Comparison of Temperature Coupling Algorithms in SMD simulations

and Mutation Design in Talin I/LWEQ Domain

M.Sc. Thesis
Magdaléna von Essen
Institute of Biomedical Technology
University of Tampere
July, 2013

MASTER'S THESIS

Place:	UNIVERSITY OF TAMPERE Institute of Biomedical Technology Protein Dynamics Group
Author:	VON ESSEN, MAGDALÉNA
Title:	Comparison of Temperature Coupling Algorithms in SMD Simulations and Mutation Design in Talin I/LWEQ Domain
Pages:	91 and 4 Appendix pages
Supervisors:	Docent Vesa Hytönen, PhD, M.Sc. Sampo Kukkurainen, M.Sc. Rolle Rahikainen
Reviewers:	Professor Matti Nykter, D.Sc, Docent Vesa Hytönen, PhD
Date:	July, 2013

Abstract

Recently, a significant number of studies regarding the activation mechanism of talin's vinculin binding site (VBS) have been published. Since the hydrophobic VBS is in natural talin buried inside the helix structure, a major conformational change prior to activation is necessary. Talin and vinculin are proteins of the focal adhesion complex which is connecting the extracellular matrix structures to the intracellular cytoskeleton. Based on these facts, it is suspected that the talin activation is force-induced. Current development of computational methods, molecular dynamics (MD) and steered molecular dynamics (SMD), allows single-molecule observations on atomic level. Such possibility widens our understanding of activation mechanisms and force-affected protein dynamics unavailable in classical experimental setting. New methods, however, bring also new challenges. For instance, the assessment of MD/SMD parameter setting and the influence on the observed protein dynamics is one of such challenges. While the parameter setting of classical MD is frequently discussed in the literature, there is no recommendation for SMD setting (thermostat, barostat, ensemble etc.) available.

In this work, two commonly used temperature coupling algorithms, Berendsen and Nosé-Hoover, and their implementation (group coupling) in the simulations were investigated. Furthermore, the influence of force magnitude on the unfolding dynamics was studied. Based on the observed unfolding trajectories, four destabilizing mutations in the talin I/LWEQ domain were proposed.

The results of the herein presented study suggest, that the choice of thermostat has insignificant impact on the unfolding dynamics. However, the implementation of the thermostat (Nosé-Hoover) influences the unfolding. The temperature control should be implemented for all parts of the simulated system, preferably in separate groups coupled to own thermostat. While unfolding with 200pN constant force did not proceed after 4ns of the simulation, gradual unfolding of the protein structure was observed with 300pN, 400pN and 500pN. Furthermore, the unfolding with 500pN exhibited C-terminal oriented unfolding, leaving H2 and VBS completely revealed yet intact between 5 - 6ns. Forces 300pN and 400pN showed comparable protein stretching from both N- and C-terminal end, suggesting fluent force penetration through the structure compared to the higher force.

Based on the results of ionic, hydrogen and hydrophobic interaction analysis of 200pN and 300pN unfolding trajectories, four bonds were identified as resisting the forced unfolding. Residues participating in these bonds were designed mutations with structure destabilizing prospect; I2316S, E2317A, D2386A and R2398A.

Keywords: mechanotransduction, mechanoactivation, mechanostability, focal adhesion, talin, protein unfolding, steered molecular dynamics, Berendsen thermostat, Nosé-Hoover thermostat.

FOREWORD

I would like to express my gratitude to Docent Vesa Hytönen, PhD, M.Sc. Sampo Kukkurainen and M.Sc. Rolle Rahikainen for their support and encouragement I received throughout the whole process. Their knowledge, experience and advice have indeed helped me to accomplish the work.

This work has been carried out in the Protein Dynamics Group, Institute of Biomedical Technology at the University of Tampere. It has been an inspiring environment and I thank to the whole group for pleasant working atmosphere.

Finally I would like to thank my family, Vaňkovi and von Essen, for supporting me during the work. At last but no least, I would like to thank Johana Kuncová-Kallio for her ever-present humor and motivation and to my loving, tolerant husband Mathias von Essen for his bottomless optimism and support.

Tampere, July 2013

Magdaléna von Essen

Table of Contents

Illustration Index

List of Abbreviations

AA	Amino Acid
ABS	Actin Binding Site
AFM	Atomic Force Microscopy
CA	Alpha Carbon
dbSNP	Single Nucleotide Polymorphism database
ECM	Extracellular Matrix
FA	Focal Adhesion
FAK	Focal Adhesion Kinase
gmx	Online GROMACS Discussion Forum
GPU	Graphical Processing Unit
H	Talin Helix
HIV	Human Immunodeficiency Virus
I/LWEQ	Conserved Sequence Module for F-actin
IAC	Integrin Adhesion Complex
IBS	Integrin Binding Site
ICM	Intracellular Matrix
Ig	Immunoglobulin
K	Kelvin
KCl	Potassium Chloride
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAYN	Layilin
MD	Molecular Dynamics
MTTK	Martyna-Tuckerman-Tobias-Klein
N	Number of Particles
NA	Nucleic Acid
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
NRAP	Nebulin
OMIM	Online Mendelian Inheritance in Man
P	Pressure
PDB id	Protein Data Bank Identification
PIP	Phosphatidylinositol 4,5-bisphosphate
SMD	Steered Molecular Dynamics
SNP	Single Nucleotide Polymorphism
T	Temperature
TAL5	Talin's Fifth Bundle
TC	Turn/Coil Protein Secondary Structure
V	Volume
VBS	Vinculin Binding Site
VH	Vinculin Helix
VMD	Visual Molecular Dynamics

Amino Acid code

Ala	A	Alanine
Cys	C	Cysteine
Asp	D	Aspartic acid
Glu	E	Glutamic acid
Phe	F	Phenylalanine
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Lys	K	Lysine
Leu	L	Leucine
Met	M	Methionine
Asn	N	Asparagine
Pro	P	Proline
Gln	Q	Glutamine
Arg	R	Arginine
Ser	S	Serine
Thr	T	Threonine
Val	V	Valine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
	U	Selenocysteine
	O	Pyrrolysine
	X	Unspecified or unknown

1 Introduction

The ability of cells to receive signals from outside environment and respond to the stimuli is fundamental to life. In fact, cells receive constant input from the membrane proteins acting as information receptors, and convert the input into a chemical process and a cellular response. Such conversion of the information input into a chemical change, i.e. signal transduction is a universal property of living cells (Geiger et al., 2009, Nelson and Cox, 2008).

It has been long known that cells can react to biochemical stimuli but also to mechanical cues. Since tissue-level biology, the signal transduction mechanisms have been studied in connection to the transmission of nerve signals, response to hormones and growth factors, sense for light, smell or taste, and control of the cell cycle (Nelson and Cox, 2008). Despite intensive research in the field of signal transduction, the mechanical stimuli conversion (mechanotransduction) did not come to attention until recently. Hence, the complete mechanisms of mechanotransduction and their connection to the biological pathways activated by force are yet not fully understood (Gingras et al., 2008, Lee et al., 2007, Hytönen and Vogel, 2008, del Rio et al., 2009).

Two diverse themes are discussed in this work; the mechanoactivation of talin and its unfolding under force and the computational methods applied throughout the experimental measurements. The literature review is therefore also discussed in two main theme areas.

In the first part of the introduction to this study, the current understanding and knowledge of mechanotransduction, mechanoactivation and mechanostability of proteins is reviewed. Special attention is given to the intensively studied focal adhesion complex; namely proteins talin and vinculin, which poses important roles in the cell proliferation, substrate adhesion, cell locomotion and cell apoptosis (Fillingham et al., 2005).

The experimental work of this study investigated the dynamics of talin unfolding engaging the steered molecular dynamics methods (SMD). The second part of the introduction is therefore dedicated to the overview of MD and SMD methods, their parameter settings and their applicability in mechanotransduction research.

1.1 Scope

The topic of mechanical signal transduction and force-induced conformational changes were studied on the proteins of the FA complex, namely talin and vinculin. Molecular, structural

and sequential protein changes available in the literature were also discussed in relevance to clinical states or pathological development.

To retain this work in reasonable scope, the practical application and use of the Molecular Dynamics and Steered Molecular Dynamics methods, rather than the detailed algorithm properties, were given attention. In the experimental part, the effect of temperature coupling algorithm choice and implementation (Berendsen vs. Nosé-Hoover), and the impact of force magnitude on the protein unfolding dynamics were investigated.

1.2 Outline

The structure of the thesis is as follows. Chapter 2 and chapter 3 provide a review of the mechanotransduction, mechanoactivation and mechanostability of protein structures, and the introduction to MD/SMD methods, respectively. In chapter 4, the objectives of the study are drawn. Chapter 5 describes the experimental setup and analysis tools engaged in the work. Results of the study are presented in chapters 6, 7 and 8; results of talin alignment and structural study, results of the effect of the temperature control choice, and results of the mutation design, respectively. Finally, chapter 9 discusses the results in connection to the literature review and provides suggestions for future research in the field, while chapter 10 concludes the whole work.

2 Introduction to Mechanotransduction

The term mechanotransduction refers to the ability of sensing mechanical stimuli and translation of the signal into a chemical/biological process which leads to a cell response (Alberts et al., 2008). Living cells respond to mechanical stimulation in a variety of ways which is shaping their phenotype in health and in disease (Johnson et al, 2007, Klein-Nulend et al., 2013).

There are several systems which are known or hypothesized to possess mechanotransduction capabilities. The most thoroughly characterized mechanotransducer systems to date are probably the mechanosensitive ion channels, which react to lateral stretching force in order to change the permeability to ions (Alberts et al., 2008). Another, more hypothetical, mechanotransduction system is the possible activation of G protein-coupled receptors by mechanically induced charge in the lipid membrane (Nelson and Cox, 2008). The recent intensive research in mechanical signal transduction, however, has focused on the focal adhesion complexes; adhesion receptors (integrin) (Tadokoro et al., 2003), the signaling molecules of intracellular proteins of focal adhesion complex (talin, vinculin, paxillin, FAK, calpain1, etc.) and the stress-bearing members of the cytoskeleton (F-actin) (Gingras et. al, 2005, 2006 and 2008, Lee et al., 2007 and 2008, Papagrigoriou et al., 2004, Smith and McCann, 2007, Fillingham et al., 2005, Bate et al., 2012).

2.1 Focal Adhesion Complex

Focal adhesion complex is a intricate mechanotransducing system involving large number of diverse components, see Illustration 1 (KEGG, accessed June 2013). The transmembrane adhesion receptors, such as integrin, mediate the interaction between the extracellular matrix (ECM) and the intracellular matrix (ICM). The ICM proteins connect to the cytoskeletal structures (actin). The cellular response to the external stress further involves passive and active cytoskeletal rearrangement (Vogel and Sheetz, 2006). In other words, the integrin based focal adhesion (FA) complexes serve as mechanosensors converting environmental mechanical cues into biological signals (Geiger et al., 2009).

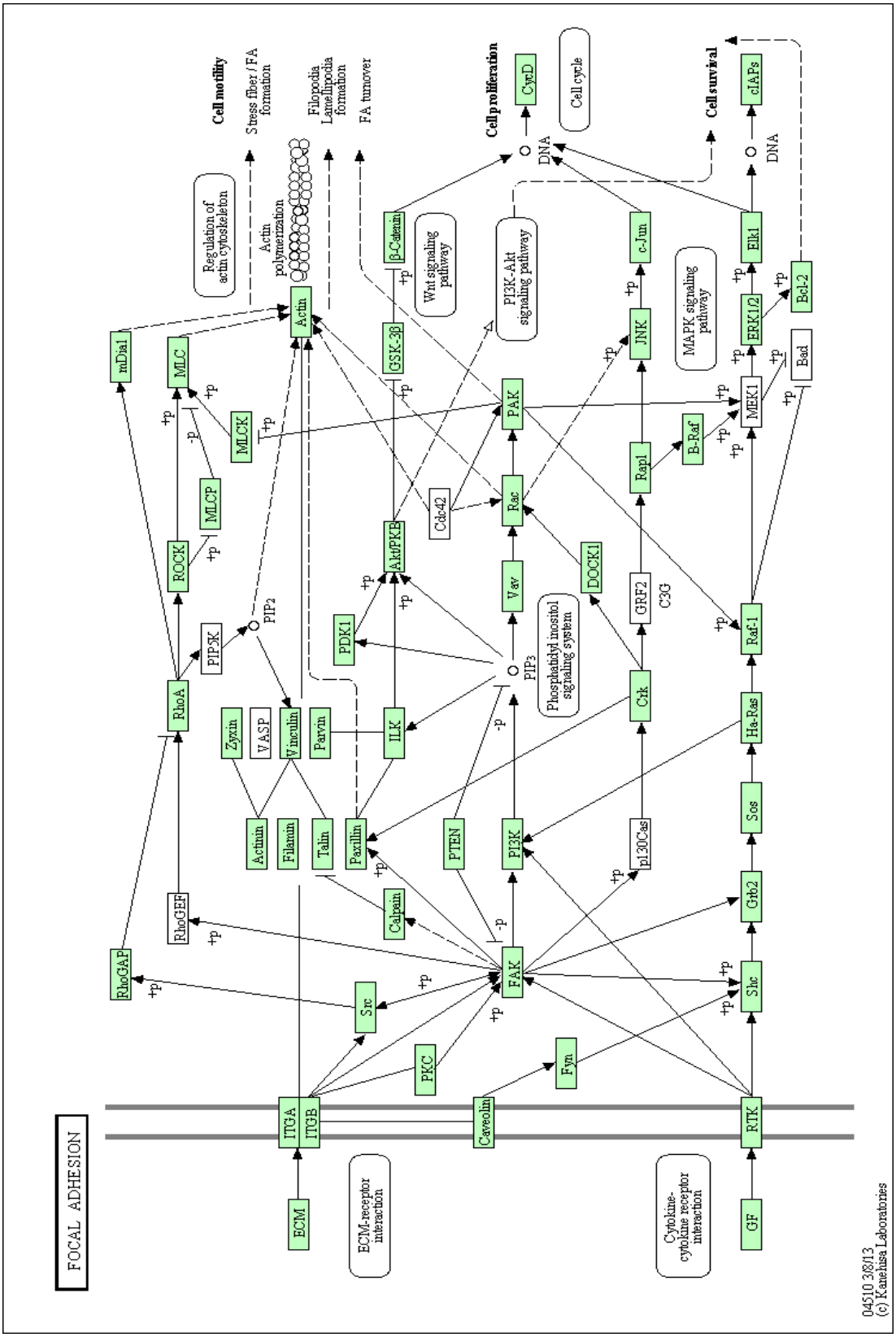


Illustration 1: Focal Adhesion pathway (source: KEGG pathway database [Accessed May, 2013]. Available at: http://www.genome.jp/kegg/pathway/hsa/hsa04510.png)

The focal adhesion complex and force sensing are vital for all cells, since cells must organize themselves in space and interact mechanically with their environment to function properly. They have to be correctly shaped, physically robust and properly structured internally (Alberts et al., 2008). During the developmental stages, for example, the adhesion complex is required for proper morphological assembly and maintenance of biological tissues and organs (Bökel and Brown 2002, Perkins et al., 2010). However, the correct embryonic development and cell morphology (Kopp et al. 2010) is not the only important role of the FA complex. The FA cascade engages in processes that require selective cell-cell interactions, such as blood clotting, cell locomotion important in immune cell functions, and cell migration important in wound healing or tumor growth and metastasis (Kumar and Weaver, 2009, Nelson and Cox, 2008).

2.2 Mechanosensitivity and Disease

For many molecular systems, a connection between failed molecular structure and disease has been discovered (Buehler and Yung, 2009). Since mechanotransduction and mechanosensitivity are vital for living cells, the altered function of abundant mechanosensitive components might be crucial in the development of pathological states. Proteins of the FA complex, as well as other mechanosensitive molecules, have been long studied in connection to disease development. To date, several links between the mechanosensing components and pathophysiological development have been reported (Zhmurov et al., 2011, Ma et al., 2009, Goel et al., 2012, Liu et al., 2007, Brown et al., 2011, Johnson et al., 2007).

2.2.1 Mechanosensitive Proteins and Disease

A link between altered protein mechanostability and disease has been identified, for example, for spectrin; a major cytoskeleton component of the red blood cell. A single destabilizing mutation in the spectrin structure affects the ability to withstand the large mechanical stress on the cell membrane while passing through narrow blood vessels (Johnson et al., 2007). There are several identified point mutations causing hereditary diseases affecting the shape, elasticity and vitality of erythrocytes (Johnson et al., 2007).

Fibrin and fibrinogen are major scaffold blocks of blood clots (thrombi) important in the maintenance of hemostasis. A study by Zhmurov et al. (2011) reported different mechanosensitivities for monomeric and oligomeric fibrin(ogen) units. The changed

sensitivity towards force and pressure increases the thrombi formation and eventually leads to heart attacks or strokes.

The large protein filamin binds to cytoskeletal actin and allows slow crawling movement of cells. When its function or concentration in cells is modified, the cells do not form tissue properly and create disorganized spots. Such behavior was observed, for instance, for melanoma (Alberts, et al., 2008). This observation suggests that also other mechanosensitive proteins of the FA complex may play a role in the metastasis formation, cancer development and spreading.

A Mutation in the α-actinin gene is connected with focal segmental glomerulosclerosis (OMIM id #603278), a cause of nephrotic syndrome and kidney failure. The disease manifests by accumulation of collagen in the ECM and/or by increased cellularity (Thomas et al., 2006).

Upon increased pressure, mechanosensitive cochlin is found to form mechanostable multimeric forms which deposit in collector channels in glaucomatous eyes (Goel et al., 2012). Cochlin might indeed be involved in the regulation of intraocular pressure as well as anywhere, where fluid is involved.

2.2.2 Talin/Vinculin and Disease

Several studies have described a connection between integrin-talin-actin complex formation and disease (Liu et al., 2007). Great attention has been given to the talin and vinculin components of the FA complex (Kopp et al, 2010, Xu et al., 1998). The study of Franco-Cea et al. (2010) illustrates that talin is essential in the initial assembly of the integrin adhesion complex (IAC) as well as in the reinforcement of the initial rather weak complex. In other words, talin is essential for the transient adhesive connections as well as for the stable adhesive contacts. Therefore, the direct binding of actin to talin is essential in tissue development and morphogenesis (Franco-Cea et al., 2010). Vinculin activation is central to the processes of focal adhesion and the reinforcement of the talin-actin binding (Golji and Mofrad, 2010). It is clear that any failure in talin or vinculin function or activation may lead to diseases associated with the failure of focal adhesion formation.

Two paralogs of talin molecules, talin-1 and talin-2, are expressed in normal cells, both of which possess overlapping functions and can therefore compensate for each other (Kopp et al., 2010). However, the exact principle of talin-1 and talin-2 compensation is still not clear. A study by Kopp at al. (2010) showed that talin-2 is overexpressed when talin-1 gene is

downregulated and as a result, compensation for talin-1 function is observed. Although the function is partially compensated, talin-1-knockout mice embryos die in gastrulation (Monkley et al., 2000). The talin-2 knockout causes mild muscular dystrophies in mice (Monkley et al., 2000), while knockout of both genes is perinatal lethal (Conti et al., 2009).

Knockout of either of the talin genes in cell cultures results often in only a mild phenotype exhibiting reduced cellular spreading and focal adhesion assembly (Franco-Cea et al., 2010)

Talin and vinculin proteins have been investigated in connection to retroviral infections (Brown et al., 2011), their infection control ability and spreading limitation. The study by Brown et al. (2011) shows that overexpression of talin and vinculin decreases the susceptibility of human cells to HIV-1 and other retroviral infections by blocking the route of entry and initial steps of the viral life-cycle. Transient knockdown of talin or vinculin increases the susceptibility of human cells to HIV-1 infection. At the same time, paxillin might function as a facilitator and positive regulator of HIV-1 infection (Brown et al., 2011).

2.2.3 Mechanosensitivity and Other Applications

Apart from the pathological relevance, mechanosensitive structures have been recently studied also in connection to cell-growing scaffolds and their design (Sheets et al., 2013). Furthermore, studies of the FA complex have been involved in the design of mechanically stable proteins (Sheets et al., 2013). The importance of mechanosensation and influence of mechanotransduction has been investigated in connection to the physiological adaptation of bone cell mass and structure to stress and pressure (Klein-Nulend et al., 2013).

2.3 Mechanoactivation

The term mechanoactivation refers to protein structural changes induced by mechanical force resulting in revealing target binding site and increasing the binding affinity to the ligand/substrate. Provided is a short overview of protein mechanoactivation knowledge in current research. In the two following paragraphs, mechanoactivation of talin and vinculin is introduced.

2.3.1 Mechanoactivation in Research

Titin is perhaps the most studied protein structure in the field of unfolding protein dynamics. The titin unfolding has been studied *in vitro* in atomic force microscopy (AFM) experiments,

as well as *in silico* single-molecule steered molecular dynamics (SMD) setting (Best et al., 2003, Isralewitz et al., 2001, Lu et al., 1998). Titin is a giant muscle/cardiac protein (Alberts et al., 2008) consisting of long series of Immunoglobulin (Ig) domains. The individual domains unfold when external stress is applied (Isralewitz et al., 2001, Best et al., 2003). Partial unfolding of the titin Ig units is observed after stretching the muscle cells. The Ig domain unfolding generates passive tension which allows the realignment of the muscle filament after stress release.

Stretching force in the titin protein also causes structural rearrangements in the titin kinase domain (Gräter et al., 2005). The study of the titin kinase domain showed significant difference in mechanical stability between the C- and N-termini of the titin kinase domain (tested in force induced protein unfolding by MD and *in vitro*). Such a property is essential for proper mechanical function of the domain activation and deactivation. In more detail, the core of the titin kinase, in particular its active site, was found extremely stable and able to resist large forces compared to other titin Ig domains. Based on such observation, titin kinase is a suspected force sensor. This theory, however, has not been tested experimentally yet (Gräter et al., 2005).

As previously mentioned, the cytoskeletal spectrin in red blood cells is most probably a mechanosensitive protein, and it has been studied in connection to possible mechanoactivation (Alberts et al., 2008, Johnson et al., 2007). Kinesin (Alberts et al., 2008), fibronectin (Gao et al., 2002), cochlin (Goel et al., 2012), filamin (Alberts et al., 2008), fibrin (Zhmurov et al., 2011), calpain (Bate et al., 2012) and proteins of the FA assembly have been studied to determine their sensitivity towards force or pressure and to uncover possible force-induced protein activation.

Furthermore, α-actinin with an amphipathic α-helix structure similar to talin, is possibly another force-sensitive protein. α-actinin is a member of the spectrin protein family containing several spectrin repeats and an actin-binding site (Beggs et al., 1992). It is suspected that the α-actinin actin-binding site activation is similar to the force-induced VBS rotation in talin (Lee et al., 2007).

2.3.2 Mechanoactivation of Talin

Talin is a large cytoplasmic protein which is highly concentrated at cell-substrate and cell-cell contacts. Talin plays an important role in integrin-mediated cell adhesion, cell spreading and migration, cell proliferation and apoptosis (Fillingham et al., 2005). In other words, talin is the

scaffold carrying signals in both directions across the plasmatic membrane between ECM and ICM. Talin connects the large intracellular integrin adhesion complex (IAC) to the cytoskeletal F-actin (Kaufmann et al., 1992) and microtubule network, and to the ECM structures (Brown et al., 2002) including collagen, fibrinogen and fibronectin (Critchley, 2009, Nelson and Cox, 2008).

Even though not fully known, there are several hypothesis about FA complex functioning including talin activation. A study by Vogel (2006) presents a theory that the acting force causes conformational changes to the individual FA protein components and alters their binding affinity to other ICM proteins in the FA pathway. The altered binding affinity can subsequently initiate a biochemical signaling cascade or produce more immediate and local structural changes (Vogel, 2006). Studies supporting this theory are already available. For example, several integrin adhesion site-associated proteins (talin, p130Cas, fibronectin) are known to change conformation upon mechanical stimulation and thereby expose their cryptic binding or phosphorylation sites (Lee et al., 2007, Swanda et al., 2006, Zhong et al., 1998).

Due to the locations of the integrin and actin binding sites in the talin molecule, it is speculated that the force vector is transmitted along the long axis of the talin molecule, hence through the whole H1-H12 rod domain (Hytönen and Vogel, 2008). Talin's integrin-binding site (IBS) is located at the N-terminal end of the talin molecule (Calderwood, 2004), while the actin-binding sites (ABS) are placed further in the sequences closer to the C-terminal end buried inside the α-helix bundles H1-H12 of the talin rod (Smith and McCann, 2007, Lee et al., 2004, Gingras et al., 2008). The force vector direction in relation to the orientation of the anti-parallel helix structures is unfortunately not yet fully known. One possible complete talin structure has been recently published by Goult et al. (2013(acc.man.)). The talin structure proposed by Goult et al. (2013(acc.man.)) comprises of an elongated FERM domain, an unstructured linker (approx. 80 residues) and a flexible rod (62 amphipathic α-helices in 13 domains (R1 – R13) of 4- or 5-helix bundles). The rod terminates with a single helix responsible for talin dimerization, which is essential for cell spreading and focal adhesion assembly.

The transduction of the mechanical signal from the extracellular matrix through the cell membrane and focal adhesion (FA) complexes is reinforced at the plasma membrane by vinculin (Gallant et al., 2005, Carisey et al., 2013). The vinculin binding sites (VBS) in native talin are buried in the interior of the folded α-helix bundles causing considerably lower affinity

to vinculin. This suggests that a major structural change, such as unfolding of talin polypeptide, is required for vinculin recruitment (Papagrigoriou et al., 2004, Vogel, 2006).
It has previously been proven that recruitment of vinculin by talin is force-dependent (Golji and Mofrad, 2010, Margadant et al., 2011). While the binding of vinculin inducing strong cell adhesion is force-regulated, the role of force in the talin structural changes is still only partially understood (Hytönen and Vogel, 2008). Work by Lee et al. (2007) suggests that the activation of VBS1 in the talin domain TAL5 is triggered though rotational movement of the VBS relative to TAL5 while forming strong hydrogen bonds with the talin H1 domain. Such rotation is a consequence of torque applied via hydrogen and ionic bonds between the H1 and the VBS (Lee et al., 2007). The movement reported here was rather small (less than 2Å extension) and can be perhaps described as "loosening" of the helix bundle prime to the VBS rotation, since no domain unfolding was observed. Nevertheless, the rotation of the VBS was observed only in MD simulations.

2.3.3 Mechanoactivation of Vinculin

An MD study by Golji and Mofrad (2010) on the activation mechanism of vinculin suggests that vinculin requires both of its bonding ligands (talin and actin) simultaneously, rather than first binding talin and afterwards actin. The concurrent binding is necessary to produce the internal stretching causing vinculin activation. Furthermore, it has been shown that talin VBS insertions alone are insufficient for vinculin activation (Golji and Mofrad, 2010, Diez et al., 2011). The study indicates a role of mechanical force in vinculin activation, based on the results of the MD simulation (Golji and Mofrad, 2010) and experimental data (Mierke et al., 2008). In addition, recent research on the activation of vinculin proposes several mechanisms of vinculin activation, which can happen simultaneously or one after another. One of the mechanisms is the activation by mechanical force (Golji et al., 2011) and the second is the activation by phosphorylation (Mierke et al., 2008).

The intracellular connections between integrin and talin can be further affected by α-actinin, paxillin, nebulin etc. (Nelson and Cox, 2008), see Illustration 1 (KEGG, accessed June 2013).

2.4 Mechanostability and Protein Unfolding

Proteins and other biomolecules undergo structural and conformational changes when force is applied (Vogel, 2006, Kumar and Li, 2010). These conformational changes can be observed as

unfolding of the natural protein structure. Every protein has a different ability to withstand the unfolding forces. The review on protein unfolding under mechanical force by Kumar and Li (2010) reports on the experimental protein unfolding force ranging between 15pN (Calmodulin) to 1100pN (bovine carbonic anhydrase II). To provide a closer notion of the acting forces, the hydrogen bonds, for example, are interrupted at 10pN force magnitude (review by Kumar and Li, 2010). Fibrinogen unfolds under the force range 125pN – 165pN in SMD simulations (Zhmurov et al., 2011, Liu et al., 2010) and under 90pN in AFM experiments (Zhmurov et al., 2011).

In detail, the structure and stabilizing interactions of a protein determine its resistance to deformation. When the stabilizing bonds and interactions are interrupted, the protein destabilizes and manifests conformational changes or unfolding (Krammer et al., 1999).

The main factors affecting the magnitude of unfolding force, available in the mentioned review (Kumar and Li, 2010), are the secondary structure, the tertiary structure, the geometry of the force application, highly localized hydrogen bonds or the presence of a hydrophobic core. The mechanical stability can be further affected by ligand binding, formation of disulphide bonds or presence of crowders (biomolecules such as sugar, NAs, lipids etc. with steric repulsion among themselves).

In more depth, the secondary structure holds an important role on the unfolding force; an α-helix conformation (α-spectrin) unfolds under smaller force compared to a β-sheet structure (titin I27 domain), apprpx. 30pN and 200pN respectively (Rief et al., 1999, Kumar and Li, 2010). A more complex tertiary structure may increase the resistance against the unfolding force. The six-domain protein ankyrin, for instance, is more resistant to unfolding compared to a single-domain ankyrin structure (Kumar and Li, 2010). A similar effect of greater force resistance of a multidomain structure was also observed for titin (Lu et al. 1998).

Until recently, it was believed that the hydrophobic core provides for the most of the stability of the protein. It was therefore believed that the pulling force direction has little or no impact on protein unfolding. However, recent investigations showed that the geometry of the pulling force affects the dynamics of the unfolding (Kumar and Li, 2010). When the force is acting along the hydrogen bonds, the unfolding can be described as shearing, while force perpendicular to the hydrogen bonds results in unzipping of the structure (Carrion-Vasquez et al., 2000, Kumar and Li, 2010).

The summarized review suggests that intervention to the protein structure, such as single mutation interrupting important bonds or changing the secondary structure, might destabilize the whole protein and consequently increase its mechanosensitivity (Johnson et al., 2007, Faham et al., 2004).

2.4.1 Protein Destabilizing Mutation

A study on bacteriorhodopsin investigated the impact of alanine residue mutations on the mechanostability of the protein (Faham et al., 2004). The study showed that the side chains at the most destabilizing positions tend to point toward the core of the protein structure, whereas those having less contribution to the protein stability tend to be located in the peripheral parts of the molecule (Faham et al., 2004). The work also revealed that polar and non-polar residues pose similar contribution on the protein stability.

A publication investigating diseases caused by spectrin unfolding reported a single pathogenic mutation in the spectrin molecule. The mutation Q471P is uncoupling the tandem of 3-helix bundle repeats of spectrin and causes divergence in the unfolding pathway (Johnson et al., 2007). The resulting effect of the mutation is the loss of helicity at physiological temperature, which causes lower protein mechanostability. This could mean that even a minimal change affecting the unfolding mechanisms might cause vital impact on the cell.

2.5 Talin Structure

Talin (270 kDa, 2451 amino acid (AA) residues) is a high-molecular weight actin-binding cytoskeletal protein belonging to the FERM domain family of proteins (Gingras et al., 2008, Kaufmann et al., 1992). The large talin molecule can be closer defined as the talin head (FERM and F0 domain = 1 – 432 AA residues, ~50kDa) and a long flexible rod domain (433 – 2541 AA residues, ~220kDa) (Winkler et al., 1997, Goult et al., 2013). The flexible rod domain is consists of 62 α-helices (McLachlan et al., 1994) organized in a series of 13 four- and five-helix bundles (Critchley, 2009, Goult et al., 2013). These separate helix bundles are structurally similar to each other. Even though the bundle domains are structurally similar, SMD simulations showed that five- and four-helix bundles differ in their response to mechanical stress (Hytönen and Vogel, 2008); a five-helix bundle is more resistant to stretch than a four-helix bundle. Furthermore, the bundle structure is also similar to other helix bundle proteins with right-handed topology, namely FAK or α-catenin (Fillingham et al., 2005). In

this work, the structure PDB id 2JSW was used (Gingras et al., 2008). Vinculin and VBS structure PDB id 1ZW2 (Gingras et al., 2005).

Talin N-terminal FERM domain contains a binding site for integrin β-tails (Calderwood, 2004). The binding of talin to integrin causes conformational changes in the integrin αβ extracellular domains which increases the affinity of the integrin to ECM structures (Calderwood et al., 1999, Tadokoro et al., 2003). In the binding of integrin to actin, talin can function in two ways; i.e. by direct binding integrin-talin-actin or indirectly by recruiting other actin-binding molecules such as vinculin (Ziegler et al., 2008).

Talin C-terminal end contains the I/LWEQ VBS and THATCH actin binding site (ABS) (Smith and McCann, 2007). At least two other ABSs are located in the talin head FERM domain (Lee et al., 2004) and in the center of the talin rod (Gingras et al., 2008). The talin rod also contains a second integrin binding site, the function of which is yet unknown (Gingras et al., 2009). The I/LWEQ domain contains a motif (62^{nd} helix) for talin parallel homodimerization, which is suspected to be a feature essential to the actin binding ability (Smith and McCann, 2007, Gingras et al., 2008).

Talin interacts with nebulin (NRAP), layilin (LAYN), Src, paxillin, PIP kinase γ and focal adhesion kinase (FAK) (Geiger et al., 2001). Furthermore, talin rod is domain composed of amphipathic α-helices contains ~11 vinculin binding sites (VBS). (Gingras et al., 2005, Gingras et al., 2006, Goult et al., 2010). Vinculin is a cytoplasmic protein which may function as structural reinforcement. Cells with disrupted vinculin function can still form focal adhesions but their ability to spread is reduced (Xu et al., 1998). The talin – vinculin binding aids stable and strong cell adhesion (Fillingham et al., 2005). The recruitment of vinculin is however predetermined by the activation of VBS. The secondary structure of VBS is described as four amphipathic helices with hydrophobic surface, consensus sequence of VBS is LxxAAxxVAxxVxxLIxxA (Gingras et al., 2005).

The unfolding of talin rod bundles leading to vinculin binding increases the flexibility of the talin rod (Fillingham et al. 2005). These may also explain why vinculin suppresses cell migration (Fillingham et al., 2005).

Latest publications concentrate on the activated and autoinhibited structure of complete talin (Goult et al., 2013) and suggest yet another scheme for talin activation by vinculin/RIAM complex. Moreover, calpain2 protein cleavage of talin is recognized as an important step in

FA turnover (Bate et al., 2012) and cell spreading; cleavage at the second calpain2 cleavage site in the talin rod removes talin's dimerization helix at the C-terminal end, which changes the stability/strength of the FA complex.

2.5.1 Talin Mutations

In the presented work, the term mutation stands for any variation in the DNA sequence. For the purposes of this study, only the variations corresponding to the 2JSW model and leading to a change in the AA sequence were collected, no matter on the connection to disease. In total, over 1870 SNPs were found in human genome for the talin gene. However, none of the reported SNPs has a known clinical relevance (NCBI, accessed June 2013).

In order to design point mutations in talin I/LWEQ domain decreasing the mechanical stability of talin molecule, the background information on the variation present in the I/LWEQ domain was collected from the NCBI dbSNP database.

The NCBI database (dbSNP) was investigated for available known mutations and variations in the region corresponding to the 2JWS structure sequence (Gingras et al., 2008). The found single nucleotide polymorphisms (SNP) resulting in a AA residue mutation in the protein sequence, the position in the sequence and the location in the studied structure are available in Table 3.

None of the found SNPs is reported in connection to a disease. Based on the location in the structure, only the mutation N2405S located in the H3 α-helix of the five-helix bundle could be interesting as possible structure destabilizing mutation. All the other SNPs found are located in the turn/coil connectors between the amphipathic helices outside of the core of the bundle.

Table 1: NCBI dbSNP mutation and variation in the talin sequence region from AA residue id 2294 to AA residue id 2482.

SNPs in I/LWEQ talin domain **NCBI dbSNP mutation and variation database**		
AA residue Position	**Mutation**	**Location (source reference)**
2327	K->N K2327N	**lysine to asparagine** located in the turn/coil structure between H1 and H2 α-helix (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=147951718)
2329	R->W R2329W	**arginine to tryptophan** located in the turn/coil structure between H1 and H2 α-helix (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=117039868)
2378	A->V A2378V	**alanine to valine** located in the turn/coil structure between H2 and H3 α-helix (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=144875708)
2405	N->S N2405S	**asparagine to serine** located in the H3 α-helix (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=147275602)

3 Overview of MD/SMD Methods

Recent development in single-molecule measurement techniques has greatly advanced the existing knowledge in the field of mechanical behavior of proteins. Basic classification of these methods in two main groups is represented by the experimental methods (*in vivo* and *in vitro*), and the simulation methods (*in silico*) (Kumar and Li, 2010, Lu et al. 1998, Johnson et al., 2007a). The most common experimental techniques applied are atomic force microscopy (AFM) and optical tweezers. The commonly used simulation techniques are the molecular dynamics (MD) and steered molecular dynamics (SMD) methods also referred to as AFM-pulling methods and Umbrella sampling (terminology used by GROMACS software). The following paragraphs concentrate on the overview of the setting and the application of the computational SMD methods.

Undoubtedly, MD simulations are a valid and valuable tool in protein activation investigation (Golji and Mofrad, 2010) since they provide a detailed atomic description of a biological system. Such information is inaccessible experimentally (Hünenberger, 2005). These methods have proven to be rather accurate and even predictive, and hence they provide an interplay between the theory and the experiments (Vendrusculo and Paci, 2003).

In principle, MD simulations simultaneously solve the classical Newton's equation of motion for a system of a number of interacting atoms in small steps over certain time period (van der Spoel, et al., 2012). In practice, however, there are several factors which have been limiting the use and application of MD simulations as a research tool in biology and biophysics. The first limitation is the computationally intensive simulation production demanding massive parallelization (Lindorff-Larsen et al., 2012). The simulation can be produced using either GPU (Zhmurov at al., 2011) or using massively parallel supercomputers, such as Anton (Lindorff-Larsen et al., 2012) or Sisu at CSC, Finland. The second limiting factor is the realistic representation of the biological systems and their interactions i.e. the choice of system appropriate force field (Lindorff-Larsen et.al, 2012).

Although not actively discussed in the connection of protein unfolding simulations, it is suspected that also other parameter settings, such as the temperature and pressure coupling algorithms, might affect the observed simulation unfolding trajectory (Rosta et al., 2009).

Moreover, there are two typical protocols of the SMD pulling simulations which may influence the observable variables; one is pulling at constant velocity and second is pulling at constant force (van der Spoel et al., 2012).

Finally, there are several different MD and SMD software tools used for computational modeling of protein structures. Commonly used programs are NAMD (Scalable Molecular Dynamics) developed at the Theoretical and Computational Biophysics Group at The University of Illinois, and GROMACS software (GROninger MAchine for Chemical Simulations) developed originally at the Biophysical Chemistry department of University of Groningen currently maintained and supported by university contributors across the world. An example of a commercial molecular dynamics software is CHARMM (Harvard University, Cambridge, MA).

There are vast amounts of algorithms for the production of MD/SMD simulations (van der Spoel et al., 2012, Hünenberger, 2005). In the further paragraphs of this review, common force fields of MD/SMD simulations are discussed. The temperature and pressure coupling algorithms supported by GROMACS are preferred in the later paragraphs. Unfolding force magnitude in SMD and AFM experiments is reviewed in paragraph 3.5 and an overview of exemplary unfolding simulations is available in paragraph 3.6.

3.1 Force Fields

The accuracy of the force field determines the usefulness of the simulation results (Lindorff-Larsen et.al, 2012). To date, there are a number of force fields which are similar in the mathematical function. The force fields, however, differ dramatically in the use of parameters defining the energetic components. The original force field parameters were defined based on the quantum-level calculations or on the small molecules experiments. Recently, number of new force fields have emerged with adjustments in some of the original parameters usually associated with important torsion angles.

The main force fields used are AMBER force fields (Cornell et al., 1995) with original **ff99** and its improved version **ff99SB** followed by **ff99SB*** and **ff03*** which have better energetic helix/coil balance. Side-chain potentials of four amino acids are adjusted in **ff99SB-ILDN** (Lindorff-Larsen et al., 2012). To the second group of commonly used force fields belong the CHARMM (MacKerell Jr. et al., 1998) force fields: **CHARMM19**, **CHARMM22,**

CHARMM22* and **CHARMM27** (Lindorff-Larsen et al., 2012). Additionally, **OPLS-AA** (Kaminski et al., 2001) or **GROMOS** (Scott et al., 1997) are often used force fields (Floudas et al., 2006, Lindorff-Larsen et al., 2012). The comparison tests of commonly used force-fields (Lindorff-Larsen et al., 2012) showed lower stability of folded protein during 10μs MD simulation with CHARMM22. Accurate simulation data, closely comparable to experimental NMR data, were achieved with ff99SB-ILDN and ff99SB*-ILDN, and with CHARMM27 and CHARMM22*. In the force-field performance comparison by Lindorff-Larsen et al. (2012), CHARMM27 and AMBER ff03 displayed overstabilization in formation of helical structures, whereas AMBER ff99SB-ILDN underestimated the stability of helices. Since the force fields display broad range of properties towards forming helical structures, the choice of force field in SMD may affect the folding/unfolding of the protein and its stability.

3.2 Ensembles

One of the concerns of the MD simulations is the specific type of thermodynamic boundary condition imposition; particularly the implementation of constant temperature control on isolated systems (van der Spoel et al., 2012, Hünenberger, 2005). An isolated system is characterized by a time-dependent Hamiltonian where the integration of the classical equation of motion results in a micro-canonical ensemble (constant energy) and its trajectory mapping (Hünenberger, 2005). Such an ensemble of MD simulation is, however, not comparable with real experimental conditions. For possible comparison with experiments, one or more independent variables need to be introduced to modify the mathematical definition of the ensemble (van der Spoel et al., 2012, Hünenberger, 2005).

Examples of the modified ensembles are the canonical ensemble (NVT), the isothermal-isobaric (Gibbs) ensemble (NPT), or grand-canonical ensemble (μVT) (Hünenberger, 2005). In the canonical ensemble, the temperature is specified while the total energy of the system can fluctuate. In the Gibbs ensemble, the temperature and pressure have specified values, while the volume can fluctuate. Finally, the grand-canonical ensemble has constant temperature and volume, while the particles can be exchanged with the surrounding bath. The most common target ensemble in MD simulations is the canonical ensemble.

3.3 Temperature Coupling

Most of the target observations of molecular behavior during a MD simulation are carried out from a constant temperature ensemble; i.e. the canonical ensemble (NVT) (van der Spoel et al., 2012, Hünenberger, 2005). In order to pursue a constant temperature ensemble in MD simulations, weak coupling thermostat, extended coupling algorithm or stochastic coupling methods allowing fluctuation in the specified variable, are applied. The instantaneous (kinetic) temperature is in MD simulations calculated from total kinetic energy by partitioning (Hünenberger, 2005).

GROMACS supports the weak coupling scheme of Berendsen (Berendsen et al., 1984), the extended ensemble of Nosé-Hoover scheme (Nosé, 1984, Hoover, 1985), velocity rescaling scheme (Bussi et al., 2007) and stochastic randomization through Andersen thermostat (Hünenberger, 2005).

Apart from ensuring a canonical ensemble, there are also other reasons to control the temperature of the system. Heating of the system due to external force in SMD or frictional forces (van der Spoel et al., 2012) are two examples of such reasons.

3.3.1 Berendsen Temperature Coupling

The Berendsen algorithm mimics coupling to an external bath with first-order kinetics (Berendsen et al., 1984). This means that the deviation of the system temperature from the initial value gets slowly corrected with time. In more detail, the Berendsen thermostat suppresses the changes of kinetic energy and hence does not provide a “correct” canonical ensemble (van der Spoel et al., 2012) due to the incorrect distribution of the kinetic energy and artificially narrow distribution of temperature and pressure values. The error in the ensemble scales with 1/N, therefore the error gets minimized in large systems resulting in insignificant effects. The main advantage of this algorithm is that the strength of the coupling can be adjusted by the user depending on the type of the simulation. The Berendsen algorithm is extremely efficient during heating or equilibrium for relaxing the system to the target temperature (van der Spoel, 2012). Besides the efficiency, another advantage of the Berendsen thermostat is the first order decay of temperature deviations without oscillations (such as in Nosé-Hoover).

3.3.2 Nosé-Hoover Temperature Coupling

The extended approach of Nosé-Hoover (Nosé, 1984, Hoover 1984) provides "correct" canonical ensemble (van der Spoel, 2012). This extended method reduces the effect of external systems by introducing a thermal reservoir and a non-Newtonian friction term into the equation of motion. These new parameters maintain the total kinetic energy constant. However, the Nosé-Hoover algorithm produces oscillatory relaxation, which poses a problem with the selected time constant; Nosé-Hoover demands a 4-5 times larger relaxation time constant (van der Spoel, 2012) compared to Berendsen.

3.3.3 Temperature Coupling in Practice

There are several considerations in the application of a thermostat in MD simulations. "Hot solvent – cold solute problem" (Hünenberger, 2005) is encountered when a system consistent of distinct subsets with different degrees of freedom (such as biomolecule and water) are coupled to one thermostat. The problem can be prevented by introducing group temperature coupling for coupling the temperature control for parts of the system separately. The main reason for the need of such a parameter is that the kinetic energy exchange between different components (distinct subsets of degrees of freedom) is not ideal (van der Spoel, 2012). When coupling the system to one water bath, the water part will tend to heat while the protein will cool down resulting in significant differences in the solute/solvent temperature as large as 100K.

The second problem arises from incorrect application of the thermostat to the atomic velocities which can cause a so-called "flying ice cube effect" (Hünenberger, 2005). The reason of this problem is the accumulation of kinetic energy in certain degrees of freedom which consequently cools down the internal degrees of freedom. This problem can be solved by removing the center of mass motion from the atomic velocities or application of the thermostat on the internal velocities only (Hünenberger, 2005).

Recent studies have shown that small differences in protein temperature over the simulation time may occur. These fluctuations point at artifacts of the temperature calculation implemented in the algorithm (Eastwood et al., 2010). Large differences in the protein/water temperature, on the other hand, may point out other problems in the simulation setting (Eastwood et al., 2010). If temperature control is applied only to a part of the system, the system will eventually still convert to or stabilize at NVT (Eastwood et al., 2010).

In classical MD simulations, Berendsen weak coupling is recommended for fast establishment of equilibrium. During the simulation production, however, other than Berendsen coupling should be applied to ensure canonical ensemble (van der Spoel et al., 2012). SMD simulations, which are in principle non-isolated (non-canonical) ensembles due to external force introduction to the system, do not have a recommended temperature nor pressure control setting. Still, both controls need to be implemented to ensure only small fluctuation in the otherwise constant temperature and pressure to correctly mimic physiological conditions.

3.4 Pressure Coupling

Similarly to the temperature coupling, discussed in the paragraphs above, the system can be coupled to a pressure bath (van der Spoel et al., 2012). In GROMACS, several pressure coupling algorithms are available; the Berendsen barostat (Berendsen et al., 1984), the extended Parrinello-Rahman approach, and the Martyna-Tuckerman-Tobias-Klein (MTTK) pressure control (van der Spoel et al., 2012).

3.4.1 Berendsen Pressure Coupling

The Berendsen pressure algorithm scales the coordinates and the box vectors every step of the simulation. The scaling has the effect of first-order kinetic relaxation of the pressure towards the reference pressure level (Berendsen et al., 1984).

The result of the Berendsen algorithm provides correct average pressure but not a true NPT ensemble. The impact of the approximation and the possible errors generated are however unknown (van der Spoel et al., 2012). This may cause problems in calculations of thermodynamic properties where the fluctuations in pressure or volume become important.

3.4.2 Parrinello-Rahman Pressure Coupling

Parrinello-Rahman approach is similar to the Nosé-Hoover temperature coupling algorithm, giving the exact NPT ensemble (van der Spoel et al., 2012). The coupling strength depends on the size of the simulation box. Similarly to the Nosé-Hoover, the time constant is not equivalent to the relaxation time (compared to Berendsen coupling). A 4-5 times larger time constant is needed with the Parrinello-Rahman pressure algorithm. It is also possible that for systems far from their pressure equilibrium, large box oscillations appear. The oscillations affect the simulation run and may even result in premature termination of the run (van der Spoel et al., 2012). Furthermore, the algorithm is not directly reversible (also pressure from

the previous step must be included) which may cause inaccuracies in high precision thermodynamic calculations (van der Spoel et al., 2012).

3.5 SMD and AFM Comparison

As discussed previously, SMD and AFM single molecule experiments provide valuable insight into the molecular behavior. Furthermore, SMD methods allow observation of biological mechanisms on an atomic level which is not possible in AFM - experimental setting (Golji and Mofrad, 2010). While the applicability of these methods reaches from biological mechanism predictions to clinical significance discovery, the comparison of the SMD simulations with AFM experimental results is still incomplete.

The main reason for such incompatibility is the insufficient computational efficiency resulting in large (or not sufficiently small) computational time step of the simulation (Sotomayor and Schulten, 2007, Golji and Mofrad, 2010). The activation of the protein molecule *in vivo* is expected in millisecond (ms) timescale, while in the computational MD/SMD simulations, the activation happens within several nanoseconds (ns) (Sotomayor and Schulten, 2007, Golji and Mofrad, 2010). The MD/SMD often increases the stretching speed by 6-8 orders of magnitude compared to the AFM or optical tweezer methods (Lu et al., 1998, Gräter et al., 2005, Golji and Mofrad, 2010, Golji et al., 2011). Obviously, the unfolding of protein domains is a dynamic process where the unfolding itself is influenced by stretching rates. Hence, the unfolding process is speed dependent (Rief et al., 1999).

The timescale difference between MD/SMD and AFM influencing the greater speed of pulling results in different force and energy levels between the simulations and experiments (Sotomayor and Schulten, 2007). The force magnitudes reported for the force-induced unfolding during AFM and SMD usually differ dramatically (Craig et al., 2004, Sotomayor and Schulten, 2007). Since the SMD force needed to unfold a protein structure is logarithmically dependent on the pulling velocity, significantly smaller forces may cause similar structural rearrangements at physiological timescales (Craig et al., 2004, Sotomayor and Schulten, 2007). Furthermore, for some molecules, the forces reported in literature may be comparable between *in silico* and *in vitro* experiments.

Even though the time, speed or force magnitude are not yet comparable among the simulation and experimental methods, the computational tools can be successfully used in predictions. SMD simulations have in the past correctly predicted relative mechanical stability of several

protein domains or positions of key energy barriers (Craig et al., 2004, Sotomayor and Schulten, 2007, Hytönen and Vogel, 2008).

It is clear that further development in computer science is needed to allow all atom simulations in biological relevant timescales and in conditions comparable to *in vivo* and *in vitro* experiments (Galera-Prat et al., 2010). In addition, linking the SMD force magnitude and time to the *in vitro* AFM experiments, or tracking the conformational changes *in vivo* might be useful steps towards methodology inter-connection (Galera-Prat et al., 2010).

3.5.1 Force in SMD and AFM

The timescale and pulling speed difference between SMD and AFM methods resulting in different force and energy levels were discussed in the previous paragraph. Following text provides a brief review of force magnitudes for protein unfolding obtained with AFM and SMD methods.

Spectrin protein unfolds at forces between 25 and 35pN (Rief et al., 1999). The force required to unfold titin observed by Rief et al. (1999) is 30pN, while the study by Isralewitz et al. (2001) reports partial unfolding of titin Ig domains at 50pN and complete Ig unfolding with 150-200pN force magnitude. Unfolding of titin has been carried out in SMD simulations with 750pN, however, that is not necessarily the lowest simulation protein unfolding force (Lu and Schulten, 1999 reported in Isralewitz et al., 2001). Talin rod is activated with 12pN (del Rio et al., 2009), while the SMD talin unfolding was carried out with forces 200pN, 300pN and more (Hytönen and Vogel, 2008). Activation of talin TAL5 domain was also reported when force of approx. 90pN was applied during initial 2ns simulation (Lee et al., 2007).

Furthermore, the unbinding force of a receptor-ligand complex depends logarithmically on the pulling speed over a wide range of tested pulling speeds (Rief et al., 1999). Similar logarithmic dependence was found also for the unfolding of titin Ig domains and fibronectin domains of protein tenascin (Rief et al., 1999). The binding of avidin to biotin requires the unbinding force of 160pN (experimental value - AFM) (Isralewitz et al., 2001).

Among already discussed computational insufficiency, another reason of force value differences (between the experimental vs. simulation, experimental vs. experimental, and simulation vs. simulation results) may be the pulling force application point. The force application points and the pulling vector direction greatly affect the unfolding dynamics

(Hytönen and Vogel, 2008, Golji and Mofrad, 2010, Kumar and Li, 2010). For instance, the force applied at N- and C-terminal ends usually causes complete protein unfolding (Lee et al., 2007), while force application distributed to several pulling points, such as whole helices, unfolds through domain disintegration (Hytönen and Vogel, 2008).

3.6 SMD Methods in Practice

Steered Molecular Dynamics (SMD) simulations are computational methods frequently used for force response studies in various molecular structures. According to the available literature, great interest lies in the mechanoactivation of the FA complex and machanostability of the ECM structures. The research reviewed here for the SMD parameter setting background involves the structures of talin (Hytönen and Vogel, 2008, Lee et al., 2007, Lee et al., 2008), vinculin (Golji and Mofrad, 2010, Golji et al., 2011), titin (Lu et al., 1998) and titin kinase domain (Gräter et al., 2005), fibronectin (Gao et.al, 2002) and fibrinogen (Zhmurov et al., 2011) and spectrin (Johnson et al., 2007). Bond strength investigation, of peptide and Antibody Fragment (AF) for example, is another possible application of SMD methods (Morfill et al., 2008).

3.6.1 SMD Software and Setting

The SMD parameter setting reviewed for the experimental part of this work showed largely varying use of different MD software tools, variables and algorithms. For instance, talin unfolding studied by Hytönen and Vogel (2008) used NAMD MD software with explicit TIP3 water model in solution neutralized with NaCl. 12Å cutoff of van der Waals interactions, PME algorithm for long-range electrostatic interactions and Berendsen barostat with tCouple thermostat were used. Lee et al. carried out their investigations using CHARMM MD software with CHARMM19 and CHARMM27 force fields in implicit (Effective Energy Function; EEF) as well as in explicit water model under Nosé-Hoover temperature control with SHAKE algorithm for bond length constraint and Shift truncation algorithm (cutoff 12Å) for non-bonded interactions (Lee et al., 2007, Lee et al., 2008).

CHARMM software was also used in vinculin simulations (Golji and Mofrad, 2010, Golji et al., 2011). CHARMM19, EEF implicit water solvent and SHAKE algorithm combination, similar to the previously mentioned studies by Lee et al., was applied. In this report, the information about the temperature and pressure coupling is missing.

Lu et al. (1998) used XPOR and NAMD for titin unfolding (CHARMM19, TIP3 water model). Titin kinase domain (Gräter et al., 2005) was, on the other hand, simulated in GROMACS ver. 3.1.4 (GROMOS96 force field and SPC water model, Berendsen thermostat and Berendsen barostat (LINCS algorithm, Lennard-Jones interactions (cutoff 10Å), PME for long range electrostatic interactions, and NaCl neutralized solution).

Zhmurov et al. (2011) utilized the SOP model and Langevin simulations on SOP-GPU package to study the unfolding dynamics of fibrin(ogen), while Johnson et al. (2007) used NAMD with CHARMM27 force field and explicit water model to study the mechanostability of spectrin and its mutants.

The peptide - AF bond strength study was carried out in GROMACS with OPLS-AA force field, SPC/E water model, LINCS constraint, PME and leapfrog algorithm (Morfill et al., 2008). No temperature or pressure control was used in this study.

3.6.2 Temperature

Only small differences in temperature setting are reported in the protein unfolding studies; simulations are mostly carried out at physiological temperatures.

The most common temperature of the water bath in the reviewed literature is 300K (27°C) (Lee et al. 2007, Lee et al. 2008, Golji and Mofrad, 2010, Golji et al. 2011, Hytönen and Vogel, 2008, Gräter et al., 2005, Lu et al., 1998, Zhmurov et al., 2011, Johnson et al., 2007, Morfill et al., 2008). Talin mechanoactivation was also studied at temperature of 310K (37°C) (Hytönen and Vogel, 2008). A greater range of temperatures was investigated for spectrin unfolding (Johnson et al., 2007); 290K (17°C), 300K (27°C) and 310K (37°C).

3.6.3 Pulling Mechanism

On the other hand, the pulling mechanism, the application of the pulling force in the protein geometry, and speed/force magnitude vary among the investigated reports.

Hytönen and Vogel (2008) applied constant force of 100pN, 200pN, 300pN and 400pN, and constant velocity pulling at 1.0, 10 and 100Å/ns with the 5.0kcal/mol/Å^2 springs constant for the talin unfolding. The pulling force was applied to both termini, or alternatively to all Cα atoms of the terminal helices. Lee et al. (2007 and 2008) investigated talin unfolding with both pulling mechanisms; i.e. with constant pulling force (ranging between 15pN to 50pN) and with constant velocity pulling (spring constant at 5.0kcal/mol/Å^2, pulling rate at 0.5Å/ns).

Vinculin was pulled at constant force ranging between 50pN to 200pN (Golji and Mofrad, 2010, Golji et al., 2011). To investigate all possible conformational changes leading to vinculin activation, pulling in three different directions was tested.

Titin was fixed at one of the termini, while external 0.5 Å/ps and 1.0 Å/ps speed was applied to the other terminus (Lu et al., 1998). Titin kinase was investigated during two-sided pull with constant velocity of 0.004Å/ps, 0.008Å/ps, 0.02Å/ps, 0.05Å/ps, 0.1Å/ps and 0.5Å/ps (Gräter et al., 2005).

Fibrinogen was constrained at the N-terminal end and pulled with the C-terminal end with pulling speed of 70nN/s (spring constant k=70pN/nm) (Zhmurov et al., 2011).

4 Objectives of the Study

The literature review on SMD practical application in the protein dynamics studies shows that the investigation of temperature coupling algorithm impacts on the simulation result is scarce.

1. **The first target of the herein presented work is to explore the temperature coupling algorithms and the impact on the protein unfolding dynamics.**
2. **Secondly, the target is to compare the effects of group temperature coupling implementation on the protein unfolding dynamics.**

It is clear that the mechanotransduction, protein mechanoactivation, and most importantly mechanostability have impact on the cell vitality, morphogenesis, proliferation and apoptosis. Affecting these cell characteristic functions by introducing protein destabilizing mutations might lead to pathological changes in cell, tissue or organ leading to disease. There is little information available regarding the mutation in talin structure or their clinical relevance, and talin mechanostability.

3. **The third target of the study is to design point mutations in the talin I/LWEQ domain that decrease the mechanical stability of the structure.**

5 Experimental Setting and Methodology

Experimental setting and methodology chapter presents the collection and multiple sequence alignment tools used in the variation study of talin I/LWEQ domain, paragraph 5.1. The structure analysis tools are introduced in paragraph 5.2. The principle and methodology of the the parameter setting tests with the work flow created for this study are presented in paragraph 5.3. In paragraph 5.4, the details of the I/LWEQ computation model, and the steps performed to produce simulation runs with specified setting are described. Finally, paragraphs 5.5 and 5.6 discuss the tools used for trajectory analysis and mutation design.

5.1 Variation Analysis Tools

Complete talin1 and talin2 protein sequences of different species were collected from NCBI database and aligned to a multiple sequence alignment using the ClustalW2 tool. For the purposes of the analysis, only a part of the constructed alignment, corresponding to the I/LWEQ domain, was analyzed for the present variations. A script was created for automated data collection and visualization of the species variation detected. Jalview ver. 2.8 was used for the human and rodent multiple sequence alignment visualization.

5.2 Structure Analysis Tools

PyMol (ver. 1.30) molecule visualization software package was used for the molecule structure studies and talin – vinculin structural alignment. GIMP image manipulation package and LibreOffice Draw tool were used for image manipulation and graphical presentation.

5.3 Design of Parameter Setting Tests

The parameters tested in the study were the choice of applied temperature and pressure controls during the equilibration and pulling simulation, and the magnitude of applied pulling force appropriate for talin I/LWEQ domain unfolding simulation.

Illustration 2 presents the work flow and the design of parameters tested in the work. The parameter setting tests were executed in three steps. Step 1 investigated the influence of chosen type of temperature and pressure controls. Berendsen temperature control combined with Berendsen pressure control setting applied during equilibration and pulling simulation was compared to Nosé-Hoover temperature control and Parrinello-Rahman pressure control setup applied during equilibration and pulling simulation.

Step 2 investigated the effect of temperature control on the protein unfolding process. To study such impact, the appropriate protein temperature control was turned off during the constant force pulling. Step 1 and step 2 simulations, produced at 310K and pressure 1bar with constant pulling force 300pN, were 5ns long.

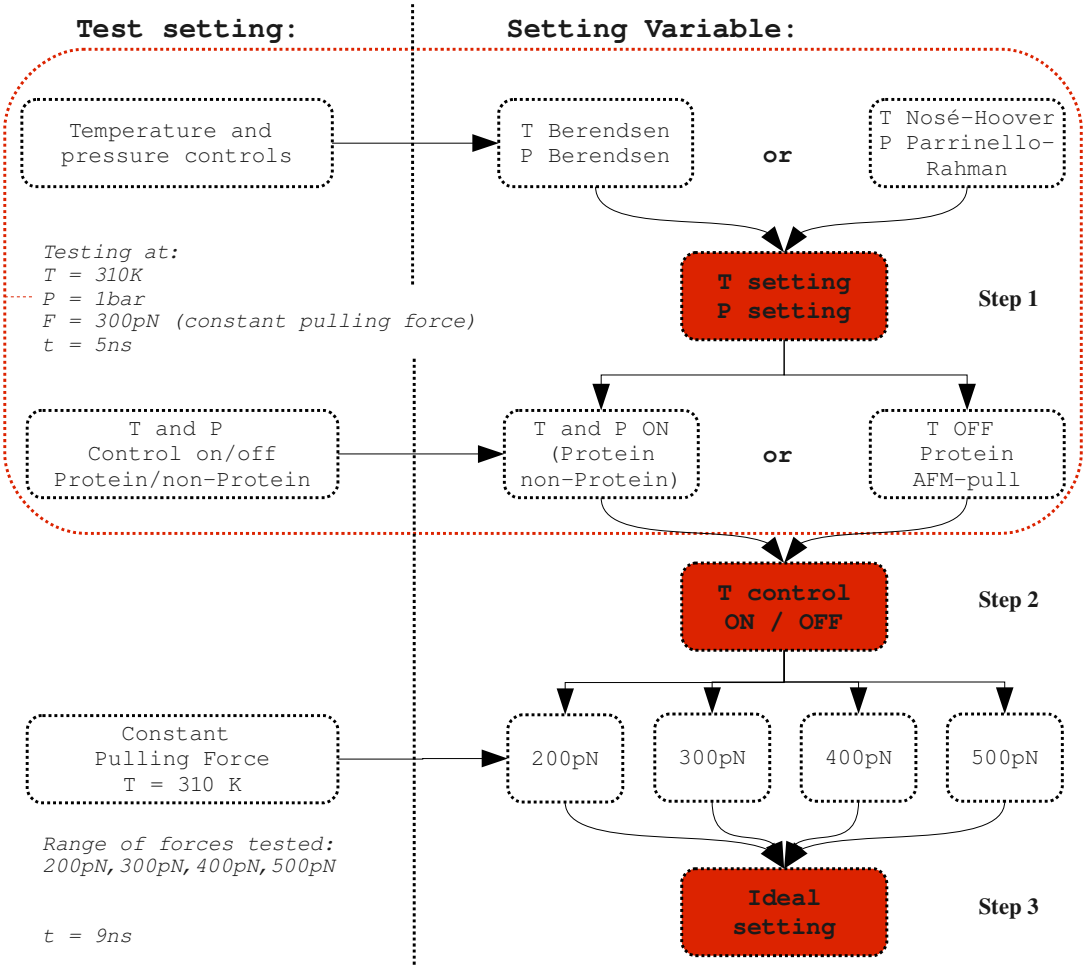


Illustration 2: Work flow of parameter setting tests.

Finally, step 3 tested the impact of range of pulling forces on the dynamics of protein unfolding. Results of pulling with 200pN, 300pN, 400pN and 500pN force during 9ns were analyzed.

The three steps of the parameter setting tests provided ideal temperature and pressure controls, ideal pulling force setting as well as information on the I/LWEQ talin domain behavior when force applied. Consequently, the obtained pulling trajectories were further used for ion bond, hydrogen bond and hydrophobic interaction analysis leading to identification of residues interesting for structure destabilizing mutation design.

5.4 Computational Model and SMD Simulation

The computational model and the SMD pulling simulations were produced using GROMACS ver. 4.5.5 software.

Following three paragraphs contain the instructions and list of GROMACS commands leading from PDB structure file to final simulation production. All details of the setting, which are not mentioned in the paragraphs bellow, are available in the simulation input files prepared for energy minimization, equilibration and simulation production, see Appendix I, II and III.

5.4.1 Talin Model and Energy Minimization

Read pdb file, create GROMACS coordinate file:

```
g_pdb2gmx -f 2JSWa.pdb -ignh -ter
             [CHARMM27, TIP3P, NH3+ and COO-]
```

Create index file with fixed atom, pulled atom, and a group of both fixed and pulled atom which describes the direction of the pulling vector:

```
g_make_ndx -f conf.gro -o index.ndx
```

Create position restraints for fixed atom:

```
g_genrestr -f conf.gro -n index.ndx -o posre_fixed.itp
```

Orientate vector between fixed and pulled atoms to be perpendicular to xy-axis plane and parallel to zx and zy-axis planes. The positions of fixed and pulled atoms were checked from manipulated coordinate file and vector calculated after each adjustment in order to find the best adjustment. The target x and y vector components equaled to 0.0:

```
g_editconf -princ -f conf.gro -n index.ndx -o princ.gro
```

```
            [in group = system, position orientation = direction of the
            vector (fixed and pulled atom group), out group = system]
g_editconf -rotate 0 -90 0 -f princ.gro -n index.ndx -o rotated.gro
            [in group = system, out group = system]
```

The impact of the coordinate file manipulation is presented in Illustration 3. Figures a, b and c show the changes in the model position in respect to xyz coordinate system. Final model was oriented so, that the pulling vector between N-terminal and C-terminal end was defined by 0.0; 0.0; 5.5 xyz vector components.

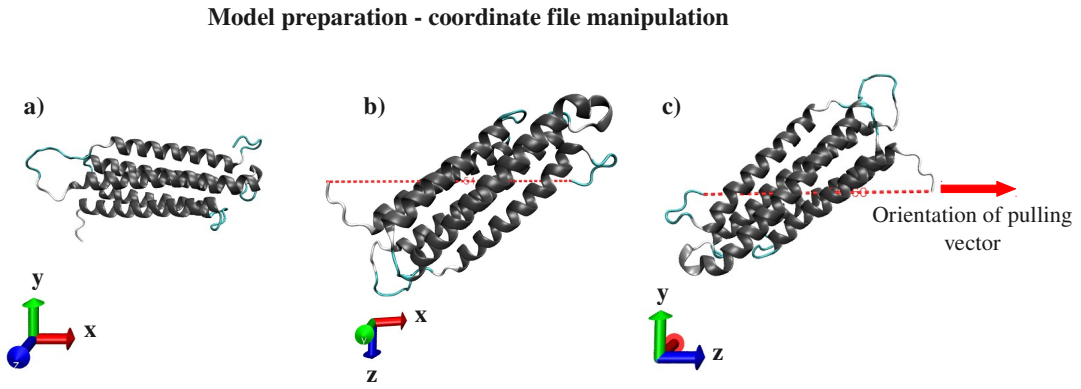


Illustration 3: Model preparation: coordinate file manipulation. Visualization with VMD.

Create a box and place talin molecule to the left hand side of the box:

```
g_editconf -f rotated.gro -o box.gro -center 3.5 3.5 7.5 -box 7.5 7.5 16
            [or -center 5 5 8 -box 10 10 40 for 9ns simulations, values in
            nm]
```

Solvate the box with water molecules and 0.15M KCl ions so that neutral solution is formed:

```
g_genbox -cp box.gro -cs -o solvated.gro -p topol.top
g_grompp -f min.mdp -n index.ndx -c solvated.gro -o min.tpr -p topol.top
g_genion -s min.tpr -o ionized.gro -p topol.top -pname K -nname CL -conc 0.15
-neutral
```

The rectangular box filled with solvent and protein are shown in Illustration 4.

Prepare new index file containing water and solvent molecules:

```
g_make_ndx -f ionized.gro -o index_fix_pull.ndx
```

Prepare production file for energy minimization step and run energy minimization:

```
g_grompp -f min.mdp -c ionized.gro -o En_min.tpr -p topol.top
g_mdrun -v -deffnm En_min
```

Model preparation – box preparation, solvation and ionization

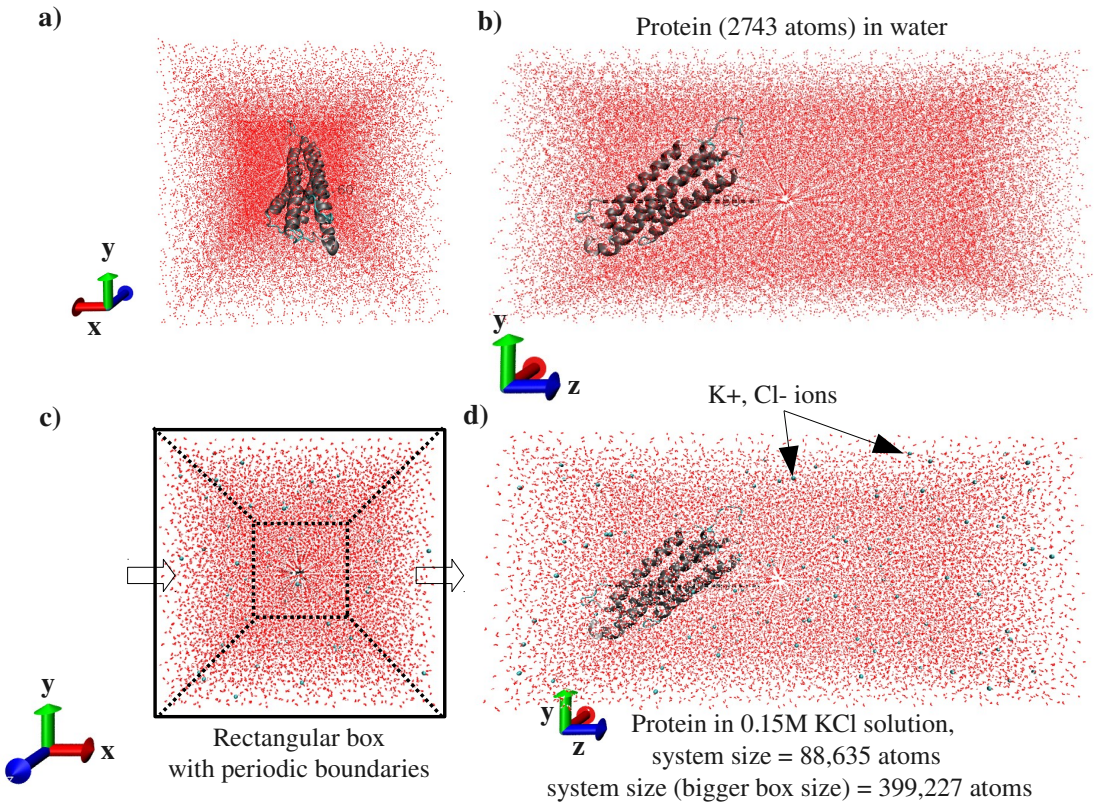


Illustration 4: Model preparation: box preparation, solvation and ionization. Visualization with VMD.

5.4.2 Equilibration

Each system was equilibrated for 1ns at temperature 310K and pressure 1bar. Temperature and pressure controls were applied for protein and non-protein part of the system separately.

Prepare production file for equilibration step and run 1ns equilibration:

```
g_grompp -f equil.mdp -n index_fix_pull.ndx -c En_min.gro -p topol.top -o
eq_pull_name.tpr -maxwarn 1

g_mdrun -mpi cpnum -noappend -deffnm eq_pull_name
```

5.4.3 SMD Pulling Simulation

For each parameter setting, two pulling simulation at temperature 310 K and pressure at 1bar were produced. As for equilibration, the temperature and pressure controls were applied for protein and non-protein parts separately. Simulations at step 1 and 2 for thermostat and barostat tests were 5ns long, while simulations at step 3 for force tests were 9ns long.

First pulling run used coordinate file produced after full 1ns equilibration. Second run used coordinate file produced after 0.8ns equilibration. The 0.8ns coordinate file needed to be created using saved trajectory step files with the following command:

```
g_trjconv -f eq_pull_name_08ns_step.cpt -s eq_pull_name.tpr -o
eq_pull_name_08ns.gro
```

In the pulling simulations, Cα of the N-terminal residue was fixed; i.e. its position was restrained. Such change was introduced manually to the topology file (topol.top) using the position restraint file (posre_fixed.itp) created earlier, see paragraph 5.4.1:

```
; Include Position restraint file
#ifdef POSRES_fixed
#include "posre_fixed.itp"
#endif
```

Prepare production file for SMD pulling simulation and run pulling simulation:

```
g_grompp -f pull_code.mdp -n index_fix_pull.ndx -c eq_pull_name.gro -p
topol.top -o pull_name.tpr -maxwarn 1

g_mdrun -mpi -cpnum -noappend -deffnm pull_name
```

5.5 Parameter Setting Analysis Tools

Following analysis methods and tools were used in order to assess the quality of the test setting and its influence on the final SMD pulling simulation. The following set of analysis was done at each step of the test setting plan using GROMACS 4.5.5 unless stated otherwise.

5.5.1 Force Penetration Analysis

The force penetration ability of the system was assessed by calculating and comparing total unfolded protein length, D1 segment length and D2 segment length (definition available in chapter 6). Additionally, the D1/D2 ratio was expressed for each simulation setting.

Measure total (N- C-terminal) protein elongation over the simulation trajectory:

```
g_polystat -f pull_name.xtc -s pull_name.tpr -n index_fix_pull.ndx -o
N_C_totalDistance.xvg
```

```
[N- C-terminal distance, total protein length]
```

Create new index file to specify D1 and D2 segments:

```
g_make_ndx -f pull_name.gro -o index_distance.ndx

[D1 and D2 CA id atoms in separate groups]
```

Measure D1 (or D2) segment length:

```
g_dist -f pull_name.xtc -s pull_name.tpr -n index_distance.ndx -o
distance_d1.xvg

[or -o distance_d2.xvg, with appropriate atom groups selected]
```

5.5.2 Temperature and Pressure Analysis

Temperature of the protein, non-protein and whole system was expressed from the energy file separately for each test setting. Pressure value of the whole system was extracted from the same energy file.

```
g_energy -f pull_name.edr -o temperature_system_part.xvg
[or pressure, calculation based on selection from available choices]
```

5.5.3 Trajectory Visual Analysis

Each trajectory as a result of SMD pulling simulation was analyzed visually (VMD visualization software ver. 1.9.1) to examine the behavior of the molecule model and the influence of the parameter setting. Images (VMD tachyon rendering) of the whole trajectory taken in 1ns intervals were compared among the different simulations.

Secondary structure changes were investigated with VMD - timeline - secondary structure changes tool which provided more objective information about the molecular structure changes during the simulation showing the unfolding progress in an easily readable graphical form. These charts are however not included in this report.

5.5.4 Viscosity Analysis

The impact of possible temperature changes in the system without protein temperature control was planned to investigate by calculating the shear and bulk viscosity of the system and its components. GROMAC g_energy tool was selected to calculate the values from energy files output during the simulation runs. The viscosity was me calculated for the whole system, for the protein and the non-protein part separately.

5.6 Mutation Design Analysis Tools

Methods described in the paragraph below were used to identify residues in the I/LWEQ domain with the structure destabilization prospect. In more detail, the target of the mutation was to faster reveal the vinculin binding site leading to vinculin binding. This means that interactions suspected to hold the five-helix bundle together were given main attention. In other words, inter-helix interactions were targeted. Finally, in the case of interactions where one of the residues was located in the H2 (VBS) helix, the residue which was not located in the vinculin binding site was chosen for the mutation design.

Trajectories to provide basis for the mutation design were chosen based on the results of the three step parameter setting simulations. Results of ion bond analysis, hydrogen bond analysis and hydrophobic interactions analysis executed on 200pN and 300pN pulling force trajectories were compared. PyMol visualization software was used to visualize the original residue, the residue as a result of the mutation and the original bonding partner in the targeted interaction. VMD visualization software was used to present the disruption of the interaction during the simulation trajectory.

The residues with the structure destabilizing potential were identified as target for the mutation. According to the review by Kumar and Li (2010), the residues with side chains pointing towards the core of the structure increase the mechanical stability of the structure. Hence, this observation was taken into account in the mutation design apart from the observed differences between 200pN and 300pN simulation trajectories.

The principle of the mutation design was to disrupt the ionic or hydrogen bond between the mutated position and its original bonding partner, or to affect the hydrophobic interactions with the residues in the proximity of the mutated residue. Furthermore, the created mutation was designed to introduce as minimal impact on the rest of the structure as possible. Residues with small uncharged side chain which do not cause repulsion or large structural changes or small polar residues were preferred in the mutation design.

5.6.1 Ionic Bond Analysis

Ionic bond analysis was done with the VMD timeline and salt bridge tool calculating and plotting the existence of salt bridges over the investigated trajectory. Default bond length with default angle setting was chosen to specify the ion bonds. Stable inter-helix bonds identified

during the slow unfolding simulation (200pN pulling force) were suspected to hold an important role in structure stability. The list of such stable inter-helical interactions was compared to a similar list obtained from the 300pN pulling force simulations. Ionic bonds that were broken during the second unfolding were identified as valuable positions for the consequent mutation design.

A small uncharged amino acid, such as alanine, was preferred in the mutation design in order to destroy the ionic bond and destabilize the structure.

5.6.2 Hydrogen Bond Analysis

Hydrogen bond analysis was done with VMD timeline and hydrogen bond tool calculating and plotting the existence of hydrogen bonds over the investigated trajectory. Default hydrogen bond length with default angle were used to specify the hydrogen bonds. The principle of identifying the interesting residues and bonds and the principle of the mutation was similar to the ion bridge analysis described above.

5.6.3 Hydrophobic Interactions Analysis

The trajectories created with 200pN and 300pN pulling force were studied with the VMD visualization software. The differences in the protein unfolding were followed on the hydrophobic and hydrophilic residues and their interactions in the structure. Important hydrophobic interactions were destroyed by a mutation of the hydrophobic residue to a small polar amino acid.

6 Results: Variation and Structural Analysis

6.1 Variation in Talin I/LWEQ Protein Sequence

Following paragraph presents the results of the variation analysis over number of different species. Illustration 5 shows the alignment of 25 ortholog sequences, while Illustration 6 presents alignment of collection of 8 human talin-1 sequences and 7 rodent talin-1 sequences (mouse and rat).

Talin protein overall shows very little variation over the 25 investigated species. Illustration 5 presents found the ortholog variation in the I/LWEQ talin domain. The multiple sequence alignment uncovered, that no variation was present in the area of VBS among the investigated species. Illustration 6 shows alignment of I/LWEQ talin-1 domain of human, mouse and rat protein sequences. The only difference between human and rodent in the domain of interest, is in position 2482, i.e. the last residue in the investigated sequence. For human, present amino acid is glutamic acid (E); in rodent sequences, aspartic acid (D) is present.

species		*Bos grunniens mutus* (yak)	*Bos taurus* (domestic cattle)	*Orcinus orca* (killer whale)	*Sus scrofa* (wild pig)	*Homo sapiens* (human)	*Pan troglodytes* (chimpanzee)	*Macaca mulatta* (rhesus macaque)	*Macaca fascicularis* (crab-eating macaque)	*Papio anubis* (olive baboon)	*Loxodonta africana* (African elephant)	*Saimiri boliviensis boliviensis* (squirrel monkey)	*Callithrix jacchus* (new world monkey)	*Trichechus manatus latirostris* (Florida mantee)	*Otolemur garnettii* (small-eared galago)	*Ceratotherium simum simum* (white rhinoceros)	*Equus caballus* (domestic horse)	*Odobenus rosmarus divergens* (Pacific walrus)	*Felis catus* (domestic cat)	*Mus musculus* (mouse)	*Rattus norvegicus* (rat)	*Cricetulus griseus* (Chinese hamster)	*Cavia porcellus* (Guinea pig)	*Sarcophilus harrisii* (Tasmanian devil)	*Gallus gallus* (chicken)	*Taeniopygia guttata* (zebra finch)
columns of alignment	position																									
	2313	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A
	2314	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	E	A	A	A	A
	2334	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	Q	Q
	2340	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	D	N
	2377	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	S	G	G	G	G
VBS	2378	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	V	A
	2379	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	X	I	I	I	I
	2384	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	V	V
	2421	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	E	E
	2448	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	H	H
	2481	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	Q	Q
	2482	E	E	E	E	E	E	E	E	E	E	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D

Illustration 5: ***I/LWEQ talin domain ortholog variation*** *over 25 species. Presented are only positions of I/LWEQ domain, where variation was detected. Gray line: location of VBS.*

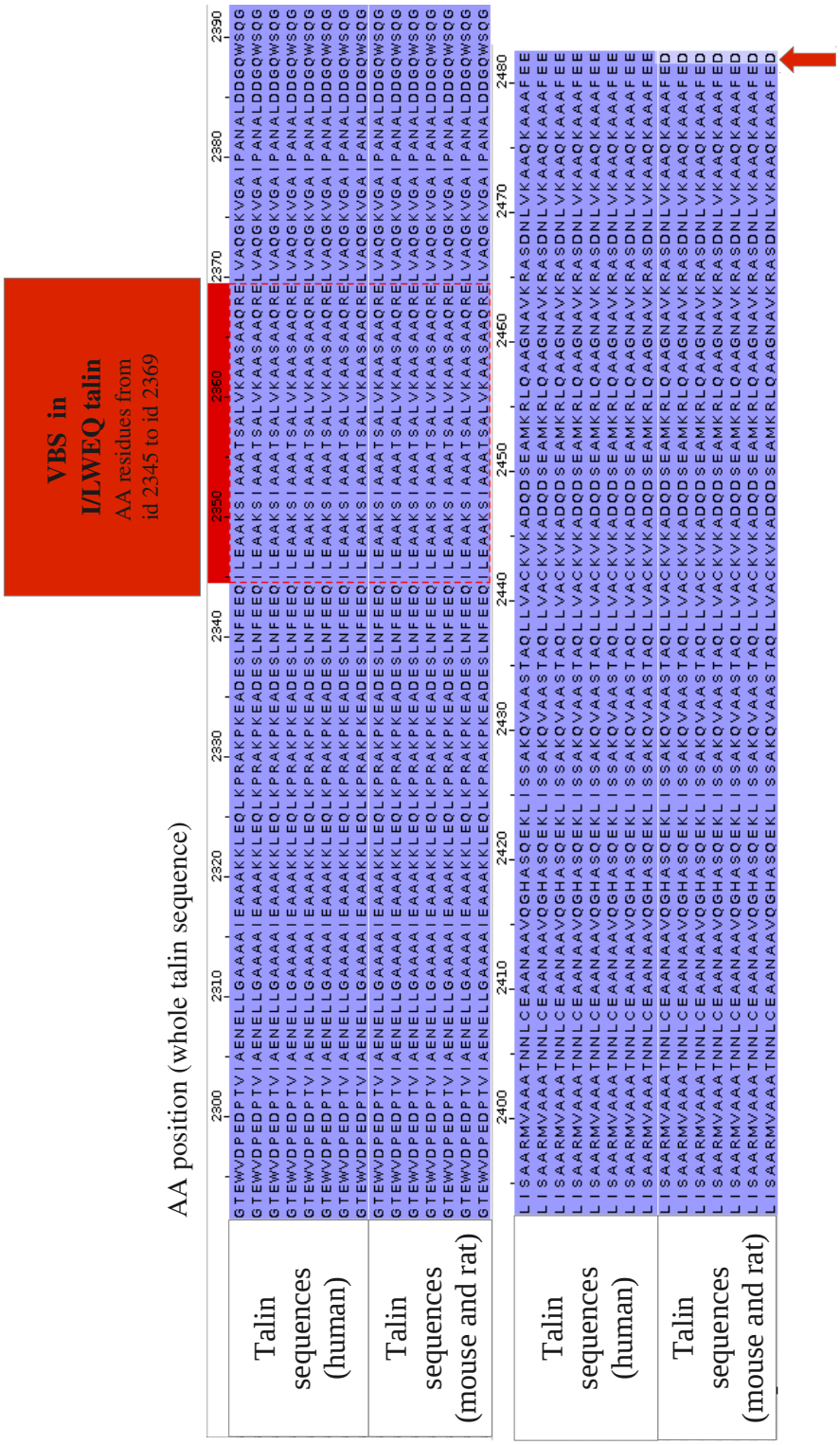


Illustration 6: ***Alignment of talin and talin-1 I/LWEQ domain (human, mouse and rat)****. Red box highlights the VBS in the I/LWEQ talin domain, white line separates the human talin sequences from the mouse and rat talin sequences, red arrow points at found variation in the alignment. (Alignment with Clustalw2 and visualization with Jalview ver. 2.8, blue coloring in the alignment presents the BLOSUM62 score.)*

The results of the background data collection and protein sequence alignment show that talin sequence is indeed highly conserved. Overall low number of reported mutation and variation in the protein sequence suggests high significance of the protein. It is highly probable that sequence conservation and correct functioning of the protein plays crucial role in cell viability.

6.2 Structure of I/LWEQ Talin Domain

In this work, the vinculin binding site of I/LWEQ domain of talin was studied. The structure used to create the model was collected from PDB database (PDB id 2JSW, Gingras et al., 2008). The talin structure study is presented in Illustration 7. The structure of a vinculin molecule forming complex with the VBS of the studied talin domain, available in Illustration 9, was found from PDB database with id 1ZW2 (Gingras et al., 2005).

The I/LWEQ talin domain (13th 5-helix bundle in talin rod) corresponds to 189 AA residues from residue G2294 to residue D2482. The original structure file contained 20 NMR solution structures of the I/LWEQ talin domain. The structure file was initially manipulated so that only the first model was kept as a set of starting coordinates for GROMACS coordinate file (further referred to as 2JSWa.pdb).

The I/LWEQ talin domain contains one vinculin binding site (AA residues 2345 to 2369) which is located in the H2 α-helix of the I/LWEQ five-helix bundle. The top view of the

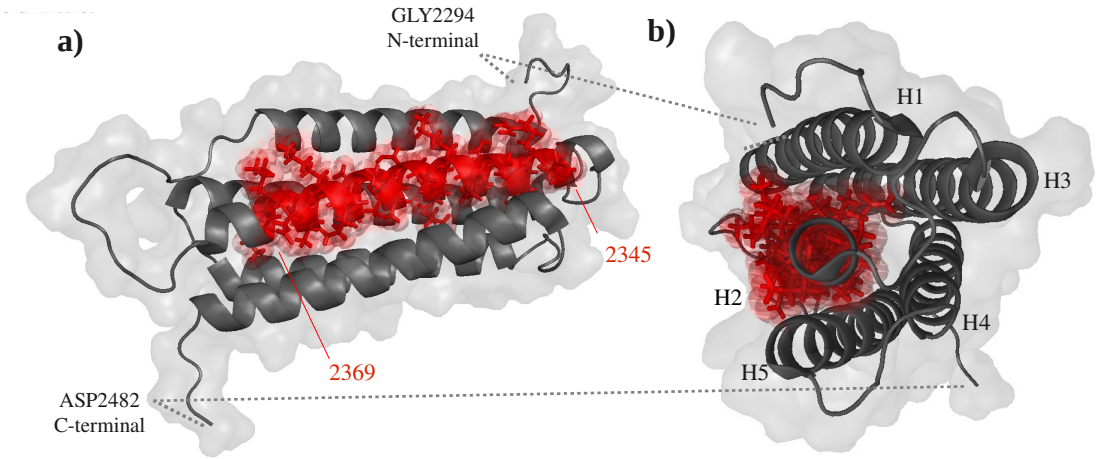


Illustration 7: ***Study of I/LWEQ talin domain structure*** *with VBS (red), GLY2294:CA at N-terminal end fixed and ASP2482:CA at C-terminal end pulled in the SMD pulling simulations.* ***a) side view:*** *amino acid residues id 2345 – id 2369 = VBS;* ***b) top view:*** *amphipathic α-helices H1-H5; H1 at N-terminal end, H5 at C-terminal end, H2 =VBS.*

I/LWEQ talin domain, Illustration 7 b), shows slight twist of the five helices present in the bundle, which further buries the hydrophobic VBS residues inside the structure.

α-carbon (CA) of glycine amino acid (residue id 2294) at the N-terminal end was specified as the fixed atom in all SMD pulling simulations. CA of aspartic acid (residue id 2482) at the C-terminal end was specified as the constant force application point; i.e. the pulled atom in all SMD pulling simulations.

For the purposes of parameter setting test analysis and mutation design, the I/LWEQ talin structure was divided into segments based on the visual examination of the secondary structure features, see Table 2 for details.

Table 2: Talin I/LWEQ secondary structure segments and labels.

Talin I/LWEQ Secondary Structure Segments				
Label	**Detail**	**Residue id**		**Description**
TC1	N-terminal	2294	2300	Turn/Coil1
H1		**2301**	**2325**	**α-Helix1**
TC2		2326	2340	Turn/Coil2
H2		**2341**	**2373**	**α-Helix2**
	VBS	**2345**	**2369**	**Vinculin binding site**
TC3		2374	2379	Turn/Coil3
3-10H		2380	2382	3-10 Helix
H3		**2383**	**2416**	**α-Helix3**
TC4		2417	2420	Turn/Coil4
H4		**2421**	**2443**	**α-Helix4**
TC5		2444	2450	Turn/Coil5
H5		**2451**	**2476**	**α-Helix5**
TC6	C-terminal	2477	2482	Turn/Coil6

Additionally, for the assessment of the force penetration during SMD pulling, the original I/LWEQ domain was specified D1 and D2 molecule segments. **D1** segment was used to describe the unfolding at the N-terminal fixed end as a distance between CA atoms of residues G2294 and A2319. Similarly, **D2** segment was used to describe the unfolding at the C-terminal pulled end as a distance between CA atoms of residues Q2437 and D2482.

6.3 Talin – Vinculin Structural Alignment

One of the targets of the work was to study the unfolding of the talin molecule in order to describe and explain revealing of the vinculin binding site for the binding of vinculin molecules. The vinculin complex with VBS sequence (PDB id 1ZW2) superimposition to talin

molecule (PDB id 2JSW) was used to study the relative location of the talin's bundle helices to the vinculin's bundle helices.

A simplified graphical presentation of the alignment is shown in Illustration 8. VH2, VH3 and VH4 vinculin helices are also in the same orientation (N- and C-termini residue in the helix) to H5, H4, and H3, respectively. On the other hand, VH1 is in opposite N- and C-terminal orientation to H1.

For the purposes of the alignment, only the VBS sequence included in model 1ZW2 was structurally aligned to the corresponding sequence in talin structure, see Illustration 9. The VBSs structural alignment showed that the 4 helices of vinculin, with high affinity to talin's VBS, have comparable locations to the 4 non-VBS talin helices. In more detail, VH1 (vinculin helix 1) was found in the similar position to H1 (talin helix 1), VH2 was in similar position to H5, VH3 to H4 and VH4 to H3.

Talin bundle helices overlapping with vinculin helices as well as the "twist" of the helices in the bundle suggested that the VBS α-helix has to be completely released from the bundle to reveal the VBS hydrophobic residues. One option for the activation could be protein unfolding, i.e. stretching the bundle. Another activation mechanism could be rotation of VBS helix in order to reveal the interface hydrophobic residues as reported by Lee et al. (2007 and 2008).

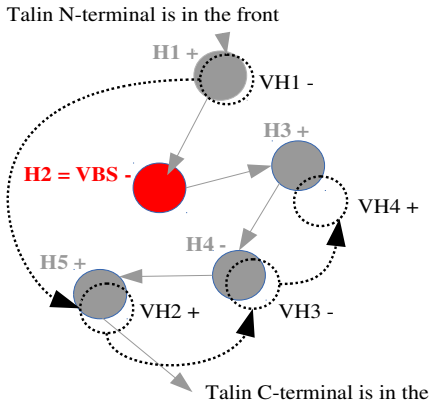


Illustration 8: ***Graphical model of vinculin and talin superimposition*** *according to the VBS sequence. Gray = talin helices (H1, H3, H4 and H5); red = VBS (H2); black = vinculin helices (VH1, VH2, VH3, VH4); + and - signs show the orientation of the amphipathic helices: + = N-end residue is close to the viewer; - = C-end residue is close to the viewer in the top view of the bundle presentation.*

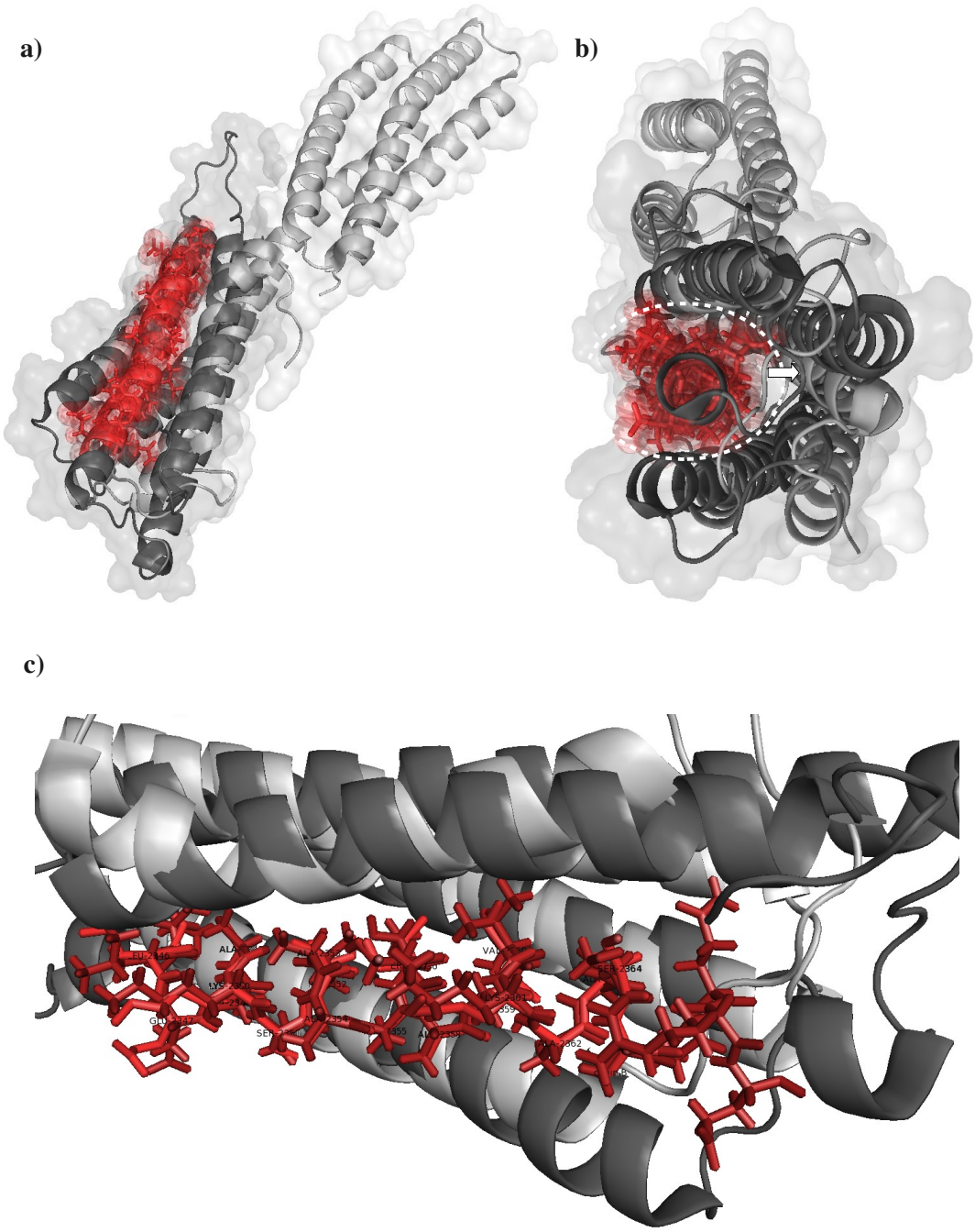


Illustration 9: ***Study of VBS superimposition in talin and vinculin molecule.*** *Red = buried vinculin binding site (VBS) light red = talin, dark red = vinculin; light gray = vinculin, dark gray = talin;* ***a) side view; b) top view*** *shows VBS buried inside the twisted α-helix structure. Hydrophobic VBS-vinculin interface (white circle and arrow) is in talin oriented towards the talin bundle core.* ***c) VBS alignment detail*** *(Talin PDB id 2JSW, vinculin PDB id 1ZW2).*

6.4 Hydrophobic Residues in Talin I/LWEQ Domain

Similar organization was observed in the vinculin structure when hydrophobic and hydrophilic residues were investigated in the structural superimposition. Illustration 10 shows a study of hydrophobic and hydrophilic residues in a simplified model. Hydrophobic residues of the VBS (V, A, I and L) are shown in red; other hydrophobic and aromatic residues are shown in white (P, M, W, Y and F); polar hydrophilic residues are shown in cyan R, K, D, E, Q and N. Other (G, S, T, C and H) are shown in black (Nelson and Cox, 2008).

As can be seen from the simplified study above, strong hydrophobic and hydrophilic residues are organized in a spiral-like organization in the H2 helix containing talin's VBS. As well as in the adjacent helices, the hydrophobic residues (the VBS itself) are mainly facing the inner parts of the molecule; i.e. the VBS is buried inside the structure. The hydrophilic residues are facing outside from the structure core. A similar study of hydrophobic residues in vinculin (not included) provided a similar arrangement of the hydrophobic and hydrophilic residues.

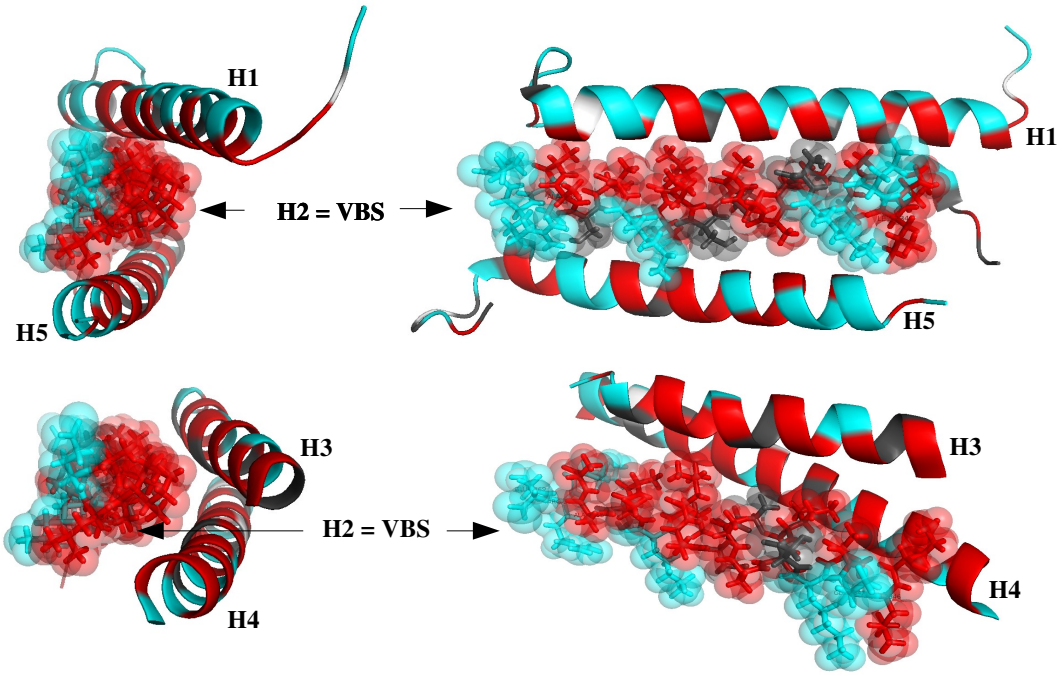


Illustration 10: ***Study of hydrophobic and hydrophilic residues in talin structure****.* ***Red:*** *hydrophobic in VBS = V, A, I and L;* ***white:*** *hydrophobic = P, M, W, Y and F;* ***cyan:*** *polar hydrophilic = R, K, D, E, Q and N;* ***black:*** *other residues G, S, T, C and H.*

7 Results: Parameter Setting Simulations

Following paragraphs describe the observations of parameter setting tests. Protein unfolding trajectories, resulting from step 1, 2 and 3, provided arguments for ideal thermostat, barostat and constant force setting for talin rod segment unfolding. Furthermore, the observations of protein unfolding dynamics was used as a ground for structure destabilizing mutations design in the I/LWEQ domain.

7.1 Temperature and Pressure Controls; Step 1

Step 1 simulations were created to test the impact of the chosen thermostat and barostat on the protein unfolding dynamics. The resulting 5ns trajectories of run 1 and 2 of SMD pulling simulations with 300pN constant pulling force (Illustration 11, left) with Berendsen thermostat and Berendsen barostat were compared to the results obtained with Nosé-Hoover thermostat and Parrinello-Rahman barostat (Illustration 11, right).

The choice of thermostat and barostat did not show large impact on the protein unfolding dynamics. Nosé-Hoover thermostat with Parrinello-Rahman barostat showed slightly faster unfolding in one of the two runs performed, and greater movement around the vertical axis. The other repetition run of Nosé-Hoover/Parrinello-Rahman studied on the protein elongation was almost identical to one of the elongation curves obtained with Berendsen temperature/Berendsen pressure coupling setup, see Illustration 12. The final protein length reached 21nm with partially unfolded H1 and H4 α-helix and completely unfolded H5 C-terminal α-helix.

The average length of D1 segment at the N-terminal was comparable between both thermostat/barostat settings. The difference was only seen in the D2 segment elongation at the C-terminal pulled end, see Illustration 13. The average D2 length of Nosé-Hoover/Parrinello-Rahman was greater compared to Berendsen/Berendsen D2 average length.

Another important feature of the unfolding that can be read from Illustration 12, is the stepwise increase in the protein length in time. Such a pattern, translated to the structure characteristics, was caused by the slow unfolding of the rigid helix structures and the fast unfolding and protein elongation when turn/coil structures are present. This may also lead to

the assumption that there are strong intra-helical and inter-helical interactions acting against the pulling force, holding the structure intact.

The simulations proved to be of stochastic nature. The choice of thermostat and barostat control did not influence the model behavior significantly during the pulling simulations. The two repetitive runs showed greater differences in the unfolding patterns and protein elongation than the differences observed between the different thermostat/barostat settings. The two repetitive runs did not provide sufficient arguments for the control selection. Hence both thermostat/barostat options were tested in the consequent step of the parameter setting study.

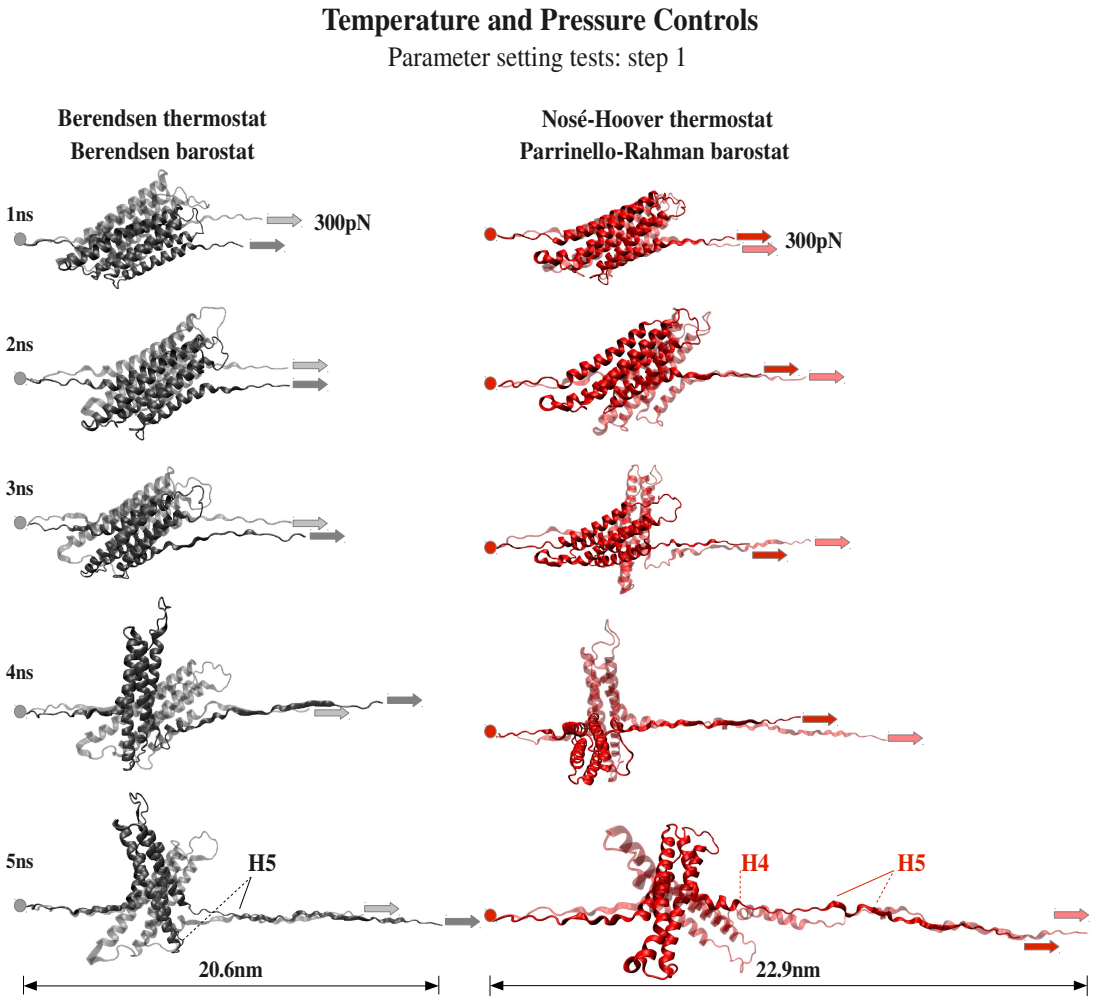


Illustration 11: ***Effect of temperature and pressure controls on the protein unfolding dynamics over 5ns.*** *Study of trajectories created during step 1 of parameter setting tests. Circles show the fixed points during pulling, arrows show the constant force application points.* ***Left****: run 1 and run 2 trajectories with Berendsen thermostat and Berendsen barostat (gray).* ***Right****: run 1 and run 2 trajectories with Nosé-Hoover thermostat and Parrinello-Rahman barostat (red).*

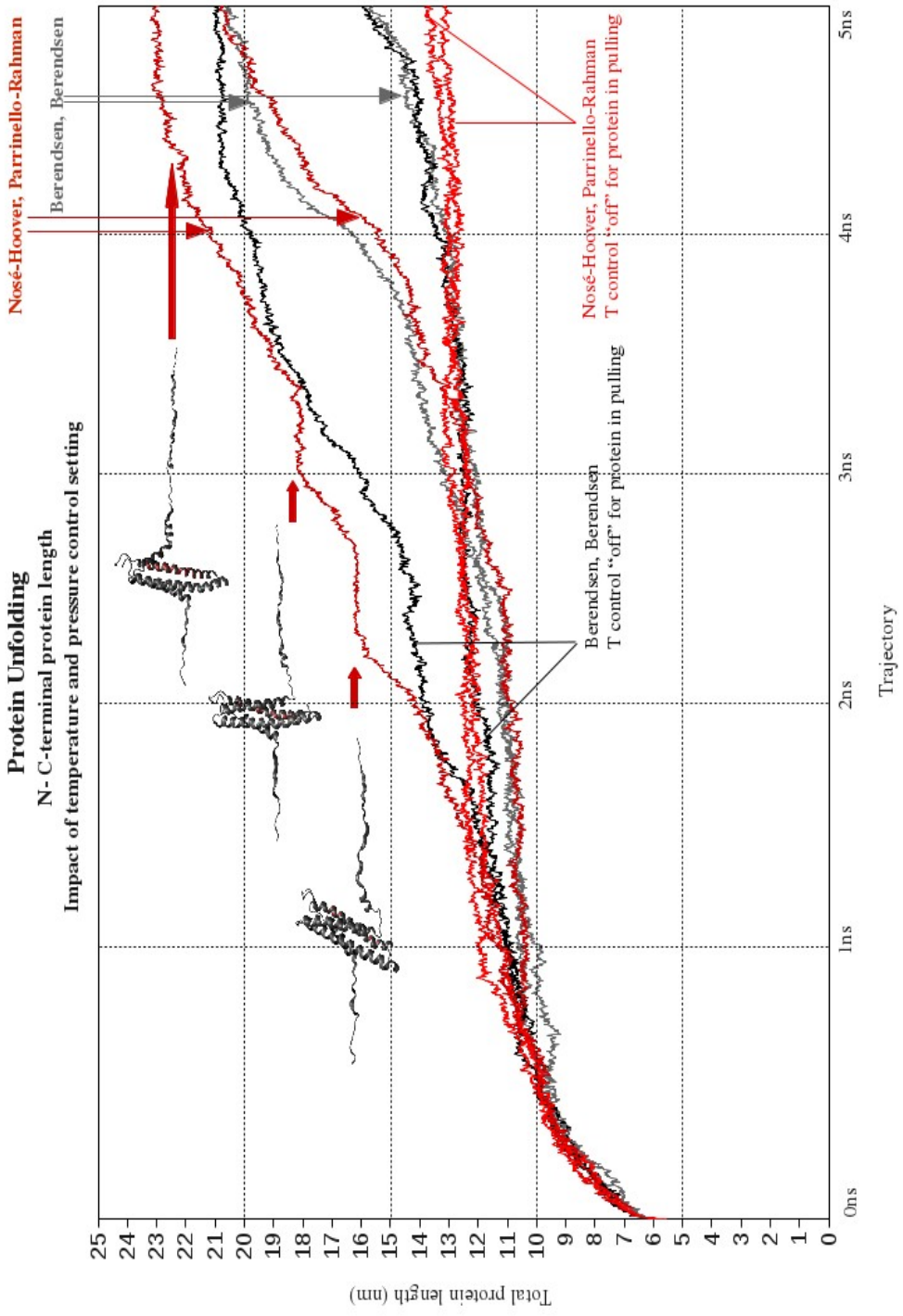


Illustration 12: ***Impact of temperature and pressure control on protein unfolding over 5ns.*** *Study of N- C-terminal protein length of Nosé-Hoover thermostat / Parrinello-Rahman barostat (dark red), Nosé-Hoover thermostat turned off for protein in pulling (light red), Berendsen thermostat / Berendsen barostat (dark gray), and Berendsen thermostat turned off for protein in pulling (light gray).*

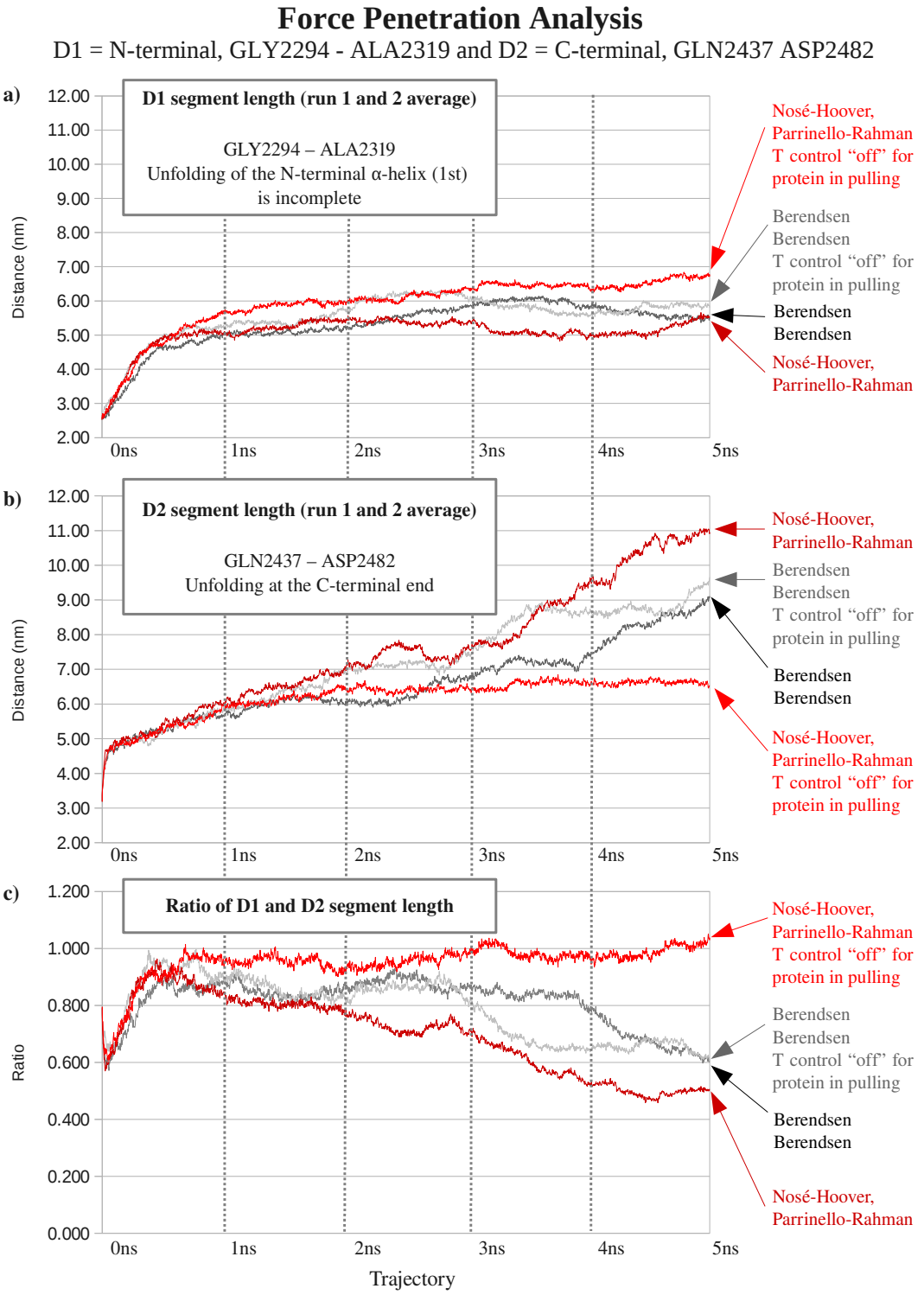


Illustration 13: ***Force penetration analysis. D1 and D2 protein segment length.*** *Comparison of D1 and D2 segment elongation when N-terminal is fixed and C-terminal is pulled.*

The results of the force penetration analysis, presented in Illustration 13, shows the elongation of the D1 and D2 segment during 5ns pulling simulation. The results present the segment elongation only for simulations where the N-terminal in D1 segment was fixed and the C-terminal in D2 segment was pulled. The comparison here is provided to present the unfolding dynamics. For the complete force penetration study, the fixed and pulled points should be switched and consequently D1 and D2 segment elongation compared to the existing charts from the existing fixed/pulled setting.

7.2 Impact of Temperature Control; Step 2

Second step in the parameter setting tests investigated the impact of chosen thermostats on the protein unfolding dynamics. Since the previous step did not give reasonable grounds for thermostat/barostat selection, both temperature controls were tested in the step 2 simulations.

The 5ns simulations were produced at 310K and 1bar pressure after 1ns and 0.8ns long equilibration. In the step 2 pulling, the Berendsen or Nosé-Hoover thermostat were turned off for protein part of the model. Each test setting simulation was repeated twice. The total protein elongation over time, the force penetration through the structure, and average temperature of the whole system were analyzed for the protein and non-protein part separately. Furthermore, the pressure fluctuation was analyzed for both setups.

Illustration 12, available in the previous paragraph, presents the total protein elongation over 5ns when temperature control was not applied, in comparison with the elongation of temperature controlled protein. As can be seen from the chart, the consistency among the two repetitive runs was very low. The only test setting that showed similar protein elongation was when Nosé-Hoover temperature control was turned off for the protein part. However, this setting resulted to very slow unfolding with final protein length of 12.8nm. For this simulation setting, neither H1 nor H5 helix was completely open. Study of the structure appearance over the simulation time is shown in Illustration 14.

As shown in Illustration 13, the protein unfolding at the fixed end of the molecule was comparable for all simulations, chart a). The total protein elongation was caused by changes in the D2 segment; i.e. changes at the pulled C-terminal end of the structure, chart b). As mentioned before, for the complete force penetration analysis, the D1/D2 relation should be investigated when the original fixed end is pulled and the original pulled end is fixed.

The on/off state of temperature control with Berendsen thermostat proved insignificant on the protein elongation in the test scale provided in this study. Protein behavior controlled with Berendsen thermostat was similar to the protein behavior when the protein temperature was not controlled.

The results of average temperature analysis for step 1 and step 2 simulations are shown in Illustration 15. Average temperature was expressed separately for protein, non-protein and the whole system. As can be seen from the chart and the statistics, the differences of non-protein and system temperature were insignificant between the system with or without temperature control.

Similar analysis was carried out for the pressure fluctuation for both systems. The pressure remained stable over the 5ns simulations, and the fluctuation among the two tested setups was comparable. The graphical presentation of the pressure analysis is not included in the report.

Impact of Protein Temperature Control

Parameter setting tests: step 2

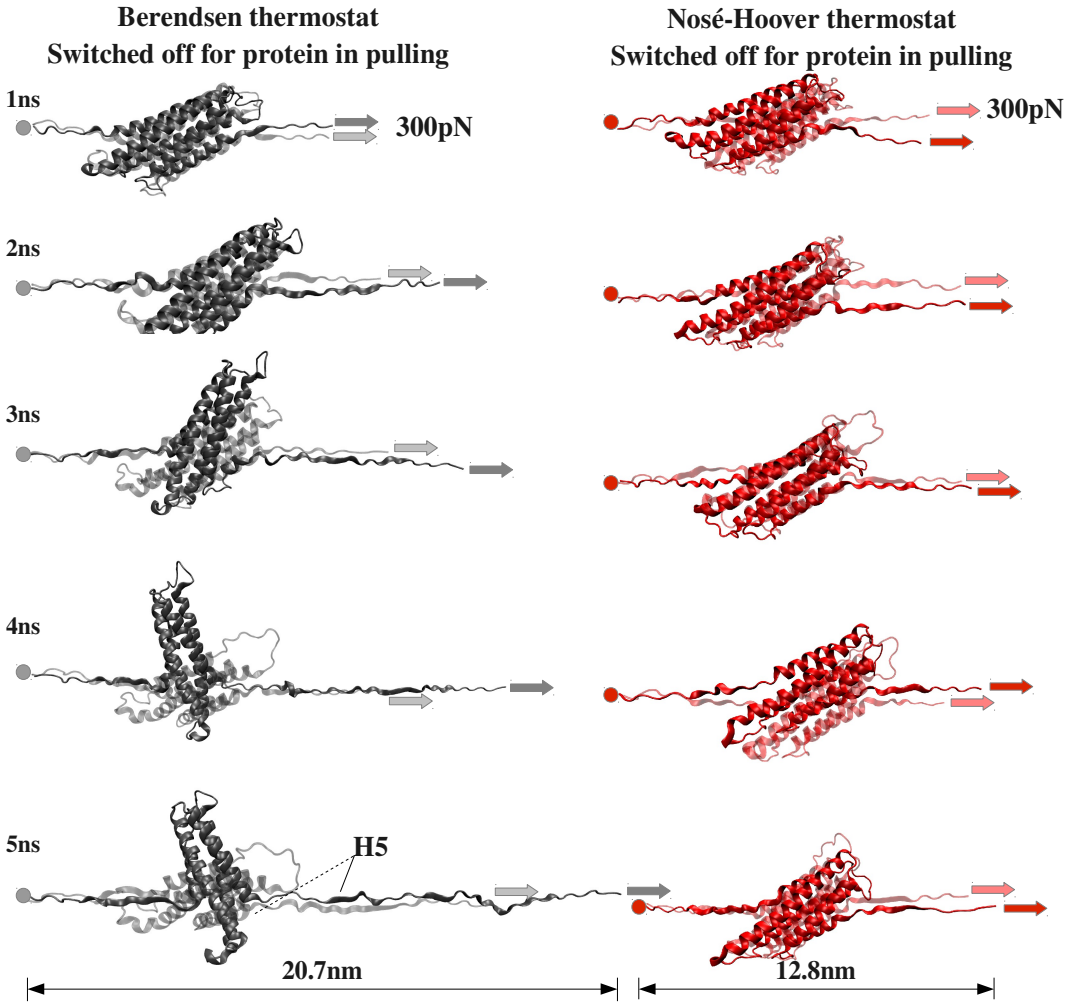
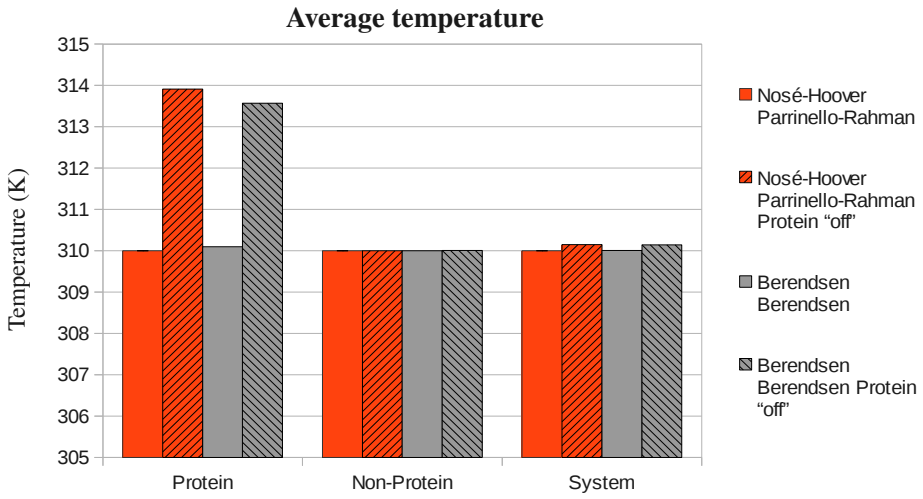


Illustration 14: ***Impact of protein temperature control on the protein unfolding dynamics over 5ns****. Study of trajectories created during step 2 of temperature setting tests. Circles show the fixed points during pulling, arrows show the constant force application points.* ***Left****: run 1 and run 2 trajectories with Berendsen thermostat turned off for protein in pulling (gray).* ***Right****: run 1 and run 2 trajectories with Nosé-Hoover thermostat turned off for protein in pulling (red).*

Average Temperature Comparison over 5ns trajectory

Berendsen Temperature coupling and Berendsen Pressure coupling vs.
Nosé-Hoover Temperature coupling and Parrinello-Rahman Pressure coupling



Temperature Analysis

Protein	Nosé-Hoover Parrinello-Rahman	Nosé-Hoover Parrinello-Rahman Protein "off"	Berendsen Berendsen	Berendsen Berendsen Protein "off"
average	**310.00**	**313.91**	**310.10**	**313.57**
max	335.96	335.11	327.69	335.56
min	306.50	290.53	294.09	294.50
range	29.46	44.58	33.60	41.07
stdev	**5.87**	**5.25**	**3.75**	**5.04**

Non-Protein	Nosé-Hoover Parrinello-Rahman	Nosé-Hoover Parrinello-Rahman Protein "off"	Berendsen Berendsen	Berendsen Berendsen Protein "off"
average	**310.00**	**310.00**	**310.00**	**310.01**
max	314.21	314.22	313.37	312.95
min	305.72	305.76	306.61	306.53
range	8.50	8.46	6.76	6.41
stdev	**1.07**	**1.07**	**0.77**	**0.77**

System	Nosé-Hoover Parrinello-Rahman	Nosé-Hoover Parrinello-Rahman Protein "off"	Berendsen Berendsen	Berendsen Berendsen Protein "off"
average	**310.00**	**310.15**	**310.01**	**310.14**
max	314.05	314.18	313.16	313.19
min	305.68	305.90	306.70	306.50
range	8.36	8.28	6.46	6.69
stdev	**1.05**	**1.05**	**0.75**	**0.77**

Illustration 15: ***Average temperature of the protein, non-protein and system*** *over 5ns simulation. Overview and comparison of the influence of thermostat/barostat setting on the temperature (Berendsen thermostat with Berendsen barostat; Nosé-Hoover thermostat with Parrinello-Rahman barostat). Overview and comparison of the influence on the temperature when Berendsen thermostat or Nosé-Hoover thermostat is not applied in pull for protein.*

The average protein temperature was approximately 4K higher in the system without temperature control in pulling. The temperature difference was, however, insignificant and consistent for both tested thermostats. The type of temperature control cannot be therefore used to explain the dramatic difference in the protein unfolding dynamics when Nosé-Hoover thermostat was not applied for protein.

To provide better understanding of the temperature development over the simulation time, protein, non-protein and total system temperature was plotted over 2ns long time period. The analysis is provided for the simulation with Berendsen thermostat applied to the whole system (step 1 simulation) in comparison to the simulation without Berendsen thermostat for the protein part (step 2 simulation). The temperature development is followed in Illustration 16. For Berendsen step 1 simulation (Illustration 16, a), the temperature was stable at approx. 310 K for all parts of the simulated system.

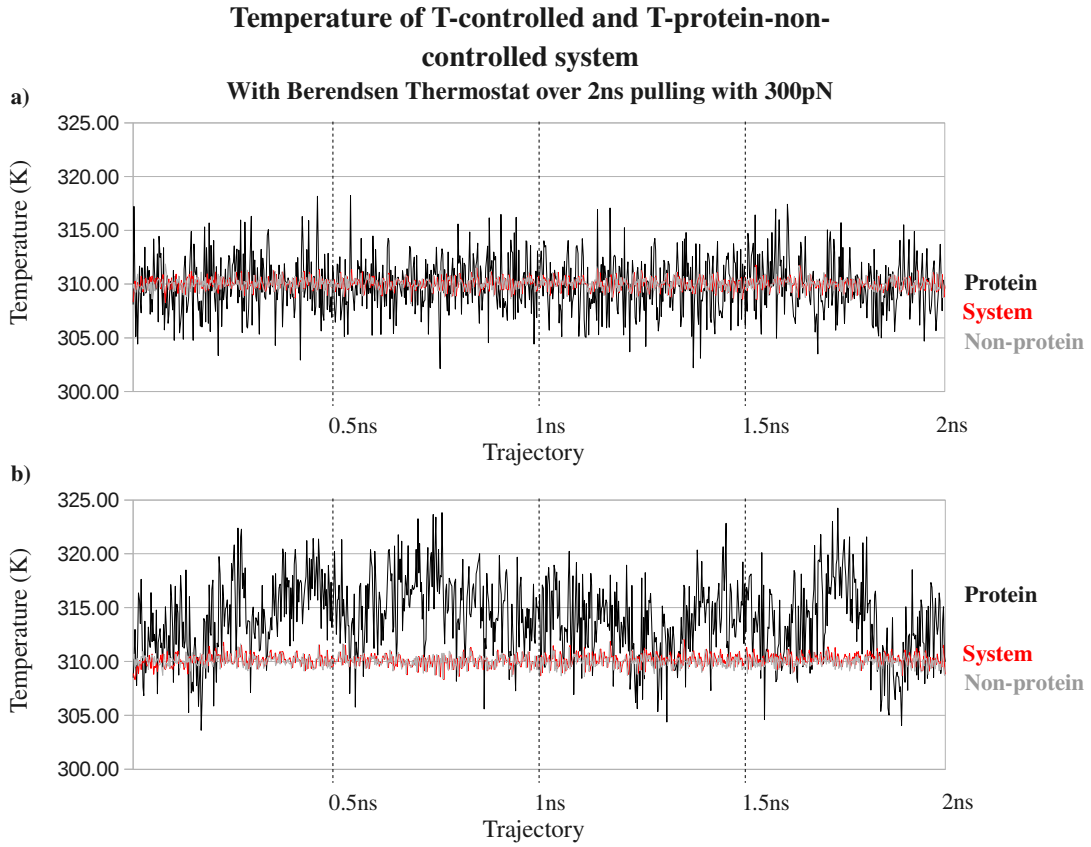


Illustration 16: Temperature evolution over 2ns of models with Berendsen temperature control applied for the whole system (a) and without protein temperature control (b).

Even though the charts follow the temperature only over 2ns long period, the temperature was stable and constant over the whole 5ns long simulation.

Greater fluctuation was observed for the protein part in both models. Furthermore, greater temperature (approx. 314K) and greater temperature fluctuation was observed for the non-controlled protein part compared to the temperature controlled protein (temperature level 310K), see Illustration 16. Similar temperature evolution was observed for the Nosé-Hoover thermostat step 1 and step 2 simulations.

To further investigate the impact of the thermostat, a 2ns test simulation was produced completely without temperature control. The temperature analysis is presented in Illustration 17. During the test simulation, the initial temperature of protein, non-protein and system was comparable to both temperature controlled systems. However, a linear increase in the temperature of all mentioned model parts was observed including large temperature fluctuation in the protein part.

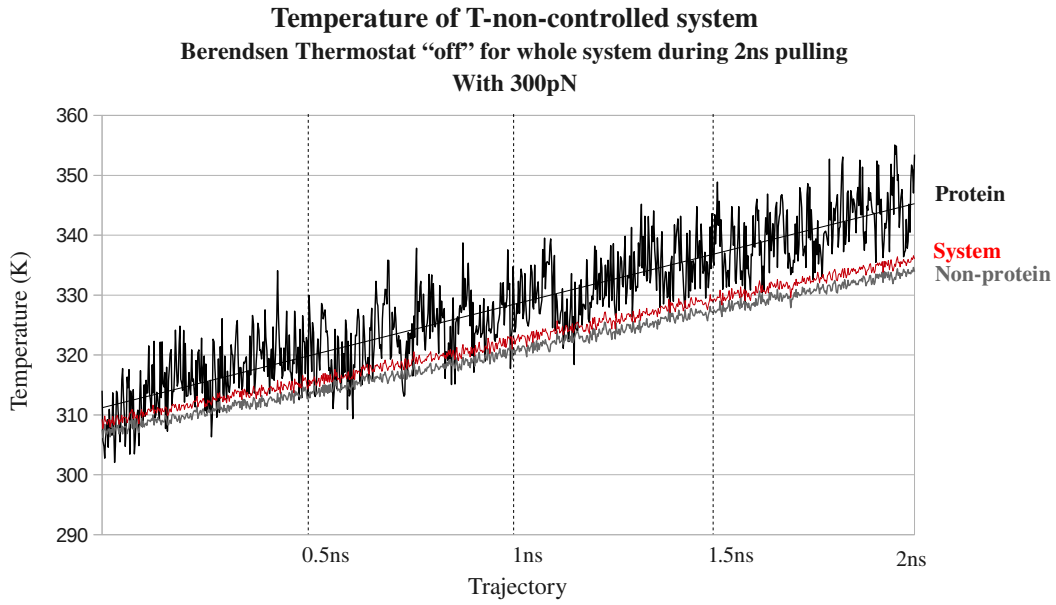


Illustration 17: Temperature evolution over 2ns of model without temperature control. Black line shows increasing linear trend in the protein part.

Viscosity analysis of energy files covering the whole trajectory of the simulation was done with GROMACS g_energy tool. Unfortunately, no reasonable cause for unfolding dynamics differences were concluded from the resulting viscosity analysis. Due to the inconclusiveness of the viscosity analysis, the results were no longer considered in this work.

Step 1 and 2 simulations showed negligible impact of chosen thermostat/barostat type on the unfolding dynamics of I/LWEQ domain. Berendsen control applied to the whole system was considered more beneficial due to lower temperature fluctuation.

One of the reasons for lacking observable differences could have been the high-quality original model of the structure and well energy minimized and equilibrated system.

Based on these observations and assumptions, Berendsen thermostat with Berendsen barostat were selected for the force tests during step 3 of the parameter setting tests. Such selection was made based solely on the results of the performed studies.

7.3 Impact of SMD Force Magnitude; Step 3

Based on the results of the previous simulations, Berendsen thermostat with Berendsen barostat with protein temperature control applied during the equilibration and pulling simulation was selected for the consequent force tests. The temperature control was set to 310K and pressure control to 1bar. The effects of constant pulling force of 200pN, 300pN, 400pN and 500pN applied to the C-terminal end were observed over 9ns simulations. These tests were repeated once. For all force testing simulations, the protein was placed in a larger elongated box (compared to the step1 and step2 tests).

The effect of the pulling force magnitude on the I/LWEQ structure and its length is summarized in Illustration 18. Pulling with 200pN showed very slow unfolding. The protein unfolding was stopped after 4ns of the simulation at final protein length equal to 12.4nm. The structure over the simulation time can be viewed in Illustration 19. The 200pN unfolding was similar to the Nosé-Hoover (thermostat turned off) simulations presented in the previous paragraph. The applied force did not cause sufficient unfolding of the five-helix bundle. As a result, the vinculin binding site stayed buried inside the structure.

Pulling with 300pN force in the elongated box showed unfolding patterns consistent with the tests performed for Berendsen thermostat/Berendsen barostat during step 1 and 2. The final protein length after 9ns was 37nm. The structure of 300pN pulling at 8 - 9ns was similar to structure of 400pN pulling at 3ns, see Illustration 19 and Illustration 20. The VBS was still not completely revealed after 9ns and contact of the VBS interface with H3 and H4 α-helix was found.

The chart curves expressing the protein length of 400pN and 500pN pulling are surprisingly close to each other; almost identical for the initial 3ns and showing similar pattern for the following 6ns, details are available in Illustration 18. The resulting protein length after 9ns for 400pN and 500pN force was approximately 54nm and 60nm respectively. The final length was stable for both pulling forces after 7ns. Even though the protein length over the simulation was comparable for both pulling forces, the unfolding dynamics differed dramatically, see Illustration 20.

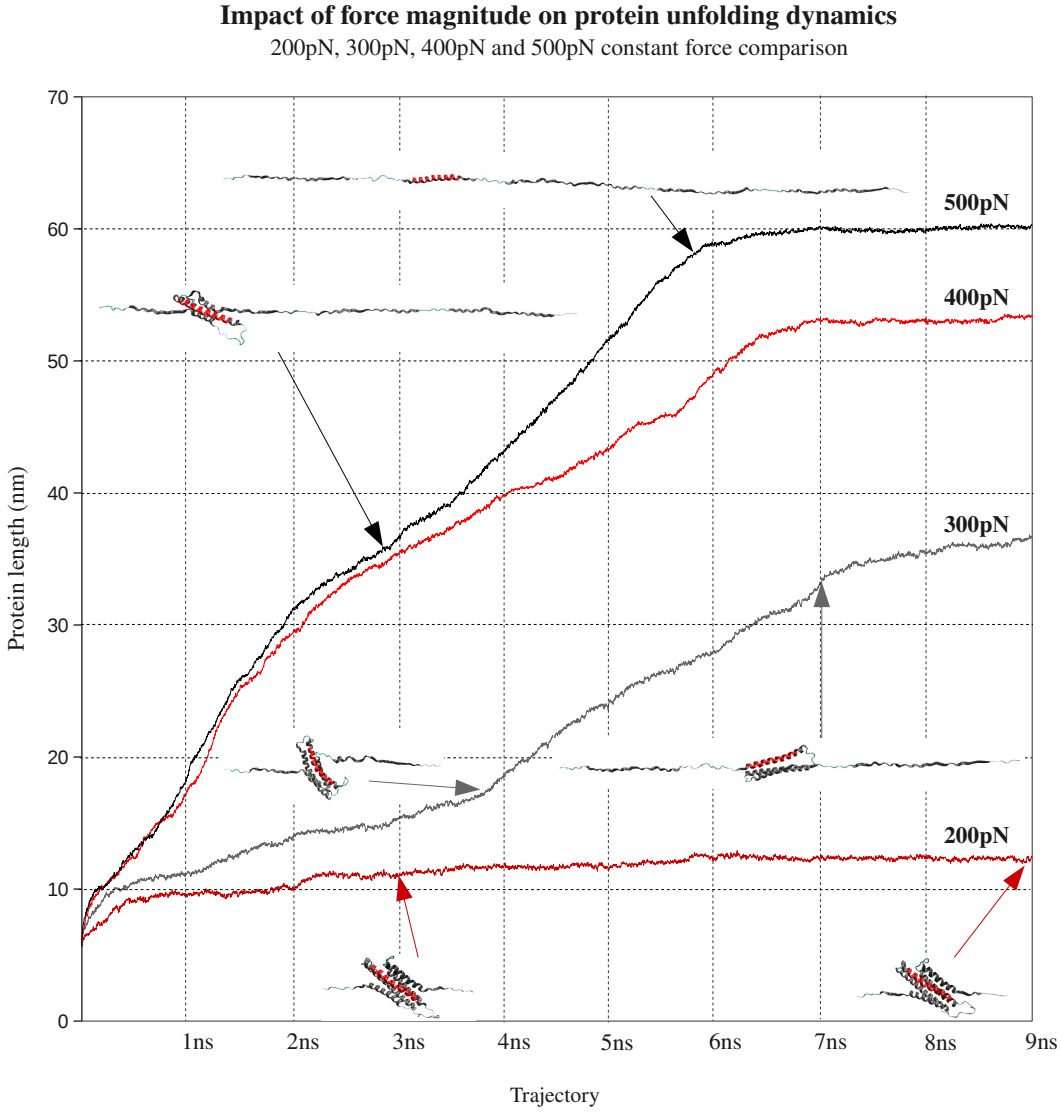


Illustration 18: ***Impact of constant pulling force magnitude on the protein unfolding dynamics over 9ns simulation .****VBS presented as red helix in the shown structures.*

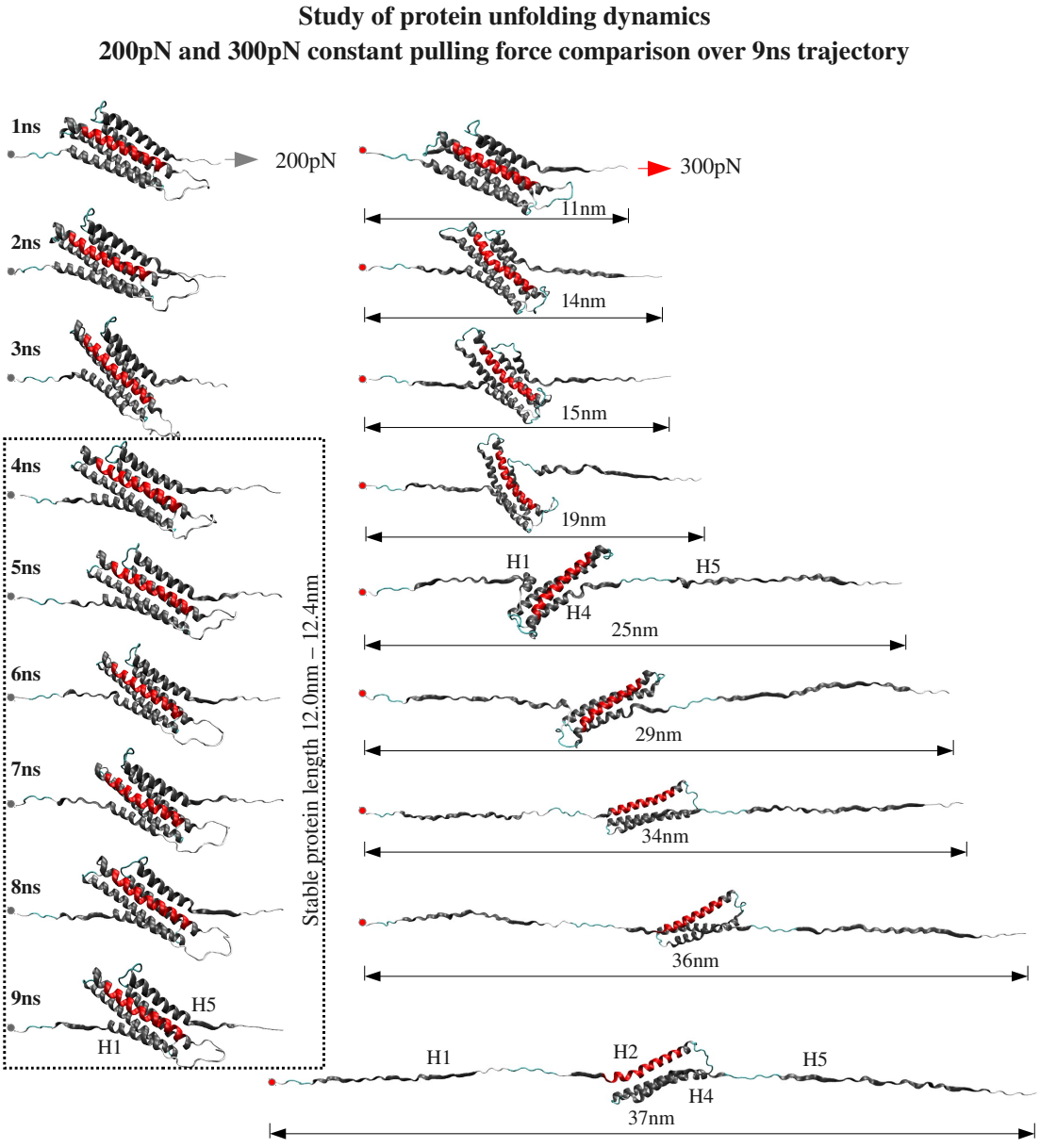


Illustration 19: ***Study of protein unfolding dynamics****; 200pN and 300pN constant pulling force. VBS presented as red helix. Circles represent fixed N-terminal end, arrows show the force application point.*

The protein unfolding when 400pN pulling force was applied occurred simultaneously from the N-terminal fixed end as from the C-terminal pulled end. After 2ns of the simulation, H1 and H5 α-helix were completely unfolded. In comparison, pulling with 500pN showed unfolding mainly at the C-terminal end. After 2ns, H1 α-helix was fairly complete whereas H4

and H5 α-helices were completely pulled apart the original bundle. The C-terminal concentrated pulling continued for 500pN force until 4ns of the simulation, H3 α-helix was completely detached from the rest of the structure, while H1 helix was still not completely pulled. Such one-end pulling resulted in the revealing of VBS intact helix structure in the H2.

From 7ns of the simulation, the protein length remained constant for 400pN and 500pN pulling force. The final protein length of 400pN pulling was 6nm shorter compared to 500pN pulling. This difference was caused by incomplete elongation of H2, H3 and H4 helix structure and slight increase in length with time was still expected.

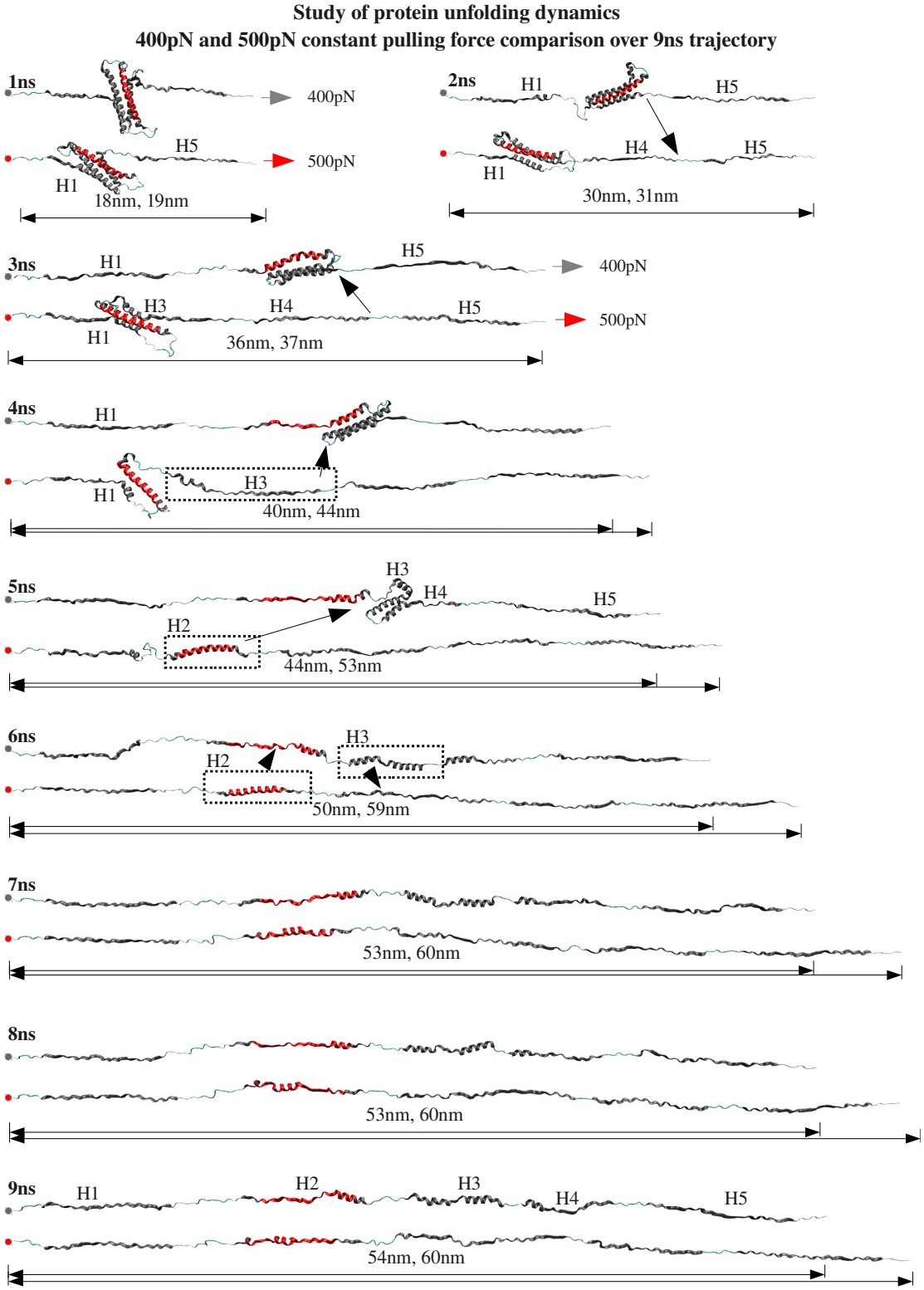


Illustration 20: ***Study of protein unfolding dynamics****; 400pN and 500pN constant pulling force. VBS presented as red helix. Circles represent fixed N-terminal end, red and gray arrows show the force application point. Black arrow high-lights similar position in the protein sequence.*

8 Results: Mutation Design

The following paragraphs describe the results of 200pN and 300pN interaction analysis in order to identify interactions responsible for stability of I/LWEQ five-helix bundle. The methods and principle of the analysis and target interaction selection is in detail described in Experimental setting chapter, paragraph 5.6.

The results of ion bond, hydrogen bond and hydrophobic interactions analysis of 200pN trajectory were compared to the combined results of 300pN trajectory analysis. Special attention was given to inter-helical interactions, which were suspected to hold a significant role in structure stability. Table 3 summarizes the final selection of suspected bonds and mutations designed to destabilize these bonds.

Table 3: ***Mutation Design Summary*** *of mutated residues, the bonding partners of the original residue, the type of the original interaction and reasoning for the residue selection.*

Mutation Design Summary			
Mutation	**Original bonding partner**	**Interaction**	**Reasoning and expected effect**
I2316S *H1*	V2360 (VBS) *H2*	**Hydrophobic interaction**	Interaction is not affected by low force (200pN) pulling. The unfolding was stopped at this point, while with greater force (300pN) the protein unfolded gradually. Mutation with non-hydrophobic residue could destabilize the structure.
E2317A *H1*	K2361(VBS) *H2*	**Ionic bond**	Stable ionic bond in equilibration and in pull. Suspected decreased stability of N-terminal end and VBS containing helix. Hence bundle decomposition and reveal of the vinculin binding site is expected.
D2386A *H3*	T2302 *H1*	**Hydrogen bond** Ionic bond	200pN force shows destruction of hydrogen bond after 4ns period. Mutation is expected to speed up unfolding. Mutated residue also takes part in several ionic bonds.
R2398A *H3*	E2308 *H1*	**Ionic bond** Hydrogen bond	Mutation affecting significant ionic and hydrogen interaction.

Following illustrations (21, 22, 23 and 24) show the studies of the original residues and the designed mutation in the structure. Furthermore, the behavior of the original bond was followed over the 300pN pulling force simulations. The original bond interruption and the time of the observation is presented in the charts below. The behavior of the corresponding bond in the 200pN force pulling simulation was then compared and discussed further in the text for the proper reasoning of the suggested mutation design.

In addition, literature shows that especially residues with side chains pointing towards the core are responsible for protein stability. The protein unfolding study reported here was carried out only on a small fraction of the talin rod domain (one bundle). Therefore it is not possible to identify properly which residues in fact fulfill this condition in the complete protein. In practice, the buried VBS in H2 helix was considered the internal core.

8.1 Mutation I2316S

I2316 – V2360 (Illustration 21) was specified as hydrophobic inter-helix (H1 - H2) interaction stable through 9ns with 200pN pulling force. 300pN pulling force disturbed the interaction after approx. 4.5ns. The bond was suspected to resist the protein unfolding. Since V2360 is located in the vinculin binding site, residue I2316 was proposed a mutation to serine. As an additional reasoning, the direction of the original bond pointing towards the structure's core was considered.

8.2 Mutation E2317A

E2317 – K2361 (Illustration 22) bond is an ionic inter-helix (H1 - H2) interaction, which is stable during first 4ns of 300pN pulling simulation. No large difference in the bond perseverance during the 200pN simulation was observed. However, the residue E2317 proposed for mutation takes part also in stable hydrogen bonding with another residue. Furthermore, the mutation to alanine was suspected to loosen up the bundle's structure close to the N-terminal end and therefore speed up the protein unfolding in the beginning of the force application.

8.3 Mutation D2386A

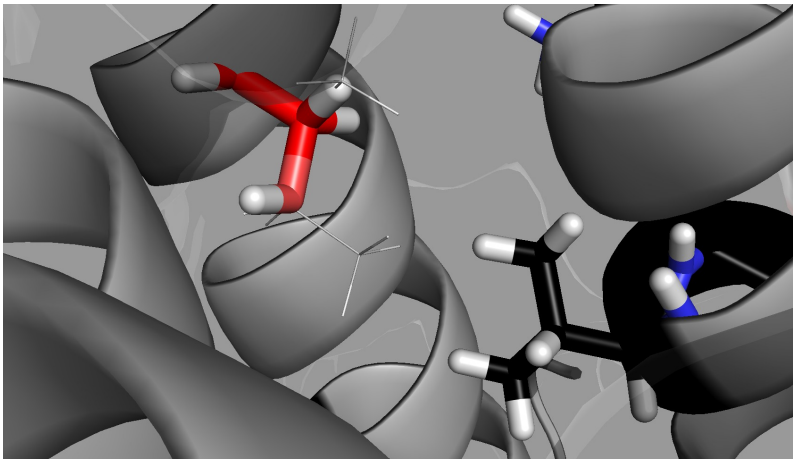
The D2386 – T2302 (Illustration 23) inter-helical (H3 – H1) hydrogen bond was destroyed after 4ns of the 200pN simulation. Even though the bond was not particularly long lasting

compared to the previous cases, a significant difference was found when compared to the 300pN pulling. Furthermore, the D2386 residue takes part in stable hydrogen bonds, and the original bond points towards the core. The suspected impact of the mutation to alanine was faster protein unfolding.

8.4 Mutation R2398A

The R2398 – E2308 inter-helical (H3 – H1) ionic bond was destroyed after 3ns with 300pN force pulling. The interaction was stable for 5ns with 200pN pulling force after which the E2308 side chain turned away as the H1 helix was pulled apart. At the same time, however, hydrogen bond between the same residue pair was formed. Mutation would prevent participation in stable ionic and hydrogen bonds. As can be seen from the structure detail in Illustration 24, the bond was causing distortion of the yet unfolded bundle. Based on the visualization, it is suspected that the interaction strongly opposes the unfolding.

Mutation I2316S and
Unfolding with original bond (I2316 - V2360)
on trajectory with 300pN pulling force



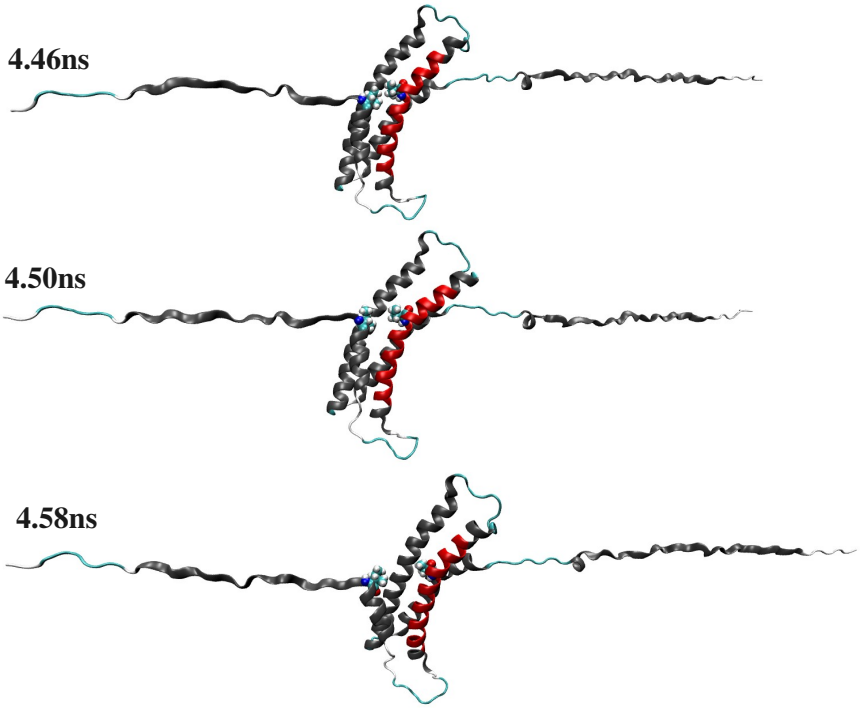


Illustration 21: ***Mutation I2316S and study of bond I2316 - V2360*** *over 9ns 300pN pulling simulation with the visualization of bond interruption. Structure:* ***gray line*** *= original residue;* ***red stick*** *= mutated residue;* ***black stick*** *= original bonding partner.*

Mutation E2317A and
Unfolding with original bond (E2317 - K2361)
on trajectory with 300pN pulling force

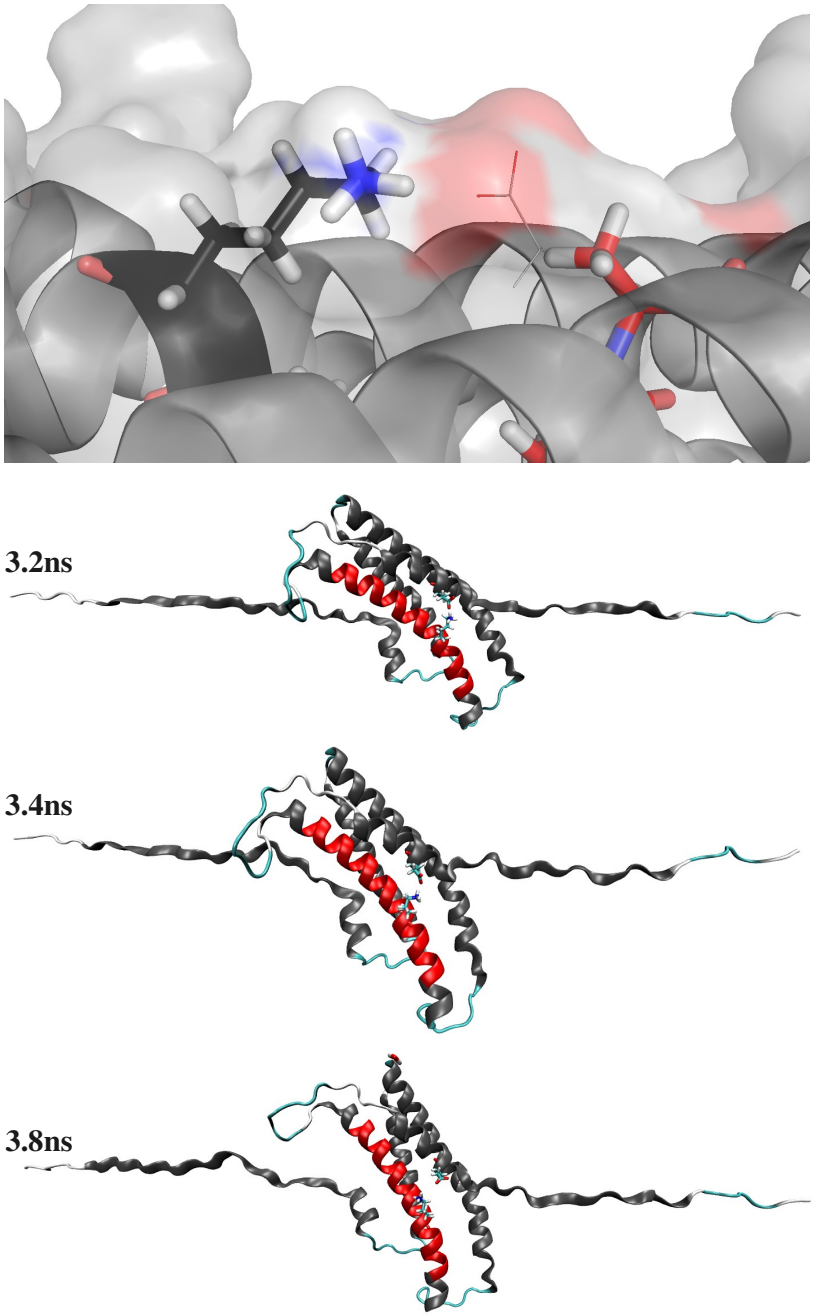
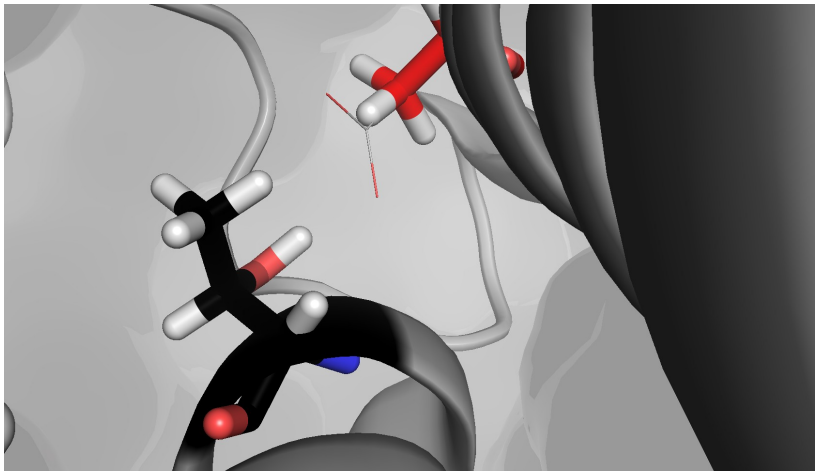


Illustration 22: ***Mutation E2317A and study of bond E2317 - K2361*** *over 9ns 300pN pulling simulation with the visualization of bond interruption. Structure:* ***gray line*** *= original residue;* ***red stick*** *= mutated residue;* ***black stick*** *= original bonding partner.*

Mutation D2386A and
Unfolding with original bond (D2386 - T2302)
on trajectory with 300pN pulling force



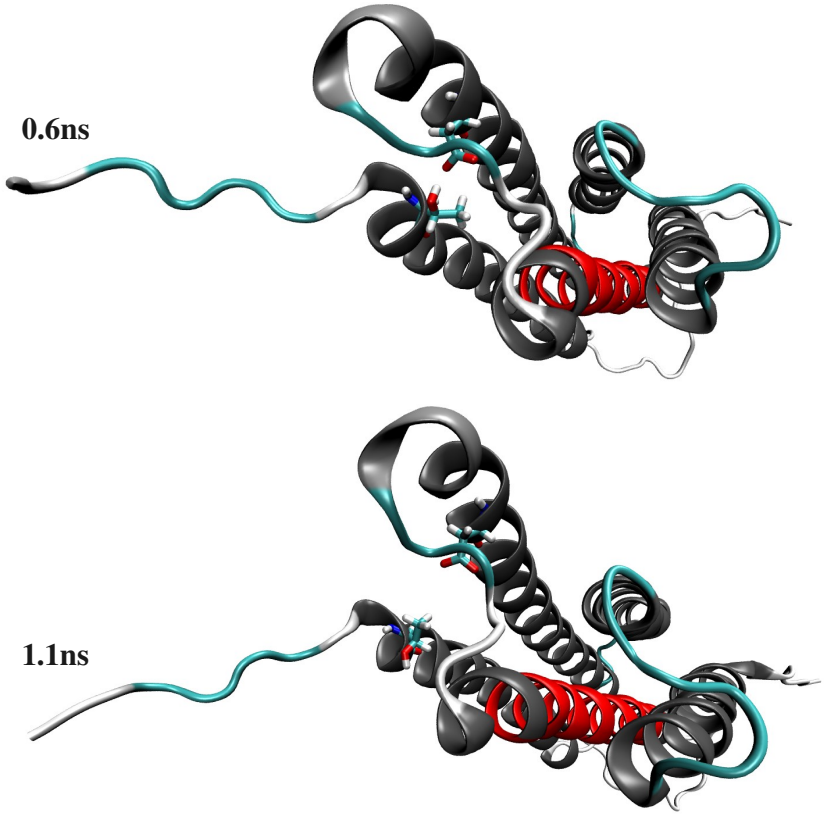


Illustration 23: ***Mutation D2386A and study of bond D2386 - T2302*** *over 9ns 300pN pulling simulation with the visualization of bond interruption. Structure:* ***gray line*** *= original residue;* ***red stick*** *= mutated residue;* ***black stick*** *= original bonding partner.*

Mutation R2398A and
Unfolding with original bond (R2398 - E2308)
on trajectory with 300pN pulling force

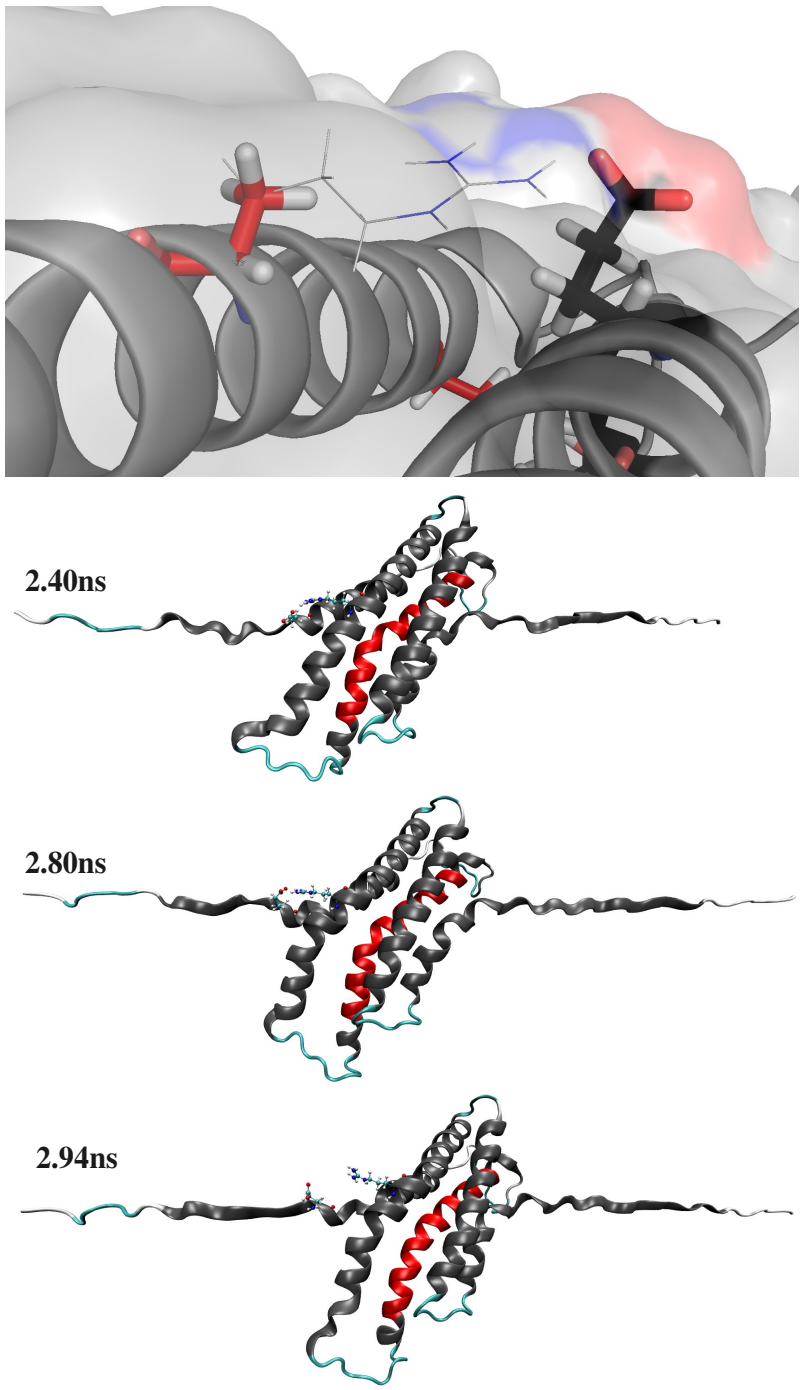


Illustration 24: ***Mutation R2398A and study of bond R2398 - E2308*** *over 9ns 300pN pulling simulation with the visualization of bond interruption. Structure:* ***gray line*** *= original residue;* ***red stick*** *= mutated residue;* ***black stick*** *= original bonding partner.*

9 Discussion and Future Work

In this chapter, the results of the herein presented work are discussed in comparison to available literature data. In addition, proposals for future work are considered and drawn.

During the last two decades, the transduction of mechanical stimuli in cells has been intensively studied. The main focus is on the proteins of the focal adhesion complex, the mechanosensitive structures in the extracellular matrix, and the structures of the cytoskeleton. In other words, the attention is on biomolecules whose structure, function or activation might be modified or induced by external force. The intensively explored mechanoactivation of FA proteins, namely talin and vinculin, provide several theoretical mechanisms of activation (Lee et al., 2007, Goult et al., 2013(acc.man.), Vogel, 2006, Golgi and Mofrad, 2010, Diez et al., 2011). However, these hypotheses have been mainly formed based on the observations on an incomplete talin structure which may have caused misinterpretations of the observed behavior.

The recent development in technology provides computational modeling and simulation methods applicable in the research of biomolecules in single-molecule experiments. Such application allows observations of molecular behavior on atomic-level, which is inaccessible by classical experimental methods (Hünenberger, 2005, Vendrusculo and Paci, 2003).

Even though MD and SMD methods are frequently used tools, the parameter setup during the simulation production (and its possible impact on the observed molecular behavior) is not given sufficient attention. This statement was deduced based on the poor documentation of MD/SMD simulation setting in the literature. To elaborate on this topic further, several reports reviewed in chapter 3 were missing crucial information on the temperature or pressure controls and their implementation, type of used force field or simply the number of repetitions produced (Zhmurov et al., 2011, Gao et al., 2002, Johnson et al., 2007). As was observed in this work (Illustration 11 and Illustration 12, pp. 45 and 46), repetitions with same simulation setting may produce significantly differing unfolding trajectories which may be easily misinterpreted. It is also suspected that different simulation setting may produce significant differences in the protein activation or unfolding dynamics (Illustration 11 (right) and Illustration 14 (right) for Nosé-Hoover thermostat).

Since reported studies implementing SMD methods show great diversity in the parameter and algorithm setting as well as varying detail of provided setting information, it is often impossible to reproduce the simulation or compare the observations. The MD and SMD settings are often case, target and molecule type specific, and hence simulation "standard setting" is not applicable. The question is, whether there should be a standard created for reporting the details of the setting to inform specifically on all details of the simulation produced and make result comparison possible.

9.1 Talin

9.1.1 Variation Analysis

The analysis of talin protein sequence data showed high conservation of the protein, which is in agreement with the published observations (Nelson and Cox, 2008). The alignment of 25 talin orthologs, presented in the Illustration 5 on page 37, shows very little variation in the IL/WEQ talin domain. Furthermore, no variation was found in the VBS among the studied species. The overall low number of single nucleotide polymorphism leading to amino acid mutation (in NCBI dbSNP and Illustration 6, p.38) and no available clinical relevance suggest that the abundant talin protein is highly significant in cell viability and vitality. Indeed, experimental investigation with talin gene knockout cells showed severe malfunctions during embryonic morphological development and spreading, leading to apoptosis and perinatal death (Franco-Cea et al., 2010, Kopp et al., 2010, Conti et al., 2009).

9.1.2 Structural Analysis

The structure of talin's VBS in the literature and the structural analysis presented in this work assign the VBS to the second helix of the I/LWEQ domain facing towards the core of the structure. For talin to be activated, a major conformational change in the structure is necessary to reveal the buried hydrophobic residues of VBS.

The extensive work of A.R. Gingras and co-workers on the proteins of FA complex speculates that talin's VBS has to be completely revealed from the talin's rod structure in order to attract and bind vinculin. The structure of I/LWEQ helix bundle studied here (Illustration 7 on page 39), displays a slight "twist" of the non-VBS helices around the VBS helix. In other words, the twisted helices prolong the VBS revealing when unfolding from N- and/or C-terminal end. Furthermore, the talin – vinculin structural alignment according to the VBS in talin and in

vinculin showed overlapping talin non-VBS helices with vinculin helices (Illustration 9 and Illustration 8, page 42). The hydrophobic/hydrophilic analysis of talin (Illustration 10 page 43) and vinculin uncovered similar organization of the polar and non-polar residues in talin and vinculin helices. Combination of these observations resulted to similar assumption of talin activation as presented by work of A.R.Gingras et al., that the talin structure needs to be completely unfolded to reveal the talin's VBS and bind vinculin.

Work by Lee et al. (2007), on the other hand, reported rotational movement of the VBS helix leading to talin activation. Such helix rotation, limited to turning the VBS helix, was not observed in any of the simulations reported here. It was speculated that the rotational movement could have been caused by the setting of the system to implicit water model and by using a slow unfolding speed. It has to be stressed, though, that the structure used in the simulations described in this study was rather small and pulling forces large. Hence, any turning or twisting movement of VBS would have been overlooked due to fast unfolding and missing comparison structure.

Since the mechanism of vinculin binding to talin and talin's VBS revealing have been so far studied only on talin's rod segments, several proposals for future talin activation studies were made. For instance, placing the vinculin molecule to the proximity of talin in one simulation system, while applying a small unfolding force to talin's N- and C- terminal helices, might uncover important mechanisms of the activation. Also following the energy levels in thermodynamic analysis in such a system could provide closer insight into vinculin binding to talin. Furthermore, until the complete talin structure is known, talin behavior under force and the VBS revealing mechanism stays unclear. Hence, another proposal for the continuation of the research is the modeling of talin's complete head-neck-rod structure, a talin dimer model and a talin-membrane model.

9.2 Energy Minimization and Equilibration Effect

Among the simulations presented in this work, surprisingly small differences in protein unfolding were observed for different simulation setup. It was expected that the weak Berendsen coupling would produce visibly different trajectories compared to the extended Nosé-Hoover system. The expected, however, did not happen over the two repetitions.

One of the presumed reasons for the observed low impact of chosen temperature coupling on the protein unfolding dynamics was the high quality of the modeled structure (coordinate file of the structure from NMR solution by Gingras et al., 2008).

In this connection, also the impact of energy minimization and equilibration of the system was discussed. It was speculated, that well minimized and equilibrated system is not probably affected significantly by the temperature coupling algorithm. The system was followed over 9ns period; the question still remains, whether changes appear during extended simulation time.

Furthermore, the error of the Berendsen algorithm causing an incorrect canonical ensemble is minimized with the increased size of the system (according to 1/N) (van der Spoel et al., 2012). The systems modeled here were fairly large; approx. 90,000 and 400,000 atoms, which probably resulted in the error minimization of the algorithm and an insignificant effect on the unfolding trajectory.

9.3 Temperature and Pressure Coupling Algorithms

As mentioned in the literature review, there is a gap in current understanding of the choice and implementation of temperature and pressure coupling algorithms during SMD simulations.

The literature (van der Spoel et al., 2012) proposes the use of extended Nosé-Hoover temperature coupling or other (v-rescale) thermostat rather than Berendsen temperature control during the classical MD simulation to ensure canonical ensemble. During the SMD simulation, external pulling force is introduced which, in principle, results in a non-canonical system. The implementation of the temperature coupling algorithm in SMD is yet not extensively discussed in the literature.

9.3.1 Thermostat/Barostat combination

The results of the performed experiments show that the temperature control in pulling simulation does not result in significant differences in unfolding trajectories when all parts of the system are controlled. Hence, the efficient Berendsen thermostat combined with the Berendsen barostat is sufficient for the SMD pulling simulation. Furthermore, the extended Nosé-Hoover thermostat combined with the Parrinello-Rahman barostat (all system parts controlled) can be used.

In the study, only two most commonly used temperature coupling algorithms were tested. Future work should concentrate on investigation of other available temperature algorithms, such as v-rescale, to provide wider overview and understanding of possible thermostat induced effects.

9.3.2 Temperature Coupling Groups

Despite insignificant temperature differences between protein controlled and non-controlled simulations in Berendsen coupled system, all parts of the system should be applied temperature control. The temperature control should be preferably coupled to separate temperature control groups to avoid "hot solvent – cold solute" problem.

The "hot solvent – cold solute" problem could have been the reason for insufficient unfolding during the Nosé-Hoover temperature coupling when T-control for protein part was switched off. Even though insignificant temperature difference was observed compared to Berendsen thermostat simulation, the friction parameter of extended Nosé-Hoover thermostat could have been affected leading to opposition towards the pulling direction and hence slow unfolding.

9.3.3 Viscosity analysis

Yet another cause for the slow unfolding in the above discussed setting, could have been a change in the viscosity of the solute. Unfortunately, the viscosity analysis tool (g_energy viscosity extraction from energy files) available in GROMACS did not provide useful results. The supposed reason of unreasonable results, was the viscosity tool implementation itself. Each simulation reported here was produced in up to 11 consequent runs by restarting the simulation production from the last output step file. This resulted into corresponding number of energy files. The viscosity values covering the whole simulation period were extracted from several energy files, which could produce incorrect value scaling and faulty error estimation and therefore unreliable results (van der Spoel et al., 2012).

In future experiments, a different viscosity analysis approach should be used in order to closer uncover the functioning of Nosé-Hoover thermostat. GROMACS provides another tool for the viscosity calculation; i.e. g_tcaf. The implementation of the tool needs to be planned before the simulations are produced since the model and run input files have to be constructed to match the algorithm.

9.3.4 Repetition Runs

The thermostat setting algorithms in SMD simulations were tested in two repetitions in this study, which proved insufficient. As discussed before, the difference in the setting did not provide grounds for a clear separation line between the implementations to closer specify or even predict the effects of the thermostat/barostat setting on the protein unfolding dynamics. For the continuation of the study, additional simulation production with the same setup is necessary to produce more conclusive thermostat/barostat comparison.

9.4 Pulling Force Magnitude

The force magnitudes tested in this study were 200pN, 300pN, 400pN and 500pN applied to C-terminal CA. The N-terminal CA was fixed in all performed simulations. The protein applied 200pN (with Berendsen temperature and pressure control) unfolded reluctantly. The protein elongation stabilized after 4ns of the simulation at approx. 12nm of total protein length. It was concluded that there are stable interactions in the talin I/LWEQ domain opposing the unfolding force.

Force of 300pN and greater unfolded the protein gradually in a step-like protein elongation which is in agreement with observations of Hytönen and Vogel (2008). The protein unfolded faster during turn/coil (TC) unfolding, compared to the elongation of helix unfolding. This observation suggests that inter- and intra-helical interactions account significantly for protein stability. Hence, these interactions should be preferably targeted in the destabilizing mutation design.

Interesting were the observations of the unfolding dynamics with 400pN and 500pN forces. Simulation with the 500pN force magnitude unfolded completely H1 helix from the N-terminal end of the I/LWEQ bundle, as well as H3, H4 and H5 from the C-terminal end, leaving only the H2 containing VBS completely revealed yet intact. The H2 helix remained unaffected for 6ns of the simulation. The unfolding dynamics of these two forces exhibited similar elongation over the simulation time but different dynamics. While the 500pN force was C-terminal end oriented, the 400pN unfolding was more equally divided through the structure. This observation might point at an important activation mechanism induced by high unfolding force. Since there was only one repetition of the force magnitude simulation tests, it is necessary to study the high force unfolding dynamics further in order to define significant connections.

Also molecular behavior at significantly lower force (under 200pN) or at 250pN is unknown and could be a subject of further investigation. For instance, the turning movement of VBS (Lee et al., 2007) discussed earlier, was observed at approx. 90pN during the initial 2ns of constant velocity pulling. However, this study was performed in implicit water, which is known to accelerate the protein unfolding compared to explicit water model under external force. Such slow unfolding in explicit water model is probably caused by limited penetration of explicit water to force-bearing bonds.

9.5 Force Application Points and Direction

The constant pulling force in this study was applied to the C-terminal CA atom while the N-terminal CA atom was fixed. Since the pulling was performed only from one end of the protein structure, the analysis of D1 and D1 segment elongation and the D1/D2 force penetration study (Illustration 13, p. 47) is only partial. At this stage of partial completeness, the analysis presents mainly the unfolding dynamics of the protein. For complete assessment of the force penetration, the pulling simulations should be performed under same conditions while the pulling and fixed ends are switched, or both ends are pulled at the same time.

Even though force application to only one atom was sufficient for the purposes of this work, it probably results in unnatural molecular behavior (Kumar and Li, 2010). The study should be extended to pulling whole helices apart at the N- and C-terminal ends.

The direction of the force vector is yet another factor which might hold an impact on the protein unfolding and activation mechanism as stated in review by Kumar and Li (2010). For instance, fixing the helix containing VBS and pulling the helix placed opposite to it (H3 or H4 in the case presented in this study) would most likely lead to structure destabilization without dramatic protein unfolding.

9.6 Mutation Design

The pulling simulations produced during temperature coupling and force magnitude testing were analyzed for ionic, hydrogen and hydrophobic interactions. Results of the analysis of slow unfolding with 200pN and gradual unfolding with 300pN were compared and stable interactions identified. Also the location of the residues taking part in the interactions were given attention. According to the review by Kumar and Li (2010), the residues (and side-chains) pointing towards the core of the structure have greater impact on the protein stability.

Based on these results, four destabilizing mutations were proposed; I2316S, E2317A, D2386A and R2398A. In this study, only the mutation design is available. Future work should proceed with computational mutagenesis and studies of the mutants in *in silico* stability tests and unfolding simulations.

Even though only four mutants are published here, several other mutation candidates causing destabilization can be found, offering yet another direction in the research and continuation of this work in the future.

The mutations were further studied *in vitro* in the Protein Dynamics group, Institute of Biomedical Technology at the University of Tampere, Finland. Providing the simulated mutant study with comparison to the results of *in vitro* experiments would provide important link between atomic-level observations and behavior in physiological system widening our understanding of talin's VBS activation mechanisms.

10 Conclusion

The proteins of focal adhesion complex, namely talin and vinculin, have been intensively studied during the last two decades. Despite the attention, the mechanisms of activation and binding dynamics is yet not fully understood. To broaden our understanding of detailed protein functioning, *in silico* single-molecule molecular dynamics and steered molecular dynamics methods are applied. MD and SMD methods are useful tool for atomic-level investigations, however setting and choice of parameters might significantly influence the observed dynamics resulting in misleading or even incorrect conclusions. For instance, the thermodynamic boundaries of the system might result in incorrect ensemble far from physical experimental conditions. While the temperature and pressure controls to ensure canonical ensemble in classical MD simulations are recommended in the literature, the "correct" setting for SMD has not been intensively discussed.

In this work, two commonly used temperature coupling algorithms were studied; Berendsen and Nosé-Hoover thermostat. The impact of the thermostat and its implementation was followed on the unfolding dynamics of talin I/LWEQ domain and assessed based in the protein elongation over the simulation period.

Based on the observations it was concluded that the choice of thermostat does not have significant impact on the unfolding dynamics when all parts of the system are coupled to own temperature control. The implementation of the thermostat may, on the other hand, affect the unfolding. Hence, the temperature of all system parts needs to be controlled.

Force magnitude affects the speed of unfolding. The force of 200pN magnitude did not unfold the protein significantly. Gradual unfolding was observed with 300pN constant force simulation. The force magnitude may also affect the unfolding dynamics and the force penetration. Unfolding with the largest tested force of 500pN was C-terminal oriented, while protein stretching with 300pN and 400pN was comparable from both N- and C- terminal ends. It was also concluded, that several repetitions of simulations is necessary to provide sufficient understanding of the specific unfolding.

Based on the results of hydrogen, ionic and hydrophobic interaction analysis performed on trajectories of 200pN and 300pN pulling, four destabilizing mutations in the talin I/LWEQ domain were designed; I2316S, E2317A, D2386A and R2398A.

The presented work provides an initial insight to the temperature algorithm use and implementation in SMD simulations. Continuation of the thermostat effects investigations proposed were:

1) testing current settings in increased number of repetitions
2) testing other available temperature coupling algorithms (v-rescale)
3) testing the impact of the model quality, energy minimization and equilibration by employment of different starting structure/model
4) follow viscosity changes

The unfolding dynamics should be further studied by:

5) inter-changing pulling and fixed points
6) pulling from both N- and C-terminal ends
7) force application to whole helices
8) studying pulling in different vector orientation/direction

And finally, to broaden our understanding of talin's VBS activation following investigations were proposed:

9) modeling whole talin structure, talin dimer and talin - membrane system
10) follow energy and thermodynamic changes in talin's activation and vinculin binding in talin – vinclulin system
11) follow mutant stability and unfolding dynamics of designed mutations

11 Bibliography

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2008. *Molecular Biology of the Cell, 5th ed.*. Garland Science. ISBN: 978-0-8153-4106-2.

Bate, N., Gingras, A.R., Bachir, A., Horwitz, R., Ye, F., Patel, B., Goult, B.T., Critchley, D.R. 2012. *Talin Contains A C-Terminal Calpain2 Cleavage Site Important In Focal Adhesion Dynamics.* PloS ONE 7(4): e 34461.

Berendsen, H. J. C., Postma, J. P. M., DiNola, A., Haak, J. R. 1984. *Molecular dynamics with coupling to an external bath*. J. Chem. Phys. 81:3684–3690

Beggs, A.H., Byers, T.J., Knoll, J.H., Boyce, F.M., Bruns, G.A., Kunkel, L.M. 1992. *Cloning and characterization of two human skeletal muscle alpha-actinin genes located on chromosomes 1 and 11.* J Biol Chem. 267(13): 9281-8.

Best, R.B., Fowler, S.B., Herrera, J.L.T., Steward, A., Paci, E., Clarke, J. 2003. *Mechanical unfolding of a Titin Ig Domain: Structure of Transition State Revealed by Combining Atomic Force Microscopy, Protein Engineering and Molecular Dynamics Simulations*. J. Mol. Biol. 330, 967-877.

Brown, N.H., Gregory, S.L., Rickoll, W.L., Fessler, L.I., Prout, M., White, R.A.H., Fristrom, J. W. 2002. *Talin is essential for integrin function in Drosophila*. Dev. Cell 3, 569–579.

Brown, C., Morham, S.G., Walsh, D., Naghavi, M.H. 2011. *Focal Adhesion Proteins Talin-1 and Vinculin Negatively Affect Paxillin Phosphorylation and Limit Retroviral Infection*. Journal of Molecular Biology 410, 761-777.

Buehler, M. J., Yung, Y. C.. 2009. *Deformation and failure of protein materials in physiologically extreme conditions and disease*. Nat. Mater. 8:175–188.

Bussi, G., Donadio, D., Parrinello, M. 2007. *Canonical sampling through velocity rescaling*. The Journal of Chemical Physics, 126, 014101, 1-7.

Bökel, C., Brown, N.H., 2002. *Integrins in development: moving on, responding to, and sticking to the extracellular matrix*. Dev. Cell 3, 311–321

Calderwood, D.A., Zent, R., Grant, R., Rees, D.J., Hynes, R.O., Ginsberg, M.H., 1999. *The Talin head domain binds to integrin beta subunit cytoplasmic tails and regulates integrin activation.* J. Biol. Chem. 274, 28071–28074.

Calderwood, D.A. 2004. *Integrin activation*. J. Cell Sci. 117, 657– 666.

Carisey, A., Tsang, R., Greiner, A.M., Nijenhuis, N., Heath, N., Nazgiewicz, A., Kemkemer, R., et al. 2013. *Vinculin regulates the recruitment and release of core focal adhesion proteins ina force-dependent manner.* Current biology: CB, 23(4), 271-81.

Carrion-Vasquez, M., Oberhauser, A.F., Fisher, T.E., Marszalek, P.E., Li, H., Fernandez, J.M. *Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering.* Prog Biophys Mol Biol. 74(1-2):63-91.

Conti, F.J., Monkley, S.J., Wood, M.R., Critchley, D.R., Muller, U., 2009. *Talin 1 and 2 are required for myoblast fusion, sarcomere assembly and the maintenance of myotendinous junctions.* Development 136, 3597–3606.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz Jr., K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A. 1995. *A second generation force field for the*

simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society 117, 5179–5197.

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

Critchley, D.R. 2000. *Focal adhesions – the cytoskeletal connections.* Current Opinion in Cell Biology, 12:133-139.

Critchley, D.R. 2009. *Biochemical and structural properties of the integrin- associated cytoskeletal protein talin*. Annu. Rev. Biophys. 38, 235–254.

del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J.M., Sheetz, M.P. 2009. *Stretching single talin rod molecules activates vinculin binding. Science*. 323:638–641.

Diez, G., Auernheimer, V., Fabry, B., Goldmann, W.H. 2011. *Head/Tail interaction of vinculin influences cell mechanical behavior.* Biochemical and Biophysical Research Communicatiions, 406, 85-88.

Eastwood, M. P., Stafford, K. A., Lippert, R. A., Jensen, M. O., Maragakis, P., Predescu, C., Dror, R. O., Shaw, D. E. 2010. *Equipartition and the calculation of temperature in biomolecular simulations*. J. Chem. Theory Comp.

Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J.P., Bowie, J.U. 2004. *Side-chain Contribution to Membrane Protein Structure and Stability*. J. Mol. Biol. 335, 297-305.

Fillingham, I., Gingras, A.R., Papagrigoriou, E., Patel, B., Emsley, J., Critchley, D.R., Roberts, G.C.K., Barsukov, I.L. 2005. *A Vinculin Binding Domain from the Talin Rod Unfolds to Form a Complex with the Vinculin Head*. Structure, Vol. 13, 65-74.

Floudas, C.A., Fung, H.K., McAllister, S.R., Mönnigmann, M., Rajgaria, R. 2006. *Advances in protein structure prediction and de novo protein design: A review.* Chemical Engineering Science 61, 966-988.

Franco-Cea, A., Ellis, S., Fairchild et al. 2010. *Distinct developmental roles for direct and indirect talin-mediated linkage to actin.* Developmental Biology, 345(1), 64-77.

Galera-Prat, A., Gómez-Sicilia, A., Oberhauser, A.O., Ciepak, M., Carrión-Vásquez, M. 2010. *Understanding biology by stretching proteins:recent progress*. Current Opinion in Structural Biology, 20:63-69.

Gallant, N. D., Michael K. E. , Garcıía A.J. 2005. *Cell adhesion strengthening: contributions of adhesive area, integrin binding, and focal adhesion assembly*. Mol. Biol. Cell. 16:4329–4340.

Gao, M., Craig, D., Vogel, V., Schulten, K., 2002. *Identifying unfolding intermediates of FN-III(10) by steered molecular dynamics*. Journal of Molecular Biology 323 (5), 939–950

Geiger, B., Bershadsky, A., Pankov, R., Yamada, K.M. 2001. *Transmembrane crosstalk between the extracellular matrix and the cytoskeleton*. Nat Rev Mol Cell Biol 2, 793-805.

Geiger, B., Spatz, J.P., Bershadsky, A.D. 2009. *Environmetnal sensing through focal adhesions*. Nat Rev Mol Cell Biol 10, 21-33.

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

Gingras, A.R., Vogel, K.P., Steinhoff, H.J., Ziegler, W.H., Patel, B., Emsley, J., Critchley, D.R., Roberts, G.C., Barsukov, I.L., 2006. *Structural and dynamic characterization of a vinculin binding site in the talin rod*. Biochemistry 45 (6), 1805–1817.

Gingras, A.R., Bate, N., Goult, B.T., Hazelwood, L., Canestrelli, I., Grossmann, J.G., Liu, H., Putz, N.S., Roberts, G.C., Volkmann, N., Hanein, D., Barsukov, I.L., Critchley, D.R., 2008. *The structure of the C-terminal actin-binding domain of talin*. EMBO J. 27, 458–469.

Gingras, A.R., Ziegler, W.H., Bobkov, A.A., Joyce, M.G., Fasci, D., Himmel, M., Rothe- mund, S., Ritter, A., Grossmann, J.G., Patel, B., Bate, N., Goult, B.T., Emsley, J., Barsukov, I.L., Roberts, G.C., Liddington, R.C., Ginsberg, M.H., Critchley, D.R., 2009. *Structural determinants of integrin binding to the talin rod*. J. Biol. Chem. 284, 8866–8876.

Goel, M., Sienkiewicz, A.E., Picciani, R., Wang, J., Lee, R.K., Bhattacharya, S.K. 2012. *Cochlin, Intraocular Pressure Regulation and Mechanosensing.* PloS ONE 7(4): e34309.

Golji, J., Mofrad, M.R.K. 2010. *A Molecular Dynamics Investigation of Vinculin Activation*. Biophysical Journal, Vol. 99, 1073-1081.

Golji, J., Lam, J., Mofrad, M.R.K. 2011. *Vinculin is Necessary for Complete Talin Binding*. Biophysical Journal, Vol 100, 332-340.

Goult, B.T., Gingras A.R., Bate, N., Barsukov, I.L., Critchley, D.R., Roberts, G.C.K. 2010. *The domain structure of talin: Residues 1815-1973 from a five-helix bindle containing a cryptic vinculin-binding site*. FEBS Letters, 584, 2237-2241.

Goult, B.T., Zacharchenko, T., Bate, N., Tsang, R., Hey, F., Gingras, A.R., Elliott, P.R., Roberts, G.C.K., Ballestrem, C., Critchley, D.R., Barsukov, I.L. 2013. *RIAM and Vinculin Binding to Talin Are Mutually Exclusive and Regulate Adhesion Assembly and Turnover.* The Journal of Biological Chemistry, 288(12), 8238-8249.

Goult, B.T., Xu, X.-P., Gingras, A.R., Swift, M., Patel, B., Bate, N., Kopp, P.M., Barsukov, I.G., Critchley, D.R., Volkmann, N., Hanein, D. 2013(acc.man.). *Structural studies on full-length talin1 reveal a compact auto-inhibited dimer:Implication for talin activation.* Journal of Structural Biology. Accepted Manusctript.

Gräter, F., Shen, J., Jiang, H., Gautel, M., Grubmüller, H. 2005. *Mechanically Induced Titin Kinase Activation Studied by Force-Probe Molecular Dynamics Simulations*. Biophysical Journal, Vol. 88, 790-804.

Hoover, W. G. 1985. *Canonical dynamics: equilibrium phase-space distributions*. Phys. Rev. A 31:1695–1697.

Hytönen, V.P., Vogel, V. 2008. *How Force Might Activate Talin's Vinculöin Binding Sites: SMD Reveals a Structural Mechanism*. PloS Computational Biology, Vol(4), Issue 2, e24.

Hüneneberger, P.H. 2005. *Thermostat Algorithms for Molecular Dynamics Simulations.* Adv. Polym. Sci. 173: 105-149.

Isralewitz, B., Gao, M., Schulten, K. 2001. *Steered molecular dynamics and mechanical functions of proteins.* Current Opinion in Structural Biology, 11:224-230.

Johnson, C.P., Gaetani, M., Ortiz, V., Bhasin, n., Harper, N., Gallagher, P.G., Speicher, D.W., Discher, D.E. 2007. *Pathogenic proline mutation in the linker between spectrin repeats: disease caused by spectrin unfolding*. Blood, Vol. 109, No. 8, 3538-3543.

Johnson, C.P., Tang, H.-Y., Carag, C., Speicher, D.W., Discher, D.E. 2007a. *Forced Unfolding of Proteins Within Cells*. Science 317, 663-666.

Kaufmann, S., Käs, J., Goldmann, W.H., Sackmann, E., Isenberg, G. 1992. *Talin anchors and nucleates actin filaments at lipid membranes. A direct demonstration.* FEBS Letters, 314(2), 203-5.

Klein-Nulend, J., Bakker, A.D., Bacabac, R.G., Vatsa, A., Weinbaum, S. 2013. *Mechanosensation and transduction in osteocytes.* Bone 54, 182-190.

Kopp, P.M., Bate, N., Hansen, T.M., Brindle, N.P.J, Praekelt, U., Debrand, E., Coleman, S., Mazzeo, D., Goult, B.T., Gingras, A.R., Pritchard, C.A., Critchley, D.R., Monkley, S. 2010. *Studies on the morphology and spreading of human endothelial cells define key inter- and intramolecular interactions for talin1.* European Journal of Cell Biology 89, 661-673.

Krammer A, Lu H, Isralewitz B, Schulten K, Vogel V. 1999. *Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch*. Proc Natl Acad Sci USA 1999, 96:1351-1356.

Kumar, S., Weaver, V.M. 2009. *Mechanics, malignancy, and metastasis: the force journey of a tumor cell*. Cancer Metast Rev;28:113–27.

Kumar, S., Li, M.S. 2010. *Biomolecules under mechanical pressure.* Phusics Reports, Vol. 886, Issue 1-2, 1-74.

Lee, H-S., Bellin, R.M., Walker, D.L., Patel, B., Powers, P., Liu, H., Garcia-Alvarez, B., de Pereda, J.M., Liddington, R.C., Volkmann, N., Hanein, D., Critchley, D.R., Robson, R.M. 2004. *Characterization of an Actin-binding Site within the Talin FERM Domain.* J. Mol. Biol. 343, 771-784.

Lee, S.E., Kamm, R., Mofrad, M. 2007. *Force-induced activation of talin and its possible role in focal adhesion mechanotransductuon*. Journal of biomechanics, 40(9), 2096-2106.

Lee, S.E, Chunsrivirot, S., Kamm, R.D., Mofrad, M.R.K. 2008. *Molecular Dynamics Study of Talin-Vinculin Binding.* Biophysical Journal, Vol. 95, 2027-2036.

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M.P., Dror, R.O., Shaw, D.E. 2012. *Systematic Validation of Protein Force Fields against Experimental Data.* PloS ONE 7(2): e32131.

Liu, M., Oberg, K., Zhou, Y. 2007. *Expression and function of vinculin in neuroendocrine tumors*. Tumour Biol. 28:196–204.

Lu, H., Isralewitz, B., Krammer, A., Vogel, V., Schulten, K. 1998. *Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation.* Biophysical Journal, Vol 75, 662-671.

Ma L, Xu M, Forman JR, Clarke J, Oberhauser AF. 2009. *Naturally occurring mutations alter the stability of polycystin-1 polycystic kidney disease (PKD) domains*. J Biol Chem, 284:32942-32949.

MacKerell Jr., A.D. et al. 1998. *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B 102, 3586–3616.

McLachlan, A. D., Stewar,t M., Hynes, R. O., Rees D. J. 1994. *Analysis of repeated motifs in the talin rod*. J. Mol. Biol. 235, 1278–1290.

Margadant, F., Chew, L.L., Hy, X., Yu, H., Bate, N., Zhang, X., Sheets, M. 2011. *Mechanotransduction In Vivo by Repeated Talin Stretch-Relaxation Events Depends upon Vinculin.* PloS Biol 9(12):

Mierke, C.T. Kollmannsberger, P., Paranhos-Zitterbart, D., Smith, J., Farby, B., Goldmann, W.H. 2008. *Mechano-coupling and regulation of contractibility by the vinculin tail domain*. Biophys. J. 94:661–670.

Morfill, J., Neumann, J., Blank, K., Steinbach, U., Puchner, E.M., Gottschalk, K-E., Gaub, H.E. 2008. *Force-based Analysis of Multidimensional Energy Landscapes: Application of Dynamic Force*

Spectroscopy and Steered Molecular Dynamics Simulation to an Antibody Fragmetn-Peptide Complex. J.Mol. Biol. 381, 1253-1266.

Monkley, S.J., Zho, X.-H., Kinston, S.J., Giblett, S.M., Hemmings, L., Priddle, H., Brown, J.E., Pritchard, C.A., Critchley, D.R., Fassler, R., 2000. *Disruption of the talin gene arrests mouse development at the gastrulation stage*. Dev. Dynamics 219, 560–574.

Nelson, D.L., Cox, M.M. 2008. *Lehninger Pricniples of Biochemistry 5th ed.* W.H. Freeman and Company, New York. ISBN-13: 978-0-7167-7108-1.

Nosé, S.A. 1984. *Molecular dynamics method for simulations in the canonical ensemble*. Mol. Phys. 52:255–268.

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

Perkins, A.D., Ellis, S.J., Asghari, P., Shamsian, A., Moore, E.D.W., Tanentzapf, G., 2010. I*ntegrin-mediated adhesion maintains sarcomeric integrity*. Dev. Biol. 338, 15–27.

Rief, M., Pascual, J., Saraste, M., Gaub, H.E. 1999. *Single Molecule Force Spectroscopy of Spectrin Repeats:Low Unfolding Forces in Helix Bundles.* J. Mol. Biol. 286, 553-561.

Rosta, E., Buchete, N.-V, Hummer, G. 2009. *Thermostat artifacts in replica exchange molecular dynamics simulations.* J Chem Theory Comput. 5(5), 1393-1399.

Sawada Y, Tamada M, Dubin-Thaler BJ, Cherniavskaya O, Sakai R, et al. 2006. *Force sensing by mechanical extension of the Src family kinase substrate p130Cas*. Cell 127: 1015–1026.

Scott, W.R.P., Hunenberger, P.H., Trioni, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Kruger, P., VanGunsteren, W.F. 1997. *The GROMOS biomolecular simulation program package.* Journal of Physical Chemistry A 103, 3596–3607.

Sheets, K., Wunsch, S., Ng, C., Nain, A.S. 2013. *Shape-dependent cell migration and focal adhesion organization on suspended and aligned nanofiber scaffolds*. Acta Biomaterialia 9, 7169-7477.

Smith, S.J., McCann, R.O., 2007. *A C-terminal dimerization motif is required for focal adhesion targeting of talin1 and the interaction of the talin1 I/LWEQ module with F-actin*. Biochemistry 46, 10886–10898.

Sotomayor, M., Schulten, K. 2007. *Single-molecule experiments in vitro and in silico*. Science. 316:1144–1148.

Tadokoro, S., Shattil, S.J., Eto, K., Tai, V., Liddington, R.C., de Per- eda, J.M., Ginsberg, M.H., Calderwood, D.A. 2003. *Talin binding to integrin beta tails: a final common step in integrin activation*. Science 302, 103–106.

Thomas, D.B., Franceschini, N., Hogan, S.L., Holder, S., Jennette, C.E., Falk, R.J., Jennette, J.C. 2006. *Clinical and pathological characteristics of focal segmental glomerulosclerosis pathologic variants.* Kidney International 69, 920-926.

van der Spoel, D., Lindahl, E., Hess, B. and the GROMACS development team. 2012. *GROMACS User Manual version 4.6*, www.gromacs.org

Vendruscolo M, Paci E. 2003. *Protein folding: bringing theory and experiment closer together*. Curr Opin Struct Biol, 13:82-87.

Vogel, V., 2006. *Mechanotransduction involving multimodular proteins: converting force into biochemical signals*. Annual Review of Biophy- sics and Biomolecular Structure 35, 459–488.

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

Winkler, J.H., Lunsdorf, H., Jockusch, B.M. 1997. *Energy-filtered electron microscopy reveals that talin is a highly flexible protein composed of a series of globular domains*. Eur. J. Biochem. 243, 430–436.

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

Ziegler, W.H., Gingras, A.R., Critchley, D.R., Emsley, J., 2008. *Integrin connections to the cytoskeleton through talin and vinculin*. Biochem. Soc. Trans. 36, 235–239.

Zhmurov, A., Brown, A.E.X., Litvinov, R.I., Dima, R.I., Weisel, J.W., Barsegov, V. 2011. *Mechanism of fibrin(ogen) Forced Unfolding.* Structure 19, 1615-1624.

Zhong C, Chrzanowska-Wodnicka M, Brown J, Shaub A, Belkin AM, et al. 1998. *Rho-mediated contractility exposes a cryptic site in fibronectin and induces fibronectin matrix assembly*. J Cell Biol 141: 539–551.

Databases and forums

NCBI (protein). [Accessed March, 2013]. Available at: http://www.ncbi.nlm.nih.gov/

dbSNP at NCBI. [Accessed January, 2013]. Available at: http://www.ncbi.nlm.nih.gov/snp/

OMIM database. [Accessed January, 2013]. Available at: http://www.ncbi.nlm.nih.gov/omim/

RCSB PDB Protein Data Bank for protein structure database. [Accessed November, 2012]. Available at: http://www.rcsb.org/pdb/home/home.do

KEGG: Kyoto Encyclopedia of Genes and Genomes, pathway database. [Accessed May, 2013]. Available at: http://www.genome.jp/kegg/

gmx-user forum. [Accessed March, 2013]. Available at: http://lists.gromacs.org/pipermail/gmx-users/2011-May/061370.html)

Appendix I Energy Minimization Input File

```
;
;      Input file
;
title                    =  Minimization
define                   =  -DPOSRES
cpp                      =  /lib/cpp
integrator               =  steep
emstep                   =  0.0001
dt                       =  0.0001
nsteps                   =  10000
nstlog                   =  1
nstenergy                =  10
nstlist                  =  1
ns_type                  =  grid
pbc                      =  xyz
coulombtype              =  PME
vdwtype                  =  Switch
rvdw                     =  1.2
rlist                    =  1.4
rcoulomb                 =  1.4
fourierspacing           =  0.12
pme_order                =  4
ewald_rtol               =  1e-5
```

Appendix II Equilibration Input File

```
;
;      Input File
;
title                       = NPT Equilibration
define                      = -DPOSRES    ; position restrain the protein
; Run parameters
integrator                  = md          ; leap-frog integrator
nsteps                      = 500000      ; 2 * 500000 = 1 ns
dt                          = 0.002       ; 2 fs
; Output control
nstxout                     = 5000        ; save coordinates every 10 ps
nstvout                     = 5000        ; save velocities every 10 ps
nstenergy                   = 5000        ; save energies every 10 ps
nstlog                      = 5000        ; update log file every 10 ps
; Bond parameters
continuation                = no          ; Initial simulation
constraint_algorithm        = lincs       ; holonomic constraints
constraints                 = hbonds      ; hydrogen bonds constrained
lincs_iter                  = 1           ; accuracy of LINCS
lincs_order                 = 4           ; related to accuracy
; Neighborsearching
ns_type                     = grid        ; search neighboring grid cels
vdwtype                     = Switch
       ; LJ Potential normal out to rvdw_switch, after it is switched off to reach zerro
rvdw_switch                 = 1.0
nstlist                     = 5           ; 10 fs
rlist                       = 1.4
       ; short-range neighbor list cutoff (nm), should be 0.1-0.3 nm larger than rvdw (source VH paper)
rcoulomb                    = 1.4         ; short-range electrostatic cutoff (nm)
rvdw                        = 1.2         ; short-range van der Waals cutoff (nm)
; Electrostatics
coulombtype                 = PME         ; Particle Mesh Ewald - long-range el.stat
pme_order                   = 4           ; cubic interpolation
fourierspacing              = 0.12        ; grid spacing for FFT
; Temperature and Pressure Controls
; Temperature coupling is on
tcoupl                      = Berendsen   ; or Nose-Hoover (Weak coupling)
tc-grps                     = Protein   Non-Protein    ; two coupling groups
tau_t                       = 0.1       0.1      ; time constant, in ps
ref_t                       = 310       310      ; reference temperature, for each group (K)
; Pressure coupling is on
pcoupl                      = Berendsen          ; or Parrinello-Rahman (Weak coupling)
pcoupltype                  = isotropic          ; uniform scaling of x-y-z box vectors
tau_p                       = 1.0  1.0           ; time constant, in ps
ref_p                       = 1.0  1.0           ; reference pressure (in bar)
compressibility             = 4.5e-5 4.5e-5      ; isothermal compressibility, bar^-1
; Periodic boundary conditions
pbc                         = xyz                ; 3-D PBC
; Dispersion correction
DispCorr                    = No                 ; account for cut-off vdW scheme
; Velocity generation
gen_vel                     = yes                ; Velocity generation is on
gen_temp                    = 310                ; temperature for velocity generation
gen_seed                    = -1                 ; random seed
; COM motion removal
comm-mode                   = None
```

Appendix III Constant Pulling Force Input File

```
;
;      Input file
;
title                    = Constant Force pulling simulation
define                   = -DPOSRES_fixed
; Run parameters
integrator               = md
dt                       = 0.002             ; 2 fs step size
tinit                    = 0                 ; starting time = 0
nsteps                   = 2500000           ; = 5 ns
; Output parameters
nstxout                  = 1000
nstvout                  = 1000
nstfout                  = 1000
nstenergy                = 100
; Bond parameters
constraint_algorithm     = lincs
constraints              = hbonds
lincs_iter               = 1
lincs_order              = 4
continuation             = yes               ; continuing from  NPT
; Neighborsearching
nstlist                  = 5
ns_type                  = grid
vdwtype                  = Switch
rvdw_switch              = 1.0
rlist                    = 1.4
rcoulomb                 = 1.4
rvdw                     = 1.2
; PME electrostatics parameters
coulombtype              = PME
fourierspacing           = 0.12
fourier_nx               = 0
fourier_ny               = 0
fourier_nz               = 0
pme_order                = 4                 ; cubic interpolation for PME
ewald_rtol               = 1e-5              ;  Relative  strength  of  the  Ewald-shifted
direct potential at rcoulomb
optimize_fft             = yes               ; Calculate the optimal FFT plan for the grid
at start-up.
; Berendsen temperature coupling is on in two groups, off for Protein
Tcoupl                   = Berendsen         ; or Nose-Hoover
tc_grps                  = Protein   Non-Protein
tau_t                    = 0.1      0.1      ; Protein group has temperature control
ref_t                    = 310       310
; Pressure coupling is on whole system in two groups
Pcoupl                   = Berendsen         ; or Parrinello-Rahman
pcoupltype               = isotropic
tau_p                    = 1.0 1.0
compressibility          = 4.5e-5 4.5e-5
ref_p                    = 1.0 1.0
; Generate velocities is off
gen_vel                  = no
; Periodic boundary conditions are on in all directions
pbc                      = xyz
; Long-range dispersion correction
DispCorr                 = no
; Pull code
pull                     = constant_force
pull_geometry            = direction         ; pull in direction of pulling vector
pull_vec1                = 0.0 0.0 5.5       ; the components of the pulling vector
pull_ngroups             = 1                 ; pull Ca atom of C-terminal residue
```

```
pull_group0                = a_2                ; fixed
pull_group1                = a_2733             ; pulled
pull_k1                    = -180.7             ; [kJ/(mol.nm)^-1], equals to 300 pN
                           ; 1kJ / (1 mol).(1 nm) = 1000 N.m / (6.022.10^23).(1.10^-9 m) = 1.6605 pN
                           ; constant_force / 1.6605 pN = pull_k1 (negative!!!)
```