Mechanotransduction by Talin: A Molecular Dynamics Study of Force-Induced Recruitment of Vinculin to a Focal Adhesion Complex

by

Seung Eun Lee

B.S. and M.S., Mechanical Engineering
University of Illinois at Chicago, 2001 and 2002

Submitted to the Department of Mechanical Engineering
on August 16, 2007 in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Mechanical Engineering

at the
Massachusetts Institute of Technology
September 2007

Signature of Author ______________________________
Department of Mechanical Engineering, MIT
August 16, 2007

Certified by ______________________________
Roger D. Kamm
Professor of Mechanical Engineering and Biological Engineering, MIT
Thesis Co-Supervisor

Certified by ______________________________
Mohammad R. K. Mofrad
Assistant Professor of Bioengineering, University of California Berkeley
Thesis Co-Supervisor

Accepted by ______________________________
Lallit Anand
Chairman, Committee on Graduate Students

Mechanotransduction by Talin: A Molecular Dynamics Study of Force-Induced Recruitment of Vinculin to a Focal Adhesion Complex

By

Seung Eun Lee

Submitted to the Department of Mechanical Engineering
on August 16, 2007 in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

It is now well established that cells can sense mechanical force, but the mechanisms by which force is transduced into a biochemical signal remain poorly understood. One example is the recruitment of vinculin to reinforce initial contacts between a cell and the extracellular matrix due to tensile force. Talin, an essential structural protein in the adhesion, contains the N-terminal five-helix bundle in the rod domain with a known cryptic vinculin binding site 1 (VBS1). The perturbation of this stable structure through elevated temperature or destabilizing mutation activates vinculin binding. Here, molecular dynamics (MD) is employed to demonstrate a force-induced conformational change that exposes the cryptic vinculin-binding-residues of VBS1 to solvent under applied forces along a realistic pulling direction. VBS1 undergoes a rigid body rotation by an applied torque transmitted through hydrogen-bonds and salt bridges. Activation was observed with mean force of 13.2±8.0pN during constant velocity simulation and with steady force greater than 18.0pN.

The crystal structure of vinculin head subdomain (Vh1) bound to the talin VBS1 implies that vinculin undergoes a large conformational change upon binding to talin, but the molecular basis for this, or the precise nature of the binding pathway remain elusive. In the second part of the thesis, MD is employed to investigate the binding mechanism of Vh1 and VBS1 with minimal constraints to facilitate the binding. One simulation demonstrates binding of the two molecules in the complete absence of external force. VBS1 makes early hydrophobic contact with Vh1 through an initial hydrophobic insertion. Then, other solvent-exposed hydrophobic residues of VBS1 gradually embed into the hydrophobic core of Vh1 further displacing helix 1 from helix 2. These highly conserved critical residues are experimentally shown to be essential in Vh1-VBS1 binding, and are also the same residues that are shown to become exposed by applied tension to talin in the first part of the thesis. Similar mechanisms are demonstrated in separate MD simulations of Vh1 binding to other VBSs both in talin and α-actinin.

Together, these results provide molecular insights, for the first time, into the early force-induced recruitment of vinculin to the mechanosensitive mechanisms of cell-matrix adhesion complex, and establish the basis for further numerical and experimental studies to fully understand the force response of focal adhesions.

Thesis supervisors: Roger D. Kamm and Mohammad R. K. Mofrad

Biography of Seung E. Lee

Education

Massachusetts Institute of Technology, Cambridge, MA
Ph.D. in Mechanical Engineering, September 2007

University of Illinois at Chicago, Chicago, IL
MS in Mechanical Engineering, December 2002

University of Illinois at Chicago, Chicago, IL
BS in Mechanical Engineering, September 2001

Published works

This dissertation is based on the following papers:

H. Karcher, S. E. Lee, M. R. Kaazempur-Mofrad, R. D. Kamm, "A Coarse-grained Model for Force-induced Protein Deformation and Kinetics," Biophys J. 2006 Apr; 90(8): 2686-97

S. E. Lee, R. D. Kamm, M. R. K. Mofrad, "Force-Induced Activation of Talin and Its Possible Role in Focal Adhesion Mechanotransduction," Journal of Biomechanics, 2007, 40(9):2096-106.

S. E. Lee, S. Chunsrivirot, R. D. Kamm, M. R. K. Mofrad, "Molecular Dynamics Study of Talin-Vinculin Binding," Submitted, 2007.

Awards and honor

Computational Science Graduate Fellowship – Department of Energy [2001–2005]

University of Illinois at Chicago Graduate Fellowship – [2001–2003]

Gates Millennium Scholars Program – Bill & Melinda Gates Foundation [2000–2007]

Korean Honor Scholarship – Korean Embassy in the USA [2005-2006]

Acknowledgements

First and foremost, I would like to express my deepest thanks to my advisors, Roger D. Kamm and Mohammad R. K. Mofrad, for their continuous help and support during the past four years and half at MIT. There were some extremely slow and stressful moments at times, but they have shown me unbelievable amount of patience and respect to get me through this process, and I feel very lucky to have them as my PhD thesis advisors. I also would like to thank my committee members, Nicolas Hadjiconstantinou, Mattew J. Lang, and Bruce Tidor for their guidance that have nudged me into the right direction and their valuable comments that have tremendously improved my dissertation.

I would like to express my thanks to my colleagues, Sid Chung, Helene Karcher, Nathan Hammond, Jeenu Kim, Hyungsuk Lee, and Aurore Zyto for their support, encouragements, and discussions. Also many thanks go to all the members of Kamm Lab for their help and company. They have provided a stimulating research atmosphere for me in the lab, and gave me supportive encouragements at many times.

I am extremely grateful for the generous financial support from Computational Science Graduate Fellowship and Gates Millennium Scholarship that got me through my study.

Finally, I would like to express heartwarming thanks to Bongjoo Kim, Sang Bum Lee, Seongwoo Kang, Yong Kwon Lee, members of KGSAME, and other friends and family for their never-ending support and encouragements during my years at MIT. They provided all the psychological support during difficult times that made the enduring much easier and motivated me to complete this PhD study.

Dedication

This thesis is dedicated to my father,

Yong Kwon Lee,

for his patience and support in making this possible.

Table of Contents

1 Introduction

1.1 Mechanotransduction

Living cells respond to mechanical stimulation in a variety of ways that shape their phenotype in health and disease. Such mechanosensing is essential for cells to probe their environment and respond accordingly to their fate of cell growth, differentiation, or death. Although the biochemical signaling pathways activated by mechanical stimuli have been extensively studied, little is known of the basic mechanisms. At least two general mechanisms of mechanotransduction have been suggested. In one, the mechanical signal is transduced into a chemical signal through protein activation leading to the upregulation of intracellular signaling proteins. Alternatively, the forces transmitted via individual proteins either at the site of cell adhesion to its surroundings or within the stress-bearing members of the cytoskeleton can cause conformational changes that alter their binding affinity to other intracellular molecules. This altered equilibrium state can subsequently initiate a biochemical signaling cascade or produce more immediate and local structural changes; see reviews (1-3). Examples of mechanotransduction include mechanosensitive ion channels, cellular response to substrate stiffness, and force regulated focal adhesion assembly. Many have investigated the signaling cascades that become activated as a consequence of mechanical stress, and these are generally well characterized (4-7). The initiating process, however, by which cells convert the applied force into a biochemical signal, termed "mechanotransduction", is much more poorly understood, and only recently have researchers begun to unravel some of these fundamental mechanisms. Several theories exist that might explain the process of mechanotransduction, but most are still in their infancy. Various mechanisms have been proposed to explain this phenomenon. Some studies have suggested that a change in membrane fluidity acts to increase receptor mobility, leading to enhanced receptor clustering and signal initiation (8,9). Stretch-activated ion channels or strain-induced activation of G proteins represent other means of mechanotransduction (10). Others have focused on the role of mechanosensing surface glycocalyx in endothelial mechanotransduction (11). Mechanical disruption of microtubules (12) or forced deformation within the nucleus has also been proposed (13). Constrained autocrine signaling is yet another mechanism whereby the strength of autocrine signaling is regulated by changes in the volume of extracellular compartments into which the receptor ligands are shed (14). Changing this volume by mechanical deformation of the tissues can increase the level of autocrine signaling. Finally, others have proposed conformational changes in intracellular proteins in the force transmission pathway connecting the extracellular matrix with the cytoskeleton through FAs as the main mechanotransduction mechanism (9,15,16). While all or a subset of these theories may contribute to mechanotransduction, little direct evidence has been presented in their support.

In the rest of chapter 1, a brief introduction will be given for mechanosensitive ion channels and cellular response to substrate stiffness in this subsection, and more detailed introduction will be given in the next subsection for the force regulation of focal adhesion, which is the main topic of this thesis.

Mechanosensitive ion channels. Organisms use mechanosensitive ion channels as means of detecting mechanical stimuli and turn it into electrical signal by opening up the gates for ion current (17). Many of the sensory functions such as hearing and touch are governed by these mechanosensitive channels (18). Even in nonsensory cells, different channels can respond to various mechanical stimuli, such as osmotic pressure, tensile stress on the lipid bilayer, etc (10,19). Propagation of Ca^{2+} waves can be triggered by mechanical stress on Ca^{2+} ion channels, which may occur through tethering of the gate to some cytoskeletal component (20). Membrane tension experienced by an animal cell is usually many orders of magnitude lower than the tension experienced by bacterial membranes (21). The tension needed to activate some of the bacterial mechanosensitive channels is near the lytic limit (11 dynes/cm) of the membrane (22). One example, the bacterial mechanosensitive channel of large conductance (MscL) transmits K^+ at a conductance of 3.2 nS when activated (23). With the available crystal structure of MscL (24), molecular dynamics studies were carried out to investigate the gate opening mechanism (25). Tension applied to membrane transmits force to an interfacial polar group, which causes helix tilting to open the pore interior of MscL (25). Mechanosensitive channel of small conductance (MscS) has a smaller conductance of 1 nS and is characterized by its voltage sensitivity (26). Applied surface tension on the numerical MscS model obtained from the crystal structure (27) provided new insights into the molecular mechanism of channel pore widening (28). Taken together, these studies provide for a molecular-level understanding of MscL and MscS which constitute the bacterial osmotic control system (29). Even in this example, however, the complexity of true biological systems becomes evident in that the osmotic balance breaks down only when both MscL and MscS genes are not expressed (29), and that the MscL knocked out exhibited no observable osmotic malfunction (30).

Probing substrate stiffness. Cell adhesion to solid substrate is essential for cell viability (31). Furthermore, cells have the ability to sense variations in substrate rigidity and respond accordingly (32). This is thought to occur as a consequence of actomyosin contractions that apply contractile forces to the substrate or extracellular matrix via focal adhesions. Only when the substrate has the rigidity to resist contractile force, does the cell form stable focal adhesions and elongated stress fibers, whereas on a soft substrate, the cell displays dynamic focal complexes and few if any stress fibers (33). This rigidity sensitivity has been observed in fibroblast, epithelial, endothelial and smooth muscle cells (33-35). The ability of a cell to sense gradients in substrate stiffness gradient and to migrate toward the stiffer substrate is called durotaxis (36). Presumably, the cell operates in a feedback loop, exerting contractile force to the matrix, measuring matrix stiffness, and responding by forming static focal adhesions when the substrate can withstand the force (32). Although myosin is necessary for the internal generation of contractile force and the subsequent formation of focal adhesion in response to stiff substrate, external forces can also be applied to the cell to promote focal adhesion formation even in the absence of myosin contractile force (37).

1.2 Force regulation of focal adhesion assembly

A focal adhesion (FA) is a protein complex forming a physical linkage between cytoskeleton and extracellular matrix (ECM) (Figure 1.1). A cell can use FAs to gain traction on the ECM surface during the process of spreading and migration. The FA protein complex contains a rich mixture of structural proteins as well as signaling proteins, and can also be used by a cell to probe the stiffness of its environment (38,39). A sequence of maturation stages leads to the formation of focal adhesion complexes. *Initial adhesions* consist of the minimum essential set of proteins to link the ECM with the cytoskeleton, and is only able to withstand tensile force on the order of 2 pN (40). When an initial contact is activated or sustains a force, it grows into a focal complex, which is a short-lived dot-like (~ 1μm) adhesion structure containing, for example, vinculin, paxillin, α-actinin, and Arp2/3 (41-43). The focal complex can eventually mature into a focal adhesion, which is characterized by its larger size (1-10μm), elongated shape, and association with stress fibers (38). Arp2/3, which is related to actin nucleation and polymerization, is absent in focal adhesion making it relatively more static in nature compare to the focal complex (41).

Focal complexes require a stiff substrate in order to mature into a stable focal adhesion (33). The myosin-mediated contractile force transmitted to the ECM and the tension applied to the adhesion complex is necessary for promoting focal adhesion development (44). However, externally applied mechanical force can also promote the formation of focal adhesions (40) even in the absence of myosin contractile force (37). Galbraith *et al.* (40) employed fibronectin-coated micro-beads of various diameters to study the force response of adhesion strengthening. When small diameter beads (1μm) are attached to lamellipodia of fibroblasts no focal complexes are formed in the submembrane beneath the bead. However, focal complexes are formed when larger diameter beads (> 3μm) are attached to the fibroblasts. Interestingly, focal complexes are formed beneath the 1 μm bead when external tensile force is applied using an optical laser trap (Figure 1.2). With the larger beads, the cell was able to apply intracellular myosin-mediated contractile force leading to focal complex formation, whereas the small beads (1 μm) could not apply intracellular contractile force, and only formed focal complexes with externally applied force.

A GTPase protein Rho is involved in the formation and regulation of focal adhesions in response to growth factors (45,46). Two downstream targets of Rho are Rho-associated kinase (ROCK) and Dia1, which are responsible for myosin II activation (47) and nucleation of actin polymerization (48), respectively. By promoting cell spreading through focal adhesion regulation, Rho may be a key regulator of cell proliferation (38). In the absence of either cell adhesion or growth factors, the cell is sent to the path of apoptosis (49). Tension applied to focal adhesions can activate a number intracellular signaling proteins including focal adhesion kinase (FAK) (50), which can in turn phosphorylate other focal adhesion proteins. FAK may be involved in promoting cell migration as it activates Rac, another Rho family GTPase, which is involved in lamellipodia formation (45). With the evidence that FAK suppresses Rho activity (51), it is likely that a cell uses this collection of signaling proteins and focal adhesions in deciding on spreading, migration, or programmed death.

1.3 Molecular basis for the early adhesion complex

Molecules involved in the early adhesion complex. Applied tension on initial adhesion between ECM and cytoskeleton allows recruitment of vinculin to reinforce the linkage to form a focal complex (40). Jiang *et al.* (52) identified that the initial adhesion consists of an ECM-integrin-talin-F-actin linkage. Separate from the force-regulated signaling pathways, the local immediate force response of adhesion re-assembly is thought to be through conformational changes in the linking proteins that enhance their binding to other reinforcing proteins (2,3). Indeed, talin1 is critical in force-dependent vinculin recruitment to adhesion sites independent of Src family kinase and focal adhesion kinase activities (53). Talin is present in both the initial adhesion and the focal complex, whereas vinculin is only found in the focal complex (40). Vinculin has binding sites for both talin (54) and F-actin (55), making it a candidate for strengthening an initial adhesion. Interestingly, talin contains 11 potential vinculin binding sites with a number of them known to be cryptic (56-58). Therefore, the proposed model for force regulation of early adhesion strengthening is (i) tension generated within initial adhesion linkage either through intracellular contractility or externally applied force gets transmitted to talin, (ii) the tensile force within talin exposes the cryptic vinculin binding site to solvent, and (iii) inactive vinculin within the cytosol gets targeted to the adhesion site and reinforces the connection (Figure 1.3).

Focal adhesion protein: vinculin. Vinculin is a highly conserved cytoplasmic protein that functions as a structural reinforcing link for cell-cell and cell-matrix junctions. It consists of a globular head domain, a proline-rich neck region, and a rod-like tail domain, which contain binding sites for numerous other cytoplasmic proteins (59,60). Vinculin head is known to bind to α-actinin (61) and talin (54), whereas vinculin tail is known to bind to paxillin (62), F-actin (55), and phosphatidylinositol 4,5 bisphosphate (PIP2) (63). The neck region binds to VASP (64), vinexin (65), and ponsin (66). In its inactive state, vinculin head binds to vinculin tail masking its cryptic binding sites for many of its ligands (67,68). A cell with vinculin disruption can still form FAs, yet it displays reduced ability to spread and increased cell motility (69). Therefore, along with its structural function, vinculin has also been suggested to be a regulatory protein in cell adhesion (70).

It has been shown that PIP2 can disrupt the autoinhibitive vinculin head and tail interaction (67). However, recent studies suggest that talin might also play a role in vinculin activation (71,72). Bakolitsa *et al.* (73) suggested a combinatorial pathway in activating vinculin, where PIP2 partially releases vinculin tail from the head exposing the talin binding site, and the head and tail interaction is severed as talin binds to vinculin head. The crystal structure of the full-length vinculin in the inactive autoinhibitory conformation provides many insights into vinculin function (73) (Figure 1.4A). Vinculin tail domain (Vt; residue 897-1066) is bound most significantly to D1 domain (residue 1-258) and relatively weakly to other domains (73). Izard *et al.* (71) reported the crystal structure of vinculin head subdomain (residues 1-258; same as D1 domain in ref (73)) associated to Vt domain. The conformation of this complex is almost identical to the vinculin head and tail structure within the full-length vinculin structure; Figure 1.4B is showing just vinculin head and tail from the full-length vinculin. In the same paper, they also reported vinculin head associated with a talin vinculin binding site, and suggested a mechanism for activation of vinculin by talin binding. Intriguingly, talin peptide binding to vinculin head induces large conformational change to vinculin, named 'helical bundle conversion,' where talin peptide is embedded in the N-terminal four-helix bundle of vinculin head forming a combined five-helix bundle (71). A model for the activated vinculin with vinculin head and tail dissociated is shown in Figure 1.4C. The actual conformational state of the activated vinculin is, however, unknown.

Focal adhesion protein: talin. Talin is another cytoplasmic protein with a globular head domain and an elongated rod domain that provides a structural link between integrin and actin cytoskeleton (70), which forms a dimer at concentrations in excess of 0.7 mg/ml (74). Talin's globular head domain has the FERM (4.1, ezrin, radixin, and moesin) motif that binds to cytoplasmic tails of certain β-integrins (75). FERM can also bind to and activate phosphatidylinositol 4,5 phosphate kinase type γ (PIPKI-γ) (72,76), which can, in turn, locally increase the production of PIP2 (72). Since PIP2 is known to activate a number of focal adhesion proteins (e.g. vinculin, talin, and paxillin), it has been suggested that talin might play an important role in rapid regulation of cell-matrix interactions (72). The talin extended rod domain contains binding sites for integrin (77) and F-actin (78). Talin rod domain is also initially known to have three high affinity vinculin binding sites where each binding sequence forms an amphipathic helix (79), however, a recent study shows that talin may have total of 11 potential vinculin binding sites (VBSs) (56). Bass *et al.* (80) showed that initially identified three vinculin binding sites are mutually exclusive, therefore suggested, given also that they are all amphipathic helices, that they bind to the same site on vinculin through an identical binding mechanism. Since the vinculin-binding sites share only a partial sequence identity (80) and α-actinin, another cytoplasmic protein, is shown to bind at the same binding site on vinculin (71), this suggests that the binding site on vinculin for vinculin-binding-sites of talin or the α-actinin is not highly specific. Unlike vinculin disrupted cells, which still formed FA (69), mouse embryonic stem cells with disrupted talin genes failed to form focal adhesions and showed spreading defects (81).

The crystal structure of full-length talin is not available, but subdomains of the talin rod (talin residues 482-789) containing the VBS1 (57) and talin rod (talin residues 755-889) containing the VBS2 (82) are reported. Although VBS1 can strongly bind to vinculin (83), the five-helix bundle containing VBS1 is unable to bind to vinculin as the binding surface is cryptic (58). The crystal structure determined by Papagrigoriou *et al.* (57) is the N-terminal nine-helix bundle of the talin rod containing the VBS1 and two more VBSs is shown in Figure 1.5 (VBSs are red helices).

Vinculin-talin binding interactions. VBSs cab bind to Vh1 at high binding affinities (83), but longer full-length talin only weakly binds to Vh1 (58) indicating that many of the VBSs are cryptic. Vh1 is a subdomain of vinculin head that contains the binding site for talin and is used in various talin-binding experiments (58,71). In a thorough mutational study on VBS1 on binding to Vh1, most of the hydrophobic residues of VBS1 that are embedded within Vh1 are shown to be important in stable binding to Vh1 (57). Interestingly, these same hydrophobic residues are also embedded within the hydrophobic core of N-terminal five-helix bundle of talin rod (TAL5) (57). Experiments have shown that isolated TAL5 has a low binding affinity for Vh1, whereas a four-helix-bundle with helix-5 (H5) removed from TAL5 (58), a mutated TAL5 with an unstable hydrophobic core (57), or the wild-type TAL5 molecule in elevated temperature solvent (58) can each disrupt TAL5 stability and strongly bind to Vh1. One working hypothesis is that the tensile force transmitted through TAL5 is the destabilizing cue that exposes the cryptic binding surface of VBS1, but direct evidence has been lacking.

1.4 Experimental and numerical studies on force response on proteins.

Molecular dynamics (MD) has been used to study the force response on various molecular structures (84-86). Mostly, unfolding pulling forces were applied between two atoms (e.g. C- and N-termini), and the results were compared with the corresponding Atomic Force Microscopy (AFM) experiments (87,88). AFM systems usually involve a protein domain or a tandem repeats of the domain anchored at the bottom and attached to the AFM probe tip (Figure 1.6A). A theoretical model representing the protein unfolding energy landscape by parabolic well is shown in Figure 1.6B. As the probe tip moves up at constant rate, the molecular extension and force applied can be measured; it has the characteristic sawtooth shape with each force drop corresponding to sequential domain unfolding (Figure 1.6C) (88-91). Corresponding MD studies, called steered molecular dynamics (SMD) (92), usually involve holding one terminus fixed and pulling on the other terminus at constant velocity or constant force (Figure 1.7) (84,85,93). Similar force drops were observed as force-bearing hydrogen bonds within the structure are broken, and the molecules displayed a number of stable intermediates during the unfolding pathways (85,86,93,94). The importance of using explicit water molecules in these unfolding simulations was emphasized as a single water molecule can play a critical role in breakage of these force-bearing hydrogen bonds (85,94). During the forced unfolding simulations, fibronectin exposed its cryptic binding site in the unfolding intermediate (85) suggesting a possible force regulation pathway to mechanotransduction.

Fibronectin contains multiple type III modular domains with considerable sequence variability, but with surprisingly high structural homology (2,88). Structural stabilities of these modular domains vary widely due to the different residue sequences as determined by both SMD and AFM experiments (88,94). Such variation may actually have a functional role in sensing different levels of force applied to fibronectin, especially if each module has different molecular recognition sites (2). The relative variation in rupture force measure using AFM (88) agrees very well with the relative variation measure using SMD (94), but the overall magnitude of the rupture force estimated by SMD is usually overestimated compared to the AFM measurement (2). The loading rates between numerical and experimental studies are many orders of magnitude different with the SMD loading being much faster. Since the rupture force for molecular unfolding is dependent on the pulling rate (95), force predicted by SMD tends therefore to be larger than those measured by AFM.

AFM experiments on a tandem repeat spectrin construct have identified more than one unfolding pathway of a single domain and simultaneous multi-domain unfolding (96,97). A SMD study from the same group has helped to identify a possible molecular mechanism for such multiple unfolding pathways, which was related to the hydration of the backbone of the linker helix (98).

SMD simulation has also been used to calculate the free energy landscape along a predetermined reaction coordinate (99,100) by utilizing the SMD and Jarzynski's equality (101). Zhang *et al.* (100) identified the most probable unbinding pathway of acetylcholine from the human alpha7 nicotinic acetylcholine receptor by evaluating the potential of mean force (PMF) along multiple trajectories and selecting the path requiring the lowest free energy cost. Umbrella sampling (102) is another method to obtain the PMF along a reaction coordinates (99,103).

1.5 Thesis objectives

Force regulation of focal adhesion assembly has been extensively studied experimentally (38,39). Vinculin is recruited to the initial adhesion site with applied external force, and the reinforced linkage eventually matures into a focal contact (40). Although a model of force-induced activation of talin is proposed to recruit vinculin (Figure 1.3), the molecular basis of talin activation or vinculin binding to talin is still elusive. MD has been employed in studying the force response of protein domains, but most studies focused on understanding the forced unfolding of proteins that complements AFM experiments (2), and no MD studies on force-regulated focal adhesion assembly are reported to date.

Therefore, the main goal of this thesis is to explore the possibility of MD in investigating more realistic molecular events not observable through experiments, namely, force-induced vinculin recruitment in early cell-matrix adhesion complex. Abundant structural (57,71,73) and experimental (56-58) data on talin and vinculin are available to support this numerical study. The outcome will not only provide molecular insights on this mechanotransduction pathway, but similar MD methods can be employed for studying other force-related protein complexes.

The remainder of this thesis is organized as follows. Chapter 2 describes the model setup, simulation methods, data analysis, and computer resources used for the molecular systems presented in this work. As a simplest possible protein model, an α-helix model is considered in Chapter 3 and used, along with a coarse-grained, theoretical model, to study force-induced protein deformation and conformational variation. This simple system employs many of the forcing techniques used in the subsequent chapters.

The result from the coarse-grained model is shown to be in good agreement with the result from MD simulation using this simple α-helix model. Chapter 4 presents force-induced activation of a talin subdomain involving hydrogen-bond mediated VBS1 rotation to expose the cryptic binding site to the solvent. This is the *first* direct evidence of forced activation of talin that is potentially taking place during the early adhesion assembly. A vinculin subdomain and talin subdomain binding mechanism is presented in Chapter 5 with initial hydrophobic interactions between the molecules followed by a final locking mechanism that induces a large conformational change to the vinculin subdomain. Binding of the two molecules proceeds with no or little external force, which is the first reported natural binding simulation that involves large conformational change upon binding. Finally, all the findings of these simulations are summarized and future directions are discussed in Chapter 6.

1.6 Figures

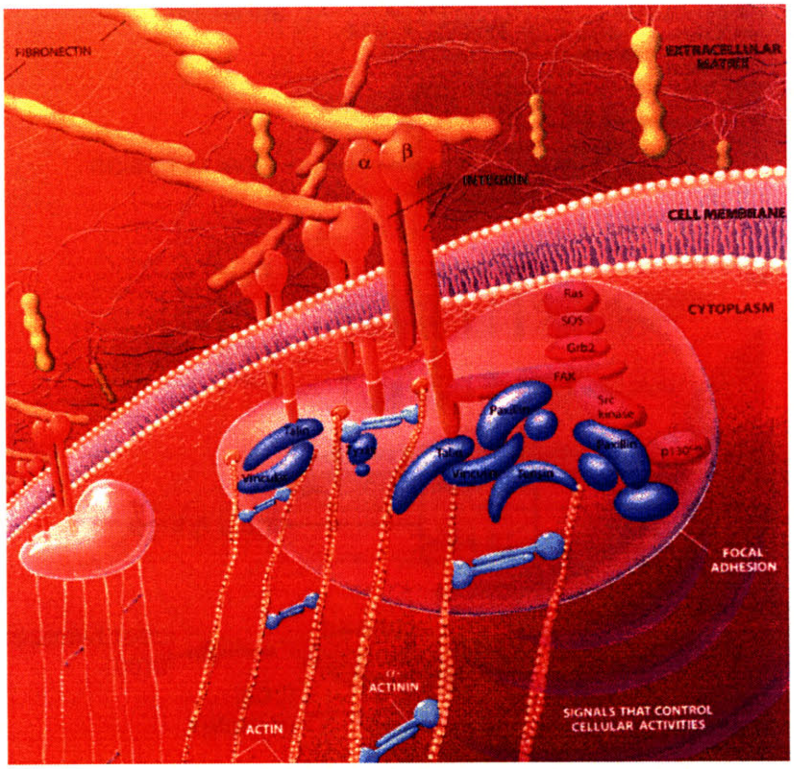

Figure 1.1. Image of focal adhesion protein complex linking the cytoskeleton with the cytoplasmic domain of the transmembrane protein, integrin. The integrin runs through the plasma membrane and is linked to the extracellular matrix. Focal adhesion protein complex consists of many structural and signaling proteins.

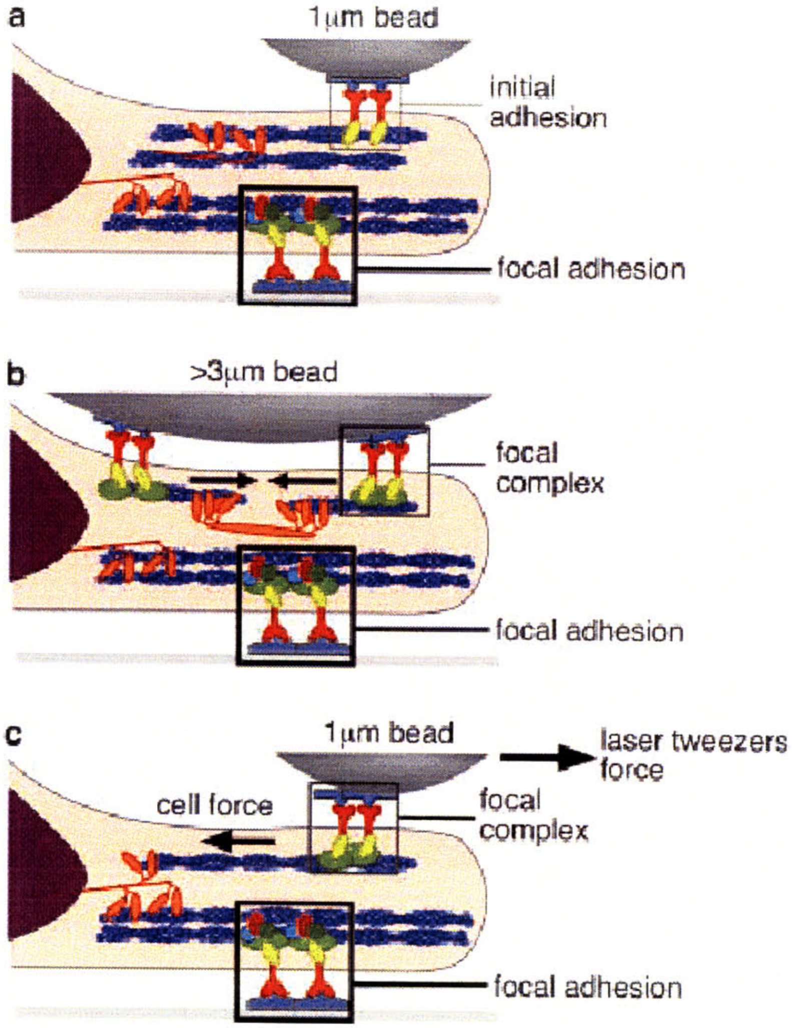

Figure 1.2. (a) Small 1-μm bead coated with fibronectin forms initial adhesion on the cell surface, but does not mature into focal contact or focal adhesion. (b) Beads of larger size (> 3-μm) can induce focal complex formation as the cell exerts myosin generated tensile forces on the bead. (c) With pulling force applied on the 1-μm bead using laser tweezer, the initial adhesion can mature into focal complex. Adapted from (40).

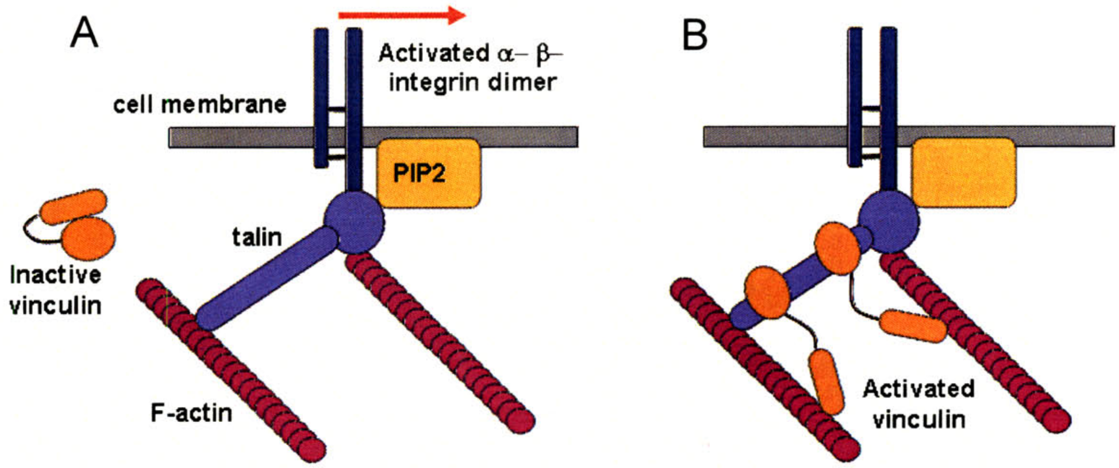

Figure 1.3. (A) Model of the initial adhesion consisting of integrin-talin-F-actin linkage. The vinculin is present in the cytosol as inactive, autoinhibitory conformation, and tensile force is applied to integrin from the outside of the cell membrane. (B) The transmitted force through the linkage alter the talin configuration, and recruits vinculin to reinforce the initial adhesion linkage.

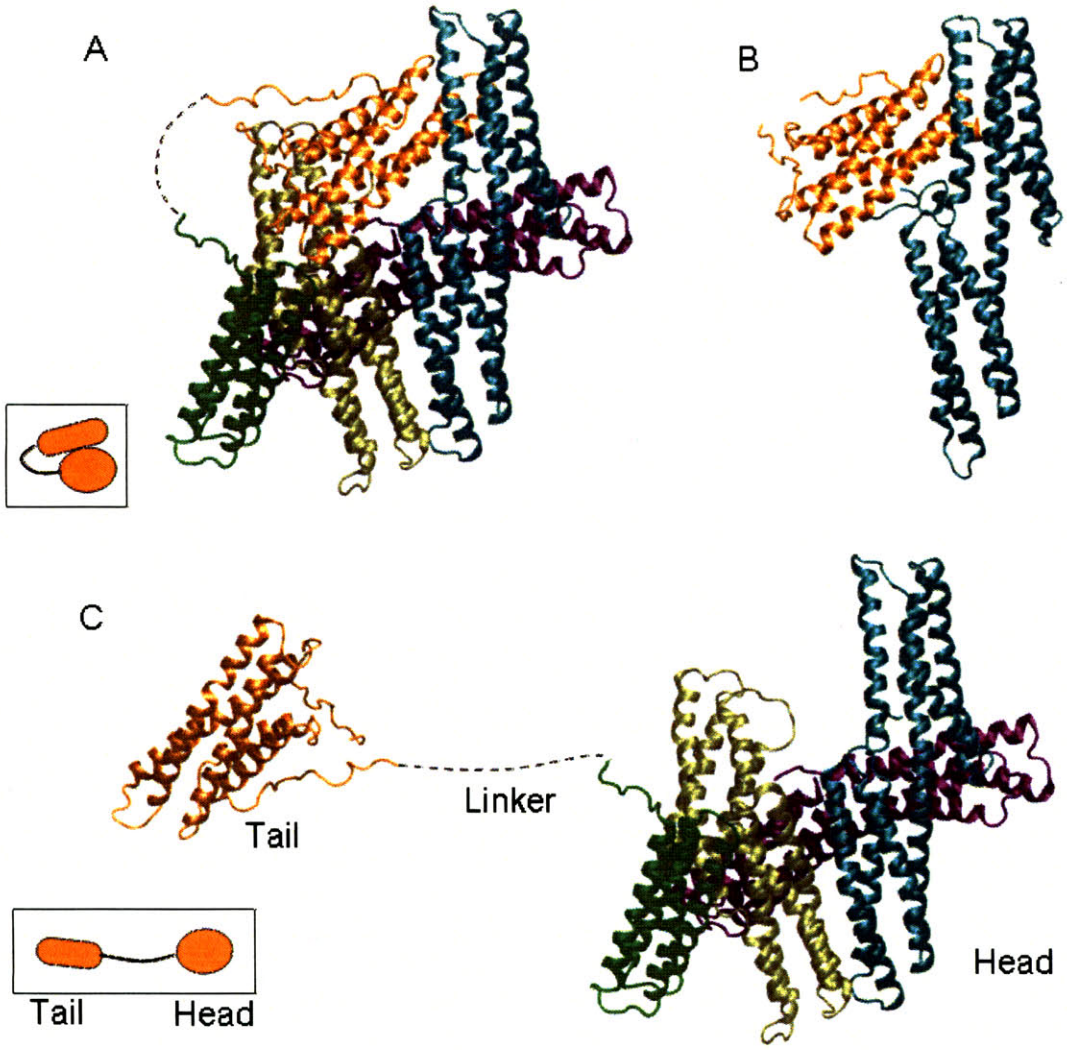

Figure 1.4. (A) Crystal structure of the full-length vinculin in its autoinhibitory conformation with D1 (cyan), D2 (purple), D3 (tan), and D4 (green) connected to Vt (orange) through a flexible linker. Inset box shows inactive vinculin model as shown in Figure 1.3. (B) Same structure as Figure 1.4A, but only showing the D1-Vt domains. This configuration has an almost identical conformation as the D1-Vt crystal structure (71). (C) Hypothetical activated vinculin when D1-Vt interaction is broken. Inset box shows the corresponding activated vinculin model similar to ones shown in Figure 1.3.

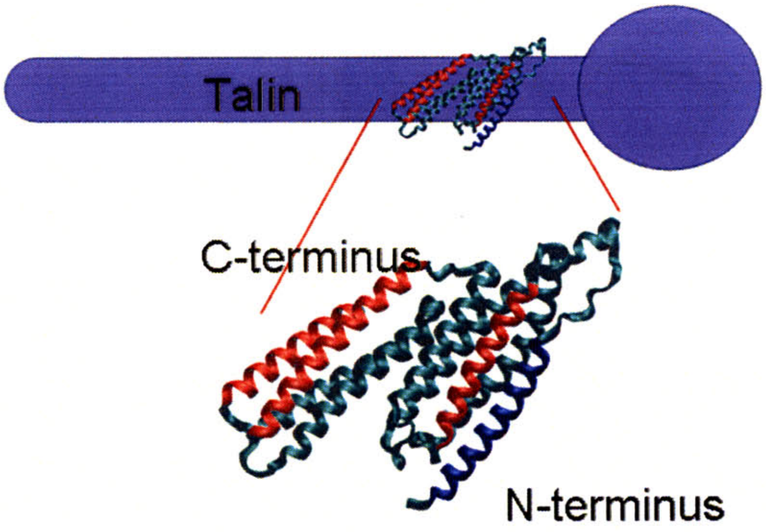

Figure 1.5. Crystal structure of the nine-helix bundle of talin rod subdomain as it is assumed to be oriented within the full-length talin model as shown in Figure 1.3. Three of the potential 11 vinculin binding sites are indicated as red helices.

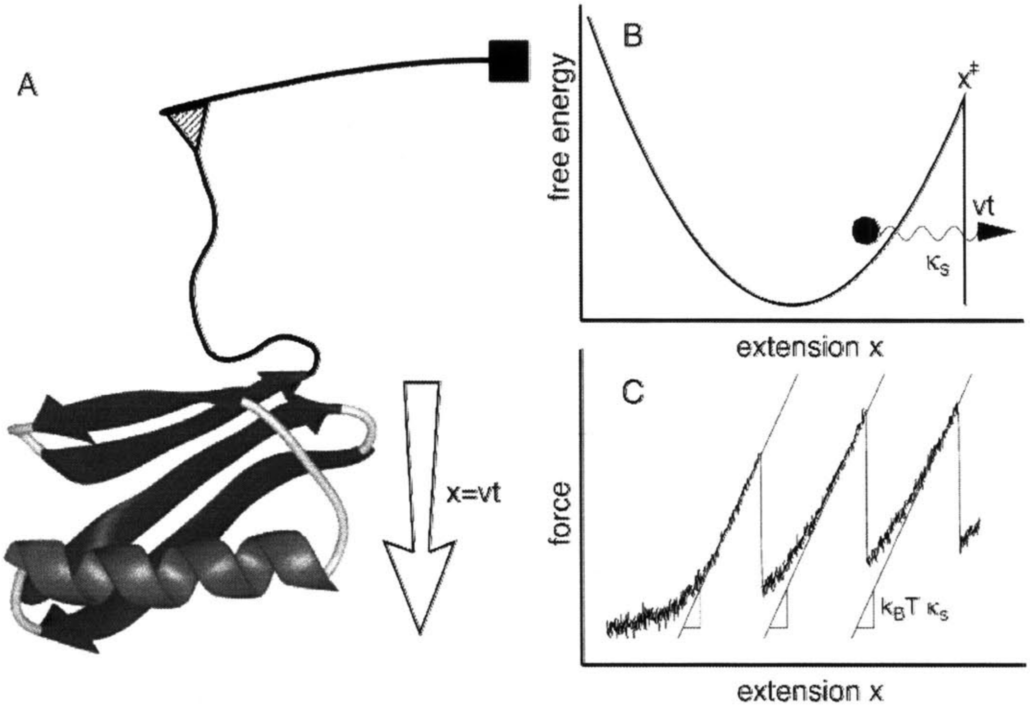

Figure 1.6. (A) A setup for AFM single molecule pulling experiment. (B) Parabolic free energy surface representing unfolding of pulled molecule. (C) A sawtooth force-extension curve obtained from AFM experiments, where each peak represents the unfolding of one molecular domain repeat. Adapted from (95).

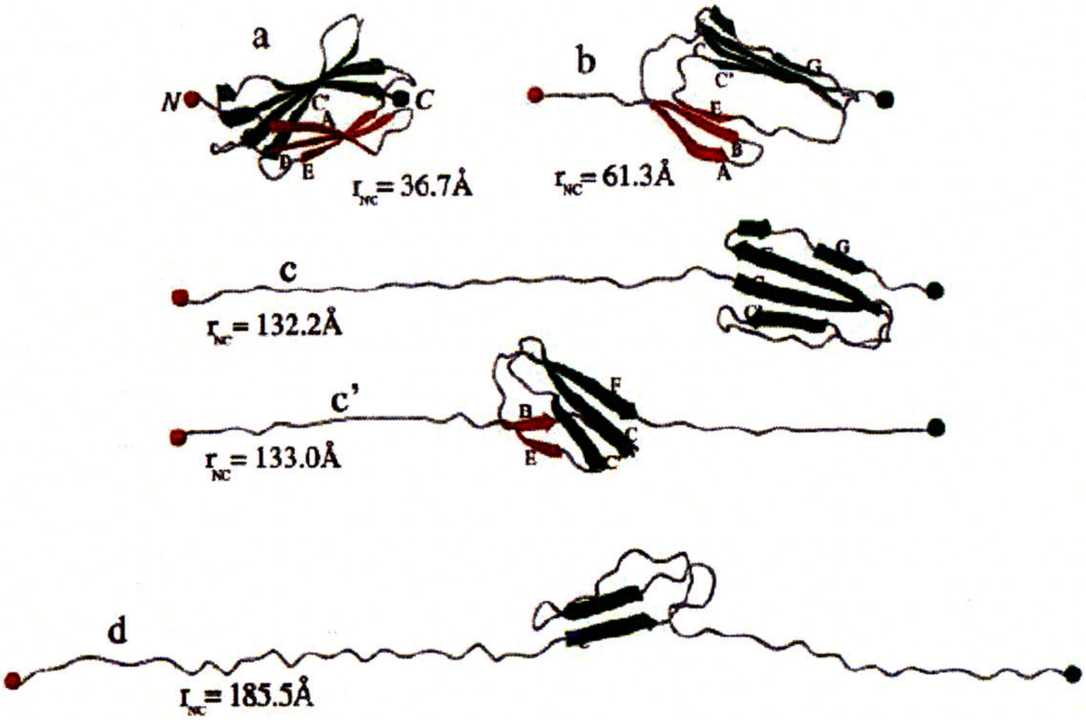

Figure 1.7. Molecular dynamics study of forced unfolding of ^{9}Fn3 titin domain. At constant force pulling of 200pN, (a) crystal structure, (b) first intermediate, (c and c') two intermediates found near r_{NC} = 140Å, and (d) final configuration at the end of 1400ps simulation. Adapted from (93).

2 Methods

The methods of employing MD for molecular system involves preprocessing of the available structural data into a molecular model, setting up and running the simulation on the molecular model, and postprocessing of the trajectory data into representative plots describing the observed mechanism. Crystal structures obtained from the PDB Databank usually have missing atom coordinates that need to be approximated and sometimes are needed to be trimmed to obtain the model for the structural domain of interest from a larger structure. Appropriate setup for the numerical experiments on the numerical model requires pre-analysis of available experimental data in terms of applying constraints in order to simulate the desired numerical experiments. On the resulting MD trajectories, creative data processing techniques are needed for pertinent interpretation of the proposed arguments. In this chapter, the general details of model preparation and model setup are described. The actual scripts used for these steps and data analysis appear in the Appendix section, which provide enough details for the reproduction of the presented work or the necessary tools to perform the subsequence future studies.

2.1 Model preparation

Commercial molecular dynamics software, CHARMM (Harvard University, Cambridge, MA) (104), was used with polar-hydrogen representation CHARMM19 force field (105) for implicit solvent simulations, and all-hydrogen representation CHARMM27 force fields (106) for explicit solvent simulations. Crystal structure of the molecule of interest is obtained from the online PDB Databank (www.rcsb.org). The PDB file may need manual alteration if some of the residue names are inconsistent with the notation in the force field. A typical example is the notation of histidine, which need to be either HIS, HSD, or HSC depending on the protonation state in CHARMM19. Residue notation for histidine other than these three using CHARMM19 will result in error in reading the PDB. Both the residue sequence and the coordinates will be read in from the PDB file. In most cases the hydrogen coordinates and also coordinates of some atoms will be missing, which can be filled in within CHARMM by using the topology file and the internal coordinates of the molecule. In cases where a short residue sequence is missing in the middle of a molecule, manually input the residue sequence inside the PDB file with default coordinate values of (x = 9999.000, y = 9999.000, z = 9999.000). This will produce the correct molecular topology, then the approximate coordinates can be incorporated by 'coordinate fill' command. This technique can be used to manipulate the total number of residues within a molecule or even to splice two separate molecules into one.

The molecular model built needs to be thoroughly energy minimized before being used for the molecular dynamics study. The energy minimization will relax high energy atomic interactions, especially when the atomic coordinates are approximately filled in by the 'coordinate fill' command. It is important to use the identical simulation parameters in the minimization as will be used for the MD simulation. For example, SHAKE constraint used to fix the bond lengths between hydrogen and heavy atoms (107) and the solvent parameters to be used in the upcoming MD simulation must be set prior to the energy minimization. By fixing the high frequency movement of hydrogen bonded to heavy atoms, SHAKE allows the use of larger time step (2 fs) in all of the simulations used in this work. The molecular structure is minimized by alternating the Steepest

Decent and Adopted Basis Newton Raphson methods with 3000 steps, and at each minimization set reducing the level of atomic constraints. This allows a thorough minimization of the given molecule without deviating it too far from the original conformation.

Sometimes it is convenient to orient the molecules in predefined directions when applying tensile force or in analyzing one molecular complex with respect to another molecular complex that share the same domain. A typical way of achieving this is to defined three points in space, usually the center of mass of three collections of atoms (e.g. helix 1, helix 3, and helix 4-6), and align the vectors defined by these points to the principle axes using vector calculus.

2.2 Solvent models and nonbonded parameters

Selection of solvent model to use and the nonbonded parameters has the most direct effect in the accuracy and the efficiency of the MD simulations. For implicit solvent simulation, the Effective Energy Function (EEF1) (108) solvent model was used with the CHARMM19 force field (105). EEF1 is an empirical method that has considerably high computational efficiency compared to other theoretical implicit solvent models (109,110). Many pulling simulations have used EEF1 and reported reasonable trajectories (108,111,112). Most importantly, equilibration simulations are performed using various implicit models and explicit model as a test, and both the EEF1 and the explicit simulation resulted in similar trajectories with most of the critical hydrophobic and hydrogen-bonding interactions intact. Therefore, EEF1 was used for most of the simulations presented in this thesis, and other related simulations and corresponding explicit simulations were performed afterward to check for consistency. Use of EEF1 needs to be invoked in CHARMM command line and the default nonbonded parameters of (CTONNB=7.0, CTOFNB=9.0, CUTNB=10.0) are used.

```
eef1 setup temp 300.0 unit 93 name solvpar.inp
update ctonnb 7.0 ctofnb 9.0 cutnb 10.0 group rdie
```

The use relatively short cutoff distance in EEF1 is to enhance its computational efficiency and the correction for the long-range effect beyond the 9Å cutoff has been hard-coded in the EEF1 method (108).

For explicit water simulation, the molecular system needs to be solvated in a water block. There are number of water blocks that can be used: spherical, orthorhombic, cubic, rhombic dodecahedral, and etc. However, in order to use periodic boundary condition, only the water blocks with flat faces can be used. The orthorhombic box can be used for elongated shaped molecule (Figure 2.1A) to minimize the total number of atoms and still have at least 10Å of solvent buffer from the edge of the molecule to the edge of the solvent block. The solvated system is obtained by overlapping the solid water block with the molecule and removing all the overlapping water molecules. There are a few disadvantages for using orthorhombic water blocks. Constant pressure simulation does not work on orthorhombic box, and the elongated molecule can rotate and have a region of it sticking out of the box and interacting with itself. This problem can be addressed by using constraints to prevent the molecule from rotation, or by using cubic or rhombic dodecahedral water box.

Rhombic dodecahedral box is created from a spherical shaped water block. Rhombic dodecahedral periodic boundary condition is applied on the surface of the spherical block and the system is minimized until the box is obtained (Figure 2.1B). It benefits of having about 1/3 fewer water molecules than a corresponding cubic box, but with uniform solvation from the center of mass of the molecule. This reduction in number of water molecules amounts to a significant computational time save, and should be ideal for running constraint-free simulations that is free to rotate. As an example, Figure 2.1B has a globular shaped five-helix bundle system for simulations with elevated temperature. It is recommended to make the total charge of the solvated system to zero, especially when using periodic boundary condition. The sodium (Na^+) or chloride (Cl^-) ions can be added in place of random water molecules in the system to neutralize the system. The system in Figure 2.1A already has a charge of zero, and no ions were added. The system in Figure 2.1B has a charge of -6, so six sodium ions (yellow spheres) were added to neutralize the system. When studying the effect of salt concentration on the molecular behavior, additional sodium and chloride ions can be added to neutralize the system and also match the salt concentration based on the volume of the water block.

An all-hydrogen representation was used with CHARMM27 force fields (106) for all the explicit water simulations. The SHAKE constraint and 2 fs time-step was used. When EWALD method is not used to treat the nonbonded interaction beyond the cutoff distance, SHIFT truncation was imposed with nonbonded parameters of (CTONNB=10.0, CTOFNB=12.0, CUTNB=14.0), which has been found to exhibit reasonable accuracy in explicit water simulations (113). The following commands should appear in CHARMM script for turning on the periodic boundary conditions for non-EWALD simulations.

```
open unit 88 read form name stream/waterbox.img
read image card unit 88
image byres sele .not. prot end
image byseg sele prot end
```

For using EWALD, SHIFT truncation was imposed with nonbonded parameters of (CTONNB=8.0, CTOFNB=10.0, CUTNB=12.0), and the following commands are used to switch on the periodic boundary condition.

```
crystal define ortho @XSIZ @YSIZ @ZSIZ 90.0 90.0 90.0
crystal build cutoff 12.0
image byres sele .not. prot end
image byseg sele prot end
energy ewald pmewald kappa 0.34 order 6 -
       fftx 64 ffty 128 fftz 64 qcor 0.0
```

2.3 Force manipulations and constraints

Harmonic constraint is used to hold an atom near a point in the Cartesian space, and applies a force, $F = k*(x_o-x)$, to the atom where k is the force constant, x_o is the reference position, and x is the current position. Nuclear-Oberhauser-Effect (NOE) constraint in CHARMM correspond to atom pair distance constraint, where the restoring force is only applied when the pair distance exceeds the specified reference distance. For example, force of $F = k_{max}*(d_{max}-d)$ is applied to the atom pair only if $d > d_{max}$, where k_{max} is the force constant, d_{max} is the maximum reference distance, and d is the current distance between the atom pair. Similarly, a NOE lower bound constraint can be applied by specifying k_{min} and d_{min}, where the atom pair coming closer than d_{min} is now pushed away from each other. NOE constraint is useful in controlling parts of the molecule from deviating too much from the original configuration or to confine two or more molecules to near each other. A good example of using NOE constraint is to constrain the backbone hydrogen bonding pairs within an α-helix to 3.5Å. This will ensure the α-helix to retain its secondary structure during the forced MD simulations. A constant force can be applied to a collection of atoms to a pre-specified direction. The specified force is applied to each atom in the collection, therefore the total force applied to the molecule is $F_{tot} = N_{applied}*F_{specified}$, where $N_{applied}$ is the number of atom selected. Constant velocity pulling can be achieved by harmonically constraining two end atoms (e.g. two termini) of a molecule to two dummy atoms, and holding one dummy atom fixed in space and moving the other at constant velocity away from the fixed atom. It would be convenient to align the molecules along the forcing direction during the system setup to simplify the force application and the data analysis. Force response on molecular systems was investigated by employing various combinations of these constraints and applied forces.

2.4 Computer resources

All the simulations in this work are performed on 8-node (2 processors per node) Linux cluster in our lab and at San Diego Supercomputer Center (SDSC) using DataStar IBM p655. Most of the implicit simulations were conducted at the Linux cluster using one node per simulation, and the explicit simulations were conducted at SDSC using 64 or 96 processors. In order to use the computer resources more efficiently, benchmark simulations were run on various solvent models with different number of processors (Table 2.1). The same molecular system was used for all four cases in Table 2.1. The differences in the number of atoms arise from the use of different solvent method. The simulation time is indicated as seconds per 1000 steps of simulation, e.g. using 2 processors solving EEF1 method needs 27 seconds to simulate 1000 time-steps. In comparing GBSW and EEF1 implicit methods using 2 processors, GBSW is more than 20-fold more computationally intensive than EEF1 simulation. Using 4 processors in solving explicit simulation at the Linux cluster and at SDSC, simulation at SDSC was about 4-fold faster. Plots of simulations at SDSC for solving explicit system at different number of processors and simulations at the Linux cluster for solving GBSW system at different number of processors are shown in Figure 2.2. The black dashed lines represent the ideal speed up with increase in the number of processors. The plots indicate that the scalability is reasonable for the explicit simulations (47,500 atoms), but poor for the GBSW simulations (4400 atoms), which is probably due to doing more calculations per processors than message passing for larger system (i.e. explicit system). Consistently, going from 2 processors to 4 processors for EEF1 simulation (2600 atoms) increased the time from 27 to 38.7 seconds per 1000-steps. For the best use of the computer resources, all the implicit simulations were carried out at the Linux cluster using one node each time and all the explicit simulations were run at SDSC.

2.5 Table and figures

# of Proc	SDSC 47,500 atoms Cutoff=10Å Exp, EWALD	Linux 47,500 atoms Cutoff=10Å Exp, EWALD	Linux 4,400 atoms Cutoff=14Å GBSW	Linux 2,600 atoms Cutoff=9Å EEF1
2	3474	-	645.5	27
4	1749.6	6017.1	370	38.7
6	-	-	280.4	-
8	959.4	-	234	-
10	-	-	208.1	-
12	-	-	193.3	-
16	621	-	-	-
32	360	-	-	-
64	207	-	-	-

Table 2.1. Time in seconds for calculating 1000-steps of corresponding molecular systems. These are all from the same protein system with different salvation methods.

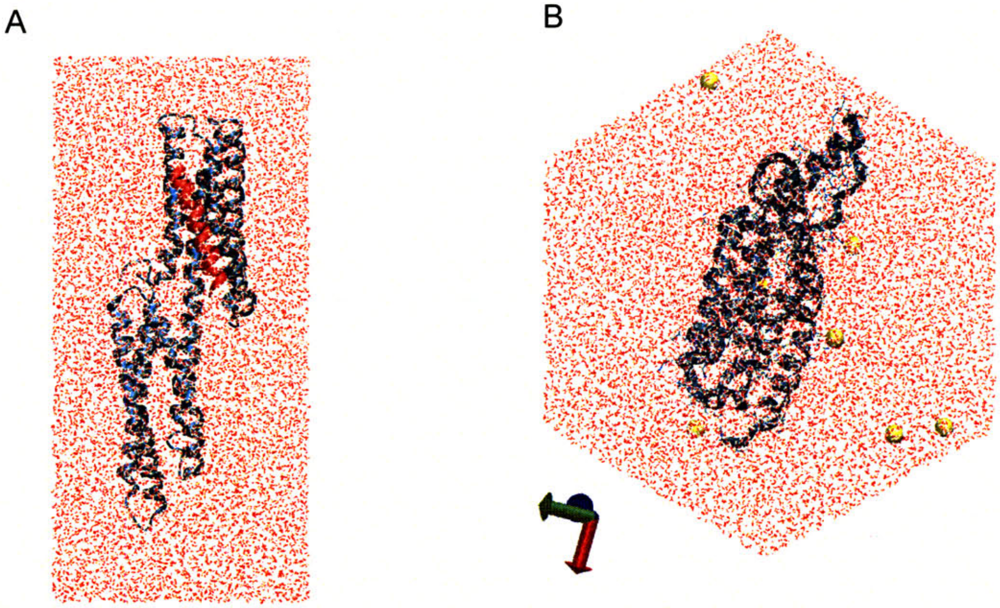

Figure 2.1. (A) Vinculin subdomain (Vh1) and talin VBS1 solvated in a orthorhombic water box. The net charge of the system is zero. (B) N-terminal five-helix bundle of talin rod containing cryptic VBS1 is solvated in rhombic dodecahedral water block. The yellow spheres represent sodium ions added to neutralize the total charge of the solvated system.

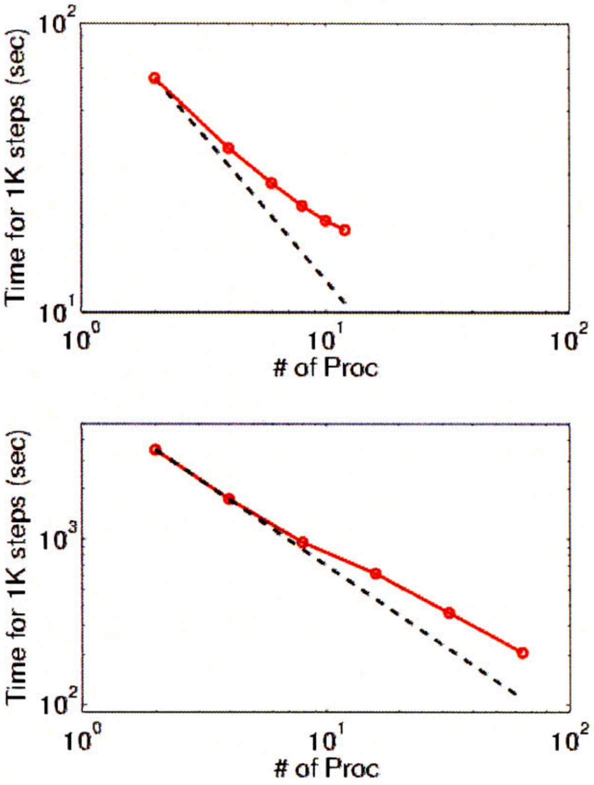

Figure 2.2. Time requirement for solving 1000 time-steps when using various number of processors. (Above) Using in-house Linux cluster for simulating GBSW system. (Below) Using SDSC for simulating explicit water system. The data points are from the Table 2.1. The dashed lines represent the ideal speed-up with increased number of processors.

3 Simple α-helix Model for Force-induced Extension and Kinetics*

3.1 Introduction

A generic coarse-grained model linking force to protein conformational change was developed and analyzed in terms of the mechanical properties of the protein states. Assuming that binding is a force-independent event and occurs preferentially in one conformation (relaxed or extended), our model links force applied to a protein to its propensity to initiate a signal. We consider a simplifying case of a protein having just two conformational states: C_1, dominating without force application, and C_2, an extended state favored by force. Our analysis is based on the simplest possible energy landscape corresponding to this situation: two harmonic wells whose minima represent the two states (Figure 3.1), connected via a one-dimensional trajectory. Even though most proteins are likely to sample several intermediary conformations (local minima between the wells) while traversing a complex reaction trajectory (114), our model accounts only for the highest energy peak, or the last one encountered before the reactive state is attained. Both the equilibrium distribution of states as well as the rates of reaction are considered.

*** The work in this chapter was in collaboration with H. Karcher (115), who has developed the presented coarse-grained model.**

Here we adopt a widely-used microscopic approach based on Smoluchowski equation to deduce mean passage times and derive kinetic rate constants of diffusion-controlled reactions (116). This approach has been successfully applied to non-forced reactions (117,118), as well as forced reactions of bond rupture by escape from a single energy well (92,95,119). Another method to account for force dependence of kinetic rate constants is to apply Bell's phenomenological exponential dependence on force for the rate of bond dissociation (120). This approach has been extended to time-dependent applied forces to find statistics on the rupture forces in AFM experiments (95).

Several methods have been proposed to extract kinetic information from single-molecule pulling experiments leading to unbinding from a substrate or unfolding. AFM experiments to unbind the avidin-biotin complex have been analyzed using mean first passage-times (92) on one-dimensional energy landscapes, similar to the approach taken in this work. Whereas unbinding was then modeled as escape from a single energy well, the present method introduce a two-well landscape to model the transition between two stable, native conformational states of a single molecule. Hummer and Szabo (95) present another method to extract rate kinetics from pulling experiments, also based on escape from a single energy well.

Most kinetic models for protein deformation or unbinding consider only the energy barrier between states, whereas the proposed model takes into account the shape of the landscape along the entire reaction path. MD offers ways to link conformational changes of specific proteins under forces applied at specified protein locations. However, such simulations require knowledge of the full atomic structure specific to the particular protein, and typically are confined, due to computational constraints, to forces large compared to those experienced *in vivo*. A simple α-helix model was designed to test the coarse-grained model. Our approach is complementary in that it only considers a single degree of freedom trajectory and a single transition between states. All intra-protein force interactions are therefore represented by the two parabolic wells to produce a simplified model for the purpose of the examining both equilibrium states and rate kinetics.

3.2 Methods

Coarse-grained energy landscape for protein extension

Consider a protein having two conformational states: C_1, preferentially populated when no force is applied, and C_2, an extended state, and acted upon by a contact force (Figure 3.1). A simple energy landscape *E(x)* describing this situation consists of two parabolic wells:

$$\begin{aligned} E(x) &= \tfrac{1}{2}\kappa_1 x^2 - Fx && \text{for } x < x_{tr} \\ E(x) &= \tfrac{1}{2}\kappa_2 (x - x_2)^2 + E_2 - Fx && \text{for } x \geq x_{tr} \end{aligned}$$

with κ_1 and κ_2 stiffness values of the first and second well, respectively, x_{tr} the position of the transition state, x_2 the position of the extended state C_2 when no force is applied, E_2 the zero-force free energy difference between C_1 and C_2, and F the force applied to the protein.

A single reaction coordinate, x, is chosen, corresponding to the direction of protein deformation and force application. Energy minima (describing C_1 and C_2 states, respectively) are located initially at $x = 0$ and $x = x_2$. The two parabolas intersect at a transition state $x = x_{tr}$. With increasing force, the transition state remains at the same reaction coordinate x_{tr} , but the minima shift to $x = x_{\min 1} = F/\kappa_1$ and $x = x_{\min 2} = x_2 + F/\kappa_2$. For simplification, the following four non-dimensional parameters were used to calculate the rate constants:

$$\Pi_F = \frac{F x_{tr}}{kT},\ \Pi_{tr} = \frac{\frac{1}{2}\kappa_1 x_{tr}^2}{kT},\ \Pi_\kappa = \frac{\kappa_1}{\kappa_2},\ \Pi_E = \frac{E_2}{kT}$$

The forward rate constant k_f for the protein to change conformation from C_1 to C_2 was approximated as the inverse of the mean first-passage time associated with the transition from C_1 to C_2: t_f, a quantity that was used before as a measure of reaction rates (117,119,121). t_f is the average time necessary for the protein extremity to diffuse from its equilibrium state C_1 (minimum of the first well) to the elongated state C_2 (minimum of the second well) (Figure 3.1). Similarly, the reverse rate constant k_r for the protein to change conformation from state C_2 to state C_1 is the inverse of the passage time t_r in the reverse direction (Figure 3.1). The details of evaluating the first-passage time for the present model can be found in reference (115).

Steered molecular dynamics simulations on a simplified protein model

For the purpose of comparison to the coarse-grained simulation, we constructed a simple α-helix (a 15-mer of poly-alanine; Figure 3.2) and analyzed it using SMD (92). One advantage of an α-helix is that the helical axis uniquely defines a uni-directional reaction coordinate, along which the external force is applied. An extensive free energy calculation using constant velocity SMD and Jarzynski's equality has recently been reported by Park *et al.* (99) on a very similar deca-alanine α-helix. Here however, rather than attempting to evaluate the potential of mean force, we applied a constant force and used distance constraints on the 15-mer of polyalanine to compare the SMD results with those from the coarse-grained model. The number of alanines and the distance constraints have been selected so as to yield a stable and simple model that exhibits two distinct conformations. Many parameters extracted from the constant force SMD of this specifically designed model can be better related to our coarse-grained model, as seen in the results section.

The poly-alanine α-helix was constructed by creating a linear polyalanine sequence and specifying all the ϕ dihedral angles to –57° and all the ψ dihedral angles to –47°, which is characteristic dihedral angles for an α-helix. The N- and C-termini were capped with an amino group and a carboxylate group, respectively, with ionic states representative of the physiological pH level. The CHARMM script for creating an α-helix is available online (75). CHARMM was used to carry out the SMD simulations with the ACE2 implicit water module (122) and SHAKE constraints for efficiency. Energy of the α-helix structure was minimized in 15000 steps, heated to 300K in 40ps, and the system was equilibrated for 120ps using a time step of 2fs. After equilibration, the helix was repositioned placing the N-terminus at the origin and the C-terminus along the x-axis. Holding the helix fixed by a harmonic constraint at the N-terminus, the C-terminus was pulled with constant force along the x-axis. After a sequence of simulations in which several polypeptides arrangements were tried, we chose an α-helical system with 11 potential H-bonds, with six forced to remain intact under force and the other five allowed to form or break due to the combined effects of electrostatic attraction and VDW repulsion. The criterion for this choice was that the system exhibits two distinct states, with no apparent intermediate states. We imposed NOE constraints to the six H-bonding pairs, out of 11 possible, starting from the N-terminus carbonyl group, by specifying a limit distance of 4.25 Å between i^{th} carbonyl carbon and $(i+4)^{th}$ amide nitrogen with a force constant of 10.0 kcal/mol-Å^2. This model leaves five H-bonding pairs near the C-terminus to simultaneously either all break or all form to yield two distinct conformations (C1 and C2). Simulations were performed for 100 ns per simulation at forces of 30pN, 65pN, 70pN, 75pN, 80pN, 85pN, and 100pN.

Thermal fluctuations caused the forced end to exhibit relatively large displacements perpendicular to the direction of force application (Figure 3.2; left end is fixed and right end fluctuates). To compare with our single-dimensional coarse-grained model, we therefore present results in terms of the time-averaged component of force acting along the helical-axis.

Parameters were extracted from SMD simulations for comparison with our coarse-grained model. End-to-end distances, defined as the distance between the two termini (Figure 3.2), were traced with respect to time (Figure 3.3) and recorded every 4ps and used to generate histograms (Figure 3.4) to identify the most frequently sampled configurations.

Forward mean passage time from the coiled to extended conformation (t_f) was determined, assuming ergodicity, as the average time the molecule resides in state C_1 before undergoing a transition to C_2, while reverse mean passage time (t_r) was determined as the time residing in the extended conformation (C_2) before returning to the coiled conformation (C_1) (Figure 3.3). These SMD-determined parameters are introduced into the coarse-grained model, and compared the forward and reverse mean passage times obtained by both methods (SMD and coarse-grained model).

3.3 Results

The end-to-end distance (l) was extracted at each time frame (4ps per frame) from all of the SMD simulations (e.g., F=78.2pN shown in Figure 3.3). Plotting the histogram of l, the molecule is seen to sample two predominant conformations (end-to-end distances with the most occurrences on Figure 3.3 and 3.4). Assuming ergodicity, these conformations correspond to energy minima of our idealized energy landscape: $x_{\min 1} = F/\kappa_1$ and $x_{\min 2} = x_2 + F/\kappa_2$. Plotting the end-to-end distance with the most occurrences ($x_{\min 1}$ and $x_{\min 2}$) as a function of force (data not shown) yields the zero-force end-to-end distance of C_1 and C_2 ($l_1 = 2.1185nm$ and $l_2 = 2.9307nm$ respectively, hence the reaction coordinate $x_2 = l_2 - l_1 = 0.8122nm$). The locations of x_{min1} and x_{min2} determined from the peaks of the histograms, follow a linear trend with applied force: $x_{\min 1} = F/\kappa_1$ and $x_{\min 2} = x_2 + F/\kappa_2$. The slope ratio of $x_{\min 1}$ and $x_{\min 2}$ from the same plot gives $\Pi_\kappa \approx 0.44$. Thermal fluctuations are greater at small forces (C_1) than at large forces (C_2) (Figure 3.3), hence $\kappa_2 > \kappa_1$, roughly by a factor of two. At *F=74pN*, the SMD simulations show that the molecule spends an equal amount of time in states C_1 and C_2. This, as well as the geometric constraints described in Methods, lead to the parameter values: $\Pi_E \approx 13.2$, $\Pi_{tr} \approx 20$ and a transition state $x_{tr} = 0.6nm$ ($0 < x_{tr} < x_2$). Finally, it follows that $\Pi_F \approx 0.14 \times F(pN)$, $\kappa_1 \approx 1070 pN/nm$ and $\kappa_2 \approx 2183 pN/nm$.

The passage time t_f decreased with applied force, and t_r increased with applied force both with lower and upper limits of zero and infinity, respectively (Figure 3.5). Hence, the coarse-grained model and SMD simulations yielded similar trends though extension rates exhibited a stronger dependence on force with the coarse-grained model. Since the extension rates are dependent upon the shape of the energy landscape, one explanation for the difference in extension rates could be that the actual shape of the wells is different from the assumed parabolic wells.

3.4 Discussion

A generic model is developed for protein extension employing the physics of diffusion under force inspired by Kramers theory. The protein is assumed to have two distinct conformational states: a relaxed state, C_1, preferred in the absence of external force, and an extended state, C_2, populated under force application. The present model takes into account the mechanical features of the protein, as influenced by the weak interactions within a single protein. Its main purpose is to mechanically characterize the behavior of a protein's force-induced deformations and kinetics using a coarse-grained, approximate method. For now, we focus on the simplest system, and present an approach based on Kramers rate theory that incorporates a two-potential well energy landscape. Equilibrium results show that transitions to an activated state can occur over a narrow range of applied force. Reaction rates initially follow the anticipated exponential dependence on force, but the behavior deviates as the energy landscape becomes increasingly distorted. When cast in dimensionless form, all these results can be expressed in terms of four dimensionless parameters.

Simulations of complete unfolding of a protein (e.g., titin in Rief *et al.* (91), fibronectin domain in Gao *et al.* (85)) or unbinding from a substrate (e.g., avidin-biotin in Izrailev *et al.* (92)) have typically used large forces (~nN) to be computationally feasible with SMD, and hence fall within a drift motion regime (92). As this probes a different regime from the thermally activated one used in our coarse-grained model (92,123), we performed new simulations with smaller, steady forces (30-90 pN), inducing small deformations (<1nm, compared to ~28nm for unfolding of a single titin domain, (91)) and slow kinetics (time scales on the order of ns rather than ps). These slower transitions with smaller displacements are perhaps of more interest in the context of mechanotransduction. Using parameter values taken from equilibrium conditions, reasonable agreement was obtained for the variation in rate constants with applied force (Figure 3.5). Values of k_f and k_r extracted from SMD do not vary as rapidly with force as those computed with the coarse-grained model. A reason for this discrepancy could be that more energy dimensions are sampled in SMD than in our one-dimensional coarse-grained model.

Interest in the fundamental mechanisms of mechanotransduction has led to an increased focus on force-induced conformational change, producing subsequent alterations in binding affinity or enzymatic activity. Progress has been slow, however, since numerous proteins are involved in the transmission of force into and throughout the cell, and only a small fraction of these are sufficiently well characterized to permit detailed analysis, either by molecular dynamics simulation or experimentally. Alternative, more approximate methods are therefore needed if progress is to be made in the near term. A simple coarse-grained model of protein conformational change is presented with the capability of simulating some of the basic characteristics of protein kinetics and conformational change. Despite its simplicity with numerous simplifications, the current model can serve as a useful starting point for more detailed models. This has been indeed demonstrated through comparison to MD results of a simple α-helix model, which resulted in a good agreement in terms of evaluated time extension (Figure 3.5). Since the solutions for the coarse-grained model are obtained numerically, features can be augmented for a more sophisticated model, such as non-harmonic potential wells with multiple minima or allowing deformations in two or three dimensions. Similarly, simulation of multiple proteins, such as those comprising a focal adhesion, becomes computationally feasible. Coarse-grained models can be an additional tool along with MD and experiments to be developed, and together they can complement the study of mechanotransduction.

3.5 Figures

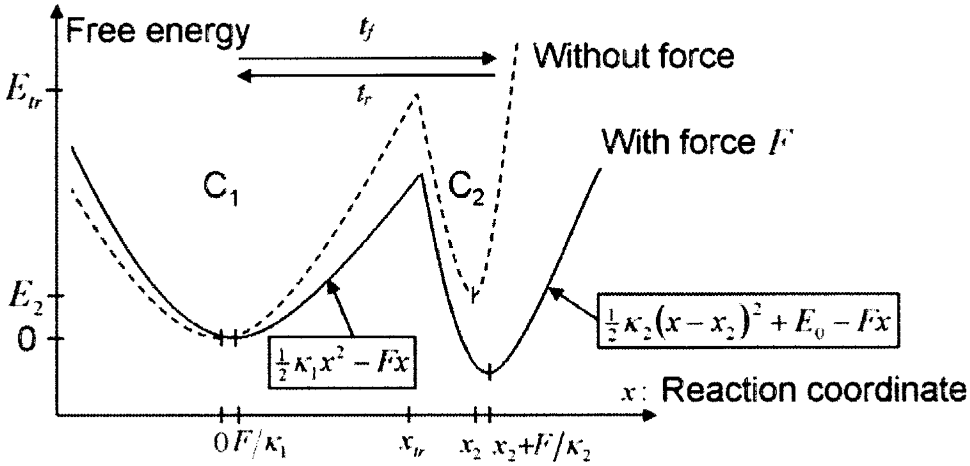

Figure 3.1. Idealized protein energy landscape when extended in the direction x. The y-axis is the Gibbs free energy G. The boxes contain the equations used to calculate the passage times and hence the protein extension rate. C_1 is an initial, relaxed state, C_2 a final, extended state. The times t_f and t_r, the first passage times to travel the distance depicted by the associated arrow, are computed to obtain the protein extension rate k_f and reverse rate k_r.

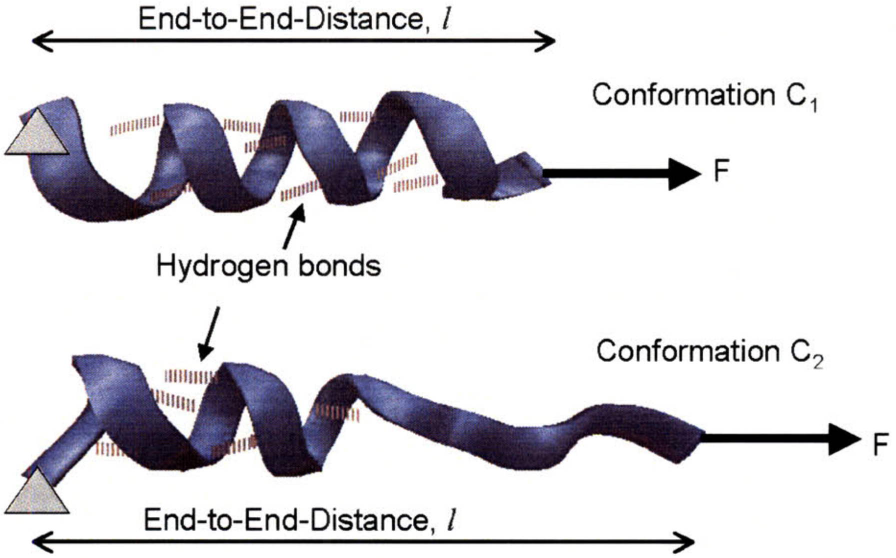

Figure 3.2. Two distinct conformations, C_1 (top) and C_2 (bottom), of the simplified protein model used in SMD. Left end of the helix is held fixed, while the right end is pulled with a constant force in the direction shown by the arrow. Six hydrogen-bonding pairs near the fixed end are constrained not to break.

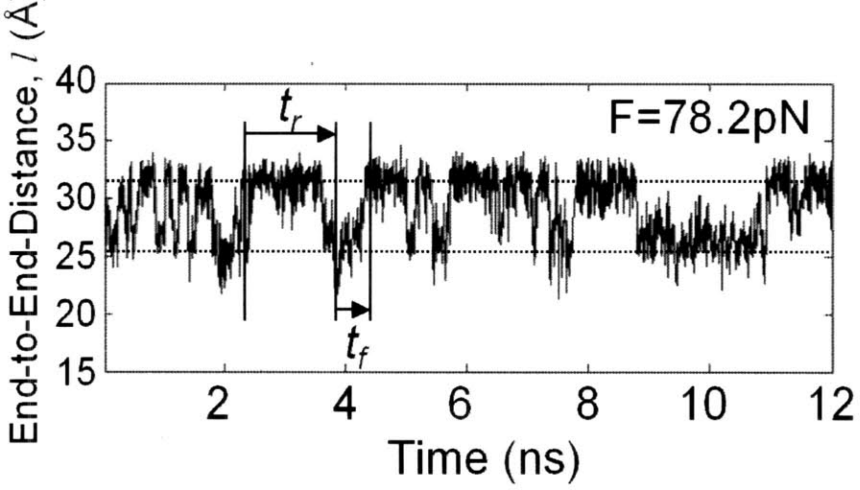

Figure 3.3. Time trace of the end-to-end distance of the helix at F=78.2pN (corrected from F=80pN). A forward passage time and a reverse passage time are shown. Mean passage times are obtained by averaging throughout the simulation.

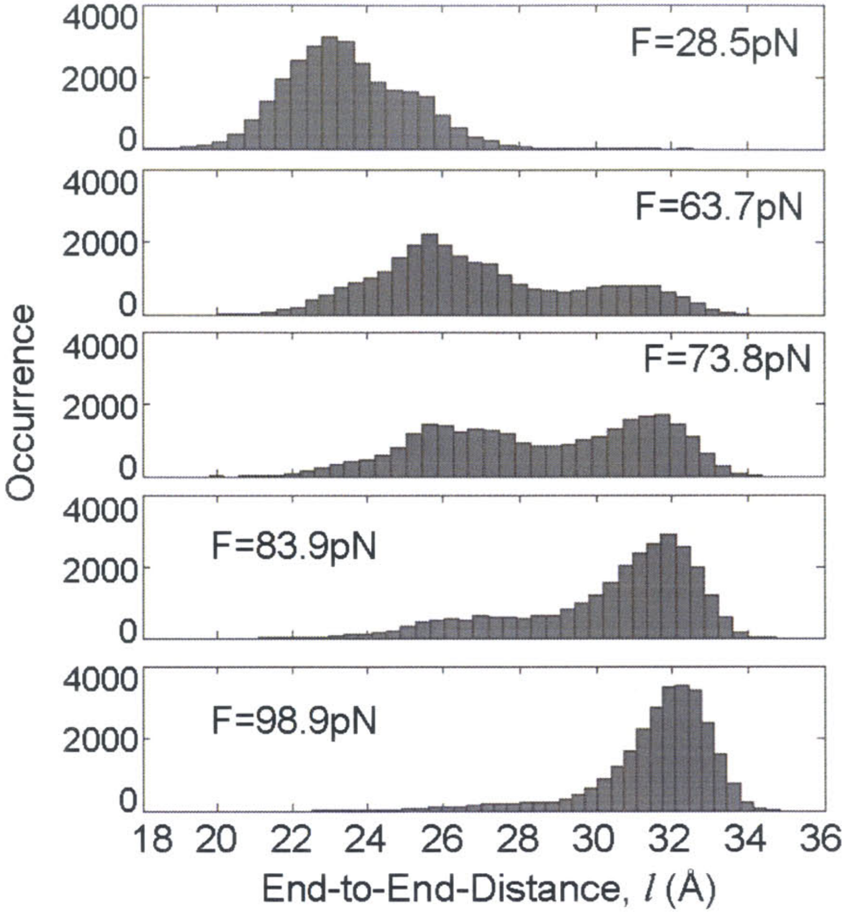

Figure 3.4. Histograms showing single and double peaks at various force magnitudes. Linear shift on the peaks are evident with varying forces.

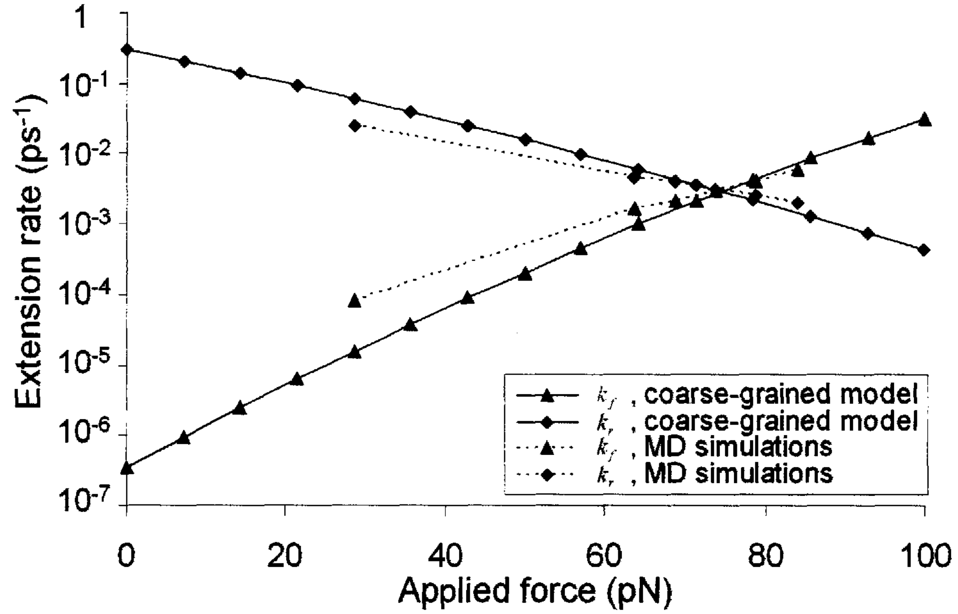

Figure 3.5. Protein extension rate from coarse-grained model and SMD model as functions of applied force along the helix axis direction. (*Dotted line*) SMD results from pulling on 15-mer of poly-alanine forming a α-helix. Kinetic rate constants are extracted as explained in Methods. (*Solid line*) results from coarse-grained model with $\Pi_\kappa = 0.44$, $\Pi_E \approx 13.2$, $\Pi_{tr} = 20$ (see text for parameter extraction). Both coarse-grained model and SMD simulations exhibit similar trends for the rates transforming the initial into the extended state (k_f) or the reverse (k_r).

4 Force-Induced Activation of Talin and Its Possible Role in Focal Adhesion Mechanotransduction

4.1 Introduction

One key mechanosensing protein in focal adhesions is talin, a cytoplasmic protein with a globular head and an elongated rod that provides an essential structural link between integrins and the actin cytoskeleton (70). The globular head of talin binds to β-integrin (75) and can also bind to and activate phosphatidylinositol 4 phosphate 5-kinase type γ (PIPKI-γ) (72,76). This, in turn, locally increases the production of phosphatidylinositol 4,5 bisphosphate (PIP2) (72), which is known to activate a number of focal adhesion proteins (e.g. vinculin and talin), hence promoting focal adhesion assembly (72). The talin rod can bind to β-integrin (77) and F-actin (78), and contains 11 vinculin binding sites (VBSs), each of which is an amphipathic α-helix (56,79). Vinculin is a cytoplasmic protein that may function as a structural reinforcement. It consists of a globular head, a proline-rich neck region, and a rod-like tail domain, which contains binding sites for many other cytoplasmic proteins (59,60). Cells with disrupted talin function fail to form focal adhesions and exhibit spreading defects (81). Cells with vinculin disruption, however, can still form focal adhesions, but display reduced ability to spread and increased cell motility (69).

Since mechanical force is needed for vinculin recruitment to focal adhesions (40), force-induced activation of cryptic VBSs on talin through conformational change may be the mechanosensing pathway leading to recruitment (2). Such recruitment could also lead to reinforcement of the focal adhesion. Indeed, talin1 is critical in force-dependent vinculin recruitment to adhesion sites independent of Src family kinase and focal adhesion kinase activities (53). Jiang *et al.* (52) identified that the initial contact that a cell makes with the extracellular matrix (ECM) consists of ECM-integrin-talin-F-actin linkages.

Some of the talin VBSs are inactive and unable to bind to the vinculin subdomain (Vh1; residues 1-258) (58). Vh1 is a subdomain of vinculin head that contains the binding site for talin and is used in various talin-binding experiments (58,71). The first vinculin-binding-site (VBS1; residues 606-636) is the fourth helix (H4) of a stable N-terminal five-helix bundle (TAL5) of talin rod (57). VBS1 has hydrophobic residues that, upon binding to Vh1, become deeply embedded in a hydrophobic core of the Vh1 (71). The same vinculin-binding residues form a tight hydrophobic core within TAL5 (57). Experiments have shown that isolated TAL5 has a low binding affinity for Vh1, whereas a four-helix-bundle with helix-5 (H5) removed from TAL5 (58), a mutated TAL5 with an unstable hydrophobic core (57), or the wild-type TAL5 molecule in elevated temperature solvent (58) can each disrupt TAL5 stability and strongly bind to Vh1.

Here, using computational methods, we demonstrate that realistically transmitted force acting on the focal adhesion protein talin leads to a conformational change that exposes the cryptic vinculin-binding-residues of VBS1. This then enables force-induced recruitment of vinculin, a critical early step in the process of focal adhesion reinforcement. Sequence homology of VBS1 with other VBSs suggests that the proposed mechanism may be a general force-induced activation mechanism of cryptic VBSs, and perhaps even be one of the general mechanotransduction mechanisms of helical bundles.

4.2 Methods

TAL5 simulation with EEF1

The structure of TAL5 was obtained by removing the C-terminal four-helix-bundle from TAL9 (PDB ID: 1SJ8) (57). The location and the assumed orientation of TAL9 within talin are shown in figure 4.1A. The longest principle length of TAL5 is aligned along the y-axis and the cross product of the vectors along H1 and H5 is aligned along the z-axis (Figure 4.1B). CHARMM was used with EEF1 solvent model and the CHARMM19 force field. The crystal structure was minimized by alternating the Steepest Decent and Adopted Basis Newton Raphson methods with 3000 steps. Bond lengths between hydrogen and heavy atoms were fixed using SHAKE constraint, and a 2fs time-step was used. Heating of the molecule to 300K occurred over 40ps, followed by a 560ps equilibration period at 300K.

Umbrella sampling (102) module of CHARMM with parabolic potential force constant of 5.0kcal/mol-Å^2 imposed on the reference reaction coordinates. One atom of each of the four residues of H5 (Q635, Q646, E650, and Q653) was harmonically constrained in space (k=0.2kcal/mol-Å^2). Forces were applied along a reaction coordinate defined as distance along a line from the center of mass of the H1 atom selection (side chain atoms of T498, S501, and S502) to a dummy atom with neutral charge and no mass located at coordinate (25.0Å, -17.0Å, 4.0Å) (Figure 4.1B). At each reference distance separated by 0.1Å, an 800ps canonical ensemble calculation was performed with Nosé-Hoover (124,125) thermostat for constant temperature control at 300K. NOE constraints were imposed to the backbone hydrogen bonding pairs within

H1 to prevent unraveling. In order to start the simulations with intact VBS1-H1 interaction, NOE constraints were imposed to polar sidechains between VBS1 and H1 during equilibration, which were removed at the beginning of the production runs. It should be noted that the sequential stepping (0.1Å) used is smaller than the fluctuation along the reaction coordinate around the reference value (~ ±1Å), and therefore, the trajectory from the umbrella sampling simulation is similar to a constant velocity MD calculation (92) with effective pulling velocity of 0.125Å/ns.

Constant force simulations with force magnitude varying between F=15.0 to 25.0pN were performed using the same TAL5 model described above. Constant force of F/(# of atoms on which the force is applied) was applied to each of the side chain atoms on H1 (T498, S501, and S502) toward the positive x-direction, such that total force applied is F.

Mutational study on TAL5

Three mutated TAL5 structures were constructed using the MMTSB toolset (126): (i) H5 residues N636 and Q639 mutated to alanines; (ii) VBS1 residues R606, Q610, K613, E621 and R624 all mutated to alanines; (iii) H1 residues N500, Q504, Q507, D514, and D515 all mutated to alanines. Umbrella sampling simulations identical to those described above are performed on each mutated structure.

Explicit water simulation on TAL5

TAL5 was solvated in an orthorhombic solvent box with each face at least 10Å away from TAL5 resulting in 23,775 atoms. Periodic boundary conditions were imposed using the image module. Electrostatic charge of the solvated system was neutralized by replacing seven water molecules with sodium ions. An all-hydrogen representation was used with CHARMM27 force fields. The SHAKE constraint and 2fs time-steps was used. SHIFT truncation was imposed with a cutoff distance of 12Å for non-bonded interactions, which has been found to exhibit reasonable accuracy in explicit water simulations (113). Long range non-bonded interactions, beyond the cutoff distance, were not taken into account in our explicit water simulations. The model was thoroughly minimized. The system was heated to 300K in 40ps and equilibrated for 960ps. An umbrella sampling potential of 5.0kcal/mol-$Å^2$ was used. H5 sidechain atoms (Q646, E650, and Q653) were harmonically constrained with force constant of 1.5kcal/mol-$Å^2$. Using a reference distance step size of 0.2Å, and 400ps simulation at each step a canonical ensemble simulation was performed with Nosé-Hoover at each reference distance, which is equivalent to 0.5Å/ns pulling rate. A constant force simulation with F=50.0pN was also performed. All explicit solvent simulations were performed on DataStar IBM p655 at San Diego Supercomputing Center (SDSC).

4.3 Results

TAL5 forms a stable structure with cryptic VBS1, which cannot bind to Vh1 in intact TAL5, but elevated temperature can effectively disrupt its stability and allow it to strongly bind to Vh1 (58). Explicit water simulations at 300K, 360K, and 420K were performed to investigate what constitutes TAL5 stability and the destabilizing effects of elevated temperature. Hydrophobic residues of H5 (L651, A647, V644, and A640) form a tight groove-fitting interaction with hydrophobic residues of VBS1 (A611, L615, A618, and L622) (Figure 4.2B). This interaction prevents VBS1 hydrophobic residues, which are the vinculin-binding-residues, from becoming exposed to the solvent. The trajectories of elevated temperature simulations (360K and 420K) did not differ much from those at room temperature (300K) other than the expected increase in thermal fluctuation. The RMSD of backbone atoms from crystal structures and average distances with fluctuations of VBS1-H1, VBS1-H2, VBS1-H3, and VBS1-H5 for three simulations are shown in Table 4.1. H5 and H3 closely interact with VBS1, and with other helices to a lesser extent (Figure 4.2B). It is likely that this interaction must be disrupted in elevated temperature or force-induced activation of VBS1.

In the TAL5 simulations displaying VBS1 activation, the hydrophobic residues of VBS1 (L608, L609, L615, V619, L622, and L623) form a tight hydrophobic core with the hydrophobic residues of H3 (V577, I580, L584, M587, V591, and L594) and the hydrophobic residues of H5 (L637, A640, V644, A647, L651, and I655) before extension. Polar and charged residues on VBS1 interact with H1 and H5 through hydrogen-bonds and salt bridges. RQK (R606, Q610, and K613) cluster on VBS1 interacts strongly with H1 (D514 and D515) by forming salt bridges (Figure 4.1B and 4.3B). ER (E621 and R624) cluster on VBS1 interacts with H5 (N636 and Q639) and more strongly with H1 (N500, Q504, and Q507) as N500, Q504, and Q507 of H1 surround and form hydrogen bonds with R624 of VBS1 (Figure 4.1B and 4.3B). As force is applied to TAL5, it is transmitted through these hydrogen-bonds. Since VBS1-H1 interaction is stronger, the transmitted force applies a torque through the RQK and ER clusters on VBS1 and the VBS1-H5 interaction is broken. As shown in figure 4.3, the hydrophobic contact formed by VBS1 with H3 and H5 eventually slips, and the hydrophobic residues of VBS1 are exposed to the solvent as VBS1 undergoes a rigid body rotation (Figure 4.3B and 4.3D). The hydrophobic residues of H5 fit into the V-shaped groove of VBS1 as one side of the 'V' (L608, L615, and L622) gets exposed to the solvent and the other side (L609, V619, and L623) forms a new hydrophobic core with H5 and H3 (Figure 4.3C and 4.3D).

Solvent accessible surface area (SASA) of the hydrophobic residues of VBS1 (Figure 4.4A, 4.4D, and 4.4G) show how much of these residues are exposed. The extent of VBS1 rotation is shown by measuring the angle made by L622 (only selected as a reference, which is within the hydrophobic core of TAL5 and gets exposed to solvent later) with the vector connecting the centers of mass of VBS1 and H5 (Figure 4.4B, 4.4E, and 4.4H; angle definition in Figure 4.5). VBS1 activation is defined by L622 angle becoming negative, since this is the clearest measure of helix rotation to expose VBS1 for possible binding. An increase in SASA is also indicative of activation, although this measure is also influenced by exposure of VBS1 residues internally, caused by H1 peeling away from VBS1. Note that the force peaks (figure 4.4C) are sharp, but the corresponding changes in rotation angle are more gradual and tend to lag behind the reduction in force. This may be due to rotation being diffusive in nature, occurring subsequent to the drop in force impeding rotation. Results from two VBS1 activated simulations and one non-activated simulation (for comparison) is shown in figure 4.4. By visual inspection on the VBS1 rotation plots displaying negative angles (e.g. Figure 4B, 4E, and 4H), 71.4% of the TAL5 simulations (n=20 out of 28) exhibited the VBS1 activation. Analyzing only the simulations with VBS1 activation, $157.5 \pm 70.9 Å^2$ of hydrophobic SASA of VBS1 was exposed to solvent, VBS1 rotated by 62.0±9.5°, and a mean force of 13.2±8.0pN was required for activation.

Activation of VBS1 follows disruption of the tight hydrophobic interaction of VBS1 with H3 and H5 rather than resulting from hydrogen-bond breakage. Rotation due to the applied torque through RQK and ER handles is opposed by the hydrophobic contacts from H3 and H5 (Figure 4.6). Non-bonded components of the interaction force on the hydrophobic sidechains of H3 and H5 experience force drops that correspond to the yielding of VBS1 to rotation (Figure 4.6B). The identified sidechains opposing VBS1 rotation exhibit simultaneous drops in force magnitude near 2ns. Time traces of the non-bonded force on A640, L651, L584, and V591 are shown in the subset of figure 4.6B.

4.4 Discussion

VBS1 activation in TAL5 is triggered by torque transmitted through the RQK and ER handles (Figure 4.3A and 4.3B). Polar side groups of H5 (N636 and Q639) oppose VBS1 activation by stabilizing the non-extended TAL5 structure. Indeed, in simulations that did not undergo VBS1 activation, the ER handle formed hydrogen-bonds with H5 after breaking the hydrogen-bonds with H1. This, however, did not cause VBS1 rotation in the opposite direction, suggesting that VBS1 activation can only occur when VBS1 forms strong hydrogen-bonds with H1. Simulations on three mutated structures were performed to verify the role of polar side chains on H1, VBS1, and H5. Simulation with H5 mutation (N636 and Q639 to alanines) enhanced VBS1 activation, whereas VBS1 mutation (RQK and ER are all mutated to alanines) and H1 mutation (N500, Q504, Q507, D514, and D515 mutated to alanines) impaired VBS1 activation, all consistent with our argument (data not shown).

Interestingly, the RQK and ER handles or similar motifs are ubiquitous in talin rod VBSs suggesting this mechanism as a general force-induced VBS activation mechanism of talin. For example, VBS2 (residues 852-876) contains sequences (**K**ILA**D** and **K**MV**E**AA**K**) (79) similar to ER (**E**LL**R**) and RQK (**R**PLL**Q**AA**K**) handles of VBS1 in terms of charges. Also, VBS3 (residues 1944-1969) contains (**KKE**LI**E**CA**RR**VS**EK**) (79). Charged and polar residues are shown in bold-face to highlight the similarity with ER and RQK sequences of VBS1. Another protein, α-actinin, localizes to cell-cell and cell-matrix junctions. Similar to talin, α-actinin has an amphipathic α-helix (αVBS; residues 731-760) that can also bind to the same binding site on Vh and contains a sequence (**R**TI**NE**) (127) similar to the ER handle of talin VBS1. As the vinculin-binding-residues of αVBS are also cryptic in an intact α-actinin structure (128), it is possible that α-actinin may be another force-sensitive protein, which gets activated by a similar mechanism, in vinculin recruitment.

The EEF1 model used in this study is empirically-based method characterized by high efficiency (108). Other implicit solvent methods (109,110) are theory-based, and they are 5-10 times slower than EEF1. EEF1 has been demonstrated to produce reasonable MD trajectories (108,111,112). To verify the validity of the EEF1 results, a constant velocity simulation with an effective pulling rate of 0.5Å/ns and a constant force simulation with F=50.0pN were performed on TAL5 with explicit solvent. In both explicit simulations, RQK and ER handles formed persistent hydrogen-bonds with H1, and the vinculin-binding surfaces of VBS1 became partially exposed in the 3.2ns constant velocity simulation or the 1.6ns constant force simulation (Figure 4.7). Explicit water molecules are known to slow diffusion-like transitions in proteins, and require much longer simulation times than the implicit simulations (8ns) to obtain similar range in motion. Full rotation of VBS1 therefore was not observed. The critical factors in VBS1 activation in EEF1 simulations are strong VBS1-H1 interaction, applied torque through RQK and ER handles, and disruption of VBS1-H5 hydrophobic interaction. In both the constant velocity simulation (10.58±0.25Å) and the constant force simulation (11.03±0.22Å), the average distances between VBS1 and H5 are clearly larger than the ones observed in the non-forced simulations (Table 4.1).

It is important to recognize that the conformational changes critically depend on the manner and direction in which the force is applied (2,111,129). Previous MD simulations have generally pulled on the N- and C-termini, and the results are often domain unfolding; e.g. (85,86). Complete unfolding is rarely observed in normal protein binding, however, so there is no reason a priori to expect that it would be necessary for force-induced reactions. We have attempted to apply a force in a realistic direction that mimics the force transmission within talin. Adjacent secondary structures of a protein commonly interact by forming hydrophobic contacts and hydrogen-bonds around the hydrophobic patch. Applied forces, therefore, are likely transmitted through the hydrogen-bonds between the secondary structures (85,86). The notable difference of pulling in this study compared to other unfolding simulations is that we apply the force on the hydrogen-bonding residues, which is a likely site of force transmission between secondary structures, rather than on the two termini. Since the pulling direction is primarily determined by the relative locations of force application sites and fixed points, in choosing to apply force to the polar residues of H1, and to fix the polar residues of H5, forces tend to be directed nearly perpendicular to the TAL5 principle axis (figure 4.1B).

VBS1 activation was found to be somewhat dependent upon the sites of force application and, consequently, on the direction of pulling (data not shown). However, given what is currently known about the structure of talin, and the probable sites of interaction with neighboring domains of the protein, these assumptions seemed reasonable.

The force extension curves have the typical sawtooth shape (Figure 4.4C, 4.4F, and 4.4I) with the force-drops corresponding to rupture of the hydrogen-bonds or slippage of hydrophobic contacts between secondary structures, which are similarly observed in other AFM and MD studies (84,87). Thus, VBS1 activation occurs through rotation of VBS1 relative to the TAL5, as a consequence of torque applied via hydrogen-bonds and salt bridges between H1 and VBS1. Thus, the conformational change required for activation is subtle, involving an extension of less than 2Å and no domain unfolding, as has been found in the activation of other cryptic sites, such as fibronectin (85).

The potential of mean force (PMF), or free energy landscape along the extension, has a monotonically increasing profile without any apparent local minima (130), suggesting that the zero force structure resides in the global minimum (Figure 4.8). There is, however, a region of decreased slope on the PMF curve near extension of x=1.4Å that corresponds to the hydrophobic slip of VBS1. Activation can be interpreted in the context of the PMF curve representing the free energy change along a specified reaction coordinate in the absence of force (130). The plateau at x=1.4Å becomes a local minimum when the molecule is exposed to a force as low as 20.0pN, allowing it to undergo the conformational change (115,130). This is reflected, for example, by an increased probability of 50% in the activated state under a force of 20.0pN compared to a zero force case. An extended state from one of the simulations with VBS1 hydrophobic residues partially exposed was used in a relaxation explicit water simulation for 1ns with all the constraints removed. In the absence of force, the exposed hydrophobic residues rotated back into the hydrophobic core indicating that the non-extended state indeed is the global minimum (Figure 4.9) and that de-activation occurs almost immediately following the release of force.

Varying magnitudes of force are applied to TAL5 during the constant velocity simulations as VBS1 undergoes activation (Figure 4.4C, 4.4F, and 4.4I). Although the peak magnitude applied through TAL5 is 55.4±19.1pN, the mean force applied throughout the simulations is 13.2±8.0pN. In order to verify this finding, constant force simulations were performed. All constant force simulations (n=8) with force magnitude $\geq$ 18.0pN underwent VBS1 activation, whereas all simulations (n=4) with forces $\leq$ 17.0pN did not (data not shown). The effective pulling rate of 0.125Å/ns is still many orders of magnitude faster than the pulling rates we might expect in vivo or with AFM experiments (~1nm/ms ~10^{-5}Å/ns). Such rapid pulling results in significantly larger force levels in bond rupture (119) or protein unfolding (95) compared to the corresponding AFM measurements. In both cases, the forces measured by AFM were ~30% of the force computed using MD (95,119). Using this value as a very rough approximation, the force needed to activate VBS1 (13.2±8.0pN) at more realistic, slower rates of pulling would lie in the range of ~4pN. This estimated lower force at slower pulling rate is on the order of (i) forces generated by a single myosin, ~4pN (131); (ii) forces needed to rupture a talin-F-actin bond, ~2pN (52); and (iii) the estimated force experienced by a single integrin linkage, based on close packing in a focal contact, ~1pN (44). On the extracellular side, the force required to break a single integrin-fibronectin bond is ~20pN (132), and a single integrin-fibrinogen bond can withstand ~100pN (133).

In conclusion, we identify a potential mechanism for VBS1 activation, involving a force-induced conformational change causing the hydrophobic vinculin-binding residues on VBS1 within TAL5 to become accessible for vinculin binding. This would then constitute the initiating event leading to force-induced focal adhesion strengthening by vinculin recruitment.

4.5 Table and figures

	Dist between VBS1-H1 (Å)	Dist between VBS1-H2 (Å)	Dist between VBS1-H3 (Å)	Dist between VBS1-H5 (Å)	RMSD (Å)
T = 300K	14.03 ± 0.26	14.38 ± 0.25	10.46 ± 0.21	8.04 ± 0.20	1.9635
T = 360K	13.79 ± 0.26	14.43 ± 0.25	10.57 ± 0.20	8.10 ± 0.20	2.3258
T = 420K	13.51 ± 0.35	14.33 ± 0.28	10.86 ± 0.26	8.44 ± 0.24	2.7919

Table 4.1. RMSD from crystal structure and average distances between helices from elevated temperature explicit water simulations.

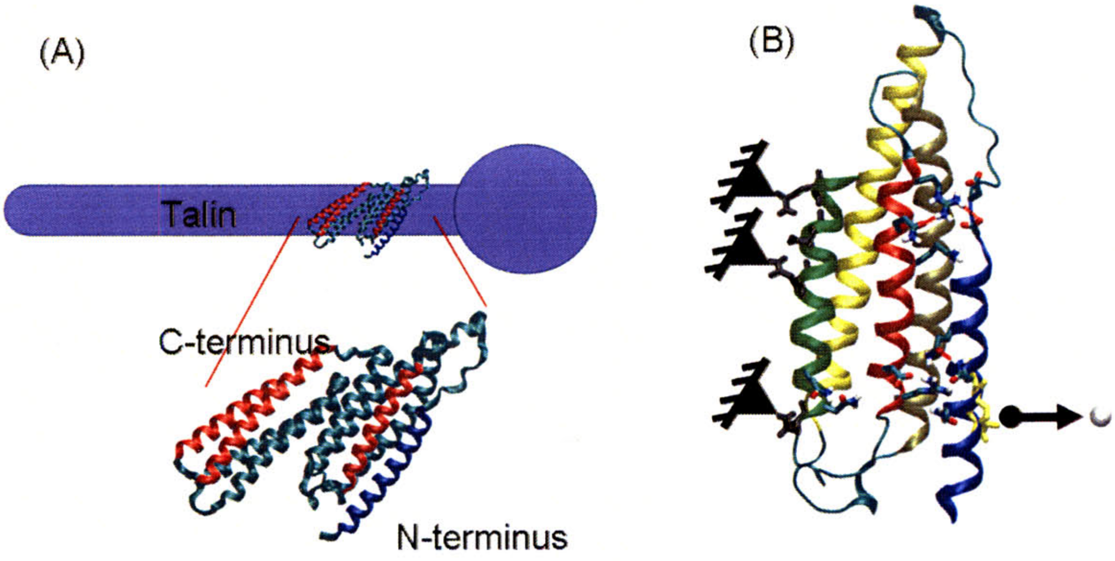

Figure 4.1. (A) Crystal structure of TAL9 (PDB ID: 1SJ8) in ribbon representation is shown superimposed on a hypothetical talin model. Three VBSs within TAL9 are shown in red, and H1 is shown in blue. Since the talin rod has tandem repeats of helical bundles, TAL9 is aligned such that the centers of mass of the two helical bundles lie on the talin rod axis. (B) Detailed view of the N-terminal five-helix bundle (TAL5) used in the TAL5 simulations. Each of the five helices is shown in a different color: H1 (blue), H2 (yellow), H3 (tan), H4 (or VBS1; red), and H5 (green). Some important polar residues are shown in stick representations. A dummy atom with no mass or charge is shown in white. H5 polar side chains (black sticks) are harmonically constrained in space (constraints shown as triangles). H1 polar side chains (yellow sticks) are pulled toward the dummy atom (effective pulling direction shown as an arrow).

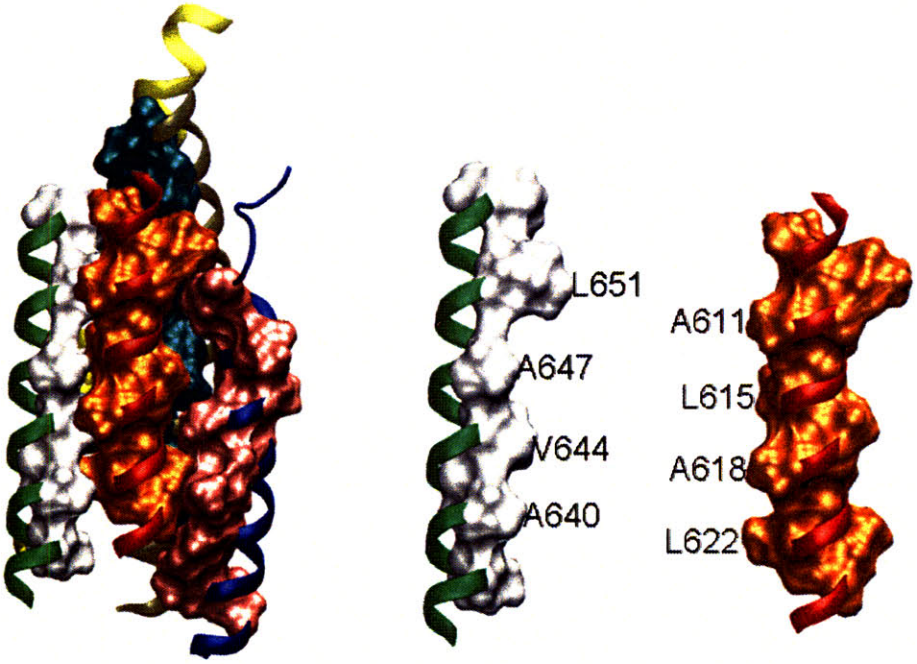

Figure 4.2. (A) TAL5 in the same orientation as Figure 4.1B with hydrophobic residues shown in surface representation for H1 (pink), H3 (cyan), VBS1 (orange), and H5 (white). Exposed ribbon sites are polar residues, whose side-chains are not shown for clarity. VBS1 and H5 form a tight groove-fitting contact, which stabilizes VBS1 in TAL5's hydrophobic core preventing VBS1 from being accessible for vinculin binding. (B) VBS1 and H5 are shown separately with labels for residues participating in the groove-fitting interaction.

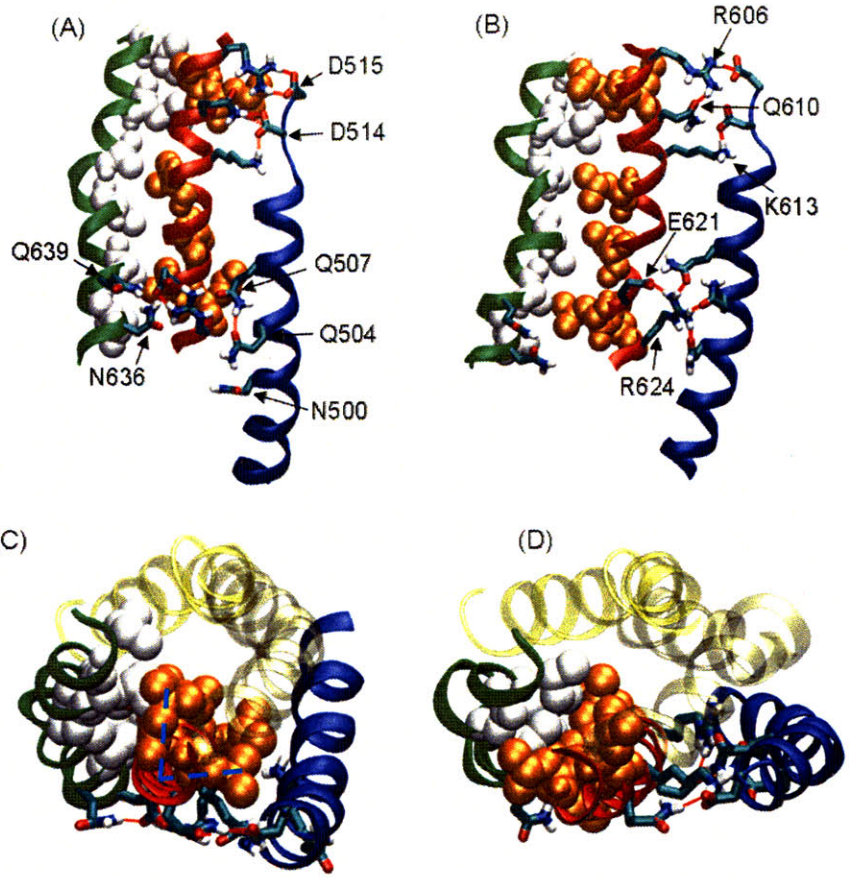

Figure 4.3. Conformations showing VBS1 activation from the TAL5 simulation: H1 (blue ribbon), H2 (transparent yellow), H3 (transparent tan), VBS1 (red ribbon), H5 (green ribbon), hydrophobic residues of VBS1 (orange VDW; also the vinculin-binding-residues), hydrophobic residues of H5 (white VDW), and some important polar residues (stick representation with color denoting the atom type). Polar residues are labeled on the figures. (A) Conformation at t=2.08ns. The hydrophobic residues of VBS1 (orange VDW) are hidden in the hydrophobic core. (B) Conformation at t=7.40ns showing the hydrophobic residues of VBS1 being exposed to solvent. Hydrogen-bonds between H5 and VBS1 are broken. The hydrophobic residues, or the vinculin-binding-residues, point into the page in (A) and point to left in (B). (C) Conformation at t=0.86ns viewed from top. The V-shaped VBS1 hydrophobic residues are packed within the hydrophobic core

of TAL5 (cyan dotted lines). (D) Conformation at t=9.24ns showing VBS1 rotation. The hydrophobic residues H5 (white VDW) fit into the 'V' of the VBS1 hydrophobic residues (orange VDW).

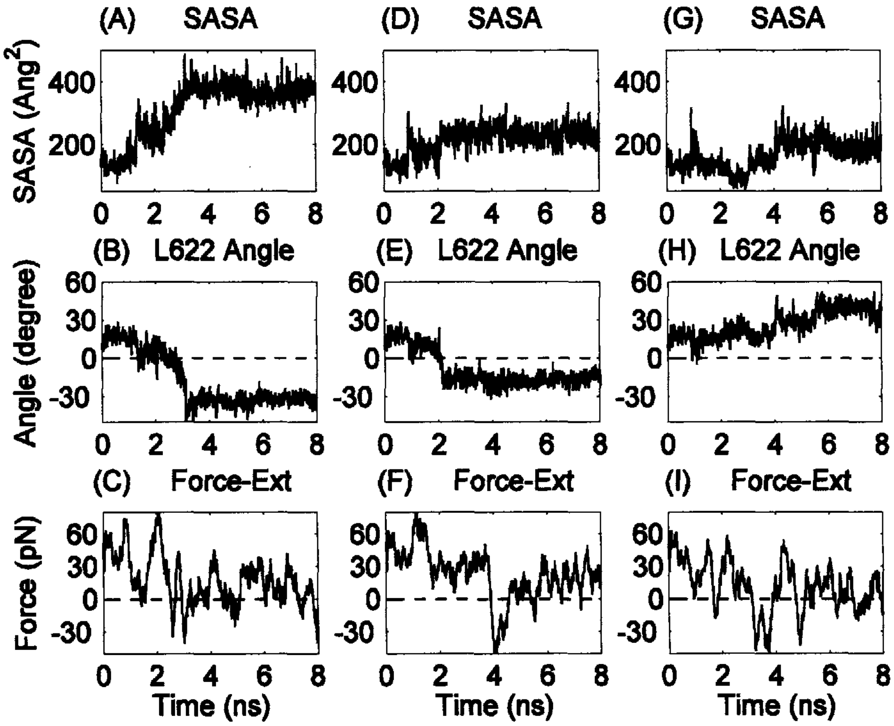

Figure 4.4. Results from three TAL5 simulations: from two simulations undergoing VBS1 activation (A)-() and (D)-(F), and also from a simulation without activation (G)-(I). (A) (D) (G) Change in SASA of hydrophobic (red) and polar (blue) residues of VBS1. (B) (E) (H) Angle of rotation of VBS1 relative to H5 (defined in Figure 4.5). Positive angle corresponds to the inactive state of VBS1, and negative angles correspond to activation. (C) (F) (I) Force applied to TAL5.

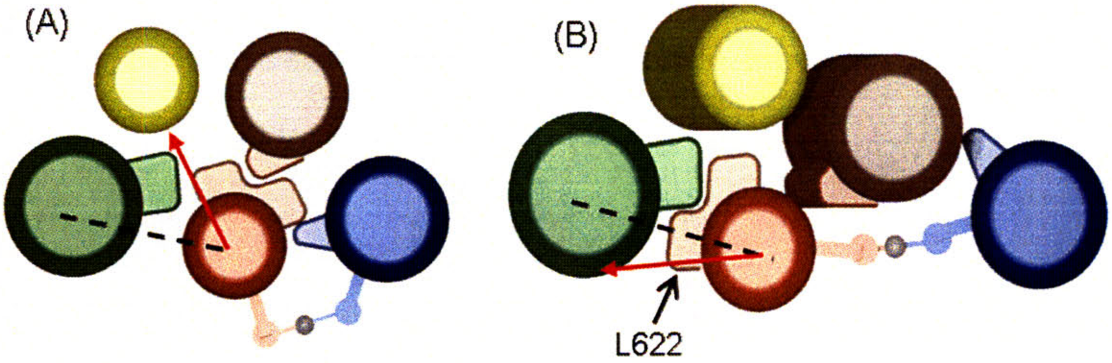

Figure 4.5. The angle of VBS1 rotation is defined by the angle formed between L622 of VBS1 and a vector connecting the centers of mass of VBS1 and H5. (A) Before force application, the angle is positive with L622 inside the hydrophobic core. (B) After forcing, the angle is negative with L622 outside the hydrophobic core.

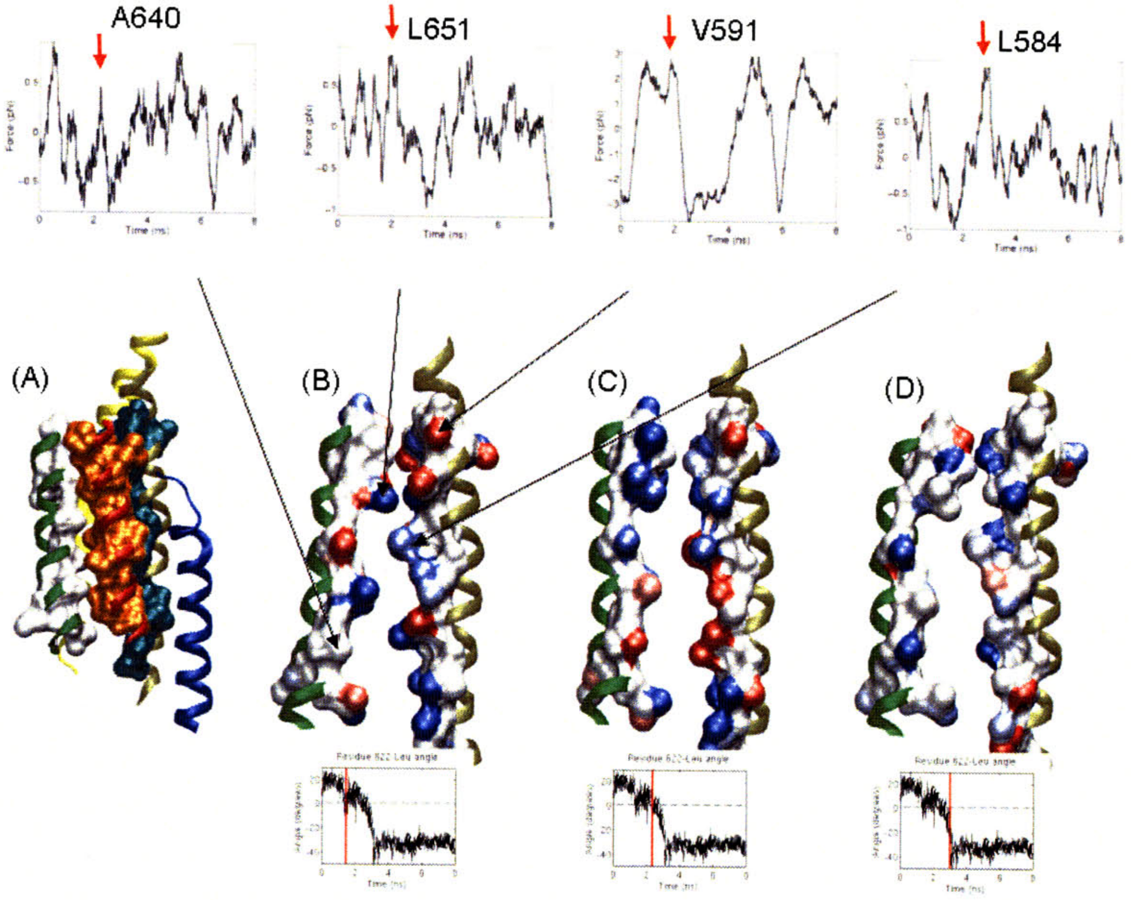

Figure 4.6. (A) TAL5 in the same orientation and color coding as in Figure 4.2A. (B) Only H3 and H5 are shown in the same orientation as in (A) at t=1.52ns to highlight the force magnitude exerted on the hydrophobic residues by the hydrophobic residues of VBS1 (not shown for clarity) with force magnitude below the average force in blue, near average in white, and above average in red. Residues shown in red are opposing VBS1 rotation. Once the hydrophobic residues slip, corresponding force drops are as shown in blue. The snap incident is indicated on the VBS1 rotation plot (Figure 4.4B) as a red vertical line. Time traces of force magnitude are shown for A640, L651, L584, and V591 in the subset. Force peaks near 2ns are indicated by red arrows, which correspond to events in which the hydrophobic contacts yield to VBS1 rotation. (C) Forces on

hydrophobic sidechains at t=2.40ns just before slip of hydrophobic residues yielding to VBS1 rotation. (D) Forces on hydrophobic side-chains at t=3.04ns after the hydrophobic slip and corresponding force drops.

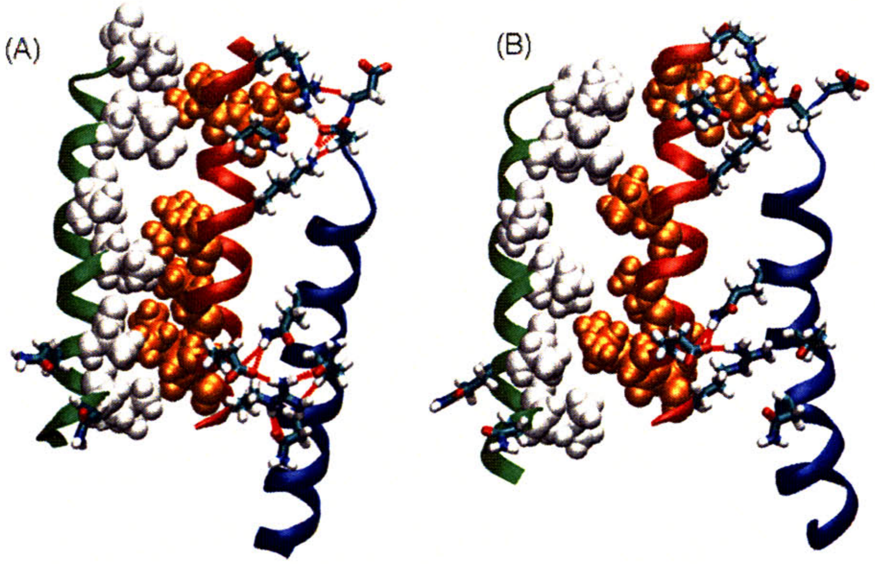

Figure 4.7. Two conformations from a constant force (F=50.0pN) explicit water simulation. (A) Conformation at t=0.112ns and (B) conformation at t=1.2ns. The simulation lasted 1.6ns. The hydrophobic residues of VBS1 seem to be beginning to expose to solvent. Although VBS1 rotation is much smaller in extent compared to those measured in the implicit solvent simulation, VBS1 still has very strong hydrogen-bonding interactions with H1, where the torque applied through transmitted force. Trajectory from constant velocity simulation in explicit solvent (3.2ns in duration) also show strong interaction between VBS1 and H1, but VBS1 rotation is not observed to the same extent as in the implicit solvent simulations.

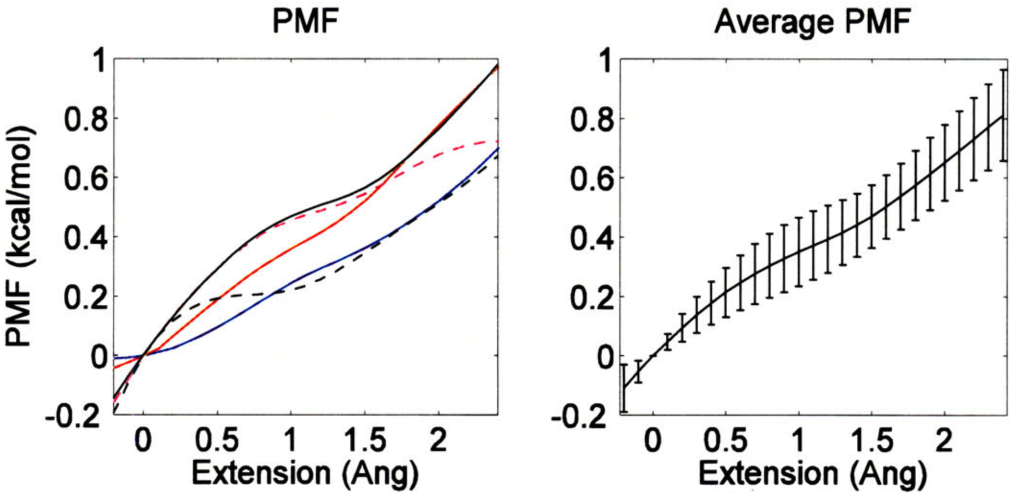

Figure 4.8. (Left) Calculated potential mean force (PMF) curves from TAL5 simulations (n=5). Each PMF is shifted vertically so that the mean value for each is zero. (Right) Averaged PMF.

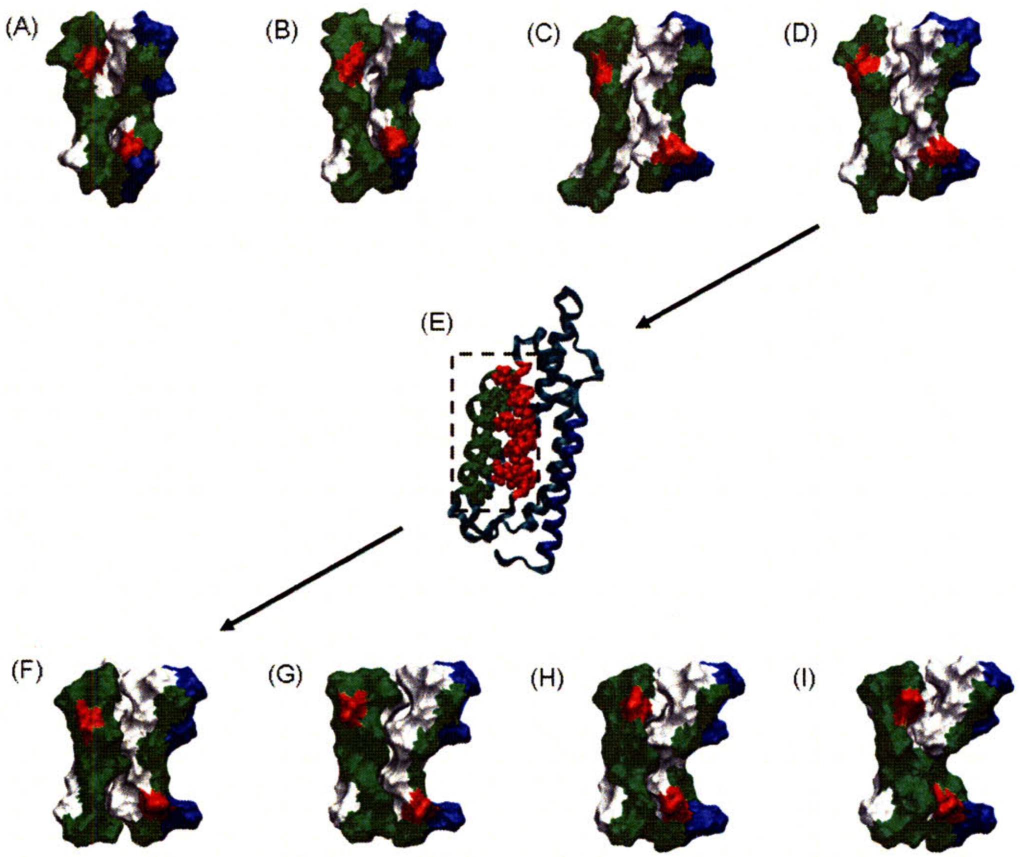

Figure 4.9. Surface representation of VBS1 and H5 from the TAL9 simulation at (A) t=0ns, (B) t=5.12ns, (C) t=10.24ns, and (D) t=15.36ns. Hydrophobic residues are shown in white, polar residues in green, negative residues in red and positive residues in blue. These views show that the hydrophobic residues of VBS1 are partially exposed to the solvent. (E) The ribbon representation of TAL5 at t=15.36ns is shown with H1 in blue ribbon, H4 in red ribbon, H5 in green ribbon, VBS1 hydrophobic residues in red VDW, and H5 hydrophobic residues in green VDW. The dotted box indicates that the surface representations are only showing VBS1 and H5. The extended TAL9 at t=15.36ns is truncated to TAL5, solvated in rhombic dodecahedron water-box, removed all external forces, and simulated for 1ns. The configurations at (F) t=0ps, (G) t=352ps, (H) t=704ps, and (I) t=1056ps of the relaxing dynamics show that the exposed hydrophobic residues return to their cryptic conformation.

5 Molecular Dynamics Study of Talin-Vinculin Binding

5.1 Introduction

Talin and vinculin are essential in forming stable focal adhesions. Talin is a cytoplasmic protein with binding sites to other focal adhesion proteins including β-integrin (77), F-actin (78), and containing 11 possible vinculin binding sites (VBSs), a number of which are cryptic (56,79). Vinculin likely provides structural reinforcement since it can simultaneously bind to talin and F-actin. It consists of a globular head, a proline-rich neck region, and a rod-like tail domain that contains binding sites for many other cytoplasmic proteins (59,60). Vinculin head is known to bind to α-actinin (61) and talin (54), whereas vinculin tail is known to bind to paxillin (62), F-actin (55), and phosphatidylinositol 4,5 bisphosphate (PIP2) (63). The neck region binds to VASP (64), vinexin (65), and ponsin (66). Vinculin forms an autoinhibitory head-tail interaction within cytosol, which masks many of its binding sites for other proteins (73,134). Recent findings show that the high affinity autoinhibition interaction in a full-length vinculin is due to cooperative effect of two low affinity binding interfaces (135). Therefore, complete vinculin activation requires a combinatory signaling pathway of vinculin interacting with one or more of its binding partners (73). Cells with disrupted talin function fail to form focal adhesions and exhibit spreading defects (81) whereas cells with vinculin disruption can form focal adhesions, but display reduced ability to spread and increased cell motility (69).

Some of the talin VBSs are inactive and unable to bind to the vinculin subdomain (Vh1; residues 1-258) (58). Vh1 is a subdomain of vinculin head that contains the binding site for talin and is used in various talin-binding experiments (58,71). The first vinculin-binding-site (VBS1; residues 606-636) is the fourth helix (H4) of a stable N-terminal five-helix bundle (TAL5) of talin rod (57). VBS1 has hydrophobic residues that, upon binding to Vh1, become deeply embedded in the hydrophobic core of Vh1 (71). Izard *et al.* (71) also demonstrated that VBS1 can bind to Vh1 of Vh1 and vinculin tail (Vt; residues 883-1066) complex and effectively sever the Vh1-Vt interaction. Purified talin, however, binds to full-length vinculin at a low affinity suggesting that talin is only one of a number of binding partners needed for full vinculin activation (73). The same vinculin-binding residues of VBS1 that gets embedded within Vh1 form a tight hydrophobic core within TAL5 (57). Experiments have shown that isolated TAL5 has a low binding affinity for Vh1, whereas a four-helix-bundle with helix-5 (H5) removed from TAL5 (58), a mutated TAL5 with an unstable hydrophobic core (57), or the wild-type TAL5 in elevated temperature solvent (58) can each disrupt TAL5 stability and strongly bind to Vh1.

Recent molecular dynamics (MD) simulations have demonstrated a mechanism by which transmitted mechanical force disrupts TAL5 stability and activates it to bind to Vh1 by a process in which a torque is applied to helix 4, causing it to rotate and making the binding site accessible (Chapter 4) (136). The hydrophobic residues exposed under applied force (Chapter 4) (136), are those that are known to be important in binding to Vh1 (56). In order for VBS1 to bind to Vh1, however, it must separate two helices and embed itself in between. How this occurs has been a matter of considerable speculation (71).

Here, MD is used to investigate the binding mechanism of Vh1 to VBS1, which is the immediate next step after force-induced talin activation (136). Together these two steps comprise the early mechanotransduction events in the force-induced recruitment of vinculin (40).

5.2 Methods

Vh1-VBS1 binding simulation with EEF1

Unbound and bound models of the Vh1-VBS1 complex (Figure 5.1) were obtained from crystal structures of Vh1-Vt (PDB ID: 1RKE) and the Vh1-VBS1 complexes (PDB ID: 1T01), respectively. When Vh1 binds to VBS1, the N-terminal four-helix bundle of Vh1 bends over and undergoes considerable conformational change, whereas the C-terminal four-helix bundle of Vh1 remains unchanged (71). Therefore, the Vh1 structures from the bound and unbound models are aligned by the backbone atoms of the C-terminal four-helix bundle of Vh1 to highlight the conformational difference between the two models. The vinculin tail domain (Vt) is removed from the Vh1-Vt complex and VBS1 is aligned with its binding site but translated 12Å away from Vh1 to obtain the unbound Vh1-VBS1 model (Figure 5.1A) whereas the known Vh1-VBS1 crystal structure is used as the bound Vh1-VBS1 model (Figure 5.1B). Views from the top show the separation of the molecules (Figure 5.1C) and the associated conformational change upon VBS1 binding toVh1 (Figure 5.1D). All of the visualizations presented in this paper were done using Visual Molecular Dynamics (137).

Commercial molecular dynamics software, CHARMM (Harvard University, Cambridge, MA) (104), was used with the Effective Energy Function (EEF1) (108) solvent model and the CHARMM19 force field (105). Energies of these models were minimized by alternating the Steepest Decent and Adopted Basis Newton Raphson methods with 3000 steps. Bond lengths between hydrogen and heavy atoms were fixed using the SHAKE constraint (107), and a 2fs time-step was used. Molecules were heated to 300K over 40ps, followed by a 560ps equilibration period at 300K. During the heating and equilibrium process, weak harmonic constraints (0.1kcal/mol-Å^2) were applied to the Cα atoms in order to minimize deviations from the original position. After equilibration, Cα constraints were removed, and the production simulations were run using the Nosé-Hoover (124,125) thermostat for constant temperature control at 300K.

Vh1-VBS1 binding simulations were performed beginning with the unbound Vh1 and VBS1 model (Figure 5.1A and 5.1C). In some simulations, VBS1 was initially rotated around the helix axis ±2 degrees to determine the effect on binding of VBS1 orientation relative to Vh1. Two types of Vh1-VBS1 binding simulations were carried out. In one, all constraints were removed after equilibration to determine how the two molecules, initially separated, might interact in the complete absence of external forces. In the other, distance constraints were imposed, where the atom pair is pulled toward each other when they are separated by the pre-specified reference distance, between residues on VBS1 (L608, L615 and L622) and Vh1 (V16, L23, V44, L116 and F126) for 800ps after equilibration in order to enhance the probability of Vh1-VBS1 binding. These constraints were removed after 800ps, and the simulations were continued for 32ns. A 30ns simulation was conducted on the Vh1-VBS1 bound complex (Figure 5.1B and 5.1D) with no constraints to characterize the binding interaction between the two molecules.

Mutational studies on Vh1-VBS1 binding, and Vh1 binding to other VBSs

A number of Vh1-VBS1 unbound models were obtained with various VBS1 mutations: (i) L622 mutated to alanine (L622A); (ii) L623 mutated to alanine (L623A); and (iii) K613 mutated to proline (K613P). Unbound models of Vh1 with other VBSs were constructed by the same method used for obtaining the Vh1-VBS1 unbound structures instead using Vh1-Vt, Vh1-VBS2 (PDB ID: 1U6H), Vh1-VBS3 (PDB ID: 1RKC) and Vh1-αVBS (PDB ID: 1YDI) crystal structures. Identical constraints as the ones applied between Vh1 and VBS1 were applied for 800ps to Vh1 and other or modified VBSs. Only in the case of Vh1-VBS3 binding simulation, for which the C-terminal end of VBS3 tended to unfold and did not bind to Vh1, were additional distance constraints on the hydrogen-bonding pairs of the VBS3 backbone helix imposed to retain helicity. As before, these constraints were removed after 800ps, and simulations continued for 32ns.

Simulation on Vh1-TAL5 binding

The structure of activated TAL5 (i.e. hydrophobic residues exposed to solvent) was obtained from the end state of force-induced activated TAL5 (136). VBS1 of TAL5 was aligned to VBS1 of the Vh1-VBS1 bound complex model (Figure 5.1B and 5.1D). After the heating and equilibration, stronger distance constraints between Vh1 and VBS1 of TAL5 were applied throughout the simulation to force the binding of Vh1 to TAL5 through VBS1. Additional helicity constraints were also applied to Vh1 and TAL5 helices to force them to retain the secondary structures.

5.3 Results

Unconstrained binding

In the first series of Vh1-VBS1 simulations with no external constraints and the initial condition of Figure 5.1A, one simulation proceeded to complete binding, for which the end configuration is very similar to the bound Vh1-VBS1 crystal structure (Figure 5.1B and 5.1D). For this binding simulation and equilibration simulation starting from the bound complex (Figure 5.1B), the average root mean square deviation (RMSD) of the Cα atoms of Vh1's N-terminal four-helix bundle and VBS1 are 2.33Å and 1.97Å, respectively (Figure 5.2A). The superimposed configurations from near the end of the binding simulation and equilibration simulation show that the two are very close in their relative orientations (Figure 5.2B and 5.2C).

In this constraint-free Vh1-VBS1 binding simulation, VBS1, initially separated by 12Å, moves toward Vh1 through hydrophobic targeting. A large hydrophobic patch is exposed to solvent on the H1 and H2 interface, and the hydrophobic residues of VBS1 become inserted between helices H1 and H2. Through this hydrophobic insertion, L608, L615, and L622 of VBS1 form contacts with V16, L23, V44, A50 and L54 of Vh1, and this contact stabilizes the interaction of the two molecules (Figure 5.3A and 5.3D). Hydrogen-bonds are formed between Q627 of VBS1 to H22 of Vh1, and also between Q610 and K613 of VBS1 to N53 and R56 of Vh1. VBS1 moves further between H1 and H2 with time as it separates H1 and H2 (Figure 5.3B and 5.3E). During this stage, VBS1 also moves closer to H4 of Vh1. Binding of Vh1 and VBS1 is complete when VBS1 rotates and effectively locks the exposed hydrophobic residues (L619 and L623) into the hydrophobic core of Vh1 (Figure 5.3C and 5.3F). When L623 moves into the hydrophobic core, R624 of VBS1 swings over to form hydrogen-bonds with Q19 and H22 of Vh1's H1. This Vh1-VBS1 binding mechanism viewed from the front and on a cross-sectional plane is shown in Figure 5.3.

Of the 20 simulations conducted with no constraints, only one progressed to the fully bound state during the 40 ns simulation window. In order to promote binding, external constraints as described above were imposed for the first 800 ps of simulation.

Constrained binding

Vh1-VBS1 binding is enhanced when VBS1 is constrained so that its hydrophobic residues are inserted between H1 and H2 of Vh1 in the beginning of the binding simulations. Even though the constraints between Vh1 and VBS1 are completely removed after 800ps, this proved sufficient to induce Vh1-VBS1 binding that occurred many nanoseconds later. The 800ps of applying constraints increased the chances of VBS1 forming the initial and necessary hydrophobic insert into Vh1 (Figure 5.3A and 5.3D), however, the separation of H1 and H2 of Vh1 did not occur during this 800ps and happened much later in the simulations. The 30ns equilibration simulation of the Vh1-VBS1 complex (Figure 5.1B) was analyzed to determine the characteristics of the Vh1-VBS1 complex. Three measures: (i) the angle formed by L623 with VBS1-H1 vector (A(623)), (ii) the distance between H1 and H2 of Vh1 (D(1-2)), and (iii) the distance between VBS1 and H4 (D(V-4)), are chosen to be the indicators of Vh1-VBS1 binding status (Figure 5.3E). Average values for A(623), D(1-2), and D(V-4) from the equilibration simulation were evaluated to be 30.1°, 20.9Å, and 12.2Å, respectively. In a given Vh1-VBS1 binding simulation, these three indicators were monitored to decide whether the molecule underwent binding. For example, the first instant when A(623) > 30.1°, D(1-2) > 20.9Å, and D(V-4) < 12.2Å simultaneously is defined as the time when the Vh1-VBS1 binding is complete. Note that this is only used to define Vh1-VBS1 binding, and the three values actually fluctuate about the threshold values shown above. Out of 12 simulations with slightly different initial conditions, six simulations (50%) underwent binding with the averages of these values to be A(623) = 36.9°±6.9°, D(1-2) = 20.2±0.8Å, and D(V-4) = 12.6±0.8Å; all values are very close to those obtained from the Vh1-VBS1 complex equilibration simulation. The average time for binding to occur was 13.9±8.0ns.

Binding of vinculin with other VBSs

Simulations between Vh1 with VBS2, VBS3, and αVBS all underwent complete binding similar to that observed with Vh1-VBS1. The plots of the three indicators defined above provide evidence to support that all VBSs bind to Vh1 through a combination of hydrophobic insertion, H1-H2 displacement, and VBS rotation (Figure 5.5). Simulations between Vh1 and VBS1 with various mutations on VBS1 (K613P, L622A and L623A) and Vh1 (A50I), however, as expected, did not bind (Figure 5.6). In order to investigate VBS1 secondary structure stability, the extent of helicity, as measured by the number of hydrogen bonds, is evaluated for (i) Vh1-VBS1 complex equilibration, (ii) Vh1-VBS1 constraint-free binding simulation, and (iii) Vh1-VBS1 with K613P mutation binding simulation (Figure 5.7).

5.4 Discussion

Simulations demonstrate that the critical early interaction between Vh1 and VBS1 is the insertion of the hydrophobic residues of VBS1 between H1 and H2 of Vh1 (Figure 5.3A). Following the hydrophobic interaction of Vh1 with VBS1, H1 and H2 in the N-terminal four-helix bundle of Vh1 are displaced to make room for VBS1 in between. By packing their hydrophobic residues in the core, the bound N-terminal four-helix bundle and VBS1 form a new five-helix bundle structure as previously suggested by Izard *et al.* (71). In one constraint-free simulation with Vh1 and VBS1 initially separated by 12Å, this binding mechanism occurred during a simulation time of 34ns. Successful binding of Vh1-VBS1 can be enhanced by forcing this initial hydrophobic insertion in the beginning of the simulations. With initial 800ps distance constraints to position VBS1 between H1 and H2 (without yet displacing H1 and H2 apart), 50% of the simulations (6 of 12) underwent the binding by the same mechanism in 13.9±8.0ns. Therefore, the critical step in the binding mechanism is the hydrophobic insertion of VBS1 into Vh1, and once this occurs, VBS1 continues to push its way into the hydrophobic core and finally snaps in by rotating the remaining exposed hydrophobic residues (L619 and L623) into the core. It is important to note that during binding, the secondary structures of both proteins remain largely intact. This observation leads us to believe that the forced activation of talin is a subtle change in conformation, and that complete unfolding is not necessary.

Only one constraint-free simulation actually underwent the Vh1-VBS1 binding mechanism out of 20 attempted constraint-free simulations. In a majority of the unsuccessful binding simulations, the hydrophobic residues of VBS1 (L608, L615 and L622) were displaced away from the H1-H2 groove of Vh1. For example, simulations in which VBS1 hydrophobic residues interacted only with the opposite side of H2 failed to displace H1 from H2, and did not bind. Nonetheless, this is the first reported simulated binding between two molecules in the complete absence of external forces that involves significant conformational change and for which the crystal structures are known for both non-bound and bound states for verification. This suggests that the key to simulating molecular binding is to ensure the correct orientation and initial contact between the two molecules. While this could be achieved without knowledge of the structure of the bound complex, it would require a large number of calculations to identify the correct approach, and this would be computationally prohibitive for most cases.

The proposed binding mechanism is further supported by successful binding simulations of Vh1 with four different VBS peptides through the identical binding mechanism: hydrophobic insertion, displacement of H1 and H2, and rotation of VBS (Figures 5.3 and 5.5). All the critical hydrophobic residues involved in the hydrophobic insertion are also found in nearby positions for VBS1, VBS2, VBS3 and αVBS (Figure 5.4). Interestingly, the residue sequence in αVBS is reversed to that of talin VBSs, but the critical hydrophobic residues are still found in the corresponding positions needed to undergo the proposed binding mechanism to Vh1. Similar to talin, α-actinin contains a cryptic VBS and is possibly subjected to mechanical force within cell-cell junction (127). Therefore, vinculin and α-actinin binding may proceed in the similar mechanism within cell-cell junctions as well as cell-matrix junctions. Also, this generality provides a critical insight into how talin, containing 11 potential VBSs (56), might modify its conformation when subjected to tensile force to recruit multiple vinculins with a concomitant increase in adhesion strength, as has been observed experimentally (138).

When Vh1 and VBS1 form the initial hydrophobic contact, a number of hydrogen-bonds are formed surrounding the hydrophobic interface. In particular, the hydrogen-bond formed between Q627 and H22 was persistent throughout the binding simulations, and was initially suspected to be a critical interaction that stabilizes the initial hydrophobic insert in place. However, this hypothesis was disproved by mutating either Q627 on Vh1 or H22 on VBS1 to alanine and demonstrating through additional simulations, that binding still occurs in these mutation simulations. The hydrogen-bond persisted because hydrophobic insertion stabilized the interaction of the two molecules and placed Q627 and H22 in close proximity, allowing the bond to remain intact, rather than the other way around. This is yet another example that supports the importance of the initial hydrophobic insertion of VBS1, and also demonstrates the power of using molecular dynamics to quickly test hypotheses.

Gingras *et al.* (56) identified the critical residues on VBS1 for binding to Vh1 through a comprehensive mutational study. The identified critical residues were mostly hydrophobic, a finding that is consistent with the present numerical study. The mutational study showed that L608, L615, L622 and L623 are each individually critical for the stable binding of VBS1 to Vh1 (56), the mechanism for which can be derived from the numerical results. L622 is apparently important in the hydrophobic insertion of VBS1, and when it is mutated to alanine (L622A), Vh1-VBS1 binding does not proceed because the hydrophobic insertion is inhibited (Figure 5.6C, 5.6H and 5.6M). In contrast, the simulation with L623A underwent the initial hydrophobic insertion and nearly completed the entire binding process, but the smaller alanine residue was insufficient to snap into the core and remain bound, hence destabilizing the Vh1-VBS1 bound complex

(Figure 5.6D, 5.6I and 5.6N). An interesting finding in the mutational study (56) was that when any one of VBS1 residues was mutated to a proline, binding was abolished. A binding simulation with K613P mutation on VBS1 shows that the proline induces a break in the α-helix, which significantly reduces VBS1 helicity and prevents VBS1 hydrophobic insertion into Vh1 (Figure 5.6B, 5.6G, 5.6L and 5.7). One mutation on vinculin, A50I, has been also identified to inhibit vinculin-talin binding by stabilizing the interaction between H1 and H2 of vinculin (73). Indeed, the A50I mutation simulation did not undergo Vh1-VBS1 binding as the bulk of Ile prevented L615 of VBS1 from inserting in between H1 and H2 (Figure 5.6E, 5.6J and 5.6O).

Interestingly, the hydrophobic residues (L608, L615, and L622) of VBS1 involved in the hydrophobic insertion between H1 and H2 of Vh1 are the exact same residues that are exposed to solvent in force-induced activation of TAL5 (136). In the activated TAL5 structure, however, the hydrophobic residues (V619 and L623) that snap into the hydrophobic core in the later stages of binding are still embedded in the hydrophobic core of TAL5 (Figure 5.8). There must therefore be a secondary conformational change in TAL5, not captured in the previous simulations, that exposes these two residues. Based on these observations, we propose that transmitted tensile force alters TAL5 structure to expose hydrophobic residues that are essential in the initial interaction with vinculin binding. In order to approximate the bound configuration of Vh1 and TAL5, binding simulations on Vh1 and force-activated TAL5 were performed with excessive distance constraints between Vh1 and VBS1 of TAL5 to force binding. The resulting configuration of the Vh1-TAL5 complex is shown in Figure 5.8C, which forms a nine-helical bundle together with VBS1 donated to the N-terminal four-helix bundle of Vh1. The binding mechanism viewed from the side is shown in Figure 5.8D-F.

All the simulations presented here are conducted without Vt, therefore, we are proposing a binding mechanism of Vh1 and VBSs. Further studies are needed to determine the vinculin activation mechanism in the presence of talin VBS (71,127). Recent evidence shows that vinculin autoinhibition and vinculin activation are achieved by cooperative efforts (73,135), that is, talin binding must be accompanied by other molecular binding, for example PIP2 (67), to vinculin for full vinculin activation.

In conclusion, a Vh1-VBS binding mechanism has been proposed, which involves hydrophobic insertion of VBS1 into Vh1, separation of H1 and H2 of Vh1, and VBS1 rotation to snap in exposed hydrophobic residues into the hydrophobic core. Results from mutational simulations and binding simulations with other VBSs suggest that the proposed mechanism may be more generally valid. This work constitutes the potential early stages of force-induced focal adhesion strengthening by vinculin recruitment immediately following the force-induced talin activation mechanism (136).

5.5 Figures

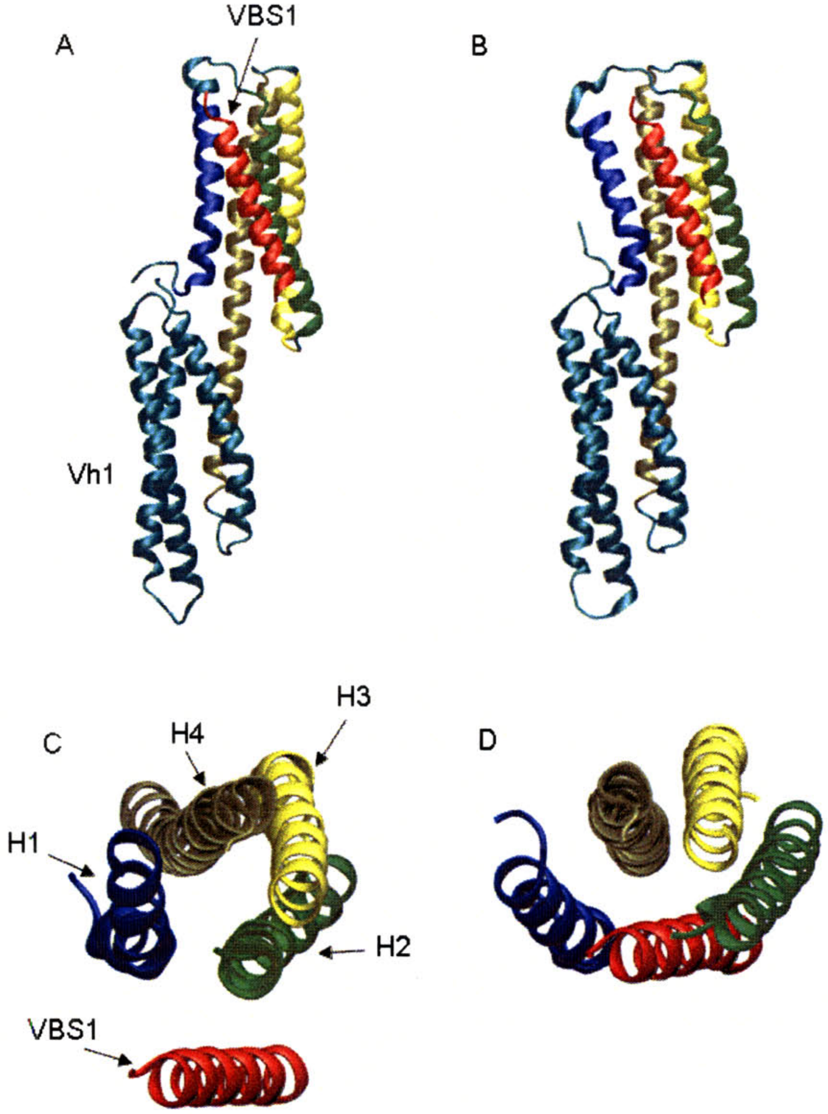

Figure 5.1. (A) Vh1 (obtained from PDB ID: 1RKE) and VBS1 (obtained from PDB ID: 1T01) unbound structures viewed from the front. VBS1 is translated by 12Å from its corresponding position within the Vh1-VBS1 complex. (B) Vh1 and VBS1 bound complex (PDB ID: 1T01) viewed from the front. (C) Vh1 and VBS1 unbound structures viewed from the top. Only the first four helices of Vh1 (a seven-helix-bundle) are shown for clarity. (D) Vh1 and VBS1 bound complex viewed from the top.

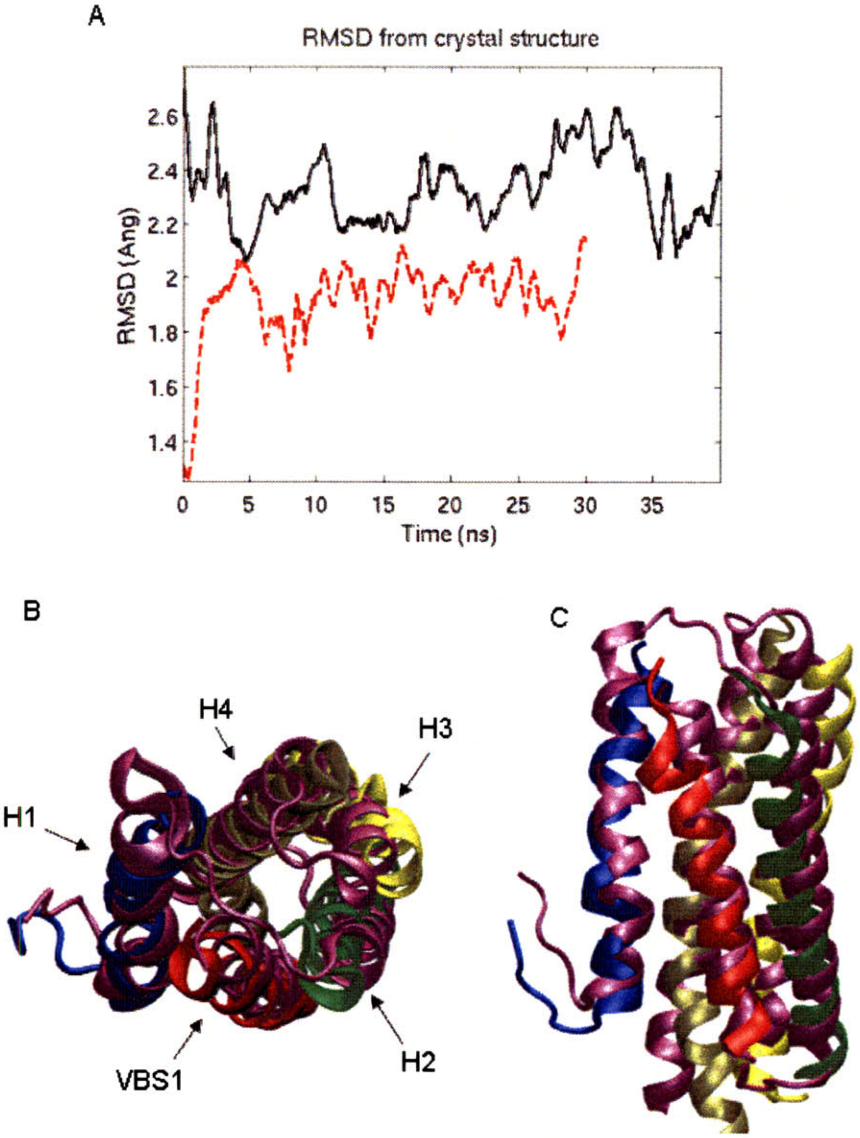

Figure 5.2. (A) Root mean square deviation (RMSD) of the equilibration simulation from the Vh1-VBS1 complex crystal structure (red dashed line) and RMSD of the Vh1 and VBS1 binding simulation from the Vh1-VBS1 complex crystal structure (black solid line). (B) One conformation from Vh1-VBS1 binding simulation (same coloring scheme as in Figure 5.1) superimposed with a conformation from Vh1-VBS1 equilibration simulation (purple) viewed from the top and (C) viewed from the front.

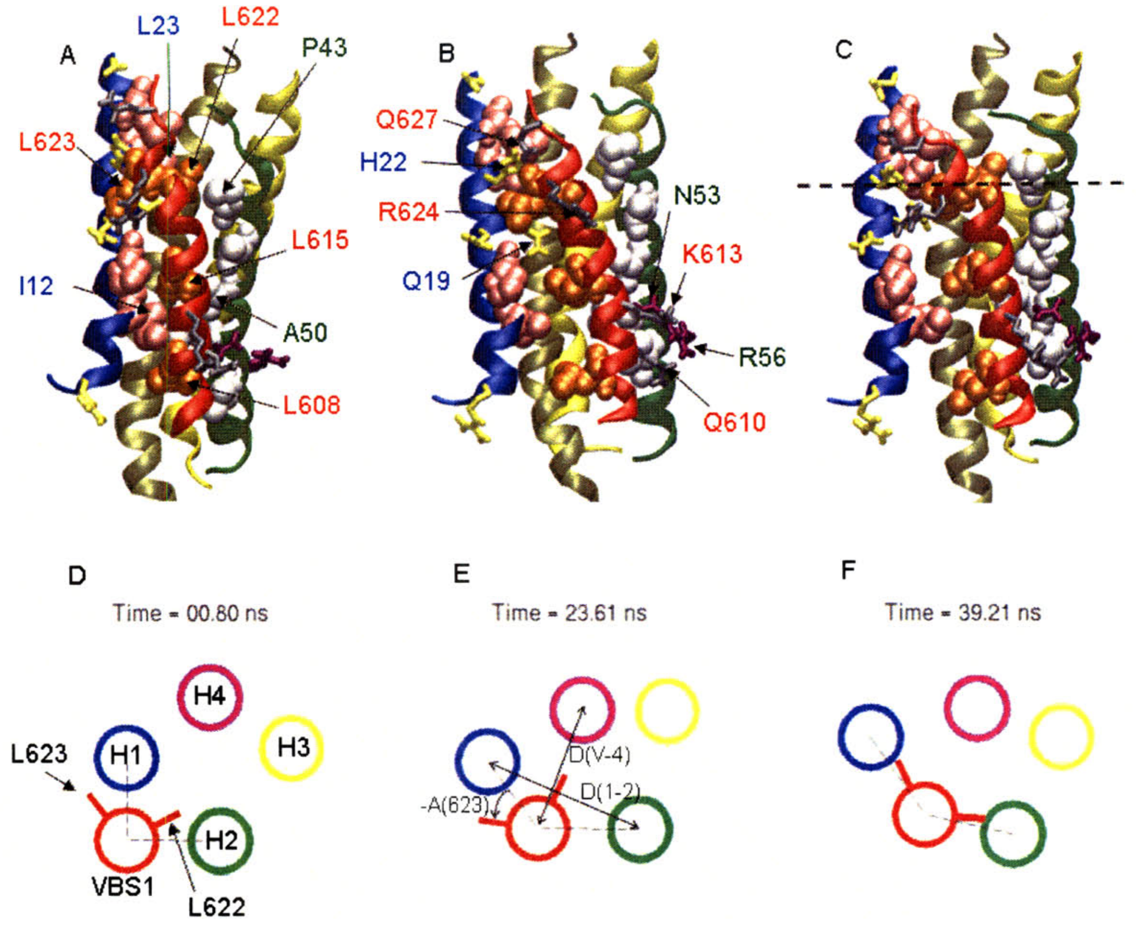

Figure 5.3. Snapshots from one Vh1-VBS1 binding simulation with ribbon representations for helical backbone, stick representations for polar and charged residues, and spherical representations for hydrophobic residues at (A) VBS1 hydrophobic insert (t=0.8ns) between the hydrophobic patch of H1 and H2, (B) Vh1's H1 and H2 separation (t=23.6ns), and (C) VBS1 rotation (t=39.2ns) to snap in the exposed hydrophobic residues, i.e., L623, into the hydrophobic core. Some of the critical residues in Vh1-VBS1 binding are labeled: residues on VBS1 (labeled red), residues on H1 (labeled blue) and residues on H2 (labeled green). (D-F) Corresponding cross-sectional views to (A)-(C) at the plane represented by the dashed line in Figure 5.3C.

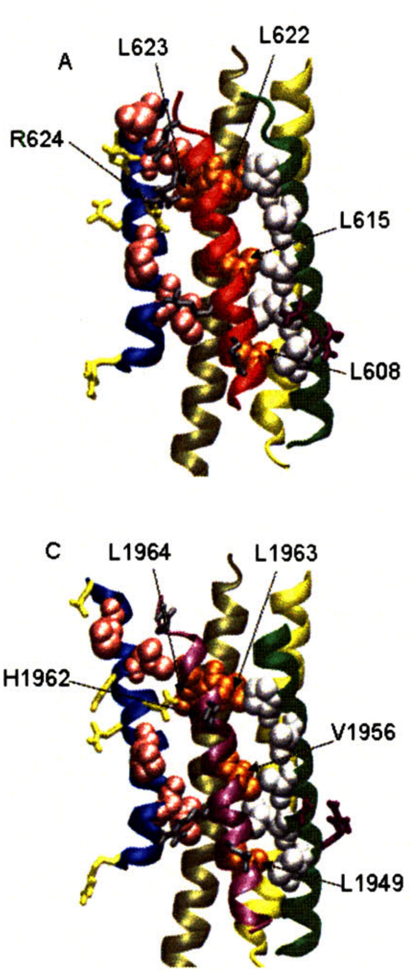

Figure 5.4. Crystal structures of Vh1 bound to various VBSs. The backbones of helical sequences are shown in ribbon representation, polar and charged residues shown in stick representation, and the hydrophobic residues are shown as spheres. (A) Vh1-VBS1 complex (PDB ID: 1T01) with critical residues important in binding of the two molecules. VBS1 is shown as red ribbon. (B) Vh1-VBS2 complex (PDB ID: 1U6H) and corresponding critical residues labeled. VBS2 is shown as a gray ribbon. (C) Vh1-VBS3 complex (PDB ID: 1RKC) and the corresponding critical residues labeled. VBS3 is shown as a purple ribbon. (D) Vh1-αVBS complex (PDB ID: 1YDI) and the corresponding critical residues are labeled. αVBS is shown as a silver ribbon.

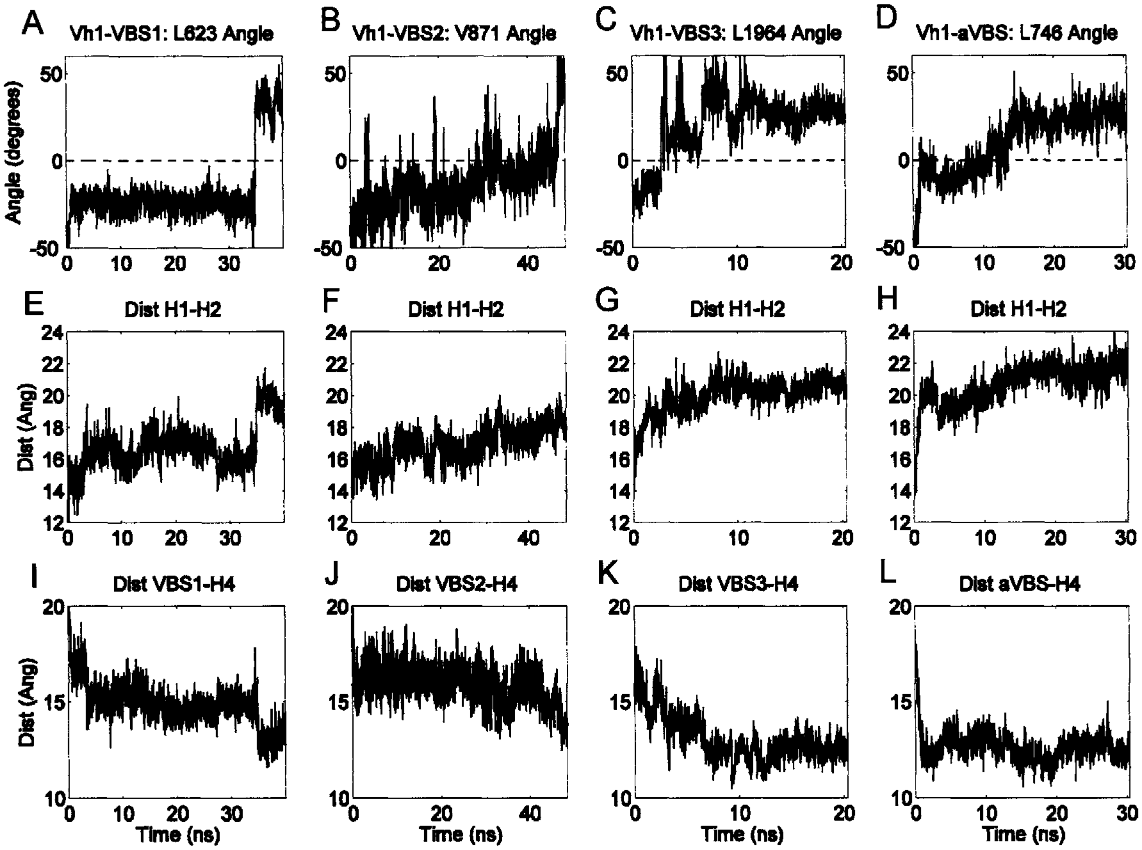

Figure 5.5. (A) Angle formed by L623 with vector connecting H1-VBS1, defined as 'A(623)' in Figure 5.3E. (B) Angle formed by V871 with vector connecting H1-VBS2. (C) Angle formed by L1964 with vector connecting H1-VBS3. (D) Angle formed by L746 with vector connecting H1-αVBS. (E-H) Distance of H1 and H2 in Vh1-VBS1, Vh1-VBS2, Vh1-VBS3, and Vh1-αVBS binding simulations. This distance is defined as 'D(1-2)' in Figure 5.3E. (I-L) Distance of VBS and H4 in Vh1-VBS1, Vh1-VBS2, Vh1-VBS3, and Vh1-αVBS binding simulations. The distance of VBS1-H4 is defined as 'D(V-4)' in Figure 5.3E.

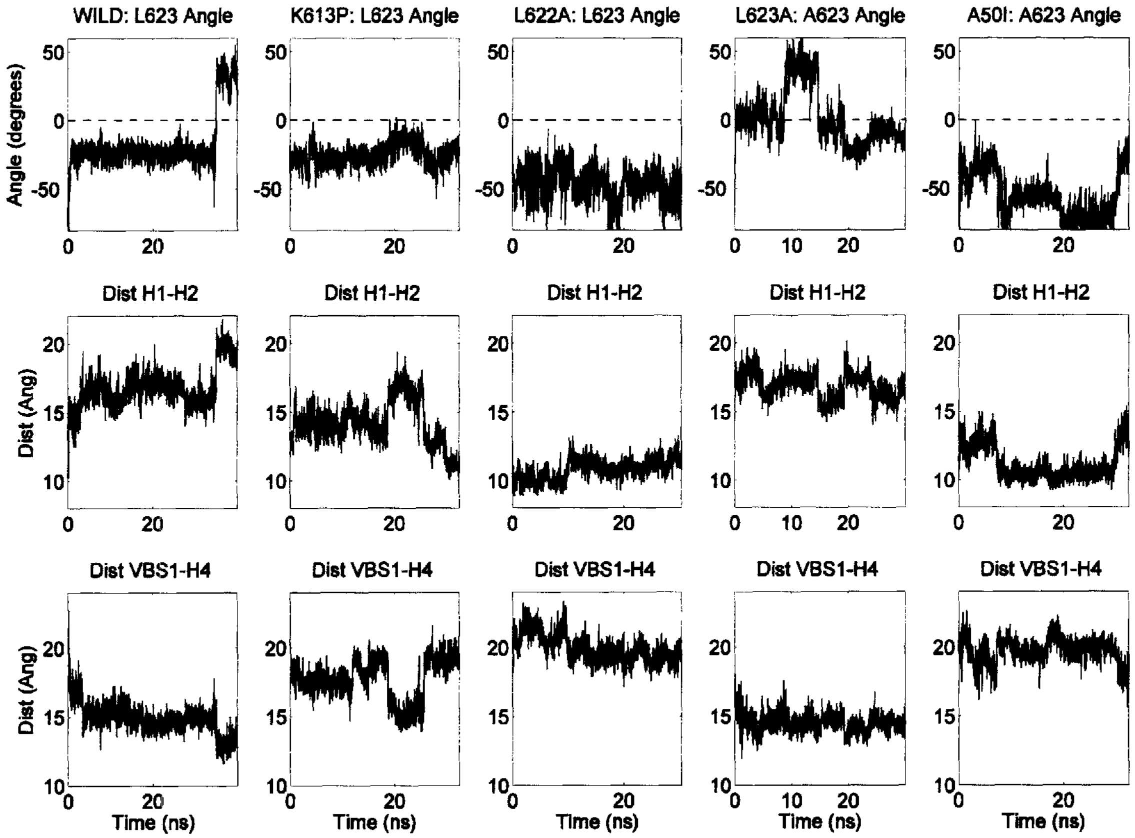

Figure 5.6. Results from the binding simulation of Vh1 with mutated VBS1 similar to Figure 3. (A) Angle formed by L623 (of wild-type VBS1) with vector connecting H1-VBS1. (B) Angle formed by L623 with vector connecting H1-VBS1, where VBS1 has K613P mutation. (C) Angle formed by L623 with vector connecting H1-VBS1, where VBS1 has L622A mutation. (D) Angle formed by A623 with vector connecting H1-VBS1, where VBS1 has L623A mutation. (E) Angle formed by A623 with vector connecting H1-VBS1, where Vh1 has I50A mutation. (F-J) Distance of H1 and H2 in Vh1-VBS1 binding simulations with wild-type, K613P, L622A, L623A mutations on VBS1, and I50A mutation on Vh1, respectively. (K-O) Distance of VBS1 and H4 in Vh1-VBS1 binding simulations with wild-type, K613P, L622A, L623A mutations on VBS1, and I50A mutation on Vh1, respectively.

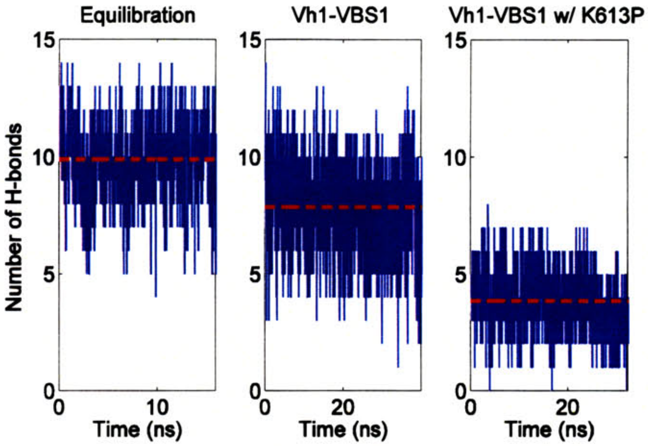

Figure 5.7. Helicity of VBS1 at Vh1-VBS1 bound complex equilibrium simulation, Vh1-VBS1 binding simulation, and Vh1-VBS1 with K613P mutation binding simulation.

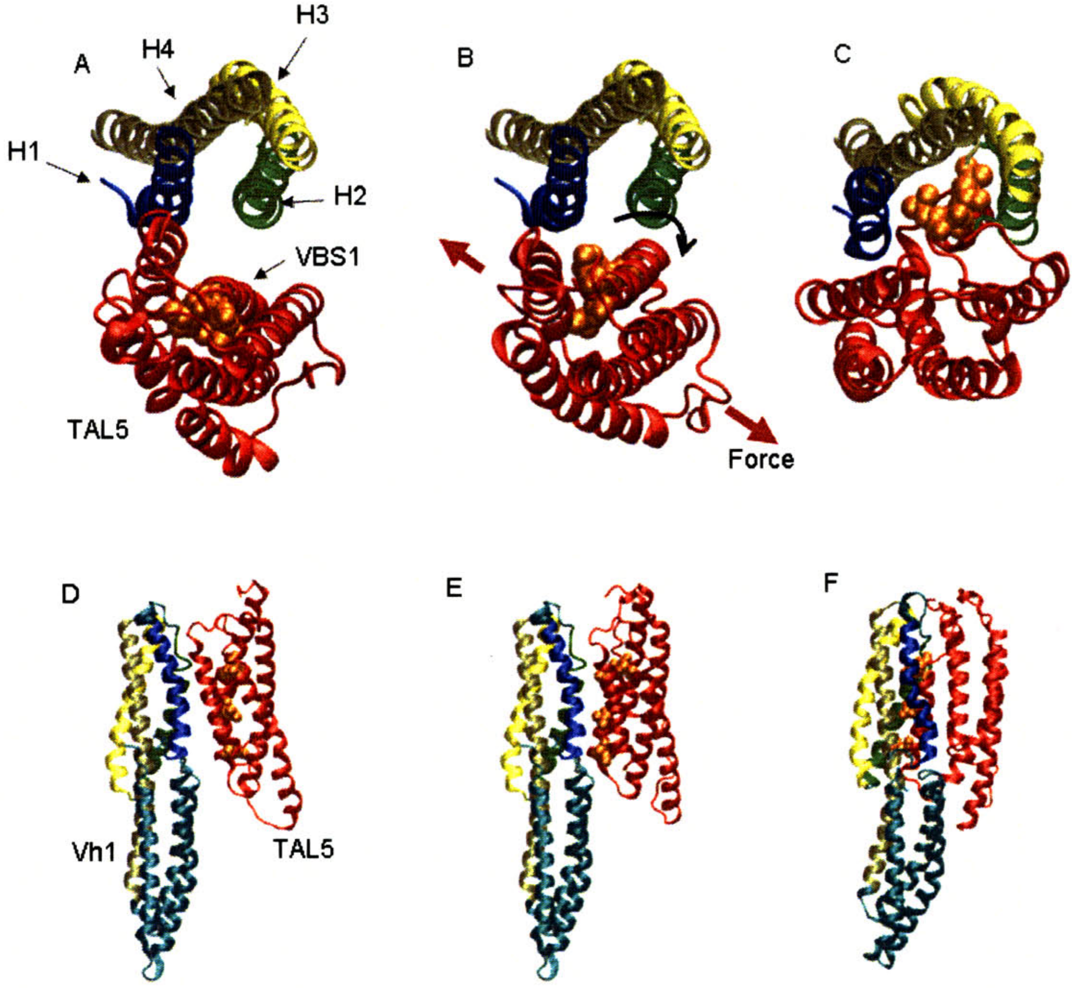

Figure 5.8. Proposed binding model for Vh1 and TAL5. (A) The N-terminal four-helix bundle of Vh1 is shown with the same coloring scheme as in Figure 5.1. Inactivated TAL5 containing VBS1 is shown in red as viewed from the top. The hydrophobic residues necessary for Vh1 binding are shown as orange spheres. (B) With application of mechanical force on TAL5, VBS1 undergoes a rigid body rotation to expose the hydrophobic residues (136). The applied force is represented by the red arrows and VBS1 rotation is indicated by the black arrow. (C) Final configuration of Vh1 with TAL5 in a binding simulation. (D-E) The corresponding configurations to (A-C) viewed from the side.

6 Conclusions and Outlook for Future Directions

Molecular dynamics provides a tool to investigate protein structural dynamics at the molecular scale not possible experimentally. However, few major advancements had been made in protein mechanics other than in the forced unfolding of proteins (2,3) due to limitations in available computational power and a lack of structural information (i.e. crystal structures or NMR structures). MD simulation of full protein unfolding, although very useful in understanding protein response to large mechanical forces, does not address the more subtle conformational changes that may be present in intracellular force-sensitive proteins. One major hurdle in the use of numerical simulation to simulate these more subtle changes is that, in most cases, we lack the essential experimental data needed for model verification. Even though many numerical pulling experiments can be carried out on available protein structures, due to approximations inherent in the models, they have little merit without the independent support of corresponding experimental results. A major problem lies in the fact that numerical and experimental studies are conducted on time scales that differ by many orders of magnitudes, and it is very difficult to make a direct comparison between the two methods. Force-regulation of focal adhesion has been in the spotlight recently, and a large number of experiments have been conducted on vinculin and talin interactions (56,58,71,73,80,82,83). In the present study, we have introduced a potential force-induced talin activation mechanism and the subsequent vinculin binding mechanism. This not only provides a big step toward understanding the force response mechanism of focal adhesions, but also demonstrates the possibility to use MD for studies on intracellular force-sensitive proteins with subtle conformational changes.

The present work progressed in three stages, chosen to systematically develop and validate methods that could be used to extend current capabilities for investigating the role of force in producing conformational changes. In the first part of this study, we used MD to model a single α-helix to investigate the kinetics of folding and unfolding and compared these results with a coarse-grained model developed to simulate the force response in terms of protein conformational change. The peptide was designed to have only two distinct conformational states, where the folded state is favored in the absence of external force. When external force was applied to the peptide, it started to sample the unfolded conformation. Increasing the magnitude of the force resulted in longer sampling of the unfolded conformation, and the characteristics of the simulations aligned well with the parameters of the coarse-grained model. Good general agreement was observed in the extension rates obtained from the MD model and the coarse-grained model. In this first study, MD was used in lieu of experimental data, justified in this instance because of the simplicity of the modeled system, for which simulation data could be viewed as reliable and accurate. This demonstrates the power of MD in visualizing and detailing the behavior of a simple protein system and using the results, in this case, to help validate a less detailed, coarse-grained model.

In the second phase of this study, we considered the effect of force on a more complicated system, one that has been suspected of playing an important role in intracellular force sensation. Considering the structural information and the protein characteristics of talin, a realistic tensile force was applied to one subdomain of the focal adhesion protein, talin and led to an entirely new concept for the force-induced activation of talin. Implementation of this model required a number of assumptions based as much as possible on available structural or experimental data. One example is the nature in which forces are transmitted along the protein. In our simulations, we assumed that force transmission occurs through the hydrogen-bonds between secondary structures (2,85,94). We focused on one of the vinculin binding sites in talin rod, VBS1 within TAL5, for which the structure was know, which has salt-bridge and hydrogen-bonding interactions with one adjacent helix but no such interaction with another helix on the opposite side. As a result of the precise nature of force application, protein extension created a torque on VBS1, causing it to undergo a rigid body rotation which, in turn, exposed its cryptic binding site to the solvent. In contrast to an unfolding process, this conformational change left all hydrogen-bonds intact, and due to the small change in conformation, relaxation of the applied force allowed the exposed hydrophobic residues to embed back to their hydrophobic core.

In the final part of the thesis, a potential binding mechanism was proposed for the binding of vinculin subdomain (Vh1) and talin peptide (VBS1). It has been experimentally shown that these two molecules bind with strong affinity (80), and their bound and unbound crystal structures are known (57,71), but the precise mechanism of their binding has not been identified. This is the *first* reported simulation work that demonstrates the binding of two molecules involving large conformational changes in the absence of external force. Briefly, VBS1 first inserts it hydrophobic residues into the hydrophobic groove between helix 1 and 2 of Vh1, helix 1 and 2 are separated as VBS1 settles in, and ultimately VBS1 rotates and its other hydrophobic residues lock into the hydrophobic core completing the binding process. Additional mutational studies and binding with different VBSs are consistent with the proposed binding mechanism, and the binding of Vh1 to α-actinin's αVBS also suggests that a similar binding mechanism occurs in other intracellular proteins with helical bundle domains.

Together, these results shed light on two of the key steps in the force-induced reinforcement of an initial contact. This is the critical first step in gaining full molecular understanding of the force-sensitive processes within the adhesion plaque, and these suggest a possible pathway leading to the next set of numerical and experimental studies to further elucidate focal adhesion maturation. One potentially fruitful direction would be to use MD to investigate the activation mechanism of vinculin. It is still not clear whether vinculin binds to talin when vinculin is already activated or if talin binding to vinculin itself activates vinculin. Evidence shows that talin may be involved in the activation of vinculin (71,73). Crystal structures of autoinhibitory vinculin head-tail complex are available for the study of potential vinculin activation through talin binding (71,73).

More experimental studies are needed for the validation of the proposed molecular mechanisms. Mutational studies in living cells would be the easiest set of experiments to verify the current findings. For example, the torque applying polar and charged residues of VBS1 in talin activation are essential for the force response of talin in vinculin recruitment. Cells transfected with vectors to produce talin with these point mutations in these critical residues, mutated to nonpolar residues, are expected to show decreased force sensitivity compared to the cells with wildtype talin. Care should be taken in isolating the role of this talin domain in the overall force sensitivity of talin, however, since talin has redundancies such as 11 total potential VBSs (56). Single molecule force spectroscopy can also be envisioned to validate the predicted mechanism of activation. The ability to apply force to a single molecule talin or just the TAL5 domain (or tandem repeats of it) would enable the mentioned mutational studies in more controlled and isolated environment. Application of force with simultaneous visualization of vinculin (e.g., by tethering a fluorescent antibody) would allow direct detection of talin activation. Collaborations are underway with another lab specializing in single-molecule experiments in order to make continuing efforts in trying to completely understand the focal adhesion mechanotransduction. With this revealing MD study as the stepping stone, numerical methods, experiments and coarse-grained modeling will be employed concurrently for the fundamental understanding of focal adhesion dynamics, which can eventually lead to a variety of clinical applications,. The future is bright, and the potential now exists to use a combined experimental/ computational approach to gain new insights into these essential phenomena in which force can regulate or activate intracellular signaling. Recognition of these "mechanical" signaling pathways holds the further potential for the use of "mechanical therapeutics" or ways of controlling cell behavior through mechanical as opposed to biochemical approaches.

7 References

1. Kamm, R. D. and M. R. Kaazempur-Mofrad. 2004. On the molecular basis for mechanotransduction. Mech Chem Biosyst 1:201-209.

2. Vogel, V. 2006. Mechanotransduction involving multimodular proteins: converting force into biochemical signals. Annu Rev Biophys Biomol Struct 35:459-488.

3. Vogel, V. and M. Sheetz. 2006. Local force and geometry sensing regulate cell functions. Nat Rev Mol Cell Biol 7:265-275.

4. Bao, G. and S. Suresh. 2003. Cell and molecular mechanics of biological materials. Nat Mater 2:715-725.

5. Chen, C. S., J. Tan, and J. Tien. 2004. Mechanotransduction at cell-matrix and cell-cell contacts. Annu Rev Biomed Eng 6:275-302.

6. Janmey, P. A. and D. A. Weitz. 2004. Dealing with mechanics: mechanisms of force transduction in cells. Trends Biochem Sci 29:364-370.

7. Kaazempur-Mofrad MR, A.-R. N., Karcher H, Mack PJ, Yap B, Kamm RD. 2005. Exploring the molecular basis for mechanosensation, signal transduction, and cytoskeletal remodeling. ActaBioMaterialia 1:281-293.

8. Haidekker, M. A., N. L'Heureux, and J. A. Frangos. 2000. Fluid shear stress increases membrane fluidity in endothelial cells: a study with DCVJ fluorescence. Am J Physiol Heart Circ Physiol 278:H1401-1406.

9. White, C. R., M. Haidekker, X. Bao, and J. A. Frangos. 2001. Temporal gradients in shear, but not spatial gradients, stimulate endothelial cell proliferation. Circulation 103:2508-2513.

10. Hamill, O. P. and B. Martinac. 2001. Molecular basis of mechanotransduction in living cells. Physiol Rev 81:685-740.

11. Yao, Y., A. Rabodzey, and C. F. Dewey, Jr. 2007. Glycocalyx modulates the motility and proliferative response of vascular endothelium to fluid shear stress. Am J Physiol Heart Circ Physiol 293:H1023-1030.

12. Odde, D. J., L. Ma, A. H. Briggs, A. DeMarco, and M. W. Kirschner. 1999. Microtubule bending and breaking in living fibroblast cells. J Cell Sci 112 (Pt 19):3283-3288.

13. Maniotis, A. J., C. S. Chen, and D. E. Ingber. 1997. Demonstration of mechanical connections between integrins, cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure. Proc Natl Acad Sci U S A 94:849-854.

14. Tschumperlin, D. J., G. Dai, I. V. Maly, T. Kikuchi, L. H. Laiho, A. K. McVittie, K. J. Haley, C. M. Lilly, P. T. So, D. A. Lauffenburger, R. D. Kamm, and J. M. Drazen. 2004. Mechanotransduction through growth-factor shedding into the extracellular space. Nature 429:83-86.

15. Helmke, B. P., A. B. Rosen, and P. F. Davies. 2003. Mapping mechanical strain of an endogenous cytoskeletal network in living endothelial cells. Biophys J 84:2691-2699.

16. Sawada, Y. and M. P. Sheetz. 2002. Force transduction by Triton cytoskeletons. J Cell Biol 156:609-615.

17. Sukharev, S. and D. P. Corey. 2004. Mechanosensitive channels: multiplicity of families and gating paradigms. Sci STKE 2004:re4.

18. Gillespie, P. G. and R. G. Walker. 2001. Molecular basis of mechanosensory transduction. Nature 413:194-202.

19. Sachs, F. and C. E. Morris. 1998. Mechanosensitive ion channels in nonspecialized cells. Rev Physiol Biochem Pharmacol 132:1-77.

20. Maroto, R., A. Raso, T. G. Wood, A. Kurosky, B. Martinac, and O. P. Hamill. 2005. TRPC1 forms the stretch-activated cation channel in vertebrate cells. Nat Cell Biol 7:179-185.

21. Sheetz, M. P. 2001. Cell control by membrane-cytoskeleton adhesion. Nat Rev Mol Cell Biol 2:392-396.

22. Anishkin, A. and S. Sukharev. 2004. Water dynamics and dewetting transitions in the small mechanosensitive channel MscS. Biophys J 86:2883-2895.

23. Sukharev, S. I., W. J. Sigurdson, C. Kung, and F. Sachs. 1999. Energetic and spatial parameters for gating of the bacterial large conductance mechanosensitive channel, MscL. J Gen Physiol 113:525-540.

24. Chang, G., R. H. Spencer, A. T. Lee, M. T. Barclay, and D. C. Rees. 1998. Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. Science 282:2220-2226.

25. Gullingsrud, J. and K. Schulten. 2003. Gating of MscL studied by steered molecular dynamics. Biophys J 85:2087-2099.

26. Sukharev, S. 2002. Purification of the small mechanosensitive channel of Escherichia coli (MscS): the subunit structure, conduction, and gating characteristics in liposomes. Biophys J 83:290-298.

27. Bass, R. B., P. Strop, M. Barclay, and D. C. Rees. 2002. Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel. Science 298:1582-1587.

28. Sotomayor, M. and K. Schulten. 2004. Molecular dynamics study of gating in the mechanosensitive channel of small conductance MscS. Biophys J 87:3050-3065.

29. Levina, N., S. Totemeyer, N. R. Stokes, P. Louis, M. A. Jones, and I. R. Booth. 1999. Protection of Escherichia coli cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. Embo J 18:1730-1737.

30. Sukharev, S. I., P. Blount, B. Martinac, and C. Kung. 1997. Mechanosensitive channels of Escherichia coli: the MscL gene, protein, and activities. Annu Rev Physiol 59:633-657.

31. Hadden, H. L. and C. A. Henke. 2000. Induction of lung fibroblast apoptosis by soluble fibronectin peptides. Am J Respir Crit Care Med 162:1553-1560.

32. Discher, D. E., P. Janmey, and Y. L. Wang. 2005. Tissue cells feel and respond to the stiffness of their substrate. Science 310:1139-1143.

33. Pelham, R. J., Jr. and Y. Wang. 1997. Cell locomotion and focal adhesions are regulated by substrate flexibility. Proc Natl Acad Sci U S A 94:13661-13665.

34. Deroanne, C. F., C. M. Lapiere, and B. V. Nusgens. 2001. In vitro tubulogenesis of endothelial cells by relaxation of the coupling extracellular matrix-cytoskeleton. Cardiovasc Res 49:647-658.

35. Engler, A., L. Bacakova, C. Newman, A. Hategan, M. Griffin, and D. Discher. 2004. Substrate compliance versus ligand density in cell on gel responses. Biophys J 86:617-628.

36. Lo, C. M., H. B. Wang, M. Dembo, and Y. L. Wang. 2000. Cell movement is guided by the rigidity of the substrate. Biophys J 79:144-152.

37. Cramer, L. P. and T. J. Mitchison. 1995. Myosin is involved in postmitotic cell spreading. J Cell Biol 131:179-189.

38. Geiger, B. and A. Bershadsky. 2001. Assembly and mechanosensory function of focal contacts. Curr Opin Cell Biol 13:584-592.

39. Zamir, E. and B. Geiger. 2001. Molecular complexity and dynamics of cell-matrix adhesions. J Cell Sci 114:3583-3590.

40. Galbraith, C. G., K. M. Yamada, and M. P. Sheetz. 2002. The relationship between force and focal complex development. J. Cell Biol. 159:695-705.

41. DeMali, K. A., C. A. Barlow, and K. Burridge. 2002. Recruitment of the Arp2/3 complex to vinculin: coupling membrane protrusion to matrix adhesion. J Cell Biol 159:881-891.

42. Laukaitis, C. M., D. J. Webb, K. Donais, and A. F. Horwitz. 2001. Differential dynamics of alpha 5 integrin, paxillin, and alpha-actinin during formation and disassembly of adhesions in migrating cells. J Cell Biol 153:1427-1440.

43. Rottner, K., A. Hall, and J. V. Small. 1999. Interplay between Rac and Rho in the control of substrate contact dynamics. Curr Biol 9:640-648.

44. Balaban, N. Q., U. S. Schwarz, D. Riveline, P. Goichberg, G. Tzur, I. Sabanay, D. Mahalu, S. Safran, A. Bershadsky, L. Addadi, and B. Geiger. 2001. Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates. Nat Cell Biol 3:466-472.

45. Nobes, C. D. and A. Hall. 1995. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell 81:53-62.

46. Ridley, A. J. and A. Hall. 1992. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. Cell 70:389-399.

47. Fukata, Y., M. Amano, and K. Kaibuchi. 2001. Rho-Rho-kinase pathway in smooth muscle contraction and cytoskeletal reorganization of non-muscle cells. Trends Pharmacol Sci 22:32-39.

48. Pruyne, D., M. Evangelista, C. Yang, E. Bi, S. Zigmond, A. Bretscher, and C. Boone. 2002. Role of formins in actin assembly: nucleation and barbed-end association. Science 297:612-615.

49. Frisch, S. M. and H. Francis. 1994. Disruption of epithelial cell-matrix interactions induces apoptosis. J Cell Biol 124:619-626.

50. Domingos, P. P., P. M. Fonseca, W. Nadruz, Jr., and K. G. Franchini. 2002. Load-induced focal adhesion kinase activation in the myocardium: role of stretch and contractile activity. Am J Physiol Heart Circ Physiol 282:H556-564.

51. Ren, X. D., W. B. Kiosses, D. J. Sieg, C. A. Otey, D. D. Schlaepfer, and M. A. Schwartz. 2000. Focal adhesion kinase suppresses Rho activity to promote focal adhesion turnover. J Cell Sci 113 (Pt 20):3673-3678.

52. Jiang, G. Y., G. Giannone, D. R. Critchley, E. Fukumoto, and M. P. Sheetz. 2003. Two-piconewton slip bond between fibronectin and the cytoskeleton depends on talin. Nature 424:334-337.

53. Giannone, G., G. Jiang, D. H. Sutton, D. R. Critchley, and M. P. Sheetz. 2003. Talin1 is critical for force-dependent reinforcement of initial integrin-cytoskeleton bonds but not tyrosine kinase activation. J. Cell Biol. 163:409-419.

54. Gilmore, A. P., P. Jackson, G. T. Waites, and D. R. Critchley. 1992. Further Characterization of the Talin-Binding Site in the Cytoskeletal Protein Vinculin. J. Cell Sci. 103:719-731.

55. Menkel, A. R., M. Kroemker, P. Bubeck, M. Ronsiek, G. Nikolai, and B. M. Jockusch. 1994. Characterization of an F-Actin-Binding Domain in the Cytoskeletal Protein Vinculin. J. Cell Biol. 126:1231-1240.

56. Gingras, A. R., W. H. Ziegler, R. Frank, I. L. Barsukov, G. C. Roberts, D. R. Critchley, and J. Emsley. 2005. Mapping and consensus sequence identification for multiple vinculin binding sites within the talin rod. J Biol Chem 280:37217-37224.

57. Papagrigoriou, E., A. R. Gingras, I. L. Barsukov, N. Bate, I. J. Fillingham, B. Patel, R. Frank, W. H. Ziegler, G. C. K. Roberts, D. R. Critchley, and J. Emsley. 2004. Activation of a vinculin-binding site in the talin rod involves rearrangement of a five-helix bundle. Embo J. 23:2942-2951.

58. Patel, B. C., A. R. Gingras, A. A. Bobkov, L. M. Fujimoto, M. Zhang, R. C. Liddington, D. Mazzeo, J. Emsley, G. C. Roberts, I. L. Barsukov, and D. R. Critchley. 2006. The activity of the vinculin binding sites in talin is influenced by the stability of the helical bundles that make up the talin rod. J Biol Chem.

59. Bakolitsa, C., J. M. de Pereda, C. R. Bagshaw, D. R. Critchley, and R. C. Liddington. 1999. Crystal structure of the vinculin tail suggests a pathway for activation. Cell 99:603-613.

60. Winkler, J., H. Lunsdorf, and B. M. Jockusch. 1996. The ultrastructure of chicken gizzard vinculin as visualized by high-resolution electron microscopy. J. Struct. Biol. 116:270-277.

61. Kroemker, M., A. H. Rudiger, B. M. Jockusch, and M. Rudiger. 1994. Intramolecular Interactions in Vinculin Control Alpha-Actinin Binding to the Vinculin Head. FEBS Lett. 355:259-262.

62. Wood, C. K., C. E. Turner, P. Jackson, and D. R. Critchley. 1994. Characterization of the Paxillin-Binding Site and the C-Terminal Focal Adhesion Targeting Sequence in Vinculin. J. Cell Sci. 107:709-717.

63. Johnson, R. P. and S. W. Craig. 1998. Vinculin tail domain is a globular monomer with cryptic oligomerization potential. Mol. Biol. Cell 9:17A-17A.

64. Brindle, N. P. J., M. R. Holt, J. E. Davies, C. J. Price, and D. R. Critchley. 1996. The focal-adhesion vasodilator-stimulated phosphoprotein (VASP) binds to the proline-rich domain in vinculin. Biochem. J. 318:753-757.

65. Kioka, N., S. Sakata, T. Kawauchi, T. Amachi, S. K. Akiyama, K. Okazaki, C. Yaen, K. M. Yamada, and S. Aota. 1999. Vinexin: A novel vinculin-binding protein with multiple SH3 domains enhances actin cytoskeletal organization. J. Cell Biol. 144:59-69.

66. Mandai, K., H. Nakanishi, A. Satoh, K. Takahashi, K. Satoh, H. Nishioka, A. Mizoguchi, and Y. Takai. 1999. Ponsin/SH3P12: An 1-afadin- and vinculin-binding protein localized at cell-cell and cell-matrix adherens junctions. J. Cell Biol. 144:1001-1017.

67. Gilmore, A. P. and K. Burridge. 1996. Regulation of vinculin binding to talin and actin by phosphatidyl-inositol-4-5-bisphosphate. Nature 381:531-535.

68. McGregor, A., A. D. Blanchard, A. J. Rowe, and D. R. Critchley. 1994. Identification of the Vinculin-Binding Site in the Cytoskeletal Protein Alpha-Actinin. Biochem. J. 301:225-233.

69. Xu, W. M., J. L. Coll, and E. D. Adamson. 1998. Rescue of the mutant phenotype by reexpression of full-length vinculin in null F9 cells; effects on cell locomotion by domain deleted vinculin. J. Cell Sci. 111:1535-1544.

70. Critchley, D. R. 2000. Focal adhesions - the cytoskeletal connection. Curr. Opin. Cell Biol. 12:133-139.

71. Izard, T., G. Evans, R. A. Borgon, C. L. Rush, G. Bricogne, and P. R. J. Bois. 2004. Vinculin activation by talin through helical bundle conversion. Nature 427:171-175.

72. Ling, K., R. L. Doughman, A. J. Firestone, M. W. Bunce, and R. A. Anderson. 2002. Type I gamma phosphatidylinositol phosphate kinase targets and regulates focal adhesions. Nature 420:89-93.

73. Bakolitsa, C., D. M. Cohen, L. A. Bankston, A. A. Bobkov, G. W. Cadwell, L. Jennings, D. R. Critchley, S. W. Craig, and R. C. Liddington. 2004. Structural basis for vinculin activation at sites of cell adhesion. Nature 430:583-586.

74. Isenberg, G. and W. H. Goldmann. 1998. Peptide-specific antibodies localize the major lipid binding sites of talin dimers to oppositely arranged N-terminal 47 kDa subdomains. FEBS Lett. 426:165-170.

75. Calderwood, D. A. 2004. Talin controls integrin activation. Biochem. Soc. Trans. 32:434-437.

76. Di Paolo, G., L. Pellegrini, K. Letinic, G. Cestra, R. Zoncu, S. Voronov, S. Chang, J. Guo, M. R. Wenk, and P. De Camilli. 2002. PI(4,5)P2 generation at sites of cell adhesion: Recruitment and regulation of PIP kinase type 1gamma by the FERM domain of talin. Mol. Biol. Cell 13:337A-337A.

77. Xing, B. D., A. Jedsadayanmata, and S. C. T. Lam. 2001. Localization of an integrin binding site to the C terminus of talin. J. Biol. Chem. 276:44373-44378.

78. Hemmings, L., D. J. G. Rees, V. Ohanian, S. J. Bolton, A. P. Gilmore, B. Patel, H. Priddle, J. E. Trevithick, R. O. Hynes, and D. R. Critchley. 1996. Talin contains three actin-binding sites each of which is adjacent to a vinculin-binding site. J. Cell Sci. 109:2715-2726.

79. Bass, M. D., B. J. Smith, S. A. Prigent, and D. R. Critchley. 1999. Talin contains three similar vinculin-binding sites predicted to form an amphipathic helix. Biochem. J. 341:257-263.

80. Bass, M. D., B. Patel, I. G. Barsukov, I. J. Fillingham, R. Mason, B. J. Smith, C. R. Bagshaw, and D. R. Critchley. 2002. Further characterization of the interaction between the cytoskeletal proteins talin and vinculin. Biochem. J. 362:761-768.

81. Priddle, H., L. Hemmings, S. Monkley, A. Woods, B. Patel, D. Sutton, G. A. Dunn, D. Zicha, and D. R. Critchley. 1998. Disruption of the talin gene compromises focal adhesion assembly in undifferentiated but not differentiated embryonic stem cells. J. Cell Biol. 142:1121-1133.

82. Fillingham, I., A. R. Gingras, E. Papagrigoriou, B. Patel, J. Emsley, D. R. Critchley, G. C. K. Roberts, and I. L. Barsukov. 2005. A vinculin binding domain from the talin rod unfolds to form a complex with the vinculin head. Structure 13:65-74.

83. Izard, T. and C. Vonrhein. 2004. Structural basis for amplifying vinculin activation by talin. J. Biol. Chem. 279:27667-27678.

84. Carrion-Vazquez, M., H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. 2003. The mechanical stability of ubiquitin is linkage dependent. Nat Struct Biol 10:738-743.

85. Gao, M., D. Craig, V. Vogel, and K. Schulten. 2002. Identifying unfolding intermediates of FN-III(10) by steered molecular dynamics. J Mol Biol 323:939-950.

86. Lu, H., B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. Biophys J 75:662-671.

87. Marszalek, P. E., H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. 1999. Mechanical unfolding intermediates in titin modules. Nature 402:100-103.

88. Oberhauser, A. F., C. Badilla-Fernandez, M. Carrion-Vazquez, and J. M. Fernandez. 2002. The mechanical hierarchies of fibronectin observed with single-molecule AFM. J Mol Biol 319:433-447.

89. Kellermayer, M. S., S. B. Smith, H. L. Granzier, and C. Bustamante. 1997. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. Science 276:1112-1116.

90. Rief, M., M. Gautel, and H. E. Gaub. 2000. Unfolding forces of titin and fibronectin domains directly measured by AFM. Adv Exp Med Biol 481:129-136; discussion 137-141.

91. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276:1109-1112.

92. Izrailev, S., S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. 1997. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J 72:1568-1581.

93. Paci, E. and M. Karplus. 1999. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol 288:441-459.

94. Craig, D., M. Gao, K. Schulten, and V. Vogel. 2004. Tuning the mechanical stability of fibronectin type III modules through sequence variations. Structure 12:21-30.

95. Hummer, G. and A. Szabo. 2003. Kinetics from nonequilibrium single-molecule pulling experiments. Biophys J 85:5-15.

96. Law, R., P. Carl, S. Harper, P. Dalhaimer, D. W. Speicher, and D. E. Discher. 2003. Cooperativity in forced unfolding of tandem spectrin repeats. Biophys J 84:533-544.

97. Law, R., S. Harper, D. W. Speicher, and D. E. Discher. 2004. Influence of lateral association on forced unfolding of antiparallel spectrin heterodimers. J Biol Chem 279:16410-16416.

98. Ortiz, V., S. O. Nielsen, M. L. Klein, and D. E. Discher. 2005. Unfolding a linker between helical repeats. J Mol Biol 349:638-647.

99. Park, S. and K. Schulten. 2004. Calculating potentials of mean force from steered molecular dynamics simulations. J Chem Phys 120:5946-5961.

100. Zhang, D., J. Gullingsrud, and J. A. McCammon. 2006. Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. J Am Chem Soc 128:3019-3026.

101. Jarzynski, C. 1997. Nonequilibrium equality for free energy differernces. Phys. Rev. Lett. 78:2690.

102. Torrie, G. a. V., J.P. 1974. Monte Carlo free energy estimates using non-Boltzmann sampling estimation: umbrella sampling. J. Comp. Phys 23:187-199.

103. Hummer, G. and A. Szabo. 2001. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. Proc Natl Acad Sci U S A 98:3658-3661.

104. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J. Comput. Chem. 4:187-217.

105. Neria, E., Fischer, S. & Karplus, M. 1996. Simulation of activation free energies in molecular dynamics system. J. Chem. Phys. 105:1902-1921.

106. MacKerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., et al. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. 102:3586-3616.

107. Krautler, V., W. F. Van Gunsteren, and P. H. Hunenberger. 2001. A fast SHAKE: Algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J. Comput. Chem. 22:501-508.

108. Lazaridis, T. and M. Karplus. 1999. Effective energy function for proteins in solution. Proteins 35:133-152.

109. Im, W., M. S. Lee, and C. L. Brooks, 3rd. 2003. Generalized born model with a simple smoothing function. J Comput Chem 24:1691-1702.

110. Schaefer, M., C. Bartels, F. Leclerc, and M. Karplus. 2001. Effective atom volumes for implicit solvent models: comparison between Voronoi volumes and minimum fluctuation volumes. J Comput Chem 22:1857-1879.

111. Brockwell, D. J., E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. 2003. Pulling geometry defines the mechanical resistance of a beta-sheet protein. Nat Struct Biol 10:731-737.

112. Paci, E. and M. Karplus. 2000. Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci U S A 97:6521-6526.

113. Beck, D. A., R. S. Armen, and V. Daggett. 2005. Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. Biochemistry 44:609-616.

114. Elber, R. and M. Karplus. 1987. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. Science 235:318-321.

115. Karcher, H., S. E. Lee, M. R. Kaazempur-Mofrad, and R. D. Kamm. 2006. A coarse-grained model for force-induced protein deformation and kinetics. Biophys J 90:2686-2697.

116. Zwanzig, R. 2001. Nonequilibrium Statistical Mechanics. New York: Oxford University Press.

117. Schulten, K., Schulten, Z., and Szabo, A. 1981. Dynamics of reactions involving diffusive barrier crossing. J. Chem. Phys. 74:4426-4432.

118. Szabo, A., Schulten, K., and Schulten, Z. 1980. First passage time approach to diffusion controlled reactions. J. Chem. Phys. 72:4350-4357.

119. Evans, E. and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. Biophys J 72:1541-1555.

120. Bell, G. I. 1978. Models for the specific adhesion of cells to cells. Science 200:618-627.

121. Kramers, H. 1940. Brownian motion in a field of force and the diffusion model of chemical reactions. Physica (Utrecht) 7:284-304.

122. Calimet, N., M. Schaefer, and T. Simonson. 2001. Protein molecular dynamics with the generalized Born/ACE solvent model. Proteins 45:144-158.

123. Grubmuller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 271:997-999.

124. Hoover, W. G. 1985. Canonical dynamics: Equilibrium phase-space distributions. Physical Review. A 31:1695-1697.

125. Nosé, S. 1984. A molecular dynamics method for simulations in the canonical ensemble. Mol. Phys. 52:255-268.

126. Feig, M., J. Karanicolas, and C. L. Brooks, 3rd. 2004. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model 22:377-395.

127. Bois, P. R., R. A. Borgon, C. Vonrhein, and T. Izard. 2005. Structural dynamics of alpha-actinin-vinculin interactions. Mol Cell Biol 25:6112-6122.

128. Ylanne, J., K. Scheffzek, P. Young, and M. Saraste. 2001. Crystal structure of the alpha-actinin rod reveals an extensive torsional twist. Structure 9:597-604.

129. Li, P. C. and D. E. Makarov. 2004. Simulation of the mechanical unfolding of ubiquitin: probing different unfolding reaction coordinates by changing the pulling geometry. J Chem Phys 121:4826-4832.

130. Kirmizialtin, S., L. Huang, and D. E. Makarov. 2005. Topography of the free-energy landscape probed via mechanical unfolding of proteins. J Chem Phys 122:234915.

131. Finer, J. T., A. D. Mehta, and J. A. Spudich. 1995. Characterization of single actin-myosin interactions. Biophys J 68:291S-296S; discussion 296S-297S.

132. Thoumine, O. and J. J. Meister. 2000. Dynamics of adhesive rupture between fibroblasts and fibronectin: microplate manipulations and deterministic model. Eur Biophys J 29:409-419.

133. Litvinov, R. I., H. Shuman, J. S. Bennett, and J. W. Weisel. 2002. Binding strength and activation state of single fibrinogen-integrin pairs on living cells. Proc Natl Acad Sci U S A 99:7426-7431.

134. Johnson, R. P. and S. W. Craig. 1995. F-actin binding site masked by the intramolecular association of vinculin head and tail domains. Nature 373:261-264.

135. Cohen, D. M., H. Chen, R. P. Johnson, B. Choudhury, and S. W. Craig. 2005. Two distinct head-tail interfaces cooperate to suppress activation of vinculin by talin. J Biol Chem 280:17109-17117.

136. Lee, S. E., R. D. Kamm, and M. R. Mofrad. 2007. Force-induced activation of Talin and its possible role in focal adhesion mechanotransduction. J Biomech 40:2096-2106.

137. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33-38, 27-38.

138. Ziegler, W. H., R. C. Liddington, and D. R. Critchley. 2006. The structure and regulation of vinculin. Trends Cell Biol 16:453-460.

Appendix

A.1 Adding missing residues in the model

The PDB file in the Databank may have missing residues, and need to be incorporated into the PDB file. This section shows an example of a PDB file with missing residues (607, 608 and 609), a modified PDB file, and a CHARMM script to read in the modified PDB file that fills in the missing coordinates. The missing coordinates can be assigned to the CHARMM default values of (9999.000, 9999.000, 9999.000), and in the CHARMM script, these are replaced based on the topology and parameter files.

A PDB file with missing residues

```
ATOM   1997  N   ARG B 606      43.037   8.025  24.093  1.00  1.00    N
ATOM   1998  CA  ARG B 606      42.486   7.894  25.433  1.00  1.00    C
ATOM   1999  C   ARG B 606      43.336   7.111  26.432  1.00  1.00    C
ATOM   2000  O   ARG B 606      42.992   7.045  27.609  1.00  1.00    O
ATOM   2001  CB  ARG B 606      41.066   7.313  25.392  1.00  1.00    C
ATOM   2002  CG  ARG B 606      40.938   5.902  24.858  1.00  1.00    C
ATOM   2003  CD  ARG B 606      39.518   5.391  25.093  1.00  1.00    C
ATOM   2004  NE  ARG B 606      39.229   4.173  24.338  1.00  1.00    N
ATOM   2005  CZ  ARG B 606      39.202   4.110  23.008  1.00  1.00    C
ATOM   2006  NH1 ARG B 606      39.448   5.197  22.286  1.00  1.00    N
ATOM   2007  NH2 ARG B 606      38.929   2.963  22.396  1.00  1.00    N
ATOM   2031  N   GLN B 610      42.752   7.477  30.352  1.00  1.00    N
ATOM   2032  CA  GLN B 610      42.481   6.371  31.264  1.00  1.00    C
ATOM   2033  C   GLN B 610      43.538   6.314  32.361  1.00  1.00    C
ATOM   2034  O   GLN B 610      43.237   5.973  33.503  1.00  1.00    O
ATOM   2035  CB  GLN B 610      42.422   5.045  30.502  1.00  1.00    C
ATOM   2036  CG  GLN B 610      41.122   4.866  29.738  1.00  1.00    C
ATOM   2037  CD  GLN B 610      41.052   3.552  29.002  1.00  1.00    C
ATOM   2038  OE1 GLN B 610      40.018   3.206  28.427  1.00  1.00    O
ATOM   2039  NE2 GLN B 610      42.156   2.810  29.006  1.00  1.00    N
```

Modified PDB file with added missing residues

```
ATOM   1997  N   ARG B 606      43.037   8.025  24.093  1.00  1.00    N
ATOM   1998  CA  ARG B 606      42.486   7.894  25.433  1.00  1.00    C
ATOM   1999  C   ARG B 606      43.336   7.111  26.432  1.00  1.00    C
ATOM   2000  O   ARG B 606      42.992   7.045  27.609  1.00  1.00    O
ATOM   2001  CB  ARG B 606      41.066   7.313  25.392  1.00  1.00    C
```

```
ATOM   2002  CG  ARG B 606      40.938   5.902  24.858  1.00  1.00     C
ATOM   2003  CD  ARG B 606      39.518   5.391  25.093  1.00  1.00     C
ATOM   2004  NE  ARG B 606      39.229   4.173  24.338  1.00  1.00     N
ATOM   2005  CZ  ARG B 606      39.202   4.110  23.008  1.00  1.00     C
ATOM   2006  NH1 ARG B 606      39.448   5.197  22.286  1.00  1.00     N
ATOM   2007  NH2 ARG B 606      38.929   2.963  22.396  1.00  1.00     N
ATOM   2008  N   PRO B 607    9999.0009999.0009999.000  1.00  1.00     N
ATOM   2009  CA  PRO B 607    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2010  C   PRO B 607    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2011  O   PRO B 607    9999.0009999.0009999.000  1.00  1.00     O
ATOM   2012  CB  PRO B 607    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2013  CG  PRO B 607    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2014  CD  PRO B 607    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2015  N   LEU B 608    9999.0009999.0009999.000  1.00  1.00     N
ATOM   2016  CA  LEU B 608    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2017  C   LEU B 608    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2018  O   LEU B 608    9999.0009999.0009999.000  1.00  1.00     O
ATOM   2019  CB  LEU B 608    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2020  CG  LEU B 608    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2021  CD1 LEU B 608    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2022  CD2 LEU B 608    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2023  N   LEU B 609    9999.0009999.0009999.000  1.00  1.00     N
ATOM   2024  CA  LEU B 609    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2025  C   LEU B 609    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2026  O   LEU B 609    9999.0009999.0009999.000  1.00  1.00     O
ATOM   2027  CB  LEU B 609    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2028  CG  LEU B 609    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2029  CD1 LEU B 609    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2030  CD2 LEU B 609    9999.0009999.0009999.000  1.00  1.00     C
ATOM   2031  N   GLN B 610      42.752   7.477  30.352  1.00  1.00     N
ATOM   2032  CA  GLN B 610      42.481   6.371  31.264  1.00  1.00     C
ATOM   2033  C   GLN B 610      43.538   6.314  32.361  1.00  1.00     C
ATOM   2034  O   GLN B 610      43.237   5.973  33.503  1.00  1.00     O
ATOM   2035  CB  GLN B 610      42.422   5.045  30.502  1.00  1.00     C
ATOM   2036  CG  GLN B 610      41.122   4.866  29.738  1.00  1.00     C
ATOM   2037  CD  GLN B 610      41.052   3.552  29.002  1.00  1.00     C
ATOM   2038  OE1 GLN B 610      40.018   3.206  28.427  1.00  1.00     O
ATOM   2039  NE2 GLN B 610      42.156   2.810  29.006  1.00  1.00     N
```

CHARMM script to read in the modified PDB file and to assign missing coordinates

```
* This script builds PSF from reading a PDB file
* Written by: Seung E. Lee
* Last modified: 1/17/07
*
!=====================================================================
==
! Parameter definition and reading in the structure information
!=====================================================================
==
  set str_dir /home/selee/stream

! READ THE RTF AND PARAM FILES
  open read unit 11 card name -
       "/home/selee/stream/toppar/toph19_eef1_add.inp"
```

```
  read rtf unit 11 card

  open read unit 12 card name -
       "/home/selee/stream/toppar/param19_eef1_add.inp"
  read parameters unit 12 card

! READING THE SEQUENCE AND COORDINATE FROM THE PDB FILE
  set pdb_in pdb/vbs1_orig
  set segname1 vbs1

  open read unit 20 card name pdb_o/@{pdb_in}.pdb
  read sequence pdb unit 20
  generate @{segname1} setup
  open read unit 14 card name @{pdb_in}.pdb
  read coordinates pdb unit 14

! FILLING IN THE MISSING COORDINATES WITH THE INFO FROM
! TOPOLOGY AND PARAM FILES
  ic purge
  ic param
  ic fill preserve
  ic build
  hbuild

! ACTIVATION OF THE IMPLICIT SOLVENT
  fast 1
  shake bonh
  set solvfile /usr/common/charmm/c29b1_i_lam/test/data/solvpar.inp
  eef1 setup temp 300.0 unit 93 name @solvfile
  update ctonnb 7.0 ctofnb 9.0 cutnb 10.0 group rdie

!=======================================================================
==
! RUNNING MINIMIZATION FOR THE STRUCTURE
!=======================================================================
==
  set nmin  1000
  set mstep 0.001
  set npr   50
  cons harm force  20.0 sele .not. type H* end
  mini sd   nstep @nmin nprint @npr step @mstep

  cons harm clear
  cons harm force  10.0 sele .not. type H* end
  mini abnr nstep @nmin nprint @npr step @mstep

  cons harm clear
  cons harm force  50.0 sele back end
  mini sd   nstep @nmin nprint @npr step @mstep

  cons harm clear
  cons harm force  25.0 sele back end
  mini abnr nstep @nmin nprint @npr step @mstep

  cons harm clear
  mini sd   nstep @nmin nprint @npr step @mstep
  mini abnr nstep @nmin nprint @npr step @mstep
```

```
! WRITING THE BUILT AND MINIMIZED STRUCTURE TO A NEW NAME
  set outfilename vbs1_new
  open write unit 17 card name @{outfilename}.psf
  write  psf unit 17 card
  open write unit 18 card name @{outfilename}.crd
  write coor unit 18 card
  open write unit 19 card name @{outfilename}.pdb
  write coor unit 19 pdb

!======================================================================
==
  stop
!======================================================================
==
```

A.2 CHARMM scripts for different simulation models and data extraction

Essential input commands for implicit model, explicit model, umbrella sampling model, and data extraction are shown in these CHARMM scripts. To avoid redundancies and to only show the essential commands, parts of the scripts are removed with dotted lines to indicate the missing lines as follows:

```
...
...
```

These examples contain most of the key commands needed to reproduce the presented simulations.

CHARMM script for running EEF1 simulation

```
* This script is for a typical EEF1 implicit simulation
* Written by: Seung E. Lee
* Last modified: 1/17/06
*
!=====================================================================
! Parameter definition and reading in the structure information
!=====================================================================
  set str_dir /home/selee/stream
  set home_dir /home/selee/ss/charmm/vhta_bind/eef

! READ THE RTF AND PARAM FILES
  open read unit 11 card name -
       "/home/selee/stream/toppar/toph19_eef1_add.inp"
  read rtf unit 11 card

  open read unit 12 card name -
       "/home/selee/stream/toppar/param19_eef1_add.inp"
  read parameters unit 12 card

! DEFINE THE NUMBER OF TIME STEPS TO USE
  set i1 3
  set nst 80
  set nt  200000
  set meq 200000
  set timestep 0.002

! READING THE STRUCTURE FOR SIMULATION
  set psf_in @{home_dir}/build_h/pdb_b/vhta_i
  set crd_in @{home_dir}/build_h/pdb_b/vhta_i
  set crd_ino @{crd_in}
```

```
  if @i1 ne 1  set crd_in crd/out02
  set segname1 vinc
  set segname2 vbs1
  set outname out

  open read unit 20 card name @{psf_in}.psf
  read psf unit 20 card
  close unit 20
  open read unit 14 card name @{crd_in}.crd
  read coordinates card unit 14
  close unit 14

! DEFINING ATOM SELECTIONS TO BE USED IN DURING AND AFTER SIMULATION
  stream @{home_dir}/stream/def.str

! ROTATE THE VBS1 BY ANGLE @PHI ABOUT ITS AXIS
  set phi -14.7
  stream @{home_dir}/stream/vbs_axis.str

  coor stat sele atom * * ca .and. segid @segname2 end
  coor trans xdir ?xave ydir ?yave zdir ?zave fact -1.0 -
       sele segid @segname2 end
  coor rotate xdir @nx ydir @ny zdir @nz -
       phi @phi sele segid @segname2 end
  coor trans xdir ?xave ydir ?yave zdir ?zave fact  1.0 -
       sele segid @segname2 end

! ACTIVATION OF THE IMPLICIT SOLVENT
  fast 1
  shake bonh
  set solvfile /usr/common/charmm/c29b1_i_lam/test/data/solvpar.inp
  eef1 setup temp 300.0 unit 93 name @solvfile
  update ctonnb 7.0 ctofnb 9.0 cutnb 10.0 group rdie

!===================================================================
! RUN HEATING AND EQUILIBRATION
!===================================================================
! RUNNING SCRIPT TO KEEP ALPHA-HELIX INTACT DURING THE SIMULATION
  noe
     reset
  end
  set h1 606
  set h2 625
  set kval 0.5
  open read unit 14 card name @{crd_ino}.crd
  read coordinates card unit 14 comp
  stream @{home_dir}/stream/hel_noe.str

! RUNNING HEATING AND EQUI FOR BEGINNING SIMULATION
! SKIPPING THIS STEP IF CONTINUING SIMULATION FROM EXISTING RST FILE
  if @i1 ne 1  goto continue
     cons harm force 0.1 sele 8hel end
     set nheat 20000
     set nequi 280000
     set timestep 0.002
     stream @{str_dir}/heat_no_image.str
  if @i1 eq 1  set i1 3
```

```
  label continue

!===================================================================
! STARTING THE LOOP FOR THE PRODUCTION RUN
!===================================================================
  set 1 1
  calc i0 @i1 - 1
  label loop

  set nnt @nt
  if @{i1} eq 3 calc nnt @nt + @meq

  set ind0 @i0
  set ind1 @i1
  if @{i0} lt 10 set ind0 0@{i0}
  if @{i1} lt 10 set ind1 0@{i1}

! APPLYING BINDING CONSTRAINTS ONLY IN THE BEGINNING OF A SIMULATION
! THE CONSTRAINTS ARE NOT APPLIED AFTER 800PS
  if @{i1} ne 3 goto skip_const
     noe
        reset
     end
     set kval 0.5
     set rval 3.0
     stream @{home_dir}/stream/pho.str

     open read unit 14 card name @{crd_ino}.crd
     read coordinates card unit 14 comp
     stream @{home_dir}/stream/hel_noe.str
  label skip_const

! RUNNING THE PRODUCTION SIMULATION
  open read unit 30 card name rst/@{outname}@{ind0}.rst
  open writ unit 31 card name rst/@{outname}@{ind1}.rst
  open writ unit 32 file name dcd/@{outname}@{ind1}.dcd
  open writ unit 34 card name ene/@{outname}@{ind1}.ene

  dyna vverlet nose qref 50.0 tref 300.0 ncyc 5 -
     restart verlet nstep @{nnt} time @timestep -
     iprfrq 2000 ihtfrq 0  ieqfrq 0 inbfrq @frq ihbfrq 0 -
     iunrea 30 iunwri 31 iuncrd 32 iunvel -1 kunit 34 ntrfrq 2000 -
     nprint 2000 nsavc 2000 -
     firstt 300.0 finalt 300.0 -
     twindh 10.0 twindl -10.0 -
     iasors 0 iscvel 0 iasvel 1 ichecw 1

  open write unit 18 card name crd/@{outname}@{ind1}.crd
  write coor unit 18 card
  noe
     reset
  end

  incr  1 by 1
  incr i0 by 1
  incr i1 by 1
  if 1 lt @{nst}.5 goto loop
```

```
!=====================================================================
  stop
!=====================================================================
```

CHARMM script for running explicit simulation

```
  ...
  ...

! READ RTF AND PARAM FILES
  open read unit 11 card name "stream/toppar/top_all27_prot_na.rtf"
  read rtf unit 11 card
  close unit 11

  open read unit 12 card name "stream/toppar/par_all27_prot_na.prm"
  read parameters unit 12 card
  close unit 12

! DEFINITION OF NONBONDED INTERACTION PARAMETERS
  set frq -1
  set von  8
  set vof 10
  set vnb 12

! SETTING UP BOX SIZE
  set xsiz 54
  set ysiz 89.3192
  set zsiz 48.708
  crystal define orthorombic @XSIZ @YSIZ @ZSIZ 90.0 90.0 90.0
  crystal build cutoff 12.0

! SETTING UP PERIODIC BOUNDARY CONDITION
  image byres sele .not. prot end
  image byseg sele prot end

! ACTIVATION OF THE EWALD FOR THE LONG-RANGE NONBONDED INTERACTION
  fast 1
  shake bonh
  energy cdie eps 1.0 fswitch vfswitch inbfrq @frq imgfrq @frq -
         ihbfrq 0 cutnb @vnb ctofnb @vof ctonnb @von cutim @vnb -
         nbscale 1.4 ewald pmewald kappa 0.34 order 6 -
         fftx 64 ffty 128 fftz 64 qcor 0.0

  ...
  ...
```

CHARMM script for running umbrella sampling simulation

```
  ...
  ...
```

```
! DEFINING ATOMS TO BE FIXED IN PULLING SIMULATION
  define hold sele ( (resid 646 .and. type OE1) .or. -
                     (resid 650 .and. type OE1) .or. -
                     (resid 653 .and. type NE2) ) end
  cons harm force 10.0 sele hold end

! DEFINING ATOMS TO BE PULLED IN PULLING SIMULATION
  define ppul sele .not. back .and. .not. (type H .or. type O) .and. -
                                  ( (resid 498) .or. -
                                    (resid 501) .or. -
                                    (resid 502) ) end

  ...
  ...

! DEFINING REACTION COORDINATES FOR UMBRELLA SAMPLING SIMULATION
! REACTION COORDINATES DEFINED BY DISTANCE BETWEEN @HOLD AND @PPUL
  rxncor: define c1 point select hold end
  rxncor: define c2 point mass select ppul end
  rxncor: define d12 distance c1 c2
  rxncor: set d12

! EVALUATING THE STARTING DISTANCE BETWEEN @HOLD AND @PPUL
  coor stat sele hold end
  set x1 ?xave
  set y1 ?yave
  set z1 ?zave

  coor stat sele ppul end
  set x2 ?xave
  set y2 ?yave
  set z2 ?zave

  calc xx @x2 - @x1
  calc yy @y2 - @y1
  calc zz @z2 - @z1

  set wsiz 5.0
  calc rxnd (@xx*@xx + @yy*@yy + @zz*@zz) ** 0.5
  calc rxnl @rxnd - @wsiz
  calc rxnh @rxnd + @wsiz

  ...
  ...

!=====================================================================
! START THE LOOP
!=====================================================================
  set deld -0.1
  label loop

  set nnt @nt
  if @{i1} eq 3 calc nnt @nt + @meq

  set ind0 @i0
  set ind1 @i1
  if @{i0} lt 10 set ind0 0@{i0}
```

```
  if @{i1} lt 10 set ind1 0@{i1}

  calc delta (@i1-2-1) * @deld
  calc refd @rxnd + @delta
  calc refl @rxnl + @delta
  calc refh @rxnh + @delta

! SETTING UP THE UMBRELLA SAMPLING RUN
  rxncor: umbrella kumb 5.0 del0 @refd form 1
  rxncor: statistics lowdelta @refl hidelta @refh -
          deldel 0.1 start 20000

! RUNNING THE PRODUCTION SIMULATION
  open read unit 30 card name rst/@{outname}@{ind0}.rst
  open writ unit 31 card name rst/@{outname}@{ind1}.rst
  open writ unit 32 file name dcd/@{outname}@{ind1}.dcd
  open writ unit 34 card name ene/@{outname}@{ind1}.ene

  dyna vverlet nose qref 50.0 tref 300.0 ncyc 5 -
     restart verlet nstep @{nnt} time @timestep -
     iprfrq 2000 ihtfrq 0  ieqfrq 0 inbfrq @frq ihbfrq 0 -
     iunrea 30 iunwri 31 iuncrd 32 iunvel -1 kunit 34 ntrfrq 2000 -
     nprint 2000 nsavc 2000 -
     firstt 300.0 finalt 300.0 -
     twindh 10.0 twindl -10.0 -
     iasors 0 iscvel 0 iasvel 1 ichecw 1

! WRITING FREE-ENERGY VS. REACTION COORDINATE DISTANCE
  rxncor: write unit 21
  close unit 21

  incr  1 by 1
  incr i0 by 1
  incr i1 by 1
  if 1 lt @{nst}.5 goto loop

!=====================================================================
  stop
!=====================================================================
```

CHARMM script for extracting data

```
* Script to extract data from the finished simulation
* Written by: Seung E. Lee
* Last modified: 02/09/06
*
!-------------------------------------------------------------------
! PARAMETER DEFINITION AND READING IN THE STRUCTURE INFO
!-------------------------------------------------------------------
  stream /home/selee/stream/iofile19_eef_add.str
  set home_dir /home/selee/ss/charmm/vhta_bind/eef

! Definition of input parameters and segid parameters
  set inp_dir /home/selee/inp_bank/anal_bind
  set psf_in @{home_dir}/build_h/pdb_b/vhta_i
```

```
  set crd_in @{home_dir}/build_h/pdb_b/vhta_i
  set crd_ac @{home_dir}/build_h/pdb_b/vhta_a
  set segname1 vinc
  set segname2 vbs1
  set dcd_in dcd/out

  set if   3                ! Index for file number
  set nst  80               ! Number of files
  set nsk  10000            ! Number of skip steps

!------------------------------------------------------------------
! READING STRUCTURE FROM THE PDB FILE
!------------------------------------------------------------------
! Read the PSF file for orientated structure
  open read unit 20 card name @{psf_in}.psf
  read psf unit 20 card
  close unit 20

! Read the crd coordinates
  open read unit 14 card name @{crd_in}.crd
  read coordinates card unit 14
  close unit 14
  coor copy comp

  ...
  ...

!------------------------------------------------------------------
! OPEN MULTIPLE DCD OUTFILES TO READ
!------------------------------------------------------------------
  set 1    1                ! counter
  set fu   51               ! First unit
  set iu   @fu              ! Index for unit number

  label loop
     set ind0 @iu
     set ind1 @if
     if @{iu} lt 10 set ind0 0@{iu}
     if @(35) lt 10 set ind1 0@(35)
     open read unit @{ind0} unformat name @{dcd_in}@{ind1}.dcd
  incr  1 by 1
  incr iu by 1
  incr if by 1
  if 1 lt @{nst}.5 goto loop

!------------------------------------------------------------------
! CREATE MERGED DCD FILE FROM MULTIPLE DCD FILES
!------------------------------------------------------------------
  open writ unit 301 unformat name mdcd/o.dcd
  merge coor format firstu @fu nunit @nst skip @nsk outputu 301 -
        sele atom * * * end

  open writ unit 302 unformat name mdcd/m.dcd
  merge coor format firstu @fu nunit @nst skip @nsk outputu 302 -
        sele atom * * * end orie mass sele ca end

!------------------------------------------------------------------
```

```
! PERFORM THE FOLLOWING DATA EXTRACTION FOR EACH TIME FRAME
!
!    heli.str   - EXTRACT HELICITY INFO
!    rms.str    - EXTRACT RMS INFO
!    dist.str   - EXTRACT DIST BETWEEN SELECTED ATOM PAIRS W TIME
!    force.str  - EXTRACT FORCE APPLIED AT EACH TIMESTEP
!    cont.str   - EXTRACT CONTACT INFO BETWEEN SECONDARY STRUCTURES
!    rot.str    - EXTRACT DATA TO EVALUATE VBS1 ANGLE OF ROTATION
!    coor.str   - EXTRACT COORDINATE INFO
!-------------------------------------------------------------------
  trajectory firstu @fu nunit @nst skip @nsk

  set 2 1
  set ntot 99999

  open writ unit 351 format name out/helicity.dat
  open writ unit 401 format name out/rms_hel.dat
  open writ unit 451 format name out/nonbf.dat
  stream @{inp_dir}/stream/copen.str
  open writ unit 701 format name out/4h.dat
  open writ unit 702 format name out/vbs.dat
  open writ unit 801 format name out/coor.dat
  open writ unit 901 format name out/dist.dat

  label loop2
     traj read comp
     coor copy

     stream @{inp_dir}/stream/heli.str
     stream @{inp_dir}/stream/rms.str
     stream @{inp_dir}/stream/dist.str
     stream @{inp_dir}/stream/force.str
     stream @{inp_dir}/stream/cont.str
     stream @{inp_dir}/stream/rot.str
     stream @{inp_dir}/stream/coor.str
  incr 2 by 1
  if 2 lt @ntot.5 goto loop2

!-------------------------------------------------------------------
  stop
!-------------------------------------------------------------------
```

A.3 Sub-scripts used by the CHARMM input scripts

CHARMM commands that are used repeatedly can be saved in a separate file to be used as sub-scripts. These files have extension of .str and can be called from a CHARMM script by '`stream test.str.`'

Sub-script 'def.str'

```
* DEFINE ATOM SELECTIONS
*
! DEFINITION OF SOME USEFUL ATOM SELECTIONS FOR LATER USE
  define prot sele segid @segname1 .or. segid @segname2 end
  define back sele prot .and. (type N .or. type CA .or. type C) end
  define ca sele prot .and. type CA end

  define hel1 sele prot .and. resid   6:29  end
  define hel2 sele prot .and. resid  39:65  end
  define hel3 sele prot .and. resid  67:97  end
  define hel4 sele prot .and. resid 101:149 end
  define hel4a sele prot .and. resid 101:128 end
  define hel5 sele prot .and. resid 159:180 end
  define hel6 sele prot .and. resid 184:218 end
  define hel7 sele prot .and. resid 222:251 end
  define vbs  sele prot .and. segid @segname2 end
  define 3hel sele hel1 .or. hel2 .or. vbs end
  define 4hel sele hel1 .or. hel2 .or. hel3 .or. hel4 end
  define 4hela sele hel1 .or. hel2 .or. hel3 .or. hel4a end
  define 7hel sele hel1 .or. hel2 .or. hel3 .or. hel4 .or. hel5 -
                      .or. hel6 .or. hel7 end
  define 8hel sele hel1 .or. hel2 .or. hel3 .or. hel4 .or. hel5 -
                      .or. hel6 .or. hel7 .or. vbs end
  define hphobic sele prot .and. resname ala .or. resname leu .or. -
                               resname val .or. resname ile .or. -
                               resname pro .or. resname phe .or. -
                               resname met .or. resname trp end

  return
```

Sub-script 'heat.str'

```
* This script contains the heating and equilibrating commands for
implicit.
* Need to specify:
* set nheat 20000  - number of timesteps for heating
* set nequi 40000  - number of timesteps for equilibrating
* set timestep 0.1 - time step size in ps
```

```
* set outname blah - name of outfile
*

! HEATING THE STRUCTURE
  open write unit 31 card name rst/@{outname}01.rst
  open write unit 32 file name dcd/@{outname}01.dcd
  open write unit 34 card name ene/@{outname}01.ene

  dyna strt verlet nstep @nheat time @timestep -
     iprfrq 1000 ihtfrq 50 ieqfrq 0 inbfrq @frq ihbfrq 0 -
     iunrea -1 iunwri 31 iuncrd 32 iunvel -1 kunit 34 ntrfrq 1000 -
     nprint 500 nsavc 1000 -
     firstt 0.0 finalt 300.0 teminc 3.0 -
     twindh 10.0 twindl -10.0 -
     iasors 1 iasvel 1 ichecw 0

  open write unit 18 card name crd/@{outname}01.crd
  write coor unit 18 card

! EQUILIBRATING THE STRUCTURE
  open read  unit 30 card name rst/@{outname}01.rst
  open write unit 31 card name rst/@{outname}02.rst
  open write unit 32 file name dcd/@{outname}02.dcd
  open write unit 34 card name ene/@{outname}02.ene

  dyna restart verlet nstep @nequi time @timestep -
     iprfrq 1000 ihtfrq 0 ieqfrq 100 inbfrq @frq ihbfrq 0 -
     iunrea 30 iunwri 31 iuncrd 32 iunvel -1 kunit 34 ntrfrq 1000 -
     nprint 500 nsavc 1000 -
     firstt 300.0 finalt 300.0 -
     twindh 10.0 twindl -10.0 -
     iasors 0 iscvel 0 ichecw 1

  open write unit 18 card name crd/@{outname}02.crd
  write coor unit 18 card

  return
```

Sub-script ‘pho.str’

```
* DISTANCE CONSTRAINTS APPLIED IN THE BEGINNING OF BINDING SIMULATION
* ONLY APPLIED FOR 800PS IN THE BEGINNING AND REMOVED AFTERWARD
*
  noe
    assign  sele segid @segname1 .and. resid 126 .and. type CE1 end -
            sele segid @segname2 .and. resid 608 .and. type CD1 end -
            kmax @kval rmax @rval
    assign  sele segid @segname1 .and. resid  16 .and. type CG2 end -
            sele segid @segname2 .and. resid 615 .and. type CD2 end -
            kmax @kval rmax @rval
    assign  sele segid @segname1 .and. resid 116 .and. type CD2 end -
            sele segid @segname2 .and. resid 615 .and. type CD2 end -
            kmax @kval rmax @rval
  end
```

```
  return
```

Sub-script ‘hel_noe.str’

```
* APPLYING DISTANCE CONSTRAINTS TO ALPHA-HELIX H-BONDING PAIRS
* TO RETAIN THE HELICAL CONFIGURATION
*
  set  r1 @h1
  calc r2 @h1 + 4

! START THE LOOP
  label loop

  coor stat sele resid @r1 .and. type O  end
  set x1 ?xave
  set y1 ?yave
  set z1 ?zave

  coor stat sele resid @r2 .and. type N  end
  set x2 ?xave
  set y2 ?yave
  set z2 ?zave

  calc xx @x2 - @x1
  calc yy @y2 - @y1
  calc zz @z2 - @z1

  calc dd ((@xx*@xx + @yy*@yy + @zz*@zz) ** 0.5)

  noe
    assign  sele segid @segname2 .and. resid  @r1 .and. type O end -
            sele segid @segname2 .and. resid  @r2 .and. type N end -
            kmax @kval rmax @dd
  end

  incr r1 by 1
  incr r2 by 1
  if r2 le @h2 goto loop

  return
```

Sub-script ‘vbs_axis.str’

```
* DETERMINING THE VECTOR REPRESENTING THE AXIS OF A HELIX
*
  coor stat sele vbs .and. resid 608:616 .and. ca end
  set x1 ?xave
  set y1 ?yave
  set z1 ?zave

  coor stat sele vbs .and. resid 617:625 .and. ca end
  set x2 ?xave
  set y2 ?yave
  set z2 ?zave
```

```
calc xx @x2 - @x1
calc yy @y2 - @y1
calc zz @z2 - @z1

calc norm (@xx*@xx + @yy*@yy + @zz*@zz) ** 0.5
calc nx @xx / @norm
calc ny @yy / @norm
calc nz @zz / @norm

return
```

A.4 Extracted data files

Although CHARMM has means of analyzing the results, it has limitations in doing detailed analyses. Therefore, CHARMM scripts are used to extract the coordinates and other data, and the extracted data are read into Matlab to do further analysis. The extracted data files from CHARMM have their own data format, which must be filtered through in order to get the needed values for Matlab. This section shows examples of data files created by CHARMM, which contains the coordinates at different time frame of the molecular trajectory and distance time trace between specified atom pairs, and one example of analyzed data file obtained from Matlab. The file obtained from Matlab reads in the distance information created by CHARMM and reports the atom pairs that stay close to each other throughout the simulation. This file is useful in understanding the critical stabilizing atom pairs at a glance.

Extracted CHARMM file with atom coordinate information at each time step

```
         COORDINATE FILE MODULE
TITLE>  * 832.784 1034.21 168.63 327.19 -4.62 -5.37
TITLE>  *
   24
2436  255 GLY  CA    18.20063   4.18328  21.68251 VBS1 605    56.22499
2441  256 ARG  CA    16.95080   7.49486  22.89267 VBS1 606     2.21000
2458  257 PRO  CA    16.57340   8.90384  19.31336 VBS1 607     6.84091
2465  258 LEU  CA    13.34876   6.78124  19.34623 VBS1 608     5.08552
2474  259 LEU  CA    11.87169   8.27222  22.36288 VBS1 609     0.12306
2483  260 GLN  CA    12.77118  12.08687  22.16191 VBS1 610     0.74575
2495  261 ALA  CA    10.66101  12.16537  18.96935 VBS1 611    17.67962
2501  262 ALA  CA     7.60023  10.83480  20.71139 VBS1 612     7.88216
2507  263 LYS  CA     8.12198  13.25786  23.56445 VBS1 613     5.91553
2520  264 GLY  CA     8.66815  15.96971  20.82964 VBS1 614    32.94937
2525  265 LEU  CA     5.26983  15.23309  19.27809 VBS1 615     3.90814
2534  266 ALA  CA     3.45018  15.26344  22.58485 VBS1 616    10.25164
2540  267 GLY  CA     5.00444  18.69353  23.28445 VBS1 617    38.12208
2545  268 ALA  CA     4.44255  20.42629  19.94653 VBS1 618     1.48692
2551  269 VAL  CA     0.84319  18.92647  19.47422 VBS1 619     1.03366
2559  270 SER  CA    -0.17404  20.30557  22.93382 VBS1 620     1.57655
2567  271 GLU  CA     1.27838  23.70001  22.06115 VBS1 621     0.00000
```

```
2577   272 LEU  CA     -0.81675  23.52929  18.91935 VBS1 622     5.11069
2586   273 LEU  CA     -4.07070  22.46802  20.60320 VBS1 623     2.20533
2595   274 ARG  CA     -3.56810  25.36893  23.26451 VBS1 624    11.80998
2612   275 SER  CA     -2.42234  28.04008  20.89989 VBS1 625    12.01371
2620   276 ALA  CA     -3.22358  27.56927  17.13227 VBS1 626    21.04213
2626   277 GLN  CA     -6.80280  27.05577  17.54744 VBS1 627     0.31163
2638   278 PRO  CA     -9.36488  27.47264  14.79384 VBS1 628    17.24632

         COORDINATE FILE MODULE
TITLE>  * 827.554 1038.89 167.12 297.57 -1.5 -5.32996
TITLE>  *
  24
2436   255 GLY  CA     16.46142   2.12652  24.02084 VBS1 605    61.03071
2441   256 ARG  CA     16.12251   5.81334  23.82071 VBS1 606     1.61720
2458   257 PRO  CA     15.93096   6.63055  20.02704 VBS1 607     2.26813
2465   258 LEU  CA     12.37974   4.95546  20.00913 VBS1 608    11.24105
2474   259 LEU  CA     11.39470   7.26817  22.81620 VBS1 609     0.00000
2483   260 GLN  CA     13.05098  10.62100  21.85820 VBS1 610     0.42362

...
...
```

Extracted CHARMM file with distance time trace information

```
           DISTANCES FOR SELECTED ATOMS
2447 VBS1 ARG  606  CZ  -1638 VINC ASP  176  CG          3.9960
2459 VBS1 PRO  607  CB  - 517 VINC VAL  57   CG2         3.9137
2467 VBS1 LEU  608  CG  -1165 VINC PHE  126  CE1         3.9153
2468 VBS1 LEU  608  CD1 -1165 VINC PHE  126  CE1         3.6100
2469 VBS1 LEU  608  CD2 -1136 VINC LEU  123  CD1         3.9635
2476 VBS1 LEU  609  CG  -   84 VINC THR  8    CG2         3.8523
2496 VBS1 ALA  611  CB  - 484 VINC LEU  54   CD2         3.8424
2501 VBS1 ALA  612  CA  - 120 VINC ILE  12   CD          3.9152
2502 VBS1 ALA  612  CB  - 120 VINC ILE  12   CD          3.6267
2528 VBS1 LEU  615  CD1 - 118 VINC ILE  12   CG2         3.8037
2528 VBS1 LEU  615  CD1 - 120 VINC ILE  12   CD          3.5903
2529 VBS1 LEU  615  CD2 - 145 VINC PRO  15   CB          3.9281
2546 VBS1 ALA  618  CB  - 416 VINC ALA  46   CB          3.7647
2546 VBS1 ALA  618  CB  - 424 VINC VAL  47   CG2         3.9471
2554 VBS1 VAL  619  CG2 - 145 VINC PRO  15   CB          3.8489
2554 VBS1 VAL  619  CG2 - 147 VINC PRO  15   C           3.9936
2568 VBS1 GLU  621  CB  - 394 VINC PRO  43   CG          3.7878
2579 VBS1 LEU  622  CG  - 393 VINC PRO  43   CB          3.9192
2590 VBS1 LEU  623  CD2 - 207 VINC HSC  22   CB          3.6371
2590 VBS1 LEU  623  CD2 - 208 VINC HSC  22   CG          3.9117
2591 VBS1 LEU  623  C   - 212 VINC HSC  22   CE1         3.9409
2621 VBS1 ALA  626  CB  - 249 VINC MET  26   CE          3.7607
2622 VBS1 ALA  626  C   - 249 VINC MET  26   CE          3.7864
2627 VBS1 GLN  627  CB  - 212 VINC HSC  22   CE1         3.6551
2639 VBS1 PRO  628  CB  - 325 VINC LYS  35   CB          3.9312
2639 VBS1 PRO  628  CB  - 327 VINC LYS  35   CD          3.9490
TOTAL EXCLUSION COUNT =         0
TOTAL 1-4 EXCLUSIONS  =         0
TOTAL NON-EXCLUSIONS  =        26
```

```
              DISTANCES FOR SELECTED ATOMS
 2436 VBS1 GLY  605  CA  -1646 VINC GLU  177  CB        3.7880
 2468 VBS1 LEU  608  CD1 -1165 VINC PHE  126  CE1       3.8804
 2469 VBS1 LEU  608  CD2 - 483 VINC LEU  54   CD1       3.9802
 2469 VBS1 LEU  608  CD2 - 484 VINC LEU  54   CD2       3.6499
 2528 VBS1 LEU  615  CD1 - 120 VINC ILE  12   CD        3.5083
 2529 VBS1 LEU  615  CD2 - 448 VINC ALA  50   CB        3.7184

 ...
 ...
```

Matlab analyzed file determined by reading in the CHARMM distance file

```
  Pair          Atom1             Atom2       Occurance
-----------------------------------------------------------------------

    1     606 ARG   CZ      177   GLU  CG       39 /  200   19 %   H9-H5
    2     606 ARG   CZ      180   GLN  CD       64 /  200   32 %   H9-H5

    3     607 PRO   CB       57   VAL CG2       49 /  200   24 %   H9-H2

    4     608 LEU   CB        8   THR CG2       35 /  200   17 %   H9-H1
    5     608 LEU  CD2       54   LEU CD1       40 /  200   20 %   H9-H2
    6     608 LEU  CD2       54   LEU CD2       55 /  200   27 %   H9-H2
    7     608 LEU  CD1      126   PHE CE1       52 /  200   26 %   H9-H4
    8     608 LEU  CD1      126   PHE CE2       32 /  200   16 %   H9-H4

    9     609 LEU  CD2        8   THR CG2       42 /  200   21 %   H9-H1

   10     612 ALA   CB       12   ILE CG1       33 /  200   16 %   H9-H1

   11     615 LEU  CD1       16   VAL CG2       47 /  200   23 %   H9-H1
   12     615 LEU  CD2       16   VAL CG2       43 /  200   21 %   H9-H1
   13     615 LEU  CD2       50   ALA  CB       37 /  200   18 %   H9-H2

   14     618 ALA   CB       43   PRO  CA       54 /  200   27 %   H9-H2
   15     618 ALA   CB       46   ALA  CB      105 /  200   52 %   H9-H2
   16     618 ALA   CB       47   VAL CG2       37 /  200   18 %   H9-H2

   17     619 VAL  CG2       15   PRO  CB       78 /  200   39 %   H9-H1

   18     621 GLU   CB       43   PRO  CG       72 /  200   36 %   H9-H2

   19     622 LEU  CD2       19   GLN  CG       36 /  200   18 %   H9-H1
   20     622 LEU  CD2       23   LEU CD2       33 /  200   16 %   H9-H1
   21     622 LEU  CD2       43   PRO  CB       36 /  200   18 %   H9-H2

   22     623 LEU  CD1       18   GLN  CB       38 /  200   19 %   H9-H1
   23     623 LEU  CD2       18   GLN  CB       52 /  200   26 %   H9-H1
   24     623 LEU    C       22   HSC CE1       52 /  200   26 %   H9-H1
   25     623 LEU   CA       22   HSC CE1       32 /  200   16 %   H9-H1
   26     623 LEU  CD1       22   HSC  CB       82 /  200   41 %   H9-H1

   27     626 ALA   CB       40   LEU CD1       35 /  200   17 %   H0-H2
```

```
28      627 GLN   CD       22   HSC CE1       89 /  200   44 %   H0-H1

29      628 PRO   CD       26   MET  CG       37 /  200   18 %   H0-H1
30      628 PRO   CG       26   MET  CG       49 /  200   24 %   H0-H1
```

A.5 Scripts for data analysis

Data files extracted from CHARMM have their own data format, and must be filtered into different format that is readable for Matlab scripts. The most convenient way to do this is to write a shell script that makes such modifications. The filtered data files are then read into Matlab for plotting and statistical analysis. This section shows a few examples of the shell scripts and Matlab analysis scripts.

Shell script to modify the CHARMM out-files into Matlab readable files

```
#!/bin/tcsh -f

  grep CA vbs.dat > t.d
  awk '{print $5, $6, $7}' t.d > vbs.d

  grep VINC  4h.dat > t.d
  awk '{print $5, $6, $7}' t.d >  4h.d

  grep VBS1  helicity.dat > t2.d
  awk '{print $3}' t2.d >  hel_res.d
  awk '{print $5, $6, $7}' t2.d >  hel_coor.d

  foreach n ( 1 2 3 4 5 )
     sed 's/-/ /g' cont/hb$n.dat > cont/hb$n.d

     sed 's/-/ /g' cont/ph$n.dat > cont/ph$n.d
  end

  rm -f t.d t2.d

  exit 0
```

Matlab script that reads in CHARMM distance file and creates the close-contact information

```
% Program: onlyhb.m
% Written by: Seung E. Lee
% Last updated: 9/16/2005
%
% Read in the output from CHARMM: e.g. line
%1621 TALI GLU  948  CG    512 VH_N VAL  57   CG2       3.5553
% And sort the data and save it to var.mat to be used by onlyhb2.m
%
```

```
  function sub_cont(work_dir,filename,outfile,nn,per2);

% READ IN THE DATEFILE AND ORGANIZE
  fid = fopen([work_dir filename],'r');
  nset = nn;                  % Number of trajectories
  fac = 1000;                 % Just a factor for future use

  fid2 = fopen([work_dir outfile],'w');
  S1 = fscanf(fid,'%s',1);    % Skipping 2 lines
  S2 = fscanf(fid,'%s',3);

  for i = 1:nset,       % Loop over all the traj
    nn = 0;
    S1 = fscanf(fid,'%s',1);
    while S1(1,1) ~= 'T',     % Loop thru each interaction of traj
      nn = nn + 1;
      S2 = fscanf(fid,'%s',1);
      S5 = fscanf(fid,'%s',1);      % 1st resname
      S6 = fscanf(fid,'%g',1);      % 1st resid
      S3 = fscanf(fid,'%s',1);      % 1st res type
      S9 = fscanf(fid,'%s',2);
      S7 = fscanf(fid,'%s',1);      % 2nd resname
      S8 = fscanf(fid,'%g',1);      % 2nd resid
      S4 = fscanf(fid,'%s',1);      % 2nd res type
      S9 = fscanf(fid,'%s',1);
      S1 = fscanf(fid,'%s',1);      % First variable of next line
      ires(S6,1:3) = S5;      % resid associated w/ resname
      ires(S8,1:3) = S7;
      if length(S3) == 1, S3 = ['  ' S3]; end
      if length(S3) == 2, S3 = [' '  S3]; end
      if length(S4) == 1, S4 = ['  '  S4]; end
      if length(S4) == 2, S4 = [' '   S4]; end
      CN(nn,1) = S6*fac + S8; % resid's combined for easy sort
      atype(nn,1:6) = [S3 S4];      % restype's combined for easy
manage
    end
    N(i,1) = nn;        % Numb of interactions for each traj
    if nn ~= 0,
       CT(1:nn,i) = CN;       % Array with all resid stored
       atypet(1:nn,1:6,i) = atype;  % Array with all restype stored
    end
    clear CN
    clear atype

    L = fgets(fid); L = fgets(fid); L = fgets(fid);
    L = fgets(fid); L = fgets(fid); % Skipping some lines
  end

  E = CT(1:N(1),1);
  F = atypet(1:N(1),1:6,1);
  for i = 2:length(N),
    E = [E; CT(1:N(i),i)];
    F = [F; atypet(1:N(i),1:6,i)];
  end                   % All resid and restype in one col
  nE = size(E,1);
  for i = 1:nE,
     rT = int2str(E(i,1));
```

```
     if E(i,1) < 100000, rT = ['0' int2str(E(i,1))]; end
     if E(i,1) <  10000, rT = ['00' int2str(E(i,1))]; end
     T(i,1:12) = [rT F(i,:)];  % Both resid and restype in one col
  end

% Saving to var.mat:
% N     = number of interactions per traj
% CT    = Array w/ all resid stored
% atype = Array w/ all restype stored
% E     = All resid in one column
% F     = All restype in one column
% ires  = indexed interaction of resid and resname

% save var N CT atypet E F T ires

%=uuu===========================================================
  [TS,ind] = sortrows(T);     % Sorting based on resids and restypes
  ES = E(ind,:);
  FS = F(ind,:);

% Condensing the sorted TS and ES to remove all the overlaps
% They are stored in TSF and ESF and the number of each occurance is
% stored in M.
  M = zeros(length(ES),1);
  n3 = 1;
  TSF(n3,1:12) = TS(1,1:12);
  ESF(n3,1) = ES(1,1);
  M(n3,1) = M(n3,1) + 1;
  for ii = 2:length(ES),
    if TS(ii,1) ~= TSF(n3,1) | TS(ii,2) ~= TSF(n3,2) | TS(ii,3) ~=
TSF(n3,3) | TS(ii,4) ~= TSF(n3,4) | TS(ii,5) ~= TSF(n3,5) | TS(ii,6) ~=
TSF(n3,6) | TS(ii,7) ~= TSF(n3,7) | TS(ii,8) ~= TSF(n3,8) | TS(ii,9) ~=
TSF(n3,9) | TS(ii,10) ~= TSF(n3,10) | TS(ii,11) ~= TSF(n3,11) |
TS(ii,12) ~= TSF(n3,12),
      n1 = floor(ES(ii,1)/fac);
      n2 = ES(ii,1) - n1*fac;
      n3 = n3 + 1;
      TSF(n3,1:12) = TS(ii,1:12);
      ESF(n3,1) = ES(ii,1);
    end
    M(n3,1) = M(n3,1) + 1;
  end
  M = M(1:n3,1);

% Store everything as integer in 'fin1': [res1 res2 num_occur]
  D1 = floor(ESF(:,1)/fac);
  D2 = ESF(:,1) - D1*fac;
  fin1(1:n3,1:3)  = [D1(1:n3) D2(1:n3) M(1:n3,1)];

% Write out the statistics of the occurances to a file.
  nn = length(N);

sub_listhb(fin1(:,1),fin1(:,2),fin1(:,3),TSF(:,7:12),ires,nn,per2,fid2);
  fclose(fid2);

  return
```

Matlab script that reads in CHARMM coordinate file and evaluates various distance and angular information

```
% SCRIPT TO GET VBS1 ROTATION AND DISTANCES BETWEEN HELICES, THEN
% ANALYZE THE RESULT TO PLOT ON SCREEN
%

% SETUP THE DRAWING TOOLS FOR THE OUTPUT
  lw2 = lw + 4;
  p = [  1  24;
        25  51;
        52  82;
        83 131];

  th = 0 : pi/20 : 2*pi;
  rad = 3;
  cirx = rad * cos(th);
  ciry = rad * sin(th);

  thd = 180;
  thr = thd * pi / 180;
  Rx  = [        1         0         0;
                 0  cos(thr)  sin(thr);
                 0 -sin(thr)  cos(thr)];
  Ry  = [ cos(thr)         0 -sin(thr);
                 0         1         0;
          sin(thr)         0  cos(thr)];
  Rz  = [ cos(thr)  sin(thr)         0;
         -sin(thr)  cos(thr)         0;
                 0         0         1];

% READ IN THE COORDINATE OUTPUT DATA FROM CHARMM
  A = load([work_dir '4h.d']);
  nA = 131;
  B = load([work_dir 'vbs.d']);
  nB = 24;
  A = Rx * A'; A = A';
  B = Rx * B'; B = B';

  ct = [ 14  16;
         28  30;
         73  75;
         90  92;
         18  20];
  if ihel == 2,
     nB = 24;
     ct(5,1:2) = [ 18 20];
  elseif ihel == 3,
     nB = 26;
     ct(5,1:2) = [ 20 22];
  elseif ihel == 4,
     nB = 24;
     ct(5,1:2) = [  5  7];
  end
```

```
% EVALUATE ANGLE AND DISTANCES BETWEEN SECONDARY STRUCTURES
  ct = ct';
  sk = 1;
  k = 0;
  for i = 1:sk:nn,
     m1 = 1 + (i-1)*nA;
     m2 = m1 + nA - 1;
     j1 = 1 + (i-1)*nB;
     j2 = j1 + nB - 1;
     tA0 = A(m1:m2,:);
     tB0 = B(j1:j2,:);
     tA = [tA0(:,1) tA0(:,3)];
     tB = [tB0(:,1) tB0(:,3)];
     pt(1,1:2) = mean(tA(ct(:,1),:),1);
     pt(2,1:2) = mean(tA(ct(:,2),:),1);
     pt(3,1:2) = mean(tA(ct(:,3),:),1);
     pt(4,1:2) = mean(tA(ct(:,4),:),1);
     pt(5,1:2) = mean(tB(ct(:,5),:),1);
     p22(1,1:2) = tB(ct(1,5),:);
     p23(1,1:2) = tB(ct(1,5)+1,:);

     p22 = (p22 - pt(5,:)) / norm(p22-pt(5,:));
     p22l = [p22*rad; p22*rad*1.8];
     p23 = (p23 - pt(5,:)) / norm(p23-pt(5,:));
     p23l = [p23*rad; p23*rad*1.8];

     k = k + 1;
     r1(1,1:2) = (pt(1,:)-pt(5,:)) / norm(pt(1,:)-pt(5,:));
     r2(1,1:2) = (pt(2,:)-pt(5,:)) / norm(pt(2,:)-pt(5,:));
     r3(1,1:2) = p23;
     r1(1,1:3) = [r1(1) 0 r1(2)];
     r2(1,1:3) = [r2(1) 0 r2(2)];
     r3(1,1:3) = [r3(1) 0 r3(2)];
     ang23(k) = acos( dot(r1,r3) ) * 180 / pi;
     cv = cross(r1,r3);
     if cv(2) < 0, ang23(k) = - ang23(k); end

     ang3(k) = acos( dot(r2,r1) ) * 180 / pi;
     cv = cross(r2,r1);
     if cv(2) > 0, ang3(k) = - ang3(k); end

     dist12(k) = norm(pt(1,:)-pt(2,:));
     dist54(k) = norm(pt(5,:)-pt(4,:));
  end

% PLOT THE EVALUATED VALUES
  zz = zeros(size(time));
  ifig = ifig + 1; figure(ifig);
  subplot(2,2,1); hold off; plot(time,ang23,'k');
  subplot(2,2,1); hold on ; plot(time,zz,'k--');
  xlabel('Time (ns)'); ylabel('Angle (degrees)');
  title([kres ' angle with VBS1-H1']);
  set(gca,'fontsize',fs);
  axis([-inf inf -inf inf]);

  subplot(2,2,2); plot(time,ang3,'k');
```

```
  xlabel('Time (ns)'); ylabel('Angle (degrees)');
  title(['Angle of H1-VBS1-H2']);
  set(gca,'fontsize',fs);
  axis([-inf inf -inf inf]);

  subplot(2,2,3); plot(time,dist12,'k');
  xlabel('Time (ns)'); ylabel('Dist (Ang)');
  title(['Distance H1-H2']);
  set(gca,'fontsize',fs);
  axis([-inf inf -inf inf]);

  subplot(2,2,4); plot(time,dist54,'k');
  xlabel('Time (ns)'); ylabel('Dist (Ang)');
  title(['Distance VBS1-H4']);
  set(gca,'fontsize',fs);
  axis([-inf inf -inf inf]);

% SAVE THE EVALUATION TO A FILE FOR LATER USE
  fidnag = fopen('o-ang.out','w');
  for i = 1:length(ang23),
     fprintf(fidnag,' %8.4f  %8.4f  %8.4f  %8.4f
%8.4f\n',time(i),ang23(i),ang3(i),dist12(i),dist54(i));
  end
```