barreiro, and M. P. Jacobson. 2012. Comparing Conformational Ensembles Using the Kullback-Leibler Divergence Expansion. Journal of Chemical Theory and Computation 8:2115-2126.
303. Cortes, C., and V. Vapnik. 1995. SUPPORT-VECTOR NETWORKS. Machine Learning 20:273-297.
304. Smola, A. J., B. Scholkopf, and K. R. Muller. 1998. The connection between regularization operators and support vector kernels. Neural Networks 11:637-649.
305. Scholkopf, B., C. J. C. Burges, and A. J. Smola. 1999. Advances in kernel methods - Support vector learning - Introduction. Advances in Kernel Methods:1-15.
306. Joachims, T. 1999. Making large-scale support vector machine learning practical. Advances in Kernel Methods:169-184.
307. Cristianini, N., and J. Shawe-Taylor. 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge.
308. Chang, C.-C., and C.-J. Lin. 2011. LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology.
309. Rodier, F., R. P. Bahadur, P. Chakrabarti, and J. Janin. 2005. Hydration of protein-protein interfaces. Proteins-Structure Function and Bioinformatics 60:36-45.
310. Buchli, B., S. A. Waldauer, R. Walser, M. L. Donten, R. Pfister, N. Bloechliger, S. Steiner, A. Caflisch, O. Zerbe, and P. Hamm. 2013. Kinetic response of a photoperturbed allosteric protein. Proceedings of the National Academy of Sciences of the United States of America 110:11725-11730.
311. Leitner, D. M. 2016. Water-Mediated Energy Dynamics in a Homodimeric Hemoglobin. Journal of Physical Chemistry B 120:4019-4027.
312. Ribeiro, A. A. S. T., and V. Ortiz. 2016. A Chemical Perspective on Allostery. Chem Rev 116:6488-6502.
313. Saenger, W. 1987. STRUCTURE AND DYNAMICS OF WATER SURROUNDING BIOMOLECULES. Annual Review of Biophysics and Biophysical Chemistry 16:93-114.
314. Meyer, E. 1992. INTERNAL WATER-MOLECULES AND H-BONDING IN BIOLOGICAL MACROMOLECULES - A REVIEW OF STRUCTURAL FEATURES WITH FUNCTIONAL IMPLICATIONS. Protein Science 1:1543-1562.
315. Otting, G. 1997. NMR studies of water bound to biological molecules. Progress in Nuclear Magnetic Resonance Spectroscopy 31:259-285.

316. Bellissent-Funel, M. C. 1998. Structure and dynamics of water near hydrophilic surfaces. Journal of Molecular Liquids 78:19-28.
317. Bhattacharyya, K., and B. Bagchi. 2000. Slow dynamics of constrained water in complex geometries. Journal of Physical Chemistry A 104:10603-10613.
318. Mattos, C. 2002. Protein-water interactions in a dynamic world. Trends in Biochemical Sciences 27:203-208.
319. Guillot, B. 2002. A reappraisal of what we have learnt during three decades of computer simulations on water. Journal of Molecular Liquids 101:219-260.
320. Bizzarri, A. R., and S. Cannistraro. 2002. Molecular dynamics of water at the protein-solvent interface. Journal of Physical Chemistry B 106:6617-6633.
321. Bagchi, B. 2005. Water dynamics in the hydration layer around proteins and micelles. Chem Rev 105:3197-3219.
322. Chandler, D. 2005. Interfaces and the driving force of hydrophobic assembly. Nature 437:640-647.
323. Raschke, T. M. 2006. Water structure and interactions with protein surfaces. Current Opinion in Structural Biology 16:152-159.
324. Chaplin, M. 2006. Opinion - Do we underestimate the importance of water in cell biology? Nature Reviews Molecular Cell Biology 7:861-866.
325. Geissler, P. L. 2013. Water Interfaces, Solvation, and Spectroscopy. Annual Review of Physical Chemistry, Vol 64 64:317-337.
326. Brunne, R. M., E. Liepinsh, G. Otting, K. Wuthrich, and W. F. Vangunsteren. 1993. HYDRATION OF PROTEINS - A COMPARISON OF EXPERIMENTAL RESIDENCE TIMES OF WATER-MOLECULES SOLVATING THE BOVINE PANCREATIC TRYPSIN-INHIBITOR WITH THEORETICAL-MODEL CALCULATIONS. Journal of molecular biology 231:1040-1048.
327. Denisov, V. P., and B. Halle. 1994. DYNAMICS OF THE INTERNAL AND EXTERNAL HYDRATION OF GLOBULAR-PROTEINS. Journal of the American Chemical Society 116:10324-10325.
328. Lee, S. H., and P. J. Rossky. 1994. A COMPARISON OF THE STRUCTURE AND DYNAMICS OF LIQUID WATER AT HYDROPHOBIC AND HYDROPHILIC SURFACES - A MOLECULAR-DYNAMICS SIMULATION STUDY. Journal of Chemical Physics 100:3334-3345.
329. Muegge, I., and E. W. Knapp. 1995. RESIDENCE TIMES AND LATERAL DIFFUSION OF WATER AT PROTEIN SURFACES - APPLICATION TO BPTI. Journal of Physical Chemistry 99:1371-1374.
330. Denisov, V. P., and B. Halle. 1996. Protein hydration dynamics in aqueous solution. Faraday Discussions; Hydration processes in biological and macromolecular systems 103:227-244.
331. BellissentFunel, M. C., J. M. Zanotti, and S. H. Chen. 1996. Slow dynamics of water molecules on the surface of a globular protein. Faraday Discussions 103:281-294.
332. Settles, M., and W. Doster. 1996. Anomalous diffusion of adsorbed water: A neutron scattering study of hydrated myoglobin. Faraday Discussions 103:269-279.

333. Abseher, R., H. Schreiber, and O. Steinhauser. 1996. The influence of a protein on water dynamics in its vicinity investigated by molecular dynamics simulation. Proteins-Structure Function and Genetics 25:366-378.
334. Rocchi, C., A. R. Bizzarri, and S. Cannistraro. 1997. Water residence times around copper plastocyanin: A molecular dynamics simulation approach. Chemical Physics 214:261-276.
335. Cheng, Y. K., and P. J. Rossky. 1998. Surface topography dependence of biomolecular hydrophobic hydration. Nature 392:696-699.
336. Denisov, V. P., B. H. Jonsson, and B. Halle. 1999. Hydration of denatured and molten globule proteins. Nature Structural Biology 6:253-260.
337. Jordanides, X. J., M. J. Lang, X. Y. Song, and G. R. Fleming. 1999. Solvation dynamics in protein environments studied by photon echo spectroscopy. Journal of Physical Chemistry B 103:7995-8005.
338. Dellerue, S., and M. C. Bellissent-Funel. 2000. Relaxational dynamics of water molecules at protein surface. Chemical Physics 258:315-325.
339. Makarov, V. A., B. K. Andrews, P. E. Smith, and B. M. Pettitt. 2000. Residence times of water molecules in the hydration sites of myoglobin. Biophysical journal 79:2966-2974.
340. Sterpone, F., M. Ceccarelli, and M. Marchi. 2001. Dynamics of hydration in hen egg white lysozyme. Journal of molecular biology 311:409-419.
341. Zhong, D. P., S. K. Pal, D. Q. Zhang, S. I. Chan, and A. H. Zewail. 2002. Femtosecond dynamics of rubredoxin: Tryptophan solvation and resonance energy transfer in the protein. Proceedings of the National Academy of Sciences of the United States of America 99:13-18.
342. Merzel, F., and J. C. Smith. 2002. Is the first hydration shell of lysozyme of higher density than bulk water? Proceedings of the National Academy of Sciences of the United States of America 99:5378-5383.
343. Marchi, M., F. Sterpone, and M. Ceccarelli. 2002. Water rotational relaxation and diffusion in hydrated lysozyme. Journal of the American Chemical Society 124:6787-6791.
344. Russo, D., P. Baglioni, E. Peroni, and J. Teixeira. 2003. Hydration water dynamics of a completely hydrophobic oligopeptide. Chemical Physics 292:235-245.
345. Modig, K., E. Liepinsh, G. Otting, and B. Halle. 2004. Dynamics of protein and peptide hydration. Journal of the American Chemical Society 126:102-114.
346. Russo, D., G. Hura, and T. Head-Gordon. 2004. Hydration dynamics near a model protein surface. Biophysical journal 86:1852-1862.
347. Russo, D., R. K. Murarka, J. R. D. Copley, and T. Head-Gordon. 2005. Molecular view of water dynamics near model peptides. Journal of Physical Chemistry B 109:12966-12975.
348. Yang, C., and K. A. Sharp. 2005. Hydrophobic tendency of polar group hydration as a major force in type I antifreeze protein recognition. Proteins-Structure Function and Bioinformatics 59:266-274.
349. Dokter, A. M., S. Woutersen, and H. J. Bakker. 2006. Inhomogeneous dynamics in confined water nanodroplets. Proceedings of the National Academy of Sciences of the United States of America 103:15355-15358.

350. Ebbinghaus, S., S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, and M. Havenith. 2007. An extended dynamical hydration shell around proteins. Proceedings of the National Academy of Sciences of the United States of America 104:20749-20752.
351. Pizzitutti, F., M. Marchi, F. Sterpone, and P. J. Rossky. 2007. How protein surfaces induce anomalous dynamics of hydration water. Journal of Physical Chemistry B 111:7584-7590.
352. Sinha, S. K., S. Chakraborty, and S. Bandyopadhyay. 2008. Thickness of the hydration layer of a protein from molecular dynamics simulation. Journal of Physical Chemistry B 112:8203-8209.
353. Patel, A. J., P. Varilly, and D. Chandler. 2010. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. Journal of Physical Chemistry B 114:1632-1637.
354. Sterpone, F., G. Stirnemann, J. T. Hynes, and D. Laage. 2010. Water Hydrogen-Bond Dynamics around Amino Acids: The Key Role of Hydrophilic Hydrogen-Bond Acceptor Groups. Journal of Physical Chemistry B 114:2083-2089.
355. Khodadadi, S., J. E. Curtis, and A. P. Sokolov. 2011. Nanosecond Relaxation Dynamics of Hydrated Proteins: Water versus Protein Contributions. Journal of Physical Chemistry B 115:6222-6226.
356. Gupta, S., R. D'Mello, and M. R. Chance. 2012. Structure and dynamics of protein waters revealed by radiolysis and mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 109:14882-14887.
357. King, J. T., and K. J. Kubarych. 2012. Site-Specific Coupling of Hydration Water and Protein Flexibility Studied in Solution with Ultrafast 2D-IR Spectroscopy. Journal of the American Chemical Society 134:18705-18712.
358. Jana, B., S. Pal, and B. Bagchi. 2012. Hydration dynamics of protein molecules in aqueous solution: Unity among diversity. Journal of Chemical Sciences 124:317-325.
359. Chandler, D. 1987. Introduction to Modern Statistical Mechanics.
360. Yeh, I. C., and G. Hummer. 2004. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. Journal of Physical Chemistry B 108:15873-15879.
361. Tazi, S., A. Botan, M. Salanne, V. Marry, P. Turq, and B. Rotenberg. 2012. Diffusion coefficient and shear viscosity of rigid water models. Journal of Physics-Condensed Matter 24.
362. Mills, R. 1973. SELF-DIFFUSION IN NORMAL AND HEAVY-WATER IN RANGE 1-45 DEGREES. Journal of Physical Chemistry 77:685-688.
363. Hardy, E. H., A. Zygar, M. D. Zeidler, M. Holz, and F. D. Sacher. 2001. Isotope effect on the translational and rotational motion in liquid water and ammonia. Journal of Chemical Physics 114:3174-3181.
364. van der Spoel, D., P. J. van Maaren, P. Larsson, and N. Timneanu. 2006. Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. Journal of Physical Chemistry B 110:4393-4398.
365. Rapaport, D. C. 1983. HYDROGEN-BONDS IN WATER NETWORK ORGANIZATION AND LIFETIMES. Molecular Physics 50:1151-1162.

366. Luzar, A., and D. Chandler. 1996. Effect of environment on hydrogen bond dynamics in liquid water. Physical Review Letters 76:928-931.
367. Luzar, A., and D. Chandler. 1996. Hydrogen-bond kinetics in liquid water. Nature 379:55-57.
368. Luzar, A. 2000. Resolving the hydrogen bond dynamics conundrum. Journal of Chemical Physics 113:10663-10675.
369. Soper, A. K., and M. G. Phillips. 1986. A NEW DETERMINATION OF THE STRUCTURE OF WATER AT 25-DEGREES-C. Chemical Physics 107:47-60.
370. Hura, G., J. M. Sorenson, R. M. Glaeser, and T. Head-Gordon. 2000. A high-quality x-ray scattering experiment on liquid water at ambient conditions. Journal of Chemical Physics 113:9140-9148.
371. Varma, S., M. Teng, and H. L. Scottt. 2012. Nonintercalating Nanosubstrates Create Asymmetry between Bilayer Leaflets. Langmuir 28:2842-2848.
372. Kumar, R., J. R. Schmidt, and J. L. Skinner. 2007. Hydrogen bonding definitions and dynamics in liquid water. Journal of Chemical Physics 126.
373. Krimm, S. 1967. Hydrogen bonding of C-H...O=C in proteins. Science 158:530-531.
374. Desiraju, G. R. 1996. The C-H ... O hydrogen bond: Structural implications and supramolecular design. Accounts of Chemical Research 29:441-449.
375. Gu, Y. L., T. Kar, and S. Scheiner. 1999. Fundamental properties of the CH ... O interaction: Is it a true hydrogen bond? Journal of the American Chemical Society 121:9411-9422.
376. Horowitz, S., J. D. Yesselman, H. M. Al-Hashimi, and R. C. Trievel. 2011. Direct Evidence for Methyl Group Coordination by Carbon-Oxygen Hydrogen Bonds in the Lysine Methyltransferase SET7/9. Journal of Biological Chemistry 286:18658-18663.
377. Scheiner, S., T. Kar, and Y. L. Gu. 2001. Strength of the (CH)-H-alpha .. O hydrogen bond of amino acid residues. Journal of Biological Chemistry 276:9832-9837.
378. Fenn, E. E., D. B. Wong, and M. D. Fayer. 2009. Water dynamics at neutral and ionic interfaces. Proceedings of the National Academy of Sciences of the United States of America 106:15243-15248.
379. Zasetsky, A. Y. 2011. Dielectric Relaxation in Liquid Water: Two Fractions or Two Dynamics? Physical Review Letters 107.
380. Varma, S., S. W. Chiu, and E. Jakobsson. 2006. The influence of amino acid protonation states on molecular dynamics simulations of the bacterial porin OmpF. Biophysical journal 90:112-123.
381. Varma, S., and E. Jakobsson. 2007. The cPLA(2) C2 alpha domain in solution: Structure and dynamics of its Ca2+-activated and cation-free states. Biophysical journal 92:966-976.
382. Bonvin, A. M. 2006. Flexible protein-protein docking. Current Opinion in Structural Biology 16:194-200.
383. Abrams, J. B., and M. E. Tuckerman. 2008. Efficient and Direct Generation of Multidimensional Free Energy Surfaces via Adiabatic Dynamics without Coordinate Transformations. Journal of Physical Chemistry B 112:15742-15757.

384. Liu, P., B. Kim, R. A. Friesner, and B. J. Berne. 2005. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. Proceedings of the National Academy of Sciences of the United States of America 102:13749-13754.
385. Crespo, Y., F. Marinelli, F. Pietrucci, and A. Laio. 2010. Metadynamics convergence law in a multidimensional system. Phys Rev E 81.
386. Limongelli, V., M. Bonomi, and M. Parrinello. 2013. Funnel metadynamics as accurate binding free-energy method. Proceedings of the National Academy of Sciences of the United States of America 110:6358-6363.
387. Troussicot, L., F. Guilliere, V. Limongelli, O. Walker, and J.-M. Lancelin. 2015. Funnel-Metadynamics and Solution NMR to Estimate Protein-Ligand Affinities. Journal of the American Chemical Society 137:1273-1281.
388. Dama, J. F., M. Parrinello, and G. A. Voth. 2014. Well-Tempered Metadynamics Converges Asymptotically. Physical Review Letters 112.
389. Martin, D. R., P. Dutta, S. Mahajan, S. Varma, and S. M. Stevens, Jr. 2016. Structural and activity characterization of human PHPT1 after oxidative modification. Scientific Reports 6.
390. Ishida, T., and K. Kinoshita. 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Research 35:W460-W464.
391. Montagna, M., F. Sterpone, and L. Guidoni. 2012. Structural and Spectroscopic Properties of Water around Small Hydrophobic Solutes. Journal of Physical Chemistry B 116:11695-11700.
392. Monastyrskyy, B., A. Kryshtafovych, J. Moult, A. Tramontano, and K. Fidelis. 2014. Assessment of protein disorder region predictions in CASP10. Proteins-Structure Function and Bioinformatics 82:127-137.
393. Deng, X., J. Eickholt, and J. Cheng. 2012. A comprehensive overview of computational protein disorder prediction methods. Molecular Biosystems 8:114-121.
394. Monastyrskyy, B., K. Fidelis, J. Moult, A. Tramontano, and A. Kryshtafovych. 2011. Evaluation of disorder predictions in CASP9. Proteins-Structure Function and Bioinformatics 79:107-118.
395. Romero, P., Z. Obradovic, X. H. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. 2001. Sequence complexity of disordered protein. Proteins-Structure Function and Genetics 42:38-48.
396. Vucetic, S., C. J. Brown, A. K. Dunker, and Z. Obradovic. 2003. Flavors of protein disorder. Proteins-Structure Function and Genetics 52:573-584.
397. van der Lee, R., M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu. 2014. Classification of Intrinsically Disordered Regions and Proteins. Chem Rev 114:6589-6631.
398. Dunker, A. K., and Z. Obradovic. 2001. The protein trinity - linking function and disorder. Nature Biotechnology 19:805-806.
399. Dunker, A. K., I. Silman, V. N. Uversky, and J. L. Sussman. 2008. Function and structure of inherently disordered proteins. Current Opinion in Structural Biology 18:756-764.

400. Oldfield, C. J., and A. K. Dunker. 2014. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. Annual Review of Biochemistry, Vol 83 83:553-584.
401. Trabuco, L. G., E. Villa, K. Mitra, J. Frank, and K. Schulten. 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure 16:673-683.
402. Trabuco, L. G., E. Villa, E. Schreiner, C. B. Harrison, and K. Schulten. 2009. Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods 49:174-180.
403. Villa, E., J. Sengupta, L. G. Trabuco, J. LeBarron, W. T. Baxter, T. R. Shaikh, R. A. Grassucci, P. Nissen, M. Ehrenberg, K. Schulten, and J. Frank. 2009. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. Proceedings of the National Academy of Sciences of the United States of America 106:1063-1068.
404. Wriggers, W., E. Mehler, F. Pitici, H. Weinstein, and K. Schulten. 1998. Structure and dynamics of calmodulin in solution. Biophysical journal 74:1622-1639.
405. Zhao, G., J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, and P. Zhang. 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature 497:643-646.
406. Crick, F. H. C. 1953. THE PACKING OF ALPHA-HELICES - SIMPLE COILED-COILS. Acta Crystallographica 6:689-697.
407. Pauling, L., and R. B. Corey. 1953. COMPOUND HELICAL CONFIGURATIONS OF POLYPEPTIDE CHAINS - STRUCTURE OF PROTEINS OF THE ALPHA-KERATIN TYPE. Nature 171:59-61.
408. Holberton, D., D. A. Baker, and J. Marshall. 1988. SEGMENTED ALPHA-HELICAL COILED-COIL STRUCTURE OF THE PROTEIN GIARDIN FROM THE GIARDIA CYTOSKELETON. Journal of molecular biology 204:789-795.
409. Marshall, J., and D. V. Holberton. 1993. SEQUENCE AND STRUCTURE OF A NEW COILED-COIL PROTEIN FROM A MICROTUBULE BUNDLE IN GIARDIA. Journal of molecular biology 231:521-530.
410. Marshall, J., and D. V. Holberton. 1995. GIARDIA GENE PREDICTS A 183 KDA NUCLEOTIDE-BINDING HEAD-STALK PROTEIN. J Cell Sci 108:2683-2692.
411. Seo, J., and C. Cohen. 1993. PITCH DIVERSITY IN ALPHA-HELICAL COILED COILS. Proteins-Structure Function and Genetics 15:223-234.
412. Strelkov, S. V., and P. Burkhard. 2002. Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. Journal of Structural Biology 137:54-64.
413. Phillips, G. N. 1992. WHAT IS THE PITCH OF THE ALPHA-HELICAL COILED COIL. Proteins-Structure Function and Genetics 14:425-429.
414. Lupas, A. N., and M. Gruber. 2005. The structure of alpha-helical coiled coils. Fibrous Proteins: Coiled-Coils, Collagen and Elastomers 70:37-+.
415. Chou, K. C., G. M. Maggiora, G. Nemethy, and H. A. Scheraga. 1988. ENERGETICS OF THE STRUCTURE OF THE 4-ALPHA-HELIX BUNDLE IN PROTEINS. Proceedings of the National Academy of Sciences of the United States of America 85:4295-4299.

416. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. BASIC LOCAL ALIGNMENT SEARCH TOOL. Journal of molecular biology 215:403-410.
417. Soding, J., A. Biegert, and A. N. Lupas. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33:W244-W248.
418. Shen, M.-Y., and A. Sali. 2006. Statistical potential for assessment and prediction of protein structures. Protein Science 15:2507-2524.
419. Tanaka, S., and H. A. Scheraga. 1976. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. Macromolecules 9:945-950.
420. Miyazawa, S., and R. L. Jernigan. 1985. ESTIMATION OF EFFECTIVE INTERRESIDUE CONTACT ENERGIES FROM PROTEIN CRYSTAL-STRUCTURES - QUASI-CHEMICAL APPROXIMATION. Macromolecules 18:534-552.
421. Miyazawa, S., and R. L. Jernigan. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. Journal of molecular biology 256:623-644.
422. Miyazawa, S., and R. L. Jernigan. 1999. An empirical energy potential with a reference state for protein fold and sequence recognition. Proteins-Structure Function and Genetics 36:357-369.
423. Sippl, M. J. 1990. CALCULATION OF CONFORMATIONAL ENSEMBLES FROM POTENTIALS OF MEAN FORCE - AN APPROACH TO THE KNOWLEDGE-BASED PREDICTION OF LOCAL STRUCTURES IN GLOBULAR-PROTEINS. Journal of molecular biology 213:859-883.
424. Sippl, M. J. 1993. BOLTZMANN PRINCIPLE, KNOWLEDGE-BASED MEAN FIELDS AND PROTEIN-FOLDING - AN APPROACH TO THE COMPUTATIONAL DETERMINATION OF PROTEIN STRUCTURES. Journal of Computer-Aided Molecular Design 7:473-501.
425. Sippl, M. J. 1993. RECOGNITION OF ERRORS IN 3-DIMENSIONAL STRUCTURES OF PROTEINS. Proteins-Structure Function and Genetics 17:355-362.
426. Kim, D. E., D. Chivian, and D. Baker. 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Research 32:W526-W531.
427. Das, R., I. Andre, Y. Shen, Y. Wu, A. Lemak, S. Bansal, C. H. Arrowsmith, T. Szyperski, and D. Baker. 2009. Simultaneous prediction of protein folding and docking at high resolution. Proceedings of the National Academy of Sciences of the United States of America 106:18978-18983.
428. Andre, I., P. Bradley, C. Wang, and D. Baker. 2007. Prediction of the structure of symmetrical protein assemblies. Proceedings of the National Academy of Sciences of the United States of America 104:17656-17661.
429. Bradley, P., K. M. S. Misura, and D. Baker. 2005. Toward high-resolution de novo structure prediction for small proteins. Science 309:1868-1871.
430. Kay, B. K., M. P. Williamson, and P. Sudol. 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. Faseb Journal 14:231-241.

431. George, R. A., and J. Heringa. 2002. An analysis of protein domain linkers: their classification and role in protein folding. Protein Engineering 15:871-879.
432. Zarrinpar, A., R. P. Bhattacharyya, and W. A. Lim. 2003. The structure and function of proline recognition domains. Science's STKE : signal transduction knowledge environment 2003:RE8-RE8.
433. Morgan, A. A., and E. Rubenstein. 2013. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome. Plos One 8.
434. Macarthur, M. W., and J. M. Thornton. 1991. INFLUENCE OF PROLINE RESIDUES ON PROTEIN CONFORMATION. Journal of molecular biology 218:397-412.
435. Williamson, M. P. 1994. THE STRUCTURE AND FUNCTION OF PROLINE-RICH REGIONS IN PROTEINS. Biochemical Journal 297:249-260.
436. Varma, S., J. P. R. O. Orgel, and J. D. Schieber. 2016. Nanomechanics of type I collagen. Biophysical Journa in press.
437. Byeon, I.-J. L., X. Meng, J. Jung, G. Zhao, R. Yang, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, and A. M. Gronenborn. 2009. Structural Convergence between Cryo-EM and NMR Reveals Intersubunit Interactions Critical for HIV-1 Capsid Function. Cell 139:780-790.
438. Chapman, M. S. 1995. RESTRAINED REAL-SPACE MACROMOLECULAR ATOMIC REFINEMENT USING A NEW RESOLUTION-DEPENDENT ELECTRON-DENSITY FUNCTION. Acta Crystallographica Section A 51:69-80.
439. Roseman, A. M. 2000. Docking structures of domains into maps from cryo-electron microscopy using local correlation. Acta Crystallographica Section D-Biological Crystallography 56:1332-1340.
440. Wriggers, W., and P. Chacon. 2001. Modeling tricks and fitting techniques for multiresolution structures. Structure 9:779-788.
441. Schroeder, G. F., A. T. Brunger, and M. Levitt. 2007. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. Structure 15:1630-1641.
442. Topf, M., K. Lasker, B. Webb, H. Wolfson, W. Chiu, and A. Sali. 2008. Protein structure fitting and refinement guided by cryo-EM density. Structure 16:295-307.
443. DiMaio, F., Y. Song, X. Li, M. J. Brunner, C. Xu, V. Conticello, E. Egelman, T. C. Marlovits, Y. Cheng, and D. Baker. 2015. Atomic-accuracy models from 4.5-angstrom cryo-electron microscopy data with density-guided iterative local refinement. Nature Methods 12:361-U129.
444. Jolley, C. C., S. A. Wells, P. Frornme, and M. F. Thorpe. 2008. Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. Biophysical journal 94:1613-1621.
445. Tama, F., O. Miyashita, and C. L. Brooks. 2004. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. Journal of Structural Biology 147:315-326.
446. Orzechowski, M., and F. Tama. 2008. Flexible Fitting of High-Resolution X-Ray Structures into Cryoelectron Microscopy Maps Using Biased Molecular Dynamics Simulations. Biophysical journal 95:5692-5705.

447. Chapman, M. S., A. Trzynka, and B. K. Chapman. 2013. Atomic modeling of cryo-electron microscopy reconstructions - Joint refinement of model and imaging parameters. Journal of Structural Biology 182:10-21.
448. McGreevy, R., I. Teo, A. Singharoy, and K. Schulten. 2016. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. Methods 100:50-60.
449. Gumbart, J., L. G. Trabuco, E. Schreiner, E. Villa, and K. Schulten. 2009. Regulation of the Protein-Conducting Channel by a Bound Ribosome. Structure 17:1453-1464.
450. Becker, T., S. Bhushan, A. Jarasch, J.-P. Armache, S. Funes, F. Jossinet, J. Gumbart, T. Mielke, O. Berninghausen, K. Schulten, E. Westhof, R. Gilmore, E. C. Mandon, and R. Beckmann. 2009. Structure of Monomeric Yeast and Mammalian Sec61 Complexes Interacting with the Translating Ribosome. Science 326:1369-1373.
451. Frauenfeld, J., J. Gumbart, E. O. van der Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten, and R. Beckmann. 2011. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. Nature Structural & Molecular Biology 18:614-U127.
452. Hsin, J., J. Gumbart, L. G. Trabuco, E. Villa, P. Qian, C. N. Hunter, and K. Schulten. 2009. Protein-Induced Membrane Curvature Investigated through Molecular Dynamics Flexible Fitting. Biophysical journal 97:321-329.
453. Sener, M., J. Hsin, L. G. Trabuco, E. Villa, P. Qian, C. N. Hunter, and K. Schulten. 2009. Structural model and excitonic properties of the dimeric RC-LH1-Pufx complex from Rhodobacter sphaeroides. Chemical Physics 357:188-197.
454. Kim, H., J. Hsin, Y. Liu, P. R. Selvin, and K. Schulten. 2010. Formation of Salt Bridges Mediates Internal Dimerization of Myosin VI Medial Tail Domain. Structure 18:1443-1449.
455. Zhang, K., L. Wang, Y. Liu, K.-Y. Chan, X. Pang, K. Schulten, Z. Dong, and F. Sun. 2013. Flexible interwoven termini determine the thermal stability of thermosomes. Protein & Cell 4:432-444.
456. Pornillos, O., B. K. Ganser-Pornillos, B. N. Kelly, Y. Hua, F. G. Whitby, C. D. Stout, W. I. Sundquist, C. P. Hill, and M. Yeager. 2009. X-Ray Structures of the Hexameric Building Block of the HIV Capsid. Cell 137:1282-1292.
457. Brown, N. G., D.-C. Chow, K. E. Ruprecht, and T. Palzkill. 2013. Identification of the β-lactamase Inhibitor Protein-II (BLIP-II) Interface Residues Essential for Binding Affinity and Specificity for Class A β-lactamases. Journal of Biological Chemistry 288:17156-17166.

APPENDIX A

STUDY OF WATER DYNAMICS AT TEM–BLIP(II) INTERFACES

The choice of the protein complexes for this study was done based on the following criteria; (i) the system must have comparable number of interstitial water molecules as seen in the crystal structure of the G protein, and (ii) the nature of interactions at the protein-protein interfaces must be different from that of the G-ephrin complexes. By doing so, we eliminate any bias in the outcome of the analysis that may be develop from the type of interactions. In this regard, the BLIPII complexes were found be suitable candidates to study the parameters like diffusion coefficient and residence time. The BLIPII-TEM1 complex has a large number of water molecules resolved at the protein-protein interface as evident from the crystal structure and hence belong to the same bin as G-B2 (shown in Figure 4.2). Also, the BLIPII-TEM1 complex has a smaller binding interface area of 2187 $Å^2$ compared to the G-B2 complex (457). In accordance with the second criteria, the type of interaction at the interface is known to be different in the BLIPII complexes. In case of the G complexes the ephrin binds via an induced fit mechanism with the G-H loop pushing into the cavity at the center of the β-propeller structure of G (90). In contrast, the BLIPII, also a β–propeller structure interacts with the TEMs by blocking their catalytic site with the hot spot residues arranged in congruence with the O-ring model on its binding surface (457). We also chose a natural mutant of TEM1 that contains the mutations at the binding interface, which will most likely alter its chemistry.

MD simulations were carried out using the same parameters as discussed in Chapter 3. For the analyses, we applied the same scheme used in the G complexes to define the interstitial water as discussed in section 4.2.1. The interstitial region (Figure A1) can be identified by the inflection point at 10 Å similar to that observed in the case of the G complexes. Hence, the interstitial region is defined as a cylinder of diameter 20 Å and height determined by the average distances between the geometric centers of the BLIPII and the TEMs. The heights of the cylinders were calculated to be 34.4 ± 0.3 Å and 33.4 ± 0.3 Å for the BLIPII-TEM1 and BLIPII-TEM22 complexes respectively.

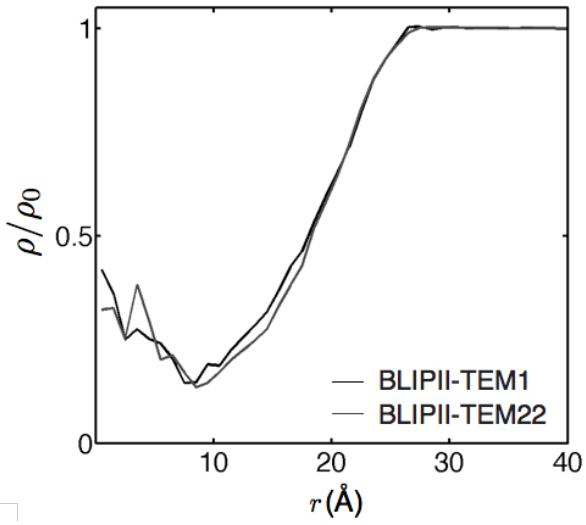


Figure A1 Water distribution in the interstitial regions of the BLIPII complexes. Normalized density of water (ρ/ρ_0) as a function of the perpendicular distance from the axes joining the geometric centers of TEM and BLIPII.

The average number of water molecules occupying the interstitial region of BLIPII-TEM1 and BLIPII-TEM22 is found to be 71.8 ± 3.5 and 69.4 ± 5.9 respectively. The values are comparable to the number of water molecules present at the interstitial region of G complexes. The diffusion coefficient values were calculated to be 0.71 ± 0.08 x 10^{-9} for BLIPII-TEM1 and 0.49 ± 0.04 x 10^{-9} m^2/s for BLIPII-TEM22 which are of the same order as that found for the G-ephrins. This implies that the residence time and

number of hydrogen bonds per water molecules should also be same as those calculated for G complexes. As expected a residence time τ_1 = 1.3 ps (shown in Figure A2) is calculated in 92% of the water molecules is, and the remaining fraction of water molecules have longer residence times of τ_2 = 54.6 ps. Consistently, a similar distribution of the fast and slow exchanging water molecules in the interstitial region of the BLIPII-TEM22 complex is calculated, yielding, *A* = 93%, τ_1 = 1.3 ps and τ_2 = 71.1 ps.

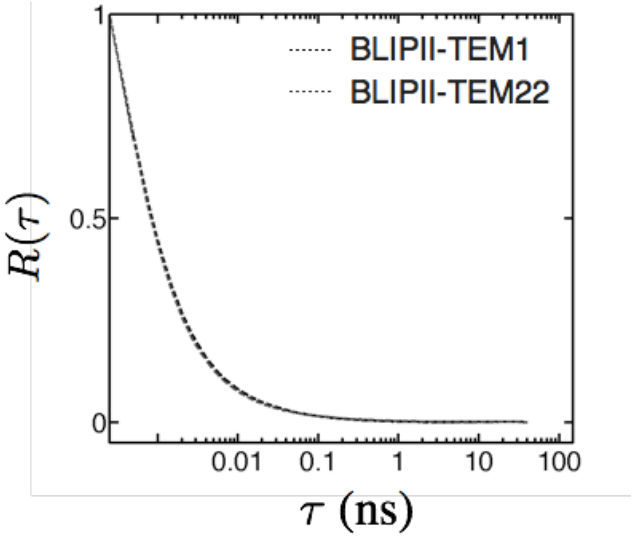


Figure A2 Residence time correlation of the water molecules, R(τ), occupying the interstitial regions in the BLIPII-TEM1 and BLIPII-TEM22 protein complexes.

Therefore, based on the results obtained from the BLIPII complexes, we can safely conclude that the qualitative nature of the dynamics of water at protein-protein interfaces is similar to that seen in the G-ephrins. Thus, we find that the qualitative dynamical behaviour of interstitial water is indeed independent of the protein system, however, further research in this direction is needed.

References

1. Brown, N. G., D.-C. Chow, K. E. Ruprecht, and T. Palzkill. 2013. Identification of the β-lactamase Inhibitor Protein-II (BLIP-II) Interface Residues Essential for Binding Affinity and Specificity for Class A β-lactamases. Journal of Biological Chemistry 288:17156-17166.
2. Bowden, T. A., A. R. Aricescu, R. J. C. Gilbert, J. M. Grimes, E. Y. Jones, and D. I. Stuart. 2008. Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. Nature Structural & Molecular Biology 15:567-572.

APPENDIX B

LIST OF LICENSE FOR REPRINT

CAMBRIDGE UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jul 02, 2016

This Agreement between Priyanka Dutta ("You") and Cambridge University Press ("Cambridge University Press") consists of your license details and the terms and conditions provided by Cambridge University Press and Copyright Clearance Center.

License Number	3900891409626
License date	Jul 02, 2016
Licensed Content Publisher	Cambridge University Press
Licensed Content Publication	Expert Reviews in Molecular Medicine
Licensed Content Title	Emerging paramyxoviruses: molecular mechanisms and antiviral strategies
Licensed Content Author	Hector C. Aguilar and Benhur Lee
Licensed Content Date	Feb 24, 2011
Licensed Content Volume Number	13
Licensed Content Issue Number	-1
Start page	0
End page	0
Type of Use	Dissertation/Thesis
Requestor type	Not-for-profit
Portion	Text extract
Number of pages requested	1
Order reference number	
Territory for reuse	World
Title of your thesis / dissertation	Computational modeling of allosteric stimulation of Nipah virus host binding protein
Expected completion date	Jul 2016
Estimated size(pages)	191
Requestor Location	Priyanka Dutta 4202 East Fowler Ave ISA 2015 TAMPA, FL 33620 United States Attn: Priyanka Dutta
Billing Type	Invoice
Billing Address	Priyanka Dutta 4202 East Fowler Ave ISA 2015 TAMPA, FL 33620 United States Attn: Priyanka Dutta

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jul 02, 2016

This Agreement between Priyanka Dutta ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3900951357827
License date	Jul 02, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Current Opinion in Virology
Licensed Content Title	Activation of paramyxovirus membrane fusion and virus entry
Licensed Content Author	Theodore S Jardetzky,Robert A Lamb
Licensed Content Date	April 2014
Licensed Content Volume Number	5
Licensed Content Issue Number	n/a
Licensed Content Pages	10
Start Page	24
End Page	33
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Figure 2
Title of your thesis/dissertation	Computational modeling of allosteric stimulation of Nipah virus host binding protein
Expected completion date	Jul 2016
Estimated size (number of pages)	191
Elsevier VAT number	GB 494 6272 12
Requestor Location	Priyanka Dutta 4202 East Fowler Ave ISA 2015

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Jul 02, 2016

This Agreement between Priyanka Dutta ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3900950178835
License date	Jul 02, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Proteins: Structure, Function and Bioinformatics
Licensed Content Title	Is allostery an intrinsic property of all dynamic proteins?
Licensed Content Author	K. Gunasekaran,Buyong Ma,Ruth Nussinov
Licensed Content Date	Jul 22, 2004
Licensed Content Pages	11
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	Computational modeling of allosteric stimulation of Nipah virus host binding protein
Expected completion date	Jul 2016
Expected size (number of pages)	191
Requestor Location	Priyanka Dutta 4202 East Fowler Ave ISA 2015 TAMPA, FL 33620 United States Attn: Priyanka Dutta
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Priyanka Dutta 4202 East Fowler Ave ISA 2015 TAMPA, FL 33620 United States Attn: Priyanka Dutta
Total	0.00 USD

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jul 02, 2016

This Agreement between Priyanka Dutta ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	3900941254350
License date	Jul 02, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature
Licensed Content Title	The ensemble nature of allostery
Licensed Content Author	Hesam N. Motlagh, James O. Wrabl, Jing Li, Vincent J. Hilser
Licensed Content Date	Apr 16, 2014
Licensed Content Volume Number	508
Licensed Content Issue Number	7496
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 2: The dynamic continuum of allosteric phenomena.
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Computational modeling of allosteric stimulation of Nipah virus host binding protein
Expected completion date	Jul 2016
Estimated size (number of pages)	191
Requestor Location	Priyanka Dutta 4202 East Fowler Ave ISA 2015 TAMPA, FL 33620 United States Attn: Priyanka Dutta
Billing Type	Invoice
Billing Address	Priyanka Dutta 4202 East Fowler Ave ISA 2015 TAMPA, FL 33620

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Water Dynamics at Protein–Protein Interfaces: Molecular Dynamics Study of Virus–Host Receptor Complexes

Author: Priyanka Dutta, Mohsen Botlani, Sameer Varma

Publication: The Journal of Physical Chemistry B

Publisher: American Chemical Society

Date: Dec 1, 2014

Logged in as:
Priyanka Dutta
Account #:
3001041779
LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

AMERICAN SOCIETY FOR MICROBIOLOGY

Title: Cysteines in the Stalk of the Nipah Virus G Glycoprotein Are Located in a Distinct Subdomain Critical for Fusion Activation

Author: Dianna Maar, Brooke Harmon, David Chu et al.

Publication: Journal of Virology

Publisher: American Society for Microbiology

Date: Jun 15, 2012

Logged in as:
Priyanka Dutta
Account #:
3001041779
LOGOUT

Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jul 07, 2016

This Agreement between Priyanka Dutta ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	3903901025449
License date	Jul 07, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature
Licensed Content Title	Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics
Licensed Content Author	Gongpu Zhao, Juan R. Perilla, Ernest L. Yufenyuy, Xin Meng, Bo Chen, Jiying Ning
Licensed Content Date	May 29, 2013
Licensed Content Volume Number	497
Licensed Content Issue Number	7451
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 1: Cryo-EM reconstruction of HIV-1 CA tubular assembly at 8 Å resolution and MDFF.
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Computational modeling of allosteric stimulation of Nipah virus host binding protein
Expected completion date	Jul 2016
Estimated size (number of pages)	191
Requestor Location	Priyanka Dutta 4202 East Fowler Ave ISA 2015 TAMPA, FL 33620 United States Attn: Priyanka Dutta
Billing Type	Invoice
Billing Address	Priyanka Dutta 4202 East Fowler Ave