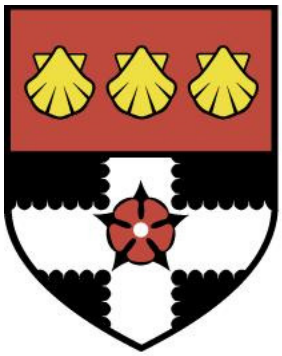
UNIVERSITY OF READING



Improving force values for the unfolding thresholds of talin's rod subdomains through force-domain simulations designed for the spatial and temporal ranges of talin.

Guy Butcher

Department of Biomedical Engineering,
School of Biological Sciences

A thesis submitted for the degree of Doctor of Philosophy

27/03/2022

Declarations

Declaration of Original Authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Guy Butcher

Funding

This work was supported by the Engineering and Physical Research Council Doctoral Training Programme (Grant Number EP/N509723/1).

Code and Data Availability

All source code and data used to generate results, tables, and figures found within this manuscript are freely available within a GitHub repository under the Apache-2.0 licence. Additional information about the source code can be found in the repository.
RRID: SCR_022116
DOI: 10.5281/zenodo.6381772
https://github.com/GuyButcher/doctoral_thesis_code_data_repository

Abstract

Most cells or proteins exist in a high force environment, meaning that they must be able to generate, withstand or transmit high levels of forces. The signalling response of talin (a core structural and signalling protein in platelet focal adhesions), for example, is the product of its rod domains unfolding and either ejecting bound ligands or allowing new ligands to bind. Modelling force effects at a molecular level is therefore helpful in understanding when talin signals other platelet processes, such as focal adhesion maturation and platelet contraction. This, in turn, may impact the understanding of diseases related to platelet binding and aggregation, as well as thrombus formation, contraction, and lysis.

Therefore, the aim of this thesis was to improve the calculated magnitude and range of force values for the unfolding thresholds of talin's rod domain. This was done by creating a novel model of talin's rod domain alpha-helices. This model was subsequently used in a newly developed force-based simulation framework to calculate intermolecular electrostatic interactions. The unique features of this method were: the new coarse-grain-like alpha-helix models derived from protein structural data, the use of a coordinate frame system to manage and manipulate the objects within the simulation, and the direct calculation of interaction forces.

This approach was chosen as other computational methods such as molecular dynamics simulations produce erroneously large unfolding force ranges that are up to $\sim$ 7500% larger than the physiologically relevant range of forces for the unfolding of talin's rod domain. Through the method developed for this thesis, a range of unfolding forces were calculated for talin's rod subdomains of $\sim$1 pN to $\sim$49 pN. This is an improvement compared to previous computational methods. Thus, the improvements in results are significant enough to warrant the effort in developing these new simulation approaches targeted at force interactions at the molecular level.

Acknowledgements

I would like to start by thanking my supervisors Prof. William Harwin and Dr. Chris Jones for their time, energy, academic enthusiasm and patience with me. William also supervised my Master's thesis and encouraged me to start this PhD. Without their guidance, this thesis would not exist and my skills in research would not have grown as they have.

I would also like to thank the Engineering and Physical Research Council for providing me with a grant through the Doctoral Training Programme (EP/N509723/1) for without this, this PhD would not have been possible.

Next I would like to thank from the bottom of my heart, my girlfriend and partner, Dr. Els Meijer. She has helped me through all the trials of this PhD. Emotionally and mentally, she has encouraged me to persevere and supported me when I needed it most. Academically, she has provided guidance and a truly unique perspective from the world of scholarly research. Linguistically, she has changed the way I think about words and language, helping me grow beyond what I once thought possible as writer, and providing invaluable care in helping to copy-edit this thesis (which is a very good thing, as this thesis would have appeared written upside-down and back-to-front otherwise). Without her, I would not have completed this PhD or thesis, and for her love and companionship during this time I cannot thank her enough.

I also want thank two close friends who have willingly put up with hours of my rantings during this PhD, and for years prior, Calum Brown and Sam Podmore. Your friendship and enjoyable distractions in gaming have helped me through this PhD more than you know.

I would also like to thank two teachers from my schooling days, Dr. Mark Howson and Mr. Anthony Spender. These two went out of their way to encourage and guide me, and to push me to improve my learning, leading me to where I am now.

Finally, I would like to thank my parents, Gary and Amanda Butcher, and my grandmother Carol Butcher. In my early years, they provided for and encouraged my education so that one day I could reach my academic potential. Their care over the years has shaped who I am now.

Table of Contens

List of Figures

List of Tables

List of Abbreviations

5HT Serotonin.

ABS Actin Binding Site.

ADP Adenosine Diphosphate.

AFM Atomic Force Microscopy.

ASA Accessible Surface Area.

ATP Adenosine Triphosphate.

CABS C-alpha, C-beta and Sidechain Representation.

CalDAGGEF1 Calcium DAG-Regulated Guanine Nucleotide Exchange Factor-1.

DAG Diacylglycerol.

DPA Dual Pseudo-Atom.

ECM Extracellular Matrix.

F-Actin Filamentous Actin.

FAK Focal Adhesion Kinase.

FV Coagulation Factor V.

G-Actin Globular Actin.

GAP RhoA GTPase Activating Protein.

GDP Guanosine Diphosphate.

GPCR Glycoprotein Coupled Receptor.

GPIb-IX-V Glycoprotein Ib-IX-V Complex.

GPVI Glycoprotein VI Complex.

GTP Guanosine Triphosphate.

IBS Integrin Binding Site.

IP3 Inositol Triphosphate.

ITAM Immunoreceptor Tyrosine-based Activation Motif.

MD Molecular Dynamics.

MLCK Myosin Light Chain Kinase.

MMT Microtubule.

NMR Nuclear Magnetic Resonance Imaging.

P-Fibrin Polymerised Fibrin.

PDB Protein Data Bank.

pdb Protein Data Bank File.

PI3K Phosphoinositide 3-Kinase.

PKC Protein Kinase C.

RIAM Rap1-GTP-Interacting Adapter Molecule.

RLC Regulatory Light Chain.

ROCK Rho-associated Coiled-Coil Kinase.

SASA Solvent Accessible Surface Area.

SFK Src Family Kinase.

SMD Steered Molecular Dynamics.

SPA Single Pseudo-Atom.

TFM Traction Force Microscopy.

TIRF Total Internal Reflection Fluorescence.

TXA2 Thromboxane A2.

UNRES United Residue Representation.

vWF von Willebrand Factor.

1 Introduction

1.1 Problem Statement

Platelets close vascular injury by aggregating, forming thrombi and then contracting [1, 2]. Mechanosensitive proteins such as talin are a key element of this process. They provide a physical interface between the external environment of the platelet, the internal cytoskeleton and the contractile machinery. They do this by sensing and adapting to forces exerted on the platelet [3–6].

Talin's rod domain contains 13 alpha-helix coiled-coils or 'bundles', which unfold if the tension applied across them exceeds a specific threshold. The surfaces of these alpha-helix bundles contain protein binding sites that are accessible while a bundle is folded. Inside the bundles, cryptic binding sites exist that are only made accessible once a bundle has unfolded. The changes in accessibility of these binding sites and the subsequent binding or ejection of ligands produces talin's signalling response [7]. Talin's signalling response controls platelet focal adhesion formation, platelet spreading and contraction. Thus, talin has a dual role, regulating force transmission within the platelet and also regulating intra-cellular signalling in response to this force.

As it is not known at what level of tensile force the alpha-helix bundles unfolding events occur, it is not possible to predict when talin signals other platelet processes, such as focal adhesion maturation and platelet contraction. Fully understanding talin's signalling response and its subsequent effect on platelets has clinical implication for diseases related to platelet binding and aggregation, and thrombus formation, contraction and lysis. Additionally, determining if it is possible to modify talin to increase or decrease platelet adhesion strength, would allow for the controllable alteration of thrombogenicity and may lead to potential treatments for thrombosis and embolisms caused by lysis.

During experiments where talin's rod domain was stretched and its unfolding dynamics measured, only three of the thirteen rod domain bundles were mapped to their unfolding force thresholds. R3 to $\sim$5 pN, R7 to $\sim$15 pN and R8 to $\sim$5 pN [8]. Molecular dynamics simulations of talin's rod domain have also been used to try to determine the forces required to unfold the rod domain bundles. However, these simulations have resulted in unfolding force values of up to 1500 pN, far larger than are physically possible for talin [9]. For reference, whole platelet contractile forces are in the range of 0.3 nN to 70 nN [10, 11] while it has been estimated that the maximum force across talin in platelets averages at 10 pN [12].

It should be noted that molecular dynamics simulations were not designed to calculate unfolding forces for protein domains. Therefore, the results from the simulations cannot be used to link more rod domain bundles to their unfolding force thresholds. To overcome the limitations of current molecular simulations to calculate unfolding forces, a new simulation methodology needs to be developed, one that specifically targets the spatial and temporal ranges of protein unfolding.

Therefore, the research question that is answered in this thesis is: can the calculated magnitude and range of force values for the unfolding thresholds of talin's rod domain be improved upon by making simulations that operate directly within the force domain and are designed for the spatial and temporal ranges that talin operates in?

The original contribution of this thesis is the use of a new force-based simulation to calculate the unfolding forces between alpha-helices. It does this by creating new, approximated structural models of alpha-helices and a simulation framework using a coordinate frame system that was designed for the spatial and temporal scales (seconds to 10s of seconds, and 0.1 pN to 25 pN) of alpha-helix unfolding. Additionally, a force-based structural optimisation was developed to improve the alpha-helix subdomain conformations.

1.2 Literature Review

Platelets are non-nucleated cells generated by megakaryocytes within bone marrow. They play a key role in haemostasis by aggregating to forming thrombi at sites of vascular damage. During normal circulation, platelets exist in an inactivated, resting state with a discoid shape and an approximate radius of 2–3 µm [13]. There are several classes of transmembrane glycoprotein receptors on the surface of platelets. One class, called integrins, is the platelets' primary method of adhering to their surrounding environment during thrombus formation. Of all the integrins, $\alpha IIb\beta_3$ occurs most commonly with approximately 80,000 copies per platelet [14, 15]. It adheres to fibrinogen and polymerised fibrin (p-fibrin) [16].

Many external factors regulate platelet behaviour, such as: the concentration of agonist and inhibitory proteins and signalling molecules present in the blood [17, 18]; changes in blood pressure [19]; sudden increases in shear stress at the vessel wall [20]; and the quantity of other activated platelets in the vicinity. Irregularities in these external signals and the platelets'response to them via chemical sensing, mechanosensing and adhesion proteins on their surface can result in serious disorders and diseases [21, 22].

Unwanted thrombus formation can result in clots which occlude blood vessels, leading to local-area cell death from lack of oxygenated blood, and potentially patient death. The unwanted thrombus formation can occur from platelet over-activation and over-adhesion at sites of vascular injury. Alternatively, it can occur from unintended platelet activation and adhesion at sites without vascular injury. The vascular occlusions formed by these clots are called occlusive thrombosis [23]. Reversely, lack of platelet activation or secure adhesion formation can lead to excessive bleeding [24–26], or to the lysis of forming thrombi that may result in an embolism forming downstream, possibly resulting in stroke or death [27]. To better understand how these disorders occur, it is important to explore the methods of platelet activation and adhesion that lead to thrombus formation.

Platelet activation is a multi-stage process that transforms a platelet from its resting, discoid shape into a globular protein that adheres and spreads itself across other platelets in forming thrombi and contracts them. Platelets exist in their resting state before interaction with platelet agonist molecules in the blood, slowing, and binding into sites of vascular injury or forming thrombus. Spreading involves the platelet's cell membrane being expanded across the surface that the platelet is adhered to and occurs after platelets have securely adhered to a site of vascular injury or thrombus. After the platelet has spread, it begins to contract. The contraction begins with alternating waves of contraction and spreading at the peripheral edge of the platelet, ending in full steady state contraction of the platelet, pulling everything the platelet is adhered to together. The collective contraction of platelets in a thrombus results in clot contraction and sealing of vascular injury.

In the following six subsections, the key biological processes of platelets are described. The chemical signalling that starts platelet activation is covered. This is followed by a description of how platelets undergo their shape change and what mechanisms drive spreading, adhesion and contraction. Finally, the contractile forces at the thrombus and platelet scales are discussed. Talin is highlighted as a key protein in platelets as it is a force sensor and force transmitter during platelet adhesion, spreading and contracting. It is also a key signalling protein that produces and activates signalling molecules in response to changing levels of forces. Talin is the focus of the research in this thesis and is explored in detail in sections 1.2.7 to 1.2.10.

1.2.1 Adhesion Receptor Platelet Activation

As the resting platelets flow by a site of vascular injury, some platelets use their glycoprotein Ib-IX-V complex receptor (GPIb-IX-V) to interact with von Willebrand factor (vWF) that is bound to exposed collagen [28, 29]. The binding between GPIb-IX-V and vWF slows the platelet down, allowing its glycoprotein VI receptor (GPVI) to bind directly to exposed collagen [13, 30]. While the binding of GPIb-IX-V to vWF results in the activation of some of the 'platelet activating' signalling pathways inside the platelet, the primary signalling response starts because of GPVI adhesion. Signalling pathways are a description of the flow of chemical signals in a biological process. Each step in a signalling pathway may promote or inhibit a subsequent step. They can also branch and interact with multiple other pathways. Additionally, signalling pathways can describe positive and negative feedback cycles. The platelet activation pathway begins with the adhesion of GPVI and GPIb-IX-V, which promotes a series of events that activate several other signalling molecules. The eventual result, is the activation of talin and integrin, the release of agonist molecules by the platelet via the secretion of granules which activates other nearby platelets, and finally the reorganisation of the internal cytoskeleton of the platelet. Below, the platelet activation signalling pathway is described in detail.

The signalling pathway activated by GPVI involves the Src family kinases (SFKs) activation of phosphoinositide 3-kinase (PI3K) [31]. SFKs also lead to the activation of tyrosine kinase Syk which induces the formation and activation of PLCγ [32]. PI3K also contributes to the activation of PLCγ and its localisation to the platelet membrane [33]. The activation of PLCγ is a key stage in platelet activation as it signals the release of inositol triphosphate (IP3) and diacylglycerol (DAG). IP3 signals the release of stored calcium ions (Ca2+) into the platelet. It also enables platelets to accept additional Ca2+ through store-controlled ion channels [34]. Together, DAG and Ca2+ activate protein kinase C (PKC) and calcium and DAG-regulated guanine nucleotide exchange factor 1 (CalDAGGEF1) [35]. CalDAGGEF1 activates Rap1 which results in the activation of Rap1-GTP-interacting adapter molecule (RIAM), which binds to and activates talin [36]. In platelets, it has been shown that the direct binding of Rap1 to talin is sufficent for talin activation, even in the absence of RIAM [37]. Activated talin alongside kindlin bind to the cytoplasmic βtail of integrin αIIbβ_3, starting the activation of integrin [38].

The signalling that leads to the partial activation of integrin αIIbβ_3 is called

inside-out signalling and is the first of two stages that define full integrin activation. The binding of these two proteins alters the conformation of integrin, causing its extracellular head domain to partially unfold from an inactive conformation to a semi-activated conformation. Once the integrin head is exposed, low-affinity binding of fibrinogen and fibrin can occur via a highly conserved RGD motif [39]. Integrin may also bind to collagen-bound vWF [40].

Fibrinogen is a glycoprotein complex that circulates in the blood. When exposed to thrombin, a coagulation factor that is produced at sites of vascular injury, fibrinogen is cleaved to form fibrin. Multiple integrin can bind to the same fibrin filament, resulting in integrin micro-clustering [41]. The binding of fibrinogen and fibrin also causes conformational changes in the cytoplasmic β-tail of integrin that enables guanine nucleotide-binding protein alpha 13 (Gα13) to bind to the integrin tail [42]. The binding of Gα13 results in the activation of the protein kinase c-Src and subsequent activation of PI3K and Syk (part of the immunoreceptor tyrosine-based activation motif (ITAM) signalling pathway) which reinforces the inside-out signalling via PLCγ activation. Additionally, c-Src activates a RhoA GTPase activating protein, p190RhoGAP, which is an inhibitor of small GTPase RhoA. RhoA itself inhibits platelet spreading as it activates the contractile machinery of the platelet.

Finally, the release of Ca2+, activation of Rap1 via CalDAGGEF1 and the activation of PKC (all of which is reinforced via outside-in signalling) together lead to the synthesis of TXA2 and the secretion of granules [43]. Platelet granules are vesicles that contain a number of molecules and membrane receptors that, when secreted, merge with the cell membrane and expel their contents [44]. Platelets contain two main granules: α-granules and dense granules. α–granules contain adhesion proteins such as fibrinogen, vWF and coagulation factor 5 (FV). Dense granules contain ADP, Ca2+ and serotonin. ADP, serotonin and TXA2 are all platelet agonists that can activate platelets via glycoprotein coupled receptors (GPCR) [45].

1.2.2 GPCR Platelet Activation

GPCRs are a class of transmembrane receptors that have a glycoprotein complex bound to their cytoplasmic domain. Upon ligand-receptor binding, the α-domain of the glycoprotein complex is phosphorylated, converting a bound molecular of guanosine diphosphate (GDP) to guanosine triphosphate (GTP). The α-domain then dissociates from the rest of the GPCR to pass on signalling to downstream

targets [46]. Six GPCR are involved in platelet activation: $P2Y_{12}$ and $P2Y_1$ which bind ADP [47], PAR1 and PAR4 which bind thrombin [48], 5HT2a which binds serotonin (5HT) [46], and finally TP which bind TXA_2 [49, 50]. Although the GPCR platelet activation signalling pathway starts at a different set receptors compared to the adhesion-receptor pathway described in the previous subsection, both signalling pathways eventually converge on the same signalling molecules. Thus leading to the activation of integrin by talin, granule secretion, and modification of the actin cytoskeleton. A high proportion of the molecules and chemicals released during granule secretion can bind to GPCRs and activate this signalling pathway. This way, platelet activation is self-reinforced. Further details of the GPCR signalling pathway are given below.

Three glycoproteins, Gi, Gq and G13 [46], are activated from these six receptors. Gi is activated by $P2Y_{12}$, Gq is activated by $P2Y_1$, PAR1, PAR4, 5HT2a and TP, and G13 is activated by PAR1, PAR4 and TP. All three glycoproteins, once activated and dissociated, activate of a signalling pathway beginning with SFKs and PI3K [51–57]. This pathway involving SFK, PI3K, NOS, sGC and PKG [58, 59] results in granule secretion and TXA_2 synthesis. G13 activation leads to the activation of RhoA via p115RhoGEF which phosphorylates RhoA by converting GDP to GTP [60]. This causes the downstream activation of the contractile machinery of the platelets, stimulating cell contraction and slowing spreading [61]. Gq activates PLCβ [62]. PLCβ results in similar downstream effects as PLCγ such as the release of Ca2+, PKC activation and Rap1 activation via CalDAGGEF1 [43]. Activation of PKC and the release of Ca2+ have a strong effect on increasing granule secretion. The activation of Rap1 increases TXA_2 synthesis [63, 64]. Activated Rap1 also leads to the activation of talin through the binding of Rap1 or Rap1/RIAM complex [37]. Activated talin binds to and activates integrin $\alpha IIb\beta_3$ via inside-out signalling, enabling integrin to bind fibrinogen and fibrin.

In summary, platelets are activated through two main signalling pathways, via adhesion receptors GPVI and GPIb-IX-V which bind collagen and collagen-bound vWF, and via GPCRs after binding with ADP, Thrombin, TXA_2 and serotonin. These signalling pathways eventually converge into a common pathway that activates PLCβ or PLCγ. The activation of the PLCs and the subsequent release of Ca2+ result in: the secretion of granules which further activate platelets through GPCRs, modification of RhoA signalling that controls platelet shape change, spreading and contraction, and the activation of Rap1. Activated Rap1 either on its own or in complex with RIAM binds to talin and activates it.

Activated talin, along with kindlin, bind to the cytoplasmic tail of integrin $\alpha IIb\beta_3$, beginning their activation through inside-out signalling. Upon integrin binding to fibrinogen or fibrin, conformational changes increase integrin's binding affinity to fibrinogen and fibrin, and start the signalling pathways that reinforce PLCγ activation. Coagulation factors in the blood interact with each other to produce thrombin from prothrombin. Thrombin cleaves fibrinogen into fibrin, allowing for the formation of fibrin filaments. These filaments bind to a platelet's integrins, leading to additional platelet activation and recruitment to the sites of injury.

1.2.3 Platelet Shape Change

In resting platelets, the discoid shape is maintained via the microtubule (MT) ring at their core, a disorganised network of actin filaments that spread outwards from the MT ring and a spectrin sub-membrane web that supports the platelets' cell membrane [65–69]. During activation, platelets undergo a significant shape change. The start of platelet shape change is characterised as the disassembly of the actin network and spectrin web, and the coiling of the microtubule-ring by dynein [70–72]. These changes in the cytoskeleton cause the platelet to lose its discoid shape, transitioning to a spherical shape [73]. This shape change is initiated via the activation of p115RhoGEF which activates RhoA [60]. RhoA is associated with activating the actomyosin contractile machinery [61]. The activation of the actomyosin contractile machinery starts the process of actin network reconfiguration and also promotes increased granule secretion [74].

After integrin has bound to fibrinogen or fibrin, conformational changes in the integrin's cytoplasmic beta-tail enable G-α-13 binding. G-α-13 binding results in the activation of the c-Src/p190RhoGAP pathway, inhibiting RhoA [75]. The inhibition of RhoA and the actomyosin machinery results in reduced platelet contraction and actomyosin shape change while promoting platelet spreading [76]. This causes actin filaments to drive the spreading of the platelet's membrane through the formation of two new cytoskeletal structures at the edge of the cell membrane: filopodia and lamellipodia [77]. In parallel, near the coiled microtubule-ring in the core of the platelet, contractile machinery called stress fibres (composing of actin filaments, myosin-9 and α-actinin) are forming that will generate the platelet's contractile forces during the platelet contraction phase [78, 79].

1.2.4 Spreading

Platelet spreading occurs because of the generation and growth of filopodia and lamellipodia that push on the cytoplasmic side of the platelet's membrane and move it outwards from the platelet core [80]. As the cell membrane spreads, stores of additional membrane in the surface of the platelet called invaginations are released as required [81]. Filopodia are long, parallel polymers of actin that are heavily cross-linked with α-actinin and fascin, extending towards the edge of the cell from the coiled MT ring [77]. The lamellipodia occupies a region near the outer edge of the platelet where actin grows as a dense, branching network and is the structural support for the cell's membrane [77]. Two proteins, arp2/3 and formin, nucleate the majority of the actin network in both the lamellipodia and filopodia [82].

Arp2/3, which is localised at the peripheral edge of the cell membrane and binds to the tip of elongated actin filaments. New actin filaments branch off arp2/3 at an approximate angle of 70° [83]. As these branched filaments grow towards the edge of the cell membrane, they continue branching. Thus, a dense branching network of actin filaments is formed. Capping proteins prevent the formation of long, unbranching actin filaments, thereby encouraging the formation of a dense network. The branching network presses against the cell membrane (preventing it from retracting as well as) applying force to extend the membrane further [84, 85].

Kage et al. [86] have argued that the main protrusive force which drives membrane extension comes from formin-mediated actin filament polymerisation. Formins are a family of proteins that are bound to the membrane and nucleate long actin filaments. Kage et al.'s study determined that deletion of FMNL formins reduced actin density and bundling in the lamellipodia without affecting the structure of branching actin filament networks. The study also showed that membrane protrusive forces were reduced and in some cases eliminated. Based on this evidence, the apr2/3 nucleated actin branching network generates a supportive structure for the lamellipodia, while formin nucleated actin filaments drive the spreading of the leading edge of the cell membrane.

As the protrusive forces from polymerising actin filaments push the membrane further outwards, the surface tension of the stretching cell membrane increases, pushing back onto the actin network of the lamellipodia. This force acting backwards from the membrane causes the whole actin network to be pushed backwards, towards the centre of the platelet as more actin filaments continue to be nucleated

at the leading edge of the cell membrane. This rearward flow of the actin network is known as actin retro-grade flow. During retro-grade flow, the rear-most actin filaments pass through the lamellum. The lamellum is the region that backs the border of the lamellipodia and the inner core of the platelet. In the lamellum, actin binding proteins such as ADF/cofilin cleave actin filaments to break them down for recycling, to be reused in growing filaments [87]. Actin filament treadmilling is the term used to describe the combined effects of: the actin network growing at the leading edge, the retro-grade flow to the actin network, and the breakdown of the old filaments to be recycled in the lamellum [88].

1.2.5 Focal Adhesions

During platelet adhesion and spreading, activated integrins interact with fibrin filaments leading to microclustering of integrins around sites of platelet adhesion [41]. These microclusters are the target for nascent focal adhesion formation [89]. Focal adhesions are a complex of proteins such as focal adhesion kinase (FAK), vinculin, paxillin, tensin and actin filaments that all grow around a core of clustered integrin and their bound talin. For example, FAK can bind and provide additional binding sites for structural and signalling proteins such as c-Src, paxillin and tensin [90]. In the formation of nascent focal adhesions, talin cross-links integrin to actin filaments. This forms a bridge between integrin and the actin network in the lamellipodia, anchoring the platelet's cytoskeleton to its surrounding environment [91]. This integrin-talin-actin linkage, illustrated in Figure 1.1, is the core force transmission pathway for the platelet [92, 93]. All contractile forces are transmitted through this pathway and thus through talin. Talin's response to these forces is crucial to a properly functioning platelet [94].

The aggregation of integrin-talin complexes and subsequent binding of talin to the lamellipodial actin network creates a force transmission pathway between the platelet's cytoskeleton and the extracellular environment. Platelets and thrombi need to withstand external forces such as shear stress from the flow of blood. These forces are transmitted through the platelet cytoskeleton to the focal adhesions. The focal adhesions maintain a platelet's adhesion to its substrate, either the site of injury or the thrombus [95, 96].

Internal forces such as those generated from retrograde flow of the actin network also pass through the focal adhesion. In the nascent focal adhesion stage, the link between talin and the actin networks is transient. This allows the retrograde flow of actin to continue as talin unbinds and rebinds to the passing actin

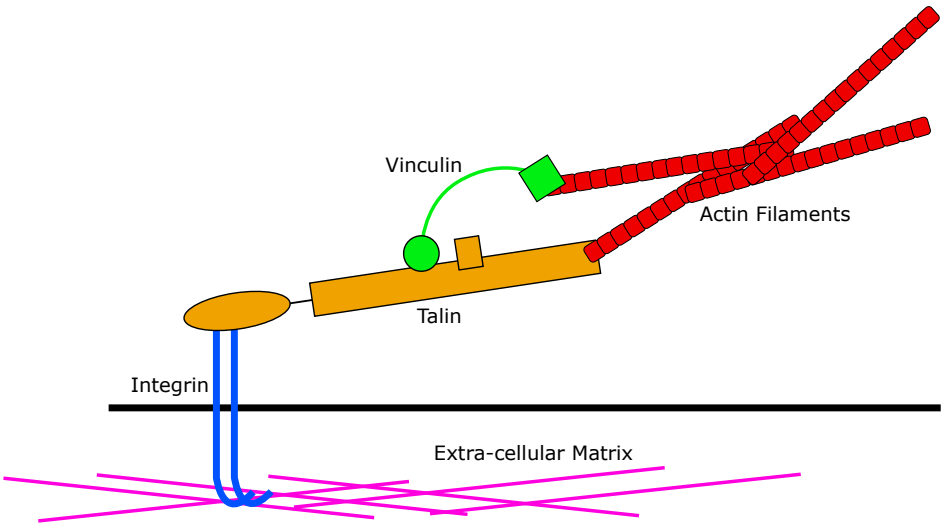


Figure 1.1: Illustration of the force transmission pathway proteins and their connections to each other. Colour code: blue, integrin; magenta, extracellular matrix (ECM); orange, talin; green, vinculin; red, actin filaments.

filaments. The retro-grade flow of actin combined with the transient binding of nascent focal adhesion leads to a friction like force on the actin network by talin and a low traction force applied to the extracellular environment via integrin [97]. Applying traction forces to the external environment through the focal adhesions results in a tensile force across the nascent focal adhesion. The tensile force is a mechanical signal for talin to begin the maturation of the focal adhesion [98]. It has been proposed that this tension requirement to form a mature focal adhesion exists in order to select nascent focal adhesions that have successfully bound to solid and stiff external substrates, allowing the efficient transfer of forces to the thrombus and site of vascular injury [99, 100].

The maturation of focal adhesions involves the binding of vinculin to talin and crosslinking to additional actin filaments [101]. Additionally, actin filaments are reinforced by the localisation of α-actinin and myosin, and by crosslinking to other nearby actin filaments. This forms a more rigid connection between the focal adhesion and actin cytoskeleton, increasing the magnitude of forces transmitted [102]. The localisation and binding of actin filaments with α-actinin and myosin also begins the integration of the focal adhesion into the contractile machinery of the platelet via the formation and growth of actin stress fibres [103].

In summary, during platelet activation, platelets change shape as their microtubule ring is coiled and the existing actin network is dismantled. Integrins cluster as they adhere the platelet to fibrinogen and fibrin filaments. New actin structures called filopodia and lamellipodia are generated that support and extend the

platelet's membrane. The extension of the membrane spreads the platelet on the site of injury or on the thrombus. Membrane spreading increases the number of integrins that can adhere as the surface area of the platelet increases. The actin network that supports the spreading of the membrane is pushed backwards towards the centre of the platelet due to the pressure of the membrane, this is called actin retrograde flow [104]. Talins connected to integrins that have formed adhesions transiently bind to this rearward flowing actin, resulting in tension forming across the talin and integrin and transmitting traction forces to the external environment. The new integrin-talin-actin link is called a nascent focal adhesion. The tension across talin is a mechanical signal which results in the recruitment of focal adhesion proteins such as vinculin and FAK. These either reinforce the connection between talin and the actin network or perform signalling functions. This recruitment of signalling proteins and reinforcement of the nascent focal adhesion to forming actin stress fibres is termed focal adhesion maturation. Focal adhesion maturation enables higher platelet contractile force transmission to the thrombus and site of injury.

1.2.6 Platelet Contraction

Activated platelets adhere and spread at sites of vascular injury and on thrombi. Spreading enables platelets to adhere to as large an area as possible, find the stiffest surfaces, and therefore apply as much contractile force to as much of their surroundings as possible. This link between substrate stiffness and the level of contractile forces measured is mentioned multiple times in literature covered in the following section.

At some point in platelet spreading, a phase transition occurs, and the platelet begins to undergo cycles of contraction and spreading [13]. Rauncher et al. have suggested that this occurs when a membrane tension limit is reached [105]. Platelets contain a motor protein known as myosin-9 or human non-muscle heavy chain myosin-IIa, or simply just myosin-IIa, that is responsible for the majority of its contractile function [106]. Platelets' contractile machinery is located mostly near the centre of the cell with filament bundles called stress-fibres stretching towards the periphery of the cell called transversal arcs. These along with ventral stress fibres connect to mature focal adhesions via dorsal stress fibres [106]. These fibres and the contractile machinery at the centre of the cell are the actomyosin complex, composed of antiparallel actin filaments cross-linked by α-actinin and myosin-IIa [107, 108]. When supplied with ATP and signalled by Mg^{2+}, the

myosin-IIa contracts, pulling parallel actin filaments past each other [109]. These contractile fibres pull through the actin branching network in the lamellipodia to the mature focal adhesions found at the lamella border. This allows the contractile filaments to pull the site of vascular injury and the thrombus towards the centre of the platelet [110].

Cai et al. [111] confirmed the connection between the contraction of actin filaments and the application of traction force to the extracellular matrix (ECM) by using total internal reflection fluorescence microscopy (TIRF) and mouse embryonic fibroblasts where the expression of myosin-IIa and myosin-IIb had been prevented. Their research showed that restricting the expression of myosin-IIa limited the cells' ability to apply traction forces to the substrate, reduced actin filament retrograde flow and caused the cells to spread at a higher rate. The connection between myosin-IIa contraction and cell motility also applies to platelet contraction. Platelet contraction can be considered a special form of cell motility as it involves the same molecular machinery, but is used for isotropic contraction of the cell and its surrounding environment [106].

The transition to full platelet contraction occurs when the traction force exerted on focal adhesions is produced primarily from the actomyosin complex rather than from actin filament retrograde flow. The force of this full contraction is transmitted through the focal adhesion. The core focal adhesion proteins — integrin, talin, vinculin and actin filaments — are therefore required to withstand and transmit these forces.

Thrombus and platelet contractile forces have been measured using several experimental methods. Early experiments by Jen et al. placed clots over an 800 µm gap between two motile plates covered in fibrin in order to measure clot contraction forces. Eighty minutes after clot activation with thrombin, the stiffness of the motile platelets and the change in distance between them were used to estimate the force from clot contraction. The clot contractile force was measured to be 0.5 nN [112]. Follow-up experiments by Carr et al. used the same experimental paradigm but with a single motile plate. They observed that clot contraction reached a steady state after 25 minutes with a contractile force between 0.3 nN to 0.4 nN [10].

A new experimental protocol was adapted by Liang et al. [113]. A grid of silicone elastomer posts were constructed with a fluorescent marker atop each post. The posts were coated in fibrinogen or fibronectin. A confocal microscope was used to determine post deflection caused by thrombus contraction. This deflection,

combined with known post stiffness was used to calculate a contractile force. Platelet were seeded over the posts and activated with thrombin. Contractile forces were measured at an average of 2.1 ± 0.1 nN.

Measuring clot contraction is useful for observing behaviour at the thrombus level. However, extrapolating those results to determine platelet level contraction is error-prone. The volume of the thrombus has to be measured and the total number of platelets estimated from that volume. Per-platelet contraction is then calculated. However, platelets do not contract uniformly throughout thrombi because of individual platelet responses to substrate stiffness and differing rates of activation [114, 115]. Therefore, studies began focusing on measuring individual platelets.

Single platelets and their contractile forces were measured by Lam et al. using atomic force microscopy (AFM) [114]. An AFM cantilever coated in fibrinogen was set at different stiffnesses. Platelets were bound between the cantilever and a solid surface, and were activated resulting in contraction. Lam et al. found that complete contraction occurred 15 minutes after contact with fibrin. The range of contractile forces measured were 1.5 nN to 79 nN, with an average of 19.0 ± 3.1 nN. Platelets were also observed producing higher contractile forces when cantilever stiffness was increased. The adhesion force of platelets to fibrin was determined to be 69.0 ± 12.7 nN. Additionally, it was noted that the magnitude of the platelet contraction force positively correlated with the adhesion strength and platelet elasticity after contraction.

Schwarz Henriques et al. used traction force microscopy (TFM) to measure single platelet contraction [116]. TFM makes use of a soft substrate with stiffnesses in the sample's physiologically relevant range. Deformations in the substrate are measured visually via markers and traction forces are calculated using the deformation displacements and substrate stiffness. Schwarz Henriques et al. used a substrate with a stiffness of 4 kPa. Steady-state contraction was recorded after 25 minutes at $\sim$24 nN with larger platelets generating larger forces. The direction of the traction forces was observed as moving towards the centre of the cell. Additionally, Schwarz Henriques observed that platelets contract nearly isotropically, meaning that the platelet contracted towards its centre with even force from every direction.

Qiu et al. performed a similar experiment to that described by Schwarz Henriques et al. where a flexible substrate of polyacrylamide gel coated in fibrinogen was used [115]. Qiu et al. recorded platelet adhesion and spreading on gels with

differing stiffness at 0.5 kPa, 5 kPa and 50 kPa. They found that stiffer substrates increase platelet adhesion and spreading. This corroborates Lam et al.'s observations on substrate stiffness altering platelet contractile forces.

Fibrin coated hydrogel of varying stiffness was used by Myers et al. to measure platelet contractile forces. Their results agreed with Qiu et al.'s regarding substrate stiffness. A gel stiffness of 25 kPa resulted in contraction forces of ∼6 nN to ∼18 nN with a median of ∼10 nN and a gel with a stiffness of 75 kPa resulted in ∼6 nN to ∼70 nN with a median of ∼30 nN. Additionally, Myers et al. identified that the Rho/ROCK pathway was linked to substrate stiffness-dependent force response. ROCK inhibited platelets produced low contractile forces on high stiffness substrate compared with a control platelet and ROCK inhibited platelet on low stiffness substrates.

In summary, full platelet contraction is myosin-IIa dependent and is activated by the combination of the RhoA/ROCK signalling pathway and local release of ADP and Mg^{2+}. Actin stress fibres, constructed from antiparallel actin filaments cross-linked with myosin and α-actinin, transmit contractile forces from the centre of the platelet to mature focal adhesions. Mature focal adhesion are sites of integrin clustering where the connection between talin and the actin network have been reinforced by focal adhesion proteins. The reinforcement is started from sustained tension across the integrin-talin-actin complex. The tensile forces are the result of actin retrograde flow, actomyosin contraction or external forces acting on the platelet. The mature focal adhesions transmit traction forces (via the actin-talin-integrin link) to the platelet adhered substrate, the thrombus, or site of vascular injury. Experimental measurements of whole clot contraction forces per-platelet were relatively low, 0.3 nN to 2.1 nN, when compared to studies of individual platelet contractile forces averaging 19 nN to 30 nN and peaking at ∼70 nN. This is likely the result of the different mechanical environments of a thrombi and hydrogel surfaces. Substrate stiffness plays a large role in modulating platelet contractile forces with higher stiffness resulting in larger contractile forces. A stiffer substrate may increase tension across talin from actin retrograde flow, signalling for increased reinforcement of focal adhesions. Focal adhesions with more reinforced links to the actin cytoskeleton could increase the magnitude of traction force transmitted to the platelet external environment.

1.2.7 Talin Structure Overview and Mechanical Properties

Talin is a structural component of platelet focal adhesions and is the core connection between integrin, the actin cytoskeleton, and the cytoskeleton's contractile machinery. Talin is a large (270 kDa) protein that binds to integrin's cytoplasmic tail through its integrin binding site (IBS) at talin's head [117, 118]. Talin also binds to the actin network primarily through the actin bind site (ABS) at its tail.

Talin contains two distinctive domains, the head domain and the rod domain, connected by an unstructured linking region. The head domain contains the F0 subdomain and a FERM region (F1-F3) which has an integrin binding site (IBS1) and sites to bind signalling proteins (such as FAK and RIAM) [119]. The talin rod domain contains 13 alpha-helix coiled coils or bundles and a single alpha-helix subdomain that allows for the formation of a talin dimer complex, where talin binds to another talin [120]. In talin's resting state (unactivated), it has an auto-inhibited conformation where the F3 domain in the head of the protein binds to the R9 domain in the rod domain. This auto-inhibition occludes a number of binding sites for actin and vinculin [121]. Therefore, talin must first be activated before integrin binding can take place. The activation of talin occurs through the binding of Rap1 or RIAM/Rap1 complex, which displaces the R9 domain from the auto-inhibiting binding site in the head domain [37]. After Rap1 or RIAM/Rap1 binds and activates talin, it localises both of them to the cytoplasmic tail of integrin $\alpha IIb\beta_3$ in order to bind to and activate it [122–124].

Talin's rod domain can make use of transient unfolding and refolding of its alpha-helix coiled-coil subdomains (rod subdomain being another term to describe a bundle) to dampen sudden external forces [125]. A sudden, large force from a platelet impacting the edge of a thrombus can temporarily unfold a number of bundles, causing an extension in talin's length. Without a continuous external force the bundles quickly refold, preventing binding of ligands to cryptic binding sites. This produces a signalling response while also preparing the structure to withstand any future force shocks. Talin's extension reduces the magnitude of the force transmitted to the cytoskeleton and thus reduces the potential damage from excessive to the cytoskeleton or the contractile machinery. This behaviour has been measured and likened to a mechanical shock absorber which prevents excessive forces from being transmitted to other platelets in a thrombus and into the internal cytoskeleton [12].

1.2.8 Rod Domain Structure and Signalling

The 13 alpha-helix rod subdomains (R1-R13) contain binding sites for several proteins such as RIAM, vinculin, actin and a second integrin binding site (IBS2) [7]. The rod domains consist of either four or five alpha-helices that are folded in a coiled-coil conformation at resting state. The 11 vinculin binding sites (shown in red in Figure 1.2) are cryptic while the bundles are folded and only accessible when fully unfolded after sufficient tension has been exerted [6, 101, 125]. The unfolding of talin's rod domains enables it to act as a mechanosensing signalling protein, where specific binding sites are available or hidden depending on the force acting on talin [3]. For example, after Rap1 or RIAM/Rap1 binds to and activates talin, it is ejected from the surface of the alpha-helix bundle once sufficient force cause the subdomain to unfold. This leaves the cryptic binding site for vinculin accessible [126].

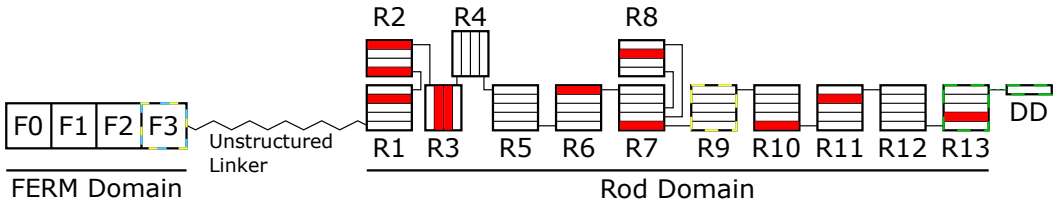


Figure 1.2: A graphical representation of the complete talin protein with labels of the FERM, unstructured linker and rod domain. All subdomains are labelled and the alpha-helices highlighted in red contain vinculin binding sites. The F3 domain highlighted in blue contains the primary integrin binding sites and the green highlighted R13 and DD domains contain the primary actin binding site. The auto-inhibition domains F3 and R9 are highlighted in yellow. Rap1/RIAM binds to F3, displacing the R9 domain and activated platelets.

During platelet spreading and even more so during platelet contraction, internal forces from actomyosin contraction and actin retro-grade flow apply force to the C-terminal end of talin [127]. The N-terminal region of talin is bound to integrin and the external environment. The pulling force from actin retro-grade flow on the C-terminal end of talin produces a tensile force across the length of talin's rod domain [128]. After an initial force threshold has been reached, the first alpha-helix bundle unfolds long enough to enable vinculin to bind to the cryptic binding site and prevent the subdomain from refolding [126]. The other end of vinculin is able to bind to the actin cytoskeleton. This additional binding increases the tensile strength and rigidity of the connection, thus enabling higher forces to be transmitted to ECM [129]. If the force across talin increases very quickly, not enough of the additional vinculin connections are made before the de-bonding force thresholds are reached and the focal adhesion dismantles. On

the other hand, if the force across talin increases very slowly or does not increase sufficiently, no additional vinculin connections are made resulting in low rigidity and low force transmission. This also leads to focal adhesion recycling [130]. The changing availability of binding sites is key to the management of the formation, maturation and dismantling of focal adhesions [7]. The force threshold at which each talin rod domain unfolds, determines the resulting signalling response such as the binding of a structural protein like vinculin, or the binding or ejection of another signalling protein. Understanding when talin produces its signalling response to forces will improve our ability to predict the dynamics of platelet activation, spreading, and contraction. Therefore, experiments that have attempted to determine these force thresholds (but only partially succeeded) are discussed in the following section.

1.2.9 Talin Stretching Experiments

Magnetic tweezers are a method for measuring forces across and extension of protein sections. Protein sections are bound between a static, flat substrate and a paramagnetic bead. Force is applied to the bead by a movable magnet and the magnitude of the applied force is controlled by altering the distance of the magnet to the bead. The position of the bead is determined using a charge-coupled device camera sensor that captures the interference pattern of light scattering off the bead from a collimated LED [131].

Magnetic tweezers were used by del Rio et al. to stretch a section of talin containing the first 12 helices of the rod domain, equating to the first three helix bundles minus R3's last helix [101]. The magnetic tweezers experiment was conducted by binding one end of a rod subdomain section to a flat substrate. The other end was attached to a magnetic bead. The distance between the substrate and the magnetic bead was controlled by manipulating the magnetic field around the bead. Unfolding events were detected through the binding of a fluorescence molecular to vinculin binding sites. At the end of the experiment, excess fluorescent molecules were washed away and TIRF microscopy was used to measure the intensity of the bound molecules. Higher intensities indictaed a higher number of unfolding events revealing vinculin binding sites occurred. Del Rio et al. used a force-clamp mode to create a displacement that resulted in a constant force through the section of talin's rod domain. Up to three unfolding events were recorded at a force of 12 pN.

Del Rio et al. then used AFM to measure the unfolding forces across the

same section of talin's rod domain. At this time magnetic tweezers had difficultly accurately measuring forces larger than 20 pN for short section of proteins such as the section of talin's rod domain used here [132]. Five unfolding peaks were detected between 29 pN to 51 pN with an average variance of 21.6 pN (Yao et al. determined that the range of force for which talin's domains unfold between is ∼5 pN to ∼25 pN [12], this will be discussed later in this subsection). The force ranges from del Rio et al.'s experiment are approximately a factor of two larger than the physiologically relevant range for talin. Additionally, the high variance in the forces makes it very difficult to determine which helix bundles were related to which force peaks. This is caused by the use of AFM to measure the forces. AFM uses a large apparatus and the cantilevers have relatively high stiffness. This results in a lower bound of the force range that can be applied by the cantilever and measured, decreasing the accuracy of low forces recorded [133].

Early magnetic tweezer experiments were unable to accurately measure forces larger than 20 pN for tether lengths shorter than 1 µm (tether length includes the length of the protein section). Unfolding protein subdomains usually require forces ranging from ∼1 pN to >100 pN [133]. Therefore, techniques had to be created to overcome the magnetic tweezer's 20 pN limitation. Chen et al. designed a novel force calibration approach to enable accurate, sub-pico newton force measurements in the range of 1 pN to 100 pN for use with short length protein segments and short strand DNA [132].

This new magnetic tweezer calibration technique was used by Yao et al. to measure the force-extension characteristics of the complete R1 through R3 rod domain segment of talin [8], compared to the incomplete R1 through R3 segment from del Rio et al. The rod domain segment was stretched at a force loading rate of 5 pN s^{-1} until fully unfolded. The segment began unfolding at ∼5 pN and was fully unfolded at ∼25 pN. Stretching was followed by a decrease of force to 0.5 pN and the rod section refolded completely within one minute. Multiple cycles of stretching and relaxing were measured on the same sample of talin and three major, independent unfolding events were observed in more than 10 cycles. At forces of ∼5 pN rapid equilibrium unfolding/refolding events were recorded. The R3 domain was identified as the domain producing these rapid unfolding/refolding events. This was done by generating a mutant form of the rod section that altered four threonine residues to hydrophobic valine and isoleucine. This stabilised the hydrophobic core of R3. After stretching experiments with this mutant form, the 5 pN unfolding event shifted to 8 pN. The result of these experiments was the

recording of three unfolding events at the ∼5 pN, ∼15 pN and ∼20 pN marks. Specifically, the ∼5 pN event was linked to the R3 domain while the remaining two could not be linked.

Yao et al. also showed that once vinculin binding sites were exposed and vinculin successfully bound to them, the alpha-helix bundles were unable to refold at lower forces, locking them in an unfolded conformation [8]. The vinculin head also dissociated from the talin rod domain at forces higher than 35 pN, possibly as a result of helix-to-coil transitions in the alpha-helices.

Further work by Yao et al. reported in a subsequent paper investigated the unfolding characteristics of the full rod domain of talin by separating the rod into four groups and stretching them [12]. All groups were stretched from 0.5 pN to 40 pN at a loading rate of $\sim 3.8\,\mathrm{pN\,s^{-1}}$. After all domains were unfolded, force loading was reduced to 0.5 pN for 60 s. This cycle of stretching and relaxing was completed multiple times for all four groups of rod domains. Yao et al. also recorded the unfolding response for the full length rod domain. A trace of the force-extension curves is shown in Figure 1.3. Twelve unfolding events were recorded between ∼5 pN to ∼24 pN for the full length talin rod. Eleven unfolding events had an extension of ∼30 nm to 40 nm and one had an extension of ∼80 nm, double the extension of the others.

It was determined that the longer 80 nm extension was caused by the simultaneous unfolding of the R7 and R8 domain. The R7 and R8 subdomains are unique in the talin rod domain because they are not in a sequential chain. Gingras et al. proposed that the four helices of the R8 subdomain are instead inserted between the third and fourth alpha-helices of the R7 subdomain [134]. The R7-8 subdomains together unfold at around 15 pN. Yao et al. took a separate measurement of the R8 domain from R7 and observed it unfolded at ∼5 pN. Combining Gingras et al.'s structural work and Yao et al.'s measurements, three of the thirteen rod domains can be linked to their corresponding unfolding events. The R3 and R8 subdomains to ∼5 pN and the R7 subdomain to ∼15 pN. This and the unfolding data for the full rod domain from Yao et al. will be used as a baseline comparison for the force results from the simulation developed for this thesis.

1.2.10 Talin Simulations

As only three of the rod domains have been linked to their unfolding events via experimentation, attempts have been made to calculate the unfolding forces via structural simulations. Below, several examples of simulations used to determine

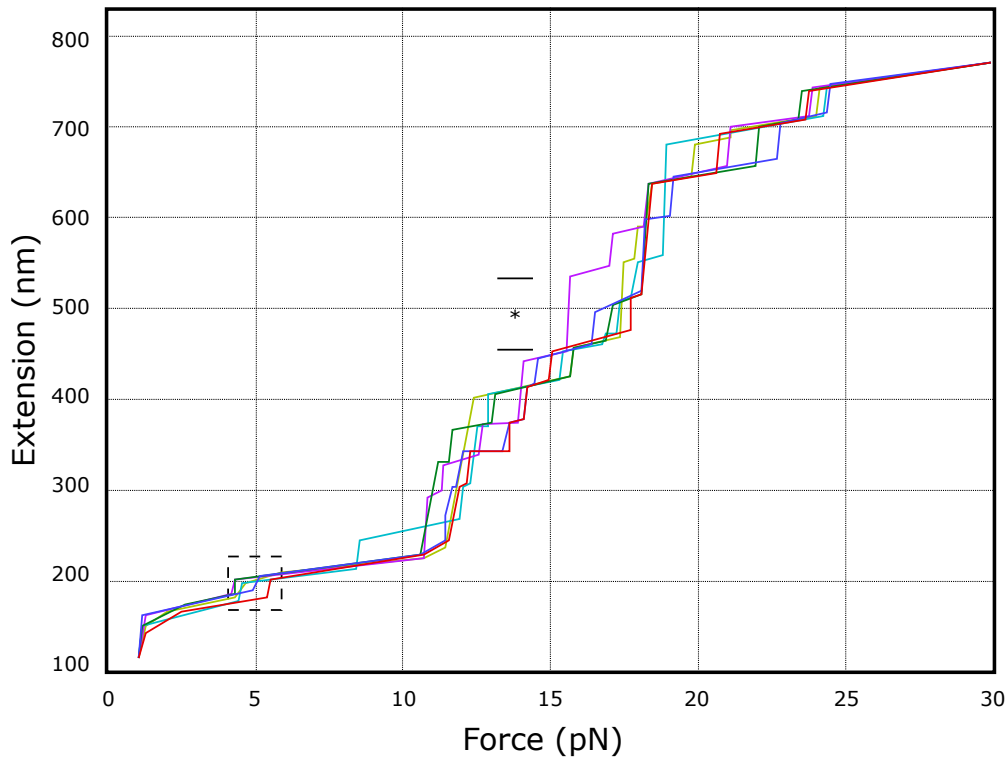


Figure 1.3: A trace of Figure 2a from of [12]. Multiple unfolding force vs. extension curves of talin rod domain at 3.8 pN s^{-1} loading rate. Each coloured curve corresponds to a seperate stretching cycle. The R3 domain unfolding event is highlighted by the box and the double length unfolding event highlighted with a '*'.

talin rod domain unfolding forces are described. The purpose of this section is to describe the current standard simulation methodology for molecular simulations of protein structures such as talin and show why this methodology is not suitable for accomplishing the goal of this thesis.

Using the data from their unfolding and refolding experiments of groups of talin rod domain bundles, Yao et al. developed a statistical model for the unfolding events of talin. The rod domain bundles were grouped into R1-2, R3, R4-6, R7-8 and R9-12. The physiological extension of talin *in vivo* is 50 nm to 350 nm [125, 135]. The unfolding and refolding rates were calculated for each group of rod domains and statistical simulations were conducted for the full rod domain. The whole talin rod domain was held at ∼100 nm and ∼200 nm. At ∼100 nm of extension, the average force across talin was estimated to be 4.73 pN. After extending to ∼200 nm, the force was 6.45 pN. Using this simulation and based on the average extension of talin, Yao et al. determined that six or seven alpha-helix bundles unfold and the mean force across talin's rod domain did not exceed 10 pN. In the simulation, the change in force across the rod domain was small as the ex-

tension was increased. This suggested that talin could unfold and refold domains in order to maintain a desired tension across its length, thereby mediating the forces transmitted through it. Unfortunately, this method cannot link additional alpha-helix bundles to their corresponding unfolding event. This is because the unfolding forces of the grouped rod domain bundles cannot be separated from each other. However, it does describe the maximum average force across talin and the number of subdomain that would unfold below that value.

Other types of simulations such as molecular dynamics have been used to try to determine the unfolding force thresholds for talin's rod domain. These simulations rely on protein structural data. Traditional all-atom molecular dynamics simulations (MD) calculate interaction energies for bonded and non-bonded interactions between the atoms or approximated structures in a specified model. At set simulation intervals, the simulation re-calculates the potential energy for all interactions to a very high level of precision. This has a large computational cost that does not scale linearly with the size of the model. Talin's rod domain is a large $\sim$210 kDa protein that is computationally expensive to simulate in all-atom molecular dynamics [136]. Additionally, talin's unfolding mechanics occur over the course of seconds [137]. Normal time scales for molecular dynamics simulations are in the order of picoseconds to nanoseconds, with femtosecond simulation intervals [138–140]. The femtosecond simulation intervals are required to avoid sampling errors (called discretisation errors) when calculating atomic positions [140]. Atom positions can fluctuate at very high frequencies and these position fluctuations need to be captured for an accurate simulation. Therefore, it is common to simulate smaller sections of the rod domain independently of each other and to use a modelling or simulation methodology that further reduces the computational complexity such as coarse-grained molecular dynamics or a modified force-field simulation.

Steered molecular dynamics (SMD) is a modification of the standard molecular dynamics approach. Rather than calculating an energy minimum for a protein model and possible ligands, SMD simulations apply an external force to one or multiple parts of a protein model and thus '*steer*' the unfolding or conformational changes in the model. This technique has been used by Haining et al. to unfold individual alpha-helix bundles of talin's rod domain by applying a constant velocity pull [9]. Two constant velocity rate experiments were conducted, one at $0.1\,\mathrm{nm\,ns^{-1}}$ and one at $2\,\mathrm{nm\,ns^{-1}}$ for 60 and 35 ns respectively. The resulting force vs. extension data for each alpha-helix bundle were grouped into three

categories: weak, intermediate and strong. Weak force bundles unfolded before 500 pN. This group contained the R3, R4, R6 and R10 bundles. The intermediate bundles R7(8), R11 and R12 unfolded between 500 pN and 1000 pN. Finally, the strong bundles of R5 and R9 unfolded above 1000 pN. Further work by Haining et al. focused on the R7 and R8 bundle, and using SMD they determined that the R7 bundle unfolds around 500 pN and the R8 bundle unfolds around 300 pN [141]. Other work using SMD by Mykuliak et al. also simulated talin rod domain alpha-helix bundles [142, 143]. At $2\,\mathrm{nm\,ns^{-1}}$ constant pull velocity, the R3, R8 and R9 bundles unfolded at 300 pN, 350 pN and 400 pN respectively.

Mykuliak et al. also used coarse-grained molecular dynamics simulations to investigate the R3 and R9 bundles. Coarse-grained molecular dynamics makes use of approximated protein structures to reduce the number of interactions that need to be calculated [144]. In the case of Mykuliak et al., united atoms were used to approximate whole amino-acids, resulting in a serial chain of united atoms that were positioned at each amino-acid's carbon-alpha. The electrostatic and chemical properties of each of the united atoms were determined based on statistical analysis of the properties of the original amino-acid. Because of the simplified protein model, a larger number of simulation intervals were used during the coarse-grained simulations. Additionally, a constant pulling velocity of 0.05 pN was chosen. The resulting forces for R3 and R9 peaked at 150 pN and 100 pN respectively. This method of using more approximated protein structures produced much lower forces than the all-atom or steered molecular dynamics approach. Still, they remain ten times larger than the physiologically relevant range of talin unfolding forces.

Compared to the physiologically relevant range of forces for talin ($\sim$5 pN to $\sim$25 pN), molecular dynamics simulations force results are erroneously large. Haining et al. reported forces in the range of 500 pN to 1500 pN, up to sixty times larger than the physiologically relevant range, and Mykuliak et al. reported forces in the range of 100 pN to 350 pN, up to 14 times larger than the physiologically relevant range. As such, they cannot be used to determine the unfolding forces of talin's rod domain or link those forces to their corresponding alpha-helix bundle. Haining et al. mentioned these inflated force values in their 2016 paper:

> "Due to the character of the simulations, calculated unfolding forces from SMD are only relative and cannot be directly compared to the experimental unfolding forces. [...] Moreover, small differences in structural changes result in large changes in the calculated force magnitude.

> Hence, SMD unfolding force is used here only for pattern analysis and talin rod domain stability classification, but not for providing predictions for the absolute unfolding force. [...] It is also good to keep in mind that the cellular processes may happen in the time frame of milliseconds to hours, not reachable by all-atom simulations." [9]

By the authors own admission, molecular dynamics simulations do not work if the goal is to calculate the absolute unfolding forces for structures such as talin's rod domain. Determining the exact unfolding force for the rod domains of talin enables the prediction of talin's signalling response to forces and therefore, the prediction of focal adhesion and platelet responses to force. Thus, in order to calculate the unfolding force values of talin's rod domain, a new simulation methodology needs to be developed that did not share the same limitations as molecular dynamics. This new methodology must: calculate directly in the force domain, rather than convert changes in potential energy to force; and have its architecture designed such that it can operate at the spatial and temporal scales of the unfolding of talin's rod domain, millisecond to second timescale and nanometre distances.

Some of the key points from this literature review for the remainder of the thesis are: integrins and talin form focal adhesions that platelets use to adhere to the ECM and thrombus. Talin is the key force transmission pathway between the platelet's contractile machinery and the thrombus/vascular injury. The unfolding of talin's rod domains determines talin's response to internal and external forces. Talin's full unfolding response has been measured through stretching experiments, but only three subdomains (R3, R7 and R8) have been linked to their corresponding unfolding event. Traditional molecular dynamics simulations of talin have resulted in calculated unfolding forces tens of times higher than the physiologically relevant range of forces for talin. Therefore, these cannot be used to link more domains to their corresponding unfolding events.

1.3 Forces and Force-Field Calculations

In the previous section, key biological concepts have been described. In this section, mathematical and engineering concepts that were used in the simulation are discussed in detail. These key concepts are: the molecular dynamics algorithm and general force-field equations; coarse-grained molecular models; and finally, intermolecular interactions. Various issues we're raised in regard to the results of

the molecular dynamics approach. However, that does not mean that all aspects of the approach were unsuitable and that there was nothing to gain by exploring these approaches further. Therefore, in this section concepts used in molecular dynamics are discussed and pitfalls to be avoided as well as usable elements for the purpose of this thesis are identified.

1.3.1 Force-Fields and Molecular Dynamics

Molecular dynamics is the computational biology field that parameterises the movement, interaction, and energy of atoms and molecules for use in computer simulation. In the late 20^{th} century, computer hardware was capable of simulating small proteins and molecular dynamics beause a useful tool for analysing biomolecules and protein structures. Some of the most popular molecular dynamics software still in use today are listed below in Table 1.1. The software were often developed alongside a custom force-field which originally were designed to replicate specific experimental findings or setups (AMBER[145], CHARMM[146]), or to improve a general aspect of molecular dynamics simulations (GROMACS[147], NAMD[148]).

Table 1.1: List of Popular Molecular Dynamics Software and Notable Differences or Goals.

Name	Created At	Original Differences or Goals
AMBER	University of California, San Francisco (1994)	Parameters Fitted for Simulation in Aquaous Solution
CHARMM	University of Harvard (1983)	Additional Improper Dyhedral Terms and Explicit Solvation
GROMACS	University of Groningen (1978)	High-speed Efficient Simulations
NAMD	University of Illinois (1995)	Highly Parallelised Computing Systems

For molecular dynamics, a force-field is characterised as a collection of equations and parameters that are used to calculate the potential energy of a system of atoms. The force-field equations in molecular dynamics are based on some of the same concepts as force-fields in classical physics. However, they describe the interatomic potential energies of the interactions within a system rather than calculating a vector-field of interaction force at several positions in space. Through the calculation of the potential energy, the position of elements in the protein structure are updated in an attempt to minimise the energy level of the system and thereby find the proteins most stable conformation.

The components of force-field equations can be separated into two groups: bonded and non-bonded potentials. The general equation for a standard force-field for bonded and non-bonded interactions is shown in Equation 1.1 following the notation of [149].

$$E_{total} = \sum_{bonds} k_r(r - r_{eq})^2 \tag{1.1a}$$

$$+ \sum_{angles} k_\theta(\theta - \theta_{eq})^2 \tag{1.1b}$$

$$+ \sum_{dihedrals} k_\phi(1 + \cos[n\phi - \gamma]) \tag{1.1c}$$

$$+ \sum_{impropers} k_\omega(\omega - \omega_{eq})^2 \tag{1.1d}$$

$$+ \sum_{i<j}^{atoms} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^{6} \right] \tag{1.1e}$$

$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{1.1f}$$

Where r is bond length, θ is atomic angle, ϕ is dihedral angle, ω is improper dihedral angle, r_{ij} is the distance between atom i and atom j, k_r, k_θ, k_ϕ and k_ω are force constants, r_{eq}, θ_{eq} and ω_{eq} equilibrium positions, the dihedral term is a periodic term characterised by a force constant k_ω, multiplicity n and phase shift γ, ϵ_{ij} is related to the Lennard-Jones well depth, r_m is the distance at which the potential reaches its minimum, q_i and q_j are the electrostatic charges of the respective atoms, and ϵ_0 is the dielectric constant. The bonded potentials include bond length (1.1a), bond angle (1.1b), proper (1.1c) and improper dihedral (1.1d). The non-bonded potentials include the electrostatic (1.1f) and van der Waal interactions (1.1e).

The terms represented in these groups scale between the world of quantum mechanics and classical physics (Figure 1.4). Because of this, the force-field parameters are often derived from multiple sources, including empirical approximation and quantum mechanics simulations [150, 151]. This thesis focuses primarily on the non-bonded, long-range electrostatic interactions. Terms 1.1a to 1.1d were ignored as their effect was smaller than the spatial scale of interest. In other words, while the effect of bonded terms above are to maintain the helices' structure, the focus of this work is on the interaction forces between alpha-helices. Therefore, at this scale, the effect of the bonded-terms on inter-helix interactions forces are

negligible. Additionally, the term in 1.1e was not considered. The Lennard-Jones potential, while it has a weak long-range effect, is an experimental relationship and its inclusion would interfere with just considering term 1.1f which is a version of Coulomb's Law.

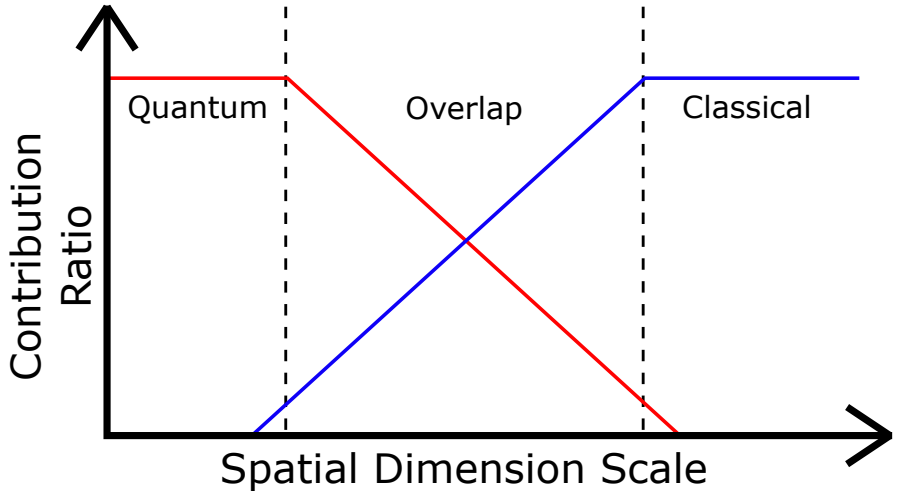


Figure 1.4: Diagram to illustrate the overlap between quantum mechanics (red line) and classical physics (blue line). The force-field bonded terms deal with atomic interactions and therefore are governed primarily by quantum mechanics. The term for the non-bonded long-range electrostatic interactions such as those between charged sidechain, is governed primarily by classical physical. Lennard-Jones, short-range electrostatic interaction and the hydrophobic effect occur within the overlap region.

The basic molecular dynamics algorithm (shown in Equation 1.2) seeks to minimise the potential energy value, E_{pot} (which includes all components from the force-field Equation 1.1), as this results in the most stable conformation of the simulated proteins. The steps of this algorithm are as follows. Where E_{pot} is potential energy, t is simulation time, dt is simulation interval. For each i^{th} atom of all N simulated atoms: x_i is the $\mathbb{R}^3$ atom coordinate, F_i is force component, a_i is acceleration, v_i is velocity and m_i is atomic mass. The trajectories are an ordered list of $\mathbb{R}^3$ atomic coordinates. For coarse-grained simulations, the atoms are simply substituted with their pseudo-atom representations. Pseudo-atoms are an approximation of positions and properties of real atoms used to combine multiple atoms into a single object.

1. Calculate the potential energy of the system of atoms using force-field equations.

$$E_{pot}(X_i) \tag{1.2a}$$

2. Determine the forces applied to each atom from the change in potential energy and position from the previous time-step.

$$F_i = -\frac{\delta E_{pot}}{\delta X_i} \tag{1.2b}$$

3. Numerically integrate the acceleration to determine the velocity and change in position of all atoms and update their trajectories. Here, the method of determining the change in velocity and position is a form of Euler Integration.

$$a_i = \frac{F_i}{m_i} \tag{1.2c}$$

$$v_i(t + dt) = v(t)_i + a_i dt \tag{1.2d}$$

$$x_i(t + dt) = x(t)_i + v_i dt \tag{1.2e}$$

As discussed in Section 1.2.10 of the literature review, when molecular dynamics simulations are used to calculate unfolding forces for proteins, the resulting values are orders of magnitude larger than expected. This indicates that something is going wrong either with the input data used, the force-field equations or the molecular dynamics algorithm. It is possibly a result of the complex interaction of quantum scale interactions and classical physics due to the spatial scales being investigated. The force-field equations shown in Equation 1.1 and the Euler integration steps in Equation 1.2 highlight this point. The first five terms of the force-field equation are classical empirical approximations for quantum behaviour, with the final term representing a fully classical law. These classical approximations are then integrated to find the resulting change in position as if they were determined by Newtonian interactions. As the research covered in this thesis is primarily interested in the interactions of force at the protein domain scale, the complex intersection of classical and quantum was avoided by choosing to operate only within the domain of classical physics. Working directly with forces also removed the required step of converting from potential energy to force values, further simplifying calculations.

1.3.2 Hydrophobicity

The six components shown in Equation 1.1 are often augmented with an additional non-bonded interaction called a solvent interaction. Solvent interactions

cover all effects of a protein or biomolecule simulated in aqueous solutions. Most importantly, this covers the hydrophobic effect which. This is an important factor of protein folding and considered equivalent in magnitude to the electrostatic interactions [152]. There are two common ways to include the effects of solvent interactions in the force-field equations: an explicit solvent where each water molecule is individually simulated in the same way as every other atom within an all-atom simulation, or via an implicit solvation component using either an accessible surface area (ASA) or a continuum electrostatic model [153].

Molecular dynamics simulations of large proteins, or over long time-scales, are in practive unable to use explicit solvation calculations due to the high computational costs. Implicit solvation is used instead. However, as the hydrophobic effect is an enthalpic effect on the whole system, it is hard to approximate it contribution to the potential energy of a system and harder still to resolve it to a force interaction. Implicit solvation calculations such as ASA where an approximated area of interaction is used to calculate a proportional energy potential, are empirical solutions that have been shown to good approximations of experimental measurements [154].

1.3.3 Structural Coarse-grained Approximations

In section 1.2.10 of the literature review, previous molecular dynamics research highlighted the difficulties of computing large protein structures at atomic spatial resolutions. These difficulties included the requirement of small simulation time-intervals to capture fast atomic motions and the impractically high computational costs of running long duration simulations, due to the large number of simulation objects. Therefore, to minimise these issues and to improve the computational efficiency of the simulations discussed in this thesis, a structural approximation similar to coarse-grained simulations was adopted. For that reason, coarse-grained approximations are briefly described here.

Coarse-grained structural approximation techniques such as C-alpha, C-beta and sidechain (CABS) [155], united residue (UNRES) [156] and Levitt-Warshel [157] simplify amino-acid residue structures to varying levels of detail. For example, the Levitt-Warshel model simplifies each residue down to two united-atoms (also called pseudo-atoms) corresponding to a backbone position and a sidechain position, and the CABS model approximates sidechains as two positions in a serial chain. The term pseudo-atom or united-atom is used to describe a single atom-like object that acts as the approximation for the positions and physical properties

of the corresponding real atoms that have been combined [158]. These approximations allow for faster simulation times or a larger simulation intervals as they reduce the number of objects in the simulation. For the sidechain approximation in this thesis, the Levitt-Warshel model was adapted. This was called the Single Pseudo Atom (SPA) approximation. The choice of structural approximation is covered in more detail in Section 5.2 of the discussion chapter.

1.3.4 Electrostatic Interactions

The key interaction in the simulation framework developed as part of this thesis is the electrostatic interaction. Using Coulomb's law, the force generated by each pair of sidechains between alpha-helices can be determined. Coulomb's law can be derived from Gauss's law, one of Maxwell's equations. A simple derivation of Coulomb's law for interaction forces between two point charges is given below.

Gauss's law in integral form states that the electric flux through an enclosed area of a surface is equal to $\frac{1}{\epsilon_0}$ times the total amount of charge contained in the region (Equation 1.3a). $\mathbf{E}$ is the electric field passing through the area $\mathbf{dS}$, $\mathbf{Q}_{total}$ is the total amount of charge and ϵ_0 is the permittivity of free space.

$$\text{The flux is: } \oint \mathbf{E}.\mathbf{dS} = \frac{\mathbf{Q}_{total}}{\epsilon_0} \tag{1.3a}$$

For a point charge q, a Gaussian sphere can be drawn around the point charge with a radius of r. According to symmetric consideration, the magnitude of the electric field E is equivalent at all points on the surface of the Gaussian sphere. For an area element $\mathbf{dS}$ on the surface of the Gaussian sphere, the angle between the electric field and the area element is defined as a face normal 0°. The flux through the area can be determined as follows:

$$\oint \mathbf{E}.\mathbf{dS} = \int \mathbf{E}\mathbf{dS}\cos 0^\circ = \int |\mathbf{E}||\mathbf{dS}| \tag{1.3b}$$

As the electric field is constant, the area element is integrated as follows:

$$\begin{aligned} \oint \mathbf{E}.\mathbf{dS} &= |\mathbf{E}| \int \mathbf{dS} \\ &= |\mathbf{E}|(4\pi r^2) \end{aligned} \tag{1.3c}$$

Therefore, substituting in Gauss's law, we get the equation below where $\mathbf{E}$ is the magnitude of the electric field at point distance r from the point charge q.

$$
\begin{aligned}
|\mathbf{E}|(4\pi r^2) &= \frac{q}{\epsilon_0} \\
\text{hence, } \Rightarrow |\mathbf{E}| &= \frac{q}{4\pi\epsilon_0 r^2}
\end{aligned}
\tag{1.3d}
$$

The direction for $\mathbf{E}$ is defined as the vector from the point charge to the position in space where E is calculated. The force applied to a point charge within an electric field is given as:

$$\mathbf{F} = q_0\mathbf{E} \tag{1.3e}$$

If a second point charge q_0 is placed at the same point where $\mathbf{E}$ was previously measured in 1.3d, then the magnitude of the force acting on the second charge is as follows, where $\mathbf{F}$ has the same direction as $\mathbf{E}$:

$$\mathbf{F} = \frac{q_0 q}{4\pi\epsilon_0 r^2} \tag{1.3f}$$

In short, molecular dynamics simulations operate in the energy domain, using force-field equations. Converting these energy potentials into output force may result in errors. Therefore, the simulation framework developed in this thesis operates directly within the force domain to bypass this issue. Coarse-grained approximations are used as they simplify simulation objects, allowing for less computationally expensive simulations. The new simulation focuses on electrostatic interactions as these are one of the largest force by magnitude in protein folding.

1.4 Literature Summary

As previously highlighted, platelets play a key role in haemostasis by closing vascular wounds. However, abnormalities in their signalling response, activation, or adhesion mechanics can lead to serious disorders and diseases. Talin activates integrin αIIbβ3, a major adhesion receptor in the platelet activation pathway that binds to exposed collagen-bound vWF or to fibrinogen and fibrin in the forming thrombus. Additionally, talin mediates platelet adhesion by signalling the development of nascent focal adhesions and their maturation. Talin-1 knockdown studies in mice platelets have shown that talin is required for integrin activation and a lack of talin results in fully impaired haemostatic function of platelets and failure to form thrombi [94, 122, 159, 160]. Therefore, talin is a major factor in platelet activation and subsequent adhesion dynamics, and to better understand platelets, our understanding of talin's response to forces and what forces result in

each signalling response need to be improved.

Previous experimental work has explored the unfolding dynamics of talin's mechanosensitive rod domain by stretching it using optical and magnetic tweezers [8, 101, 161, 162]. Because of this, the R3, R7 and R8 subdomains were linked to their unfolding force thresholds of ∼5 pN, ∼15 pN and ∼5 pN respectively. Statistical simulations have been developed using the unfolding and refolding rate kinetics of the rod domains [12]. From these results, it was determined that the average tensile forces across talin do not exceed 10 pN. Additionally, it estimated that only six or seven of the rod domains unfold at or below this force level. Therefore, only one domain has been determined to unfold in the physiologically relevant range of forces for talin, the R3 domain at ∼5 pN. Even though the R8 domain also unfolds at around the same force magnitude as R3, it is protected within the R7 domain which unfolds at forces greater than 10 pN.

Alternatively, theoretical approaches using molecular dynamics simulations have been used to study the unfolding of talin's rod domains. Molecular dynamics simulations calculate the potential energy of a protein structure through the use of parameterised force-field equations. However, due to the inherent limitations of molecular dynamics simulations when unfolding force values of protein structures, the unfolding force results have been erroneously large in magnitude. All-atom molecular dynamics models offer a way to get a detailed perspective of the atomic level of the interactions between structures within the alpha-helix bundles [141, 163]. However, they have the specific disadvantage that they can only simulate small portions of proteins over short time durations, often operating at the order of 1 fs to 2 fs to avoid sampling errors of the highest frequency motion within the simulation [164]. The computational cost of simulating protein structures at such high temporal accuracy forces the choice between the level of complexity of the structure or the length of the simulation. Additionally, when estimating rod domain unfolding forces, molecular dynamics simulations have calculated forces in the range of 0.1 nN to 1.5 nN. This is closer to the range of whole platelet contractile forces and up to 250 times larger than the estimated maximum force talin would normally experience.

To overcome the limitations of current simulation approaches, a new simulation framework and methodology were developed specifically for estimating unfolding forces in protein domains.

1.5 Thesis Overview

In order to answer the research question, the following elements will be discussed: In Chapter Two, the forces present in platelets and their core force transmission pathway are discussed, followed by a description of exploratory simulations of talin that informed the requirements for the alpha-helix model and simulation framework that were developed as part of this thesis. In Chapter Three, the alpha-helix model and its backbone and sidechain approximations are detailed, along with the generation of the models from pdb structural data. The mathematical definitions for the attachment of local coordinate frames are explained, followed by the implementation of the simulation framework and an example simulation. Chapter Four covers the validation of the alpha-helix model's backbone and sidechain components and the presentation of the initial static force simulation results. After this the structural optimisation work is explained, finishing in a comparison between experiment data and molecular dynamics data from literature to the original simulation results and the simulation results after structural optimisation. In chapter Five, a number of key issues that arose during and after the development of the simulation are covered, such as hydrophobic interaction approximation and segmenting the alpha-helix backbone model.

2 Forces in Platelets & Defining the Requirements for the Simulation Framework

In the previous chapter, platelets and their core force transmission pathway were discussed. This pathway links the internal contractile machinery of the platelet (and the cytoskeleton) to focal adhesions containing talin and integrin. The transmembrane integrins in the focal adhesions bind to exposed collagen-bound vWF and to fibrinogen and fibrin on thrombi [40, 41]. All forces that are generated by the internal contractile machinery in the actin cytoskeleton, and all the forces exerted on platelets from their external environment are transmitted through this force pathways [92, 93].

Platelets live in a force rich environment where they need to withstand, sense and produce forces. At the platelet scale, these forces range from 100 pN up to 70 nN. At the scale of proteins, such as talin and integrin, these forces range from less than 1 pN up to approximately 25 pN. Because of this, the structures that comprise the platelets' primary force transmission pathway need to be investigated.

Platelets must withstand and respond to numerous mechanical properties and forces, including: substrate stiffness, external shear forces, internal actin retrograde flow, and actomyosin contraction. There are several proteins and structures that generate, transmit and must withstand these forces. In this chapter, the force responses of the proteins and structures in this force pathway and talin's importance at the core of this pathway is discussed. First, actin filaments, myosin-IIa and actomyosin contractile fibres, integrins and talin are discussed below. This is to show that talin is the most complex and relevant protein in the force transmission pathway for this research, because of its impact on platelet signalling and structural reinforcement of focal adhesions. Then, using exploratory simulations, the key requirements of the final simulations are defined.

2.1 Exploring Force Responses of Platelet Proteins

2.1.1 Actin Filaments

Actin filaments (f-actin) are polymer chains of globular actin monomers (g-actin). ATP bound g-actin can bind together to form long actin filaments that grow in a single direction. Actin filaments are directional, they have a positive (+) end and a negative (−) end as shown in Figure 2.1. G-actins bind to the positive end to elongate the filament [165]. Actin filaments have three main roles in platelets: 1) support the cell and its membrane as part of the platelet cytoskeleton, 2) transmit contractile forces as part of the actomyosin contractile machinery, and 3) generate protrusive forces and extend the cell membrane during platelet spreading. The first two of these roles rely on actin filament's resistance to shear and tensile stresses [166]. The third, however, is more complex as the protrusive forces and resistance of the membrane to those forces affect the rate of actin filament polymerisation by making it harder for g-actin to reach the growing end of the filament.

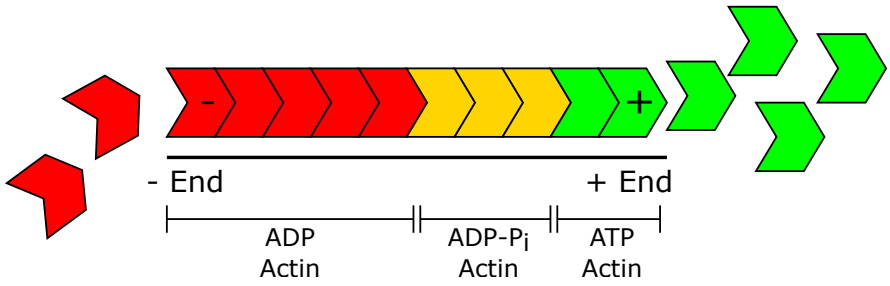


Figure 2.1: ATP-actin binds to the plus end of the growing filament. As the actin monomers within the filament age, their bound ATP hydrolyses to ADP-P_i and eventually the P_i detaches. After ATP hydrolysis, the actin monomer's bond to the actin filament is weakened and loss of its P_i weakens it further. ADP-actin dissociates from the filament as it reaches the minus end and this process may be assisted by protein such as ADF/Cofilin.

As the positive end of f-actin extends towards and presses on the cell membrane at the leading edge of a spreading platelet, g-actin has to force its way between the cell membrane and the positive end in order to bind and elongate the filament. As the membrane is stretched further, membrane tension increases and the normal force from increasing membrane tension presses back on the filament. This makes it harder for g-actin to reach the positive end of the filament and elongate it, resulting in a decrease in polymerisation rate as membrane tension increases [167, 168]. An estimate of the protrusive force per actin filament is $\sim$3.8 pN [169].

Actin filaments withstand the increasing compressive stress from the normal force caused by the membrane tension via a natural conformational change that

increases their resistance to compressive forces. As an actin filament compresses, the actin monomer components are pressed more tightly together, resulting in small shifts in their subdomain structure. These changes increase the compressive resistance of actin filaments [170]. Actin filaments also apply traction forces to focal adhesion structures during retrograde flow. This is also caused by the membrane pushing actin filaments rearwards as they polymerise. However, other than a change in their mechanical properties and alter rates of filament elongation, actin filaments do not have a large or complex response to forces, they are primarily structures that transmit force. Therefore, they are useful in understanding how forces are transmitted to other force sensing structures but do not produce a complex response in their own right.

2.1.2 Myosin Motor Proteins

Myosin-IIa (referred to as myosin from here onwards) is a motor protein that catalyses bound ATP to ADP and converts the chemical energy into mechanical force. Together with actin filaments, myosin generates platelets' contractile force. Myosin has a head and a rod domain connected by an intermediate domain called the 'lever-arm', with an ATP binding site near the connection between the head domain and the lever-arm. The rod domain binds to other parallel myosin rod domains and the head domain can bind to actin filaments. During force generation, the myosin head domain and lever arm rotate from a position in front of the rod domain to a position behind it. This is called the myosin powerstroke. If an actin filament is bound to the rod domain at this time, it is pulled past the head domain region towards the rod region before being released [171].

The average powerstroke displacement of myosin is $\sim$10 nm resulting in force in the range of 2.3–6 pN [172–175]. On average, platelets contain 12,000 myosin proteins [176]. It has previously been estimated from average platelet contractile forces that 25% of myosins contribute to contraction at any given time, which is in line with estimates of myosin contributions in other cells [114]. Myosin- requires available Mg2+ and ATP in order to carry out its powerstroke and produce forces. Additionally, myosin requires the activation of its regulatory light chains (RLCs). A number of proteins are known to activate RLCs, but the most common are myosin light chain kinase (MLCK) and Rho-associated, coiled-coil kinase (ROCK) [177, 178]. These two kinases are activated during platelet activation via the release of Ca2+ and activation of RhoA.

Once activated and supplied with Mg2+ and ATP, myosin can bind to actin

filaments and execute the powerstroke by ejecting the phosphate ion produced during ATP hydrolysis. It resets the lever-arm by ejecting ADP and binding to a new ATP molecule [173]. This continues until the myosin's RLCs are deactivated, it runs out of ATP and Mg2+, or it stalls during a powerstroke as external forces resist actin filament contractile movement. The force generated per powerstroke was measured to vary depending on the length of the power stroke [172]. Therefore, the primary purpose of myosin is to produce contractile forces. These forces can vary based on the length of the myosin powerstroke, but do not procure a more complex force response.

2.1.3 Integrins

Integrins bind platelets to their external environment after activation by the binding of talin and kindlin to its cytoplasmic tail [38]. All forces exerted on a platelet by their environment and produced inside the platelet, such as by actomyosin contraction and actin retrograde flow, are transmitted through integrins. The primary external force exerted on platelets, other than the contraction of other platelets, is from blood flow [179]. The flow of blood applies shear stress to the exposed surface of the platelet. This shear stress increases with increasing blood velocity, decreasing vessel diameter, and the closer the platelet is to the vessel wall [179]. Individually, integrins do not produce a signalling response to force, but they facilitate force transmission to the focal adhesion structures where signalling occurs [180, 181]. However, integrins do have a measured debonding force which could be considered a response to force. If forces exceed on the platelet increase boyond integrins' debonding force threshold, integrins debond from the fibrinogen and fibrin filaments, which results in platelets detaching from the surface they are adhered to.

However, integrin debonding is unlikely to happen during normal haemostasis. On their surface, platelets express $\sim 80,000$ αIIbβ3 integrins [14]. Full platelet contraction produces 19 nN of force on average [114]. Single integrin debonding forces range from 50 pN to 80 pN when bound to fibrinogen [182].

For example, to show that integrin debond is rare, let 65 pN be the integrins debonding force threshold. Given the total platelet contractile force, only 293 integrins are required to be bound in order to maintain full adhesion to withstand platelet contraction without debonding. Inversely, assuming 30% of integrins on the surface of the platelet are involved in platelet adhesion during contraction, each integrin needs to withstand 0.79 pN of force. This is much less than the

debonding force value of 65 pN. Therefore, while integrins do not actively respond to forces, they can debond if the magnitude of forces increases beyond set level. However, these debonding events take place at force levels far higher than those that occur during normal cell behaviour. Thus, the debonding of integrin does not occur during normal focal adhesion behaviour, and therefore is not relevant for the aims of this work.

2.1.4 Talin

Talin is an interesting protein as it is much less understood compared to the other proteins and structures already discussed. It has been hard to investigate it experimentally or through models as it is a large protein with a wide range of responses [9]. It has a complex mechanical response to force, acting like a shock absorber and it has a complex signalling response that is also force dependent [7]. The remainder of this subsection, describes talin's responses to force and highlights its importance in regulating the platelets' signalling response to mechanical stimuli.

Talin's response to force is the most complex and diverse of all the previously mentioned proteins and structures. Each of talin's rod domains can unfold and bind different ligands [7]. The majority of the cryptic binding sites across the rod domain are for the 11 vinculin binding sites [183]. Vinculin binding and reinforcement of focal adhesions is vital for the maturation of focal adhesions, which transmit the contractile forces generated within the platelet [93].

The dynamics of the unfolding and refolding of the rod subdomains of talin give its response to force a number of properties. These properties include: discrete force sensing per rod domain, time-dependent responses to force through ligand binding rates, and memory of past levels of force based on ligands bound to cryptic binding sites. The combination of these properties allows for very complex signalling behaviours to emerge. By determining the unfolding force thresholds for each of talin's rod subdomains, these complex responses can be predicted, and simulations of talin's responses during the haemostatic function of platelets to be developed. The details of these mechanical responses to force, and the resulting signalling responses are discussed in further detail in Section 5.5. Below is a list of responses talin exhibits and the possible signalling responses. These are listed in Table 2.1 and the transitions are represented as a state diagram in Figure 2.3. The responses are visualised in Figure 2.2.

Table 2.1: List of possible signalling responses of talin's rod subdomains. Shown as states and transition events. The letters in the first column of this table link to letter in Figure 2.2 and in Figure 2.3. Transitions 'f', 'g' and their events are marked with an asterisk as these are special cases. The rod subdomain cannot refold fully as there are ligands bound internally. As there is no unfolding force threshold, these rod subdomains do not unfold normally.

	Original State	Event	Resulting State	Signalling
a	Ext-Ligand Bound	Force Up, Unfolding	Ext-Ligand Ejected	Signal Removed
b	No Ligands Bound	Force Down, Refolding	Ext-Ligand Binds	Signal Restarts
c	Ext-Ligand Bound	Force Up, Unfolding	Int-Ligand Binds	Signal Switching
d	No Ligand Bound	Force Up, Unfolding	Int-Ligand Binds	Signal Starting
e	No Ligand Bound	Force Down, Refolding	No Ligand Binds	No Signalling. No Change
f*	Int-Ligand Bound	Force Up, Unfolding*	Int-Ligand Bound	Signalling. No Change.
g*	Int-Ligand Bound	Force Down, Refolding*	Int-Ligand Bound	Signalling. No Change.

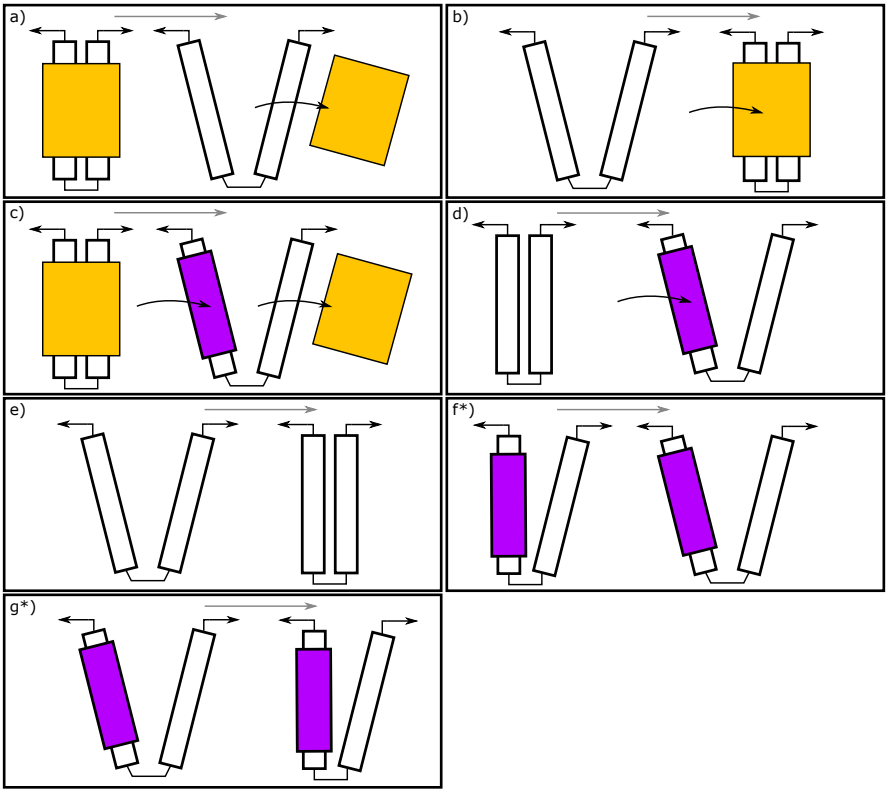


Figure 2.2: (a-g) Rod subdomain signalling diagrams. External ligand (yellow). Internal ligand (purple). Lettering matches with Table 2.1 and Figure 2.3.

The three states shown in Figure 2.3 describe the three possible signalling states of a talin rod subdomain. 1) A ligand is bound to the outside of a folded subdomain and signalling, 2) there are no ligands bound to the outside or to the exposed cryptic binding sites resulting in no active signalling, and 3) a ligand is bound to the exposed cryptic binding site and signalling. Transitions between states involve the unfolding or refolding of a subdomain and the subsequent binding, ejection or switching of bound ligands [184]. The transitions occur the unfolding or refolding force threshold of the subdomain is reached. A predictive model of the subdomains signalling response could be built if force thresholds for each rod subdomain, and the binding and unbinding rates of each ligand were known. This predictive model may be used to describe the changing signalling response of talin as force increases and decreases.

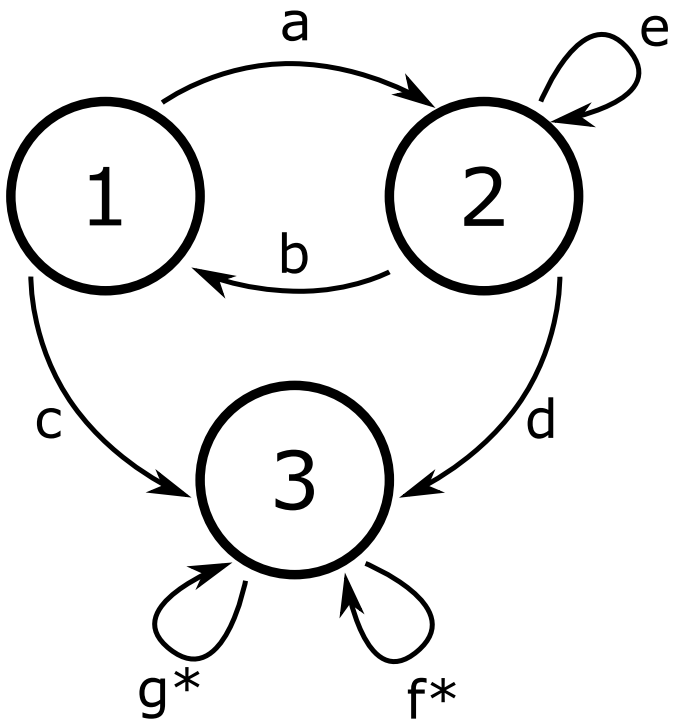


Figure 2.3: State transition diagram of a talin rod subdomain. State 1) External Ligand Bound. State 2) No External or Internal Ligands Bound. State 3) Internal Ligand Bound. Transition letters link to the state transition listed in Table 2.1 and in Figure 2.2.

The most interesting response is produced by state three, the bound cryptic ligand. Cryptic ligands like vinculin bind to a rod subdomain after it has unfolded. If the force level drops, vinculin stays bound and the alpha-helices are prevented from refolding correctly. This interrupts the normal intermolecular forces between helices that maintain the folded state of the subdomain, and results in a removal of the unfolding force threshold. Thus, if the force level increases again, the bundle

unfolds as soon as tensile force is applied to it and with the vinculin still attached. Vinculin provides structural reinforcement for the focal adhesion complex. This combination of properties of vinculin and the rod subdomain produce a memory like response. This is because the binding and reinforcement by vinculin persists after force levels drop below the unfolding threshold. Additionally, it prevents the refolding of the subdomain and the binding of potential external ligands. In this way, talin's signalling response contains a memory of past events, as they change the current signalling behaviour of talin even after the original events have occurred.

In addition to it complex signalling response, talin sits at the core of focal adhesions structures and reacts to all force generated by internal structures and transmitted through integrin [161]. Talin was selected for further mechanical study because of its impact on platelet signalling and structural reinforcement of focal adhesions. Another factor in choosing talin was the lack of data surrounding the unfolding force values for each rod subdomain, as only three rod subdomains have been linked to their unfolding force threshold as discussed in Section 1.2.9. If the unfolding force thresholds for each rod domain were linked to their corresponding signalling response, it may be possible to predict signalling responses of talin throughout platelet's haemostatic process and determine if talin can be a target of therapeutic treatments for platelet based disorders.

2.2 Exploratory Simulations and Modelling of Talin

One of talin's roles in platelet focal adhesions has been likened to that of a mechanical shock absorber [12]. A shock absorber in its simplest form is a mass-spring-damper. In the field of engineering, it is common to try to model new systems using as simple components as possible. It helps to determine the fundamental behaviour of a system and identify which components are likely driving this fundamental behaviour. This is a similar concept to using model cells such as platelets or model organisms such as zebrafish [185–187]. By recreating behaviour in simple environments, the mechanisms of the behaviour can be better understood without being disturbed. Therefore, talin was explored by its behaviour was modelled using springs and other simple mechanical components.

To start, a simple model of talin's behaviour was developed using mechanical analogues. The purpose of this model was to explore the core components of talin and its rod domain to see whether mechanical analogues could indeed be used to replicate talin's behaviour. It also served to clarify what parts of talin would be

required to build a functioning model and simulation.

In order to describe the development of the talin behavioural model, a quick summary of talin's structure and specifically its rod domain is given below. Talin's macrostructure has three main domains: the head domain, an unstructured linker domain, and finally the rod domain [119]. The head domain contains the primary integrin binding site and the talin auto-inhibition site as shown in Figure 2.4. Displacement of talin's rod domain from the inhibition site by Rap1 or Rap1/RIAM makes the integrin binding site available to other ligands [121]. The head domain is connected to talin's rod domain via a linker region consisting of an unstructured amino-acid chain. The term unstructured refers to the lack of secondary structures such as alpha-helices or beta-sheets in this region. The rod domain is the largest domain of talin and can be broken down into 13 alpha-helix coiled-coil domains and a single dimerisation domain. The dimerisation domain allows for two talins to bind together at their tail ends [120].

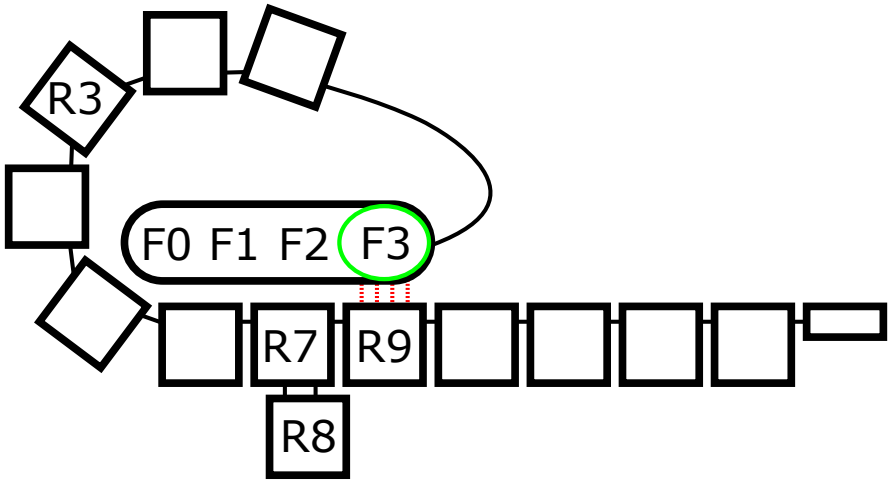


Figure 2.4: Diagram showing the auto-inhibited state of talin. The green ring highlighting F3 indicates the RIAM binding site that activates talin by interrupting the auto-inhibition site this is highlighted in red. Domains R3, R7, R8 and R9 are shown for reference.

The rod domain's 13 alpha-helix coiled-coil subdomains contain either four or five alpha-helices. The R2, R3, R4 and R8 subdomains have four helices, but the others have five helices [4]. The R8 subdomain is unique in its placement. All other rod subdomains are in a serial chain, however, the R8 subdomain is inserted between the third and fourth helices of the R7 subdomain. Therefore, the R8 subdomain only experiences tensile forces when the R7 subdomain has already unfolded [134]. The rod domain contains a number of binding sites for different proteins such as RIAM, paxillin and a second integrin binding site, but

two important binding sites for talin are the actin binding site that covers the R13 and dimerisation domains, and its vinculin binding sites [7]. All but the vinculin binding sites of talin are exposed on the surface of the alpha-helix coiled-coil domains. At this time, the only cryptic binding sites hidden inside the rod subdomains are vinculin and are only accessible after the subdomain has unfolded. There are eleven vinculin binding sites in total and a list is given in Table 2.2 along side the other key binding sites.

Table 2.2: A list of the domains of talin, the number of alpha-helices in the rod subdomains and the key binding sites for auto-inhibition, Rap1, RIAM and vinculin. This table is abridged and only contains binding sites relevent to this thesis.

Domain	Helices	Key Binding Sites
F0	–	–
F1	–	–
F2	–	–
F3	–	R9 (Auto-inhibited), Rap1, RIAM/Rap1, Integrin
R1	5	Vinculin
R2	4	Vinculin ×2
R3	4	Vinculin ×2
R4	4	–
R5	5	–
R6	5	Vinculin
R7	5	Vinculin
R8	4	Vinculin
R9	5	F3 (Auto-inhibited)
R10	5	Vinculin
R11	5	Vinculin
R12	5	–
R13	5	Actin, Vinculin
DD	1	Actin

The unfolding of talin's full rod domain has been recorded by Yao et al. where they stretched all 13 rod subdomains [12]. This experiment produced a set of force vs. extension plots of the rod domain. The tensile force across the rod domain was increased and the extension of talin was measured. The data included sudden increases in extension of approximately 30 nm to 40 nm, indicating the unfolding of rod subdomains. A manually digitised version of their results graph is shown in Figure 2.5. Each vertical segement of the graph indicates an unfolding event.

From other data collected by Yao et al. from stretching experiments of other talin rod domain constructs, the R3 and R8 subdomains unfold at ∼5 pN and the

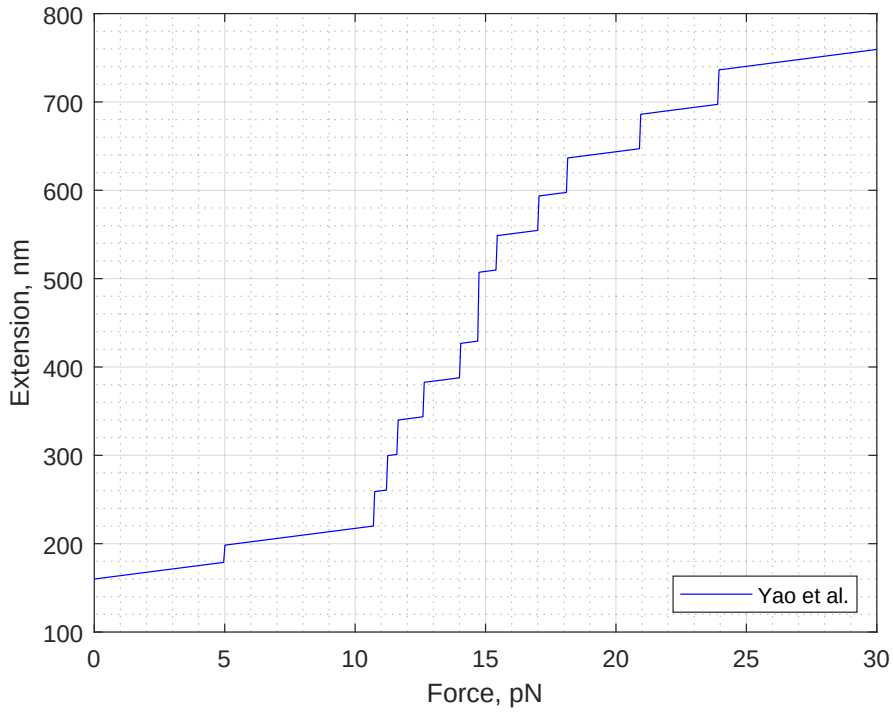


Figure 2.5: A manually digitised version of the talin rod domain force vs. extension graph from Yao et al. [12].

R7 subdomain unfolds at ∼15 pN [12]. The full force vs. extension data and the three unfolding force threshold values were used as a baseline to compare with the models and simulations developed in this thesis.

2.2.1 Modelling Talin's Behaviour as a Series of Springs with Catches.

One of the initial models of talin that was developed, was of talin's extension as tensile force was applied. This model did not predict the unfolding response but instead mimicked it using (simple mechanical component such as) springs. Each rod subdomain was modelled as an independent section. Experimental work from literature has detailed the unfolding of individual rod subdomains where they remain stably folded until their force threshold is reached [12]. After the threshold level is reached the folded alpha-helices unfold rapidly, resulting in a rapid elongation of the subdomain. Initially, attempts were made to model a rod subdomain as a simple step-wise function or as a series of linear spring-like components. The step-wise function modelled the unfolding behaviour well, but did so using the artificial bounds of step-wise functions that do not exist in mechanical analogues. The series of linear springs were unable to match the

unfolding response curve because of non-linear response of unfolding curve and an uncapped extension.

The best approximation (shown in Figure 2.6) was a combination of three components: a single linear spring followed by a non-linear spring surrounded by a catch mechanism. The linear spring component modelled the slow and constant extension that occurs as the rod subdomains are stretched but are not unfolding as seen in the flat line sections between 5 pN to 10 pN in Figures 2.5. The response of this component is shown in the top panel of Figure 2.7a. The non-linear spring and catch component set up in parallel generated the unfolding behaviour. This two component set-up allows for a minimal extension from the linear spring as force is increased. Near the force threshold, the non-linear spring begins to extend rapidly before being caught after extending 40 nm. This response is shown in the bottom panel of Figure 2.7a. The non-linear spring was designed to have an exponential growth response to force. Below the force threshold, the spring constant is very high, resulting in minimal extension. As the force nears the threshold, the extension undergoes rapid exponential growth with a very low spring constant. The extension increases rapidly before being caught by the distance catch. The combined response of the linear spring, non-linear spring and catch is shown in Figure 2.7b.

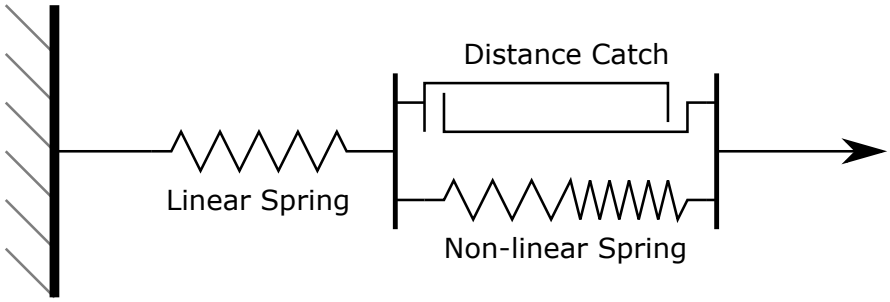
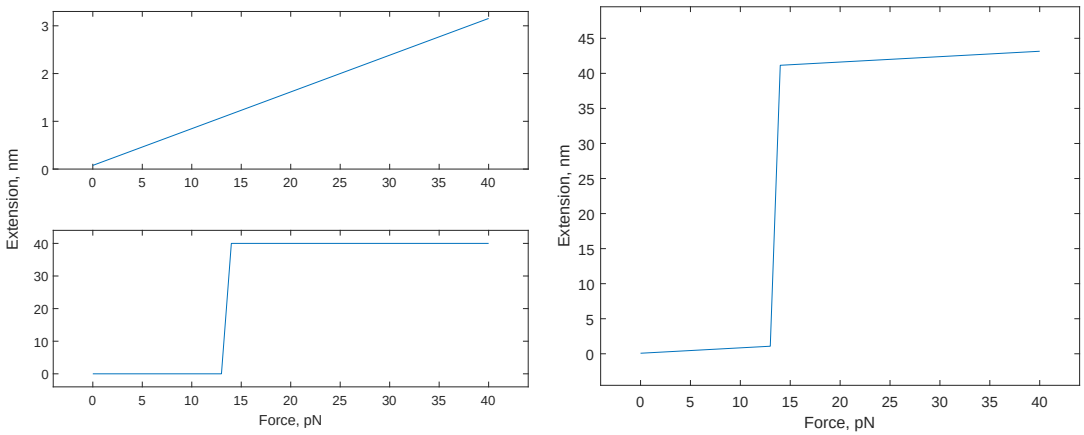


Figure 2.6: Component diagram of hybrid spring model for a single subdomain. Left to right: Immovable surface component, linked to single linear spring with high spring constant, linked to two components in parallel: 1) a non-linear spring with an exponential response 2) a catch in parallel that prevents spring extension beyond a set distance (30 nm to 40 nm).

Using the non-linear spring and catch model with the spring constant transition point tuned to the unfolding thresholds from Yao et al.'s data, a full unfolding force plot (Figure 2.8) was generated. This plot shows that talin's rod domain unfolding behaviour can be modelled by taking each rod subdomain as a single component. Once the subdomain unfolding thresholds are known, this model will be a very efficient approximation and useful for later simulations on signalling response. However, this type of modelling can not be used as a predictive tool as



(a) (b)

Figure 2.7: a) Top-panel: Linear Spring Response. Bottom-panel: Non-linear spring and Catch response. b) Full response of single hybrid spring model. The spring and catch parameters were tuned to the force and distance ranges of talin, ∼0 pN to ∼40 pN and 0 nm to ∼40 nm.

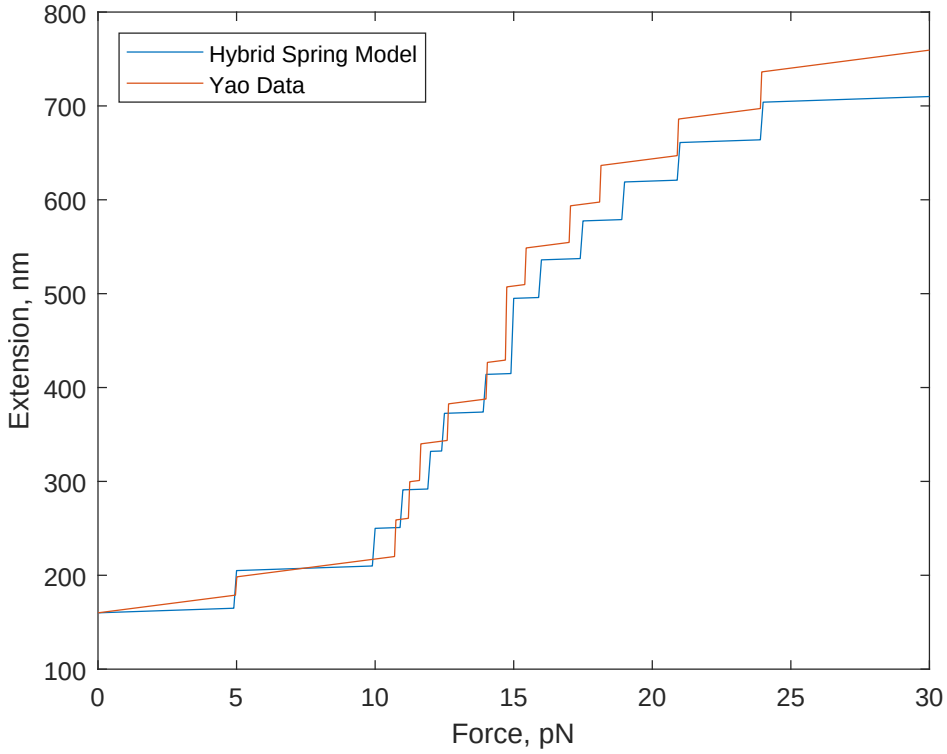


Figure 2.8: Combined hybrid spring response for each rod domain bundle. Unfolding values digitised from [12] and original curve also shown for comparison.

the unfolding behaviour is mimicked and not based on the individual properties of the rod subdomains. Therefore, other types of modelling were explored.

2.2.2 Simulation of Simple Alpha-Helix Models

One of the easily observable physical properties of the rod subdomains are the different lengths of the alpha-helices. Each helix within a bundle has a different length. Figure 2.9 shows a ribbon diagram of the first two alpha-helices from the R7 domain, where the second helix is over 50% longer than the first.

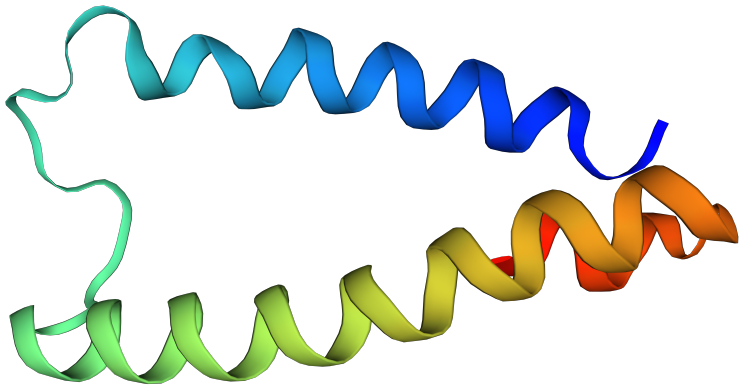


Figure 2.9: Diagram of two helices from the R7 domain using the 428p.1 PDB structure. Helix one in blue is 17 amino-acid residues long and helix two in green/orange is 27 amino-acid residues long.

The assumption was made that longer helicecs with more sidechains would be more attractive to other helices. If this assumption held true, there was a possibility that the length of the helices on their own would be enough to estimate the unfolding forces of a subdomain. However, a comparison of the lengths of helices in each rod subdomain to experimentally measured unfolding forces revealed no strong correlations. Therefore, just the lengths of the helices did not provide enough data for the estimation. However, in the process of checking this assumption, it was noted that the length of the alpha-helices did affect the number of sidechains exposed on the surface of each alpha-helix. The alpha-helix sidechains contain the atoms and compounds that produce the majority of the alpha-helices's chemical and physical properties. The longer the helix, the more sidechains exist, and the more complex and strong the chemical or physical properties can be. In order to resist the forces attempting to unfold an alpha-helix bundle, attractive forces must pull the alpha-helices together. Only after those forces are overcome does the bundle unfold. There are a number of intermolecular forces that contribute to maintaining the folded rod subdomains: electrostatic, hydrophobic, salt-bridges, dipole and van der Waals. The two largest in magnitude are the

electrostatic interactions and hydrophobic effect [188].

For the initial model, a single type of interaction was selected to simplify the implementation. Electrostatic interactions were used as the primary interaction force in the first model of alpha-helix interactions. This is because the hydrophobic effect can not currently be estimated or approximate as a direct force. Even its effects as a change in potential energy of the system are heavily approximated in the most common implicit solvent calculations. Hydrophobic interactions were explored and the development of a suitable approximation was attempted as part of this research project. However, it was determined that this work was beyond the scope of this thesis. The progress and methodology used are discussed further in Section 5.3. As will become apparent in section 4.2, just using the electrostatic interactions resulted in forces in the physiologically relevant range of talin.

In the next simulation, models that were more representative of the alpha-helices were used. While modelling using simple mechanical components is useful in understanding basic behaviours, determining how these behaviours occur requires models with more detail that replicate the structure of the system more closely. The next simulation did this by using simple models that represented individual helices with points of interaction approximating the role of sidechains, to calculate the interaction forces between them.

The models of interacting alpha-helices were simple two-dimensional representations with 'interaction points' at set unit intervals along the length of the alpha-helix representation. The simulation had two 2D helix models with the interaction points only the side facing the other helix. The interaction points always produced an attractive interaction between interaction points on the opposite helix. During the simulations, the helices were moved apart and the affect of helix length (and therefore number of interaction points) was tested. A diagram of the helix models with their interactions points is shown in Figure 2.10. The interactions were calculated using a simple Coulomb's Law implementation and each interaction point on one helix interacted with all other points on the opposite helix. Repeating pairs were excluded.

Equation 2.1 shows the formula used in the simulation where h_1 and h_2 refer to the two helices, N_{h1} and M_{h2} are the number of interaction points on each helix, and d_{ij} is the Euclidean distance between the i^{th} interaction point of helix one and the j^{th} interaction point of helix two. In the simulation, the separation distance d was started at a value of 2 and was increased to 27 in increments of 1. The starting value of 2 was chosen to prevent an infinite force value.

$$F(h_1, h_2) = \sum_{i=0}^{N_{h1}} \left(\sum_{j=0}^{M_{h2}} \left(\frac{1}{{d_{ij}}^2} \right) \right) \tag{2.1}$$

The force vs. separation curves from this simulation and their data are shown in Figure 2.11 and Table 2.3. The 'percentage interaction force' axis in the figure is scaled around the maximum force of curve *a* with helix lengths of 25 taken from the table. The remaining five curves were normalised to that value. This allowed for the comparison of the effects of different helix lengths on the maximum and minimum interaction forces.

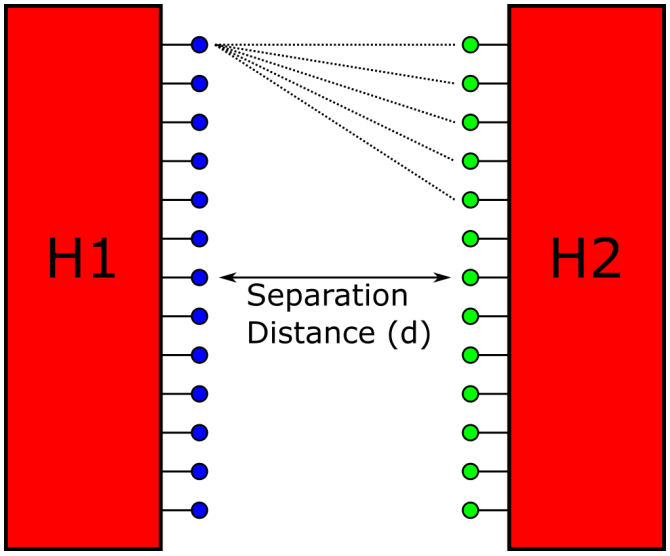


Figure 2.10: Diagram of two simple helix models. The left helix's interaction points are shown in blue and the right helix's in green. The dotted line indicates the point-to-point interaction forces. The models used in the simulation have the interaction points embedded on the edge of the helix. This was done to increase visibility in the diagrams.

Table 2.3: Result from the six simple helix model separating experiments. Curve letters link to the graph in Figure 2.11.

Curve	Helix Lengths	Max % Force	Median % Force	Min % Force
a	25:25	100	12.2	4.51
b	25:30	115	15.6	6.00
c	25:20	89.6	11.5	4.28
d	30:30	125	16.3	6.23
e	20:20	75.6	8.41	3.01
f	30:20	95.3	13.8	5.41

As expected, the response curves of the distance vs. force data matches an inverse square law. Additionally, increases in length of helices and the number

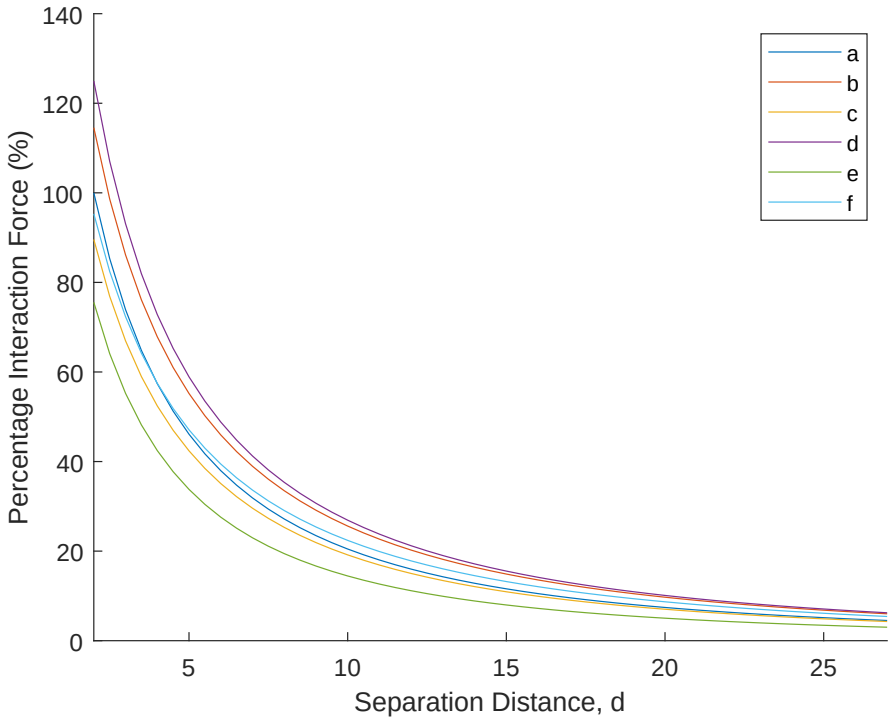


Figure 2.11: Six graphs of percentage interaction force vs. separation distance, d, for six pairs of helices with varying lengths. (a) Both helices the same length, same number of interaction points. (b) Second model is longer by 5 units of distances with 5 more interactions points. (c) Second model is shorter by 5 units of distances with 5 fewer interactions points. (d) Both models are longer than original length. (e) Both models are shorter than original length. (f) First model is longer and second model is shorter.

of interaction points resulted in higher attractive forces. Therefore, the length of each alpha-helix and the number and type of sidechains need to be modelled together as they directly determine the chemical and physical properties of the alpha-helices. The data for the sidechain positions, types, and properties was taken from the Protein Data Bank (PDB) [189, 190].

After the separating simulation, a number of simulations were developed where the alpha-helices unfolded by rotating away from each other. Various types of linkages and their effects on the unfolding motions of the alpha-helices were explored and the diagrams in Figures 2.12, 2.13 and 2.14 show three configurations of hinges and unfolding motions. These configurations are described here to illustrate the pattern where the linkage between the helices did not affect the unfolding force values, but did affect the final motion of the helices. These included linkages where the base of the alpha-helices maintained a fix distance from each other as they unfolded, or where they separated as they unfolded.

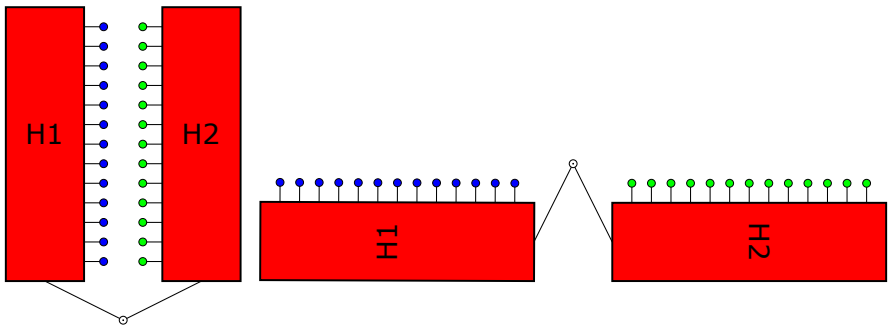


Figure 2.12: Before and after diagram of simulation using two simple helix models, both with the same length. V-shape hinge with the point of rotation at the base of the V. Axes of rotation are marked with a small circle with a dot at its centre.

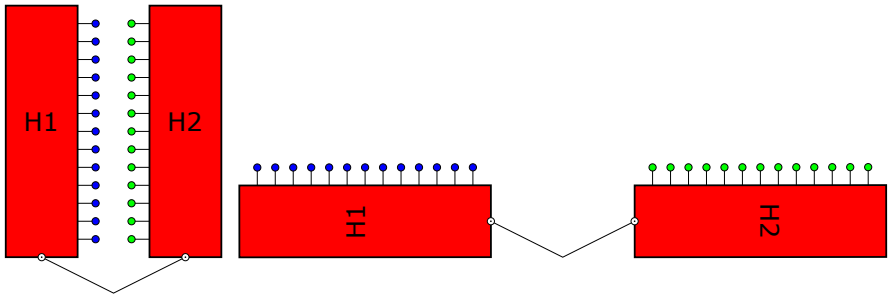


Figure 2.13: Before and after diagram of simulation using two simple helix models, both with the same length. V-shaped hinge with the point of rotation at the base of each helix. Axes of rotation are marked with a small circle with a dot at its centre.

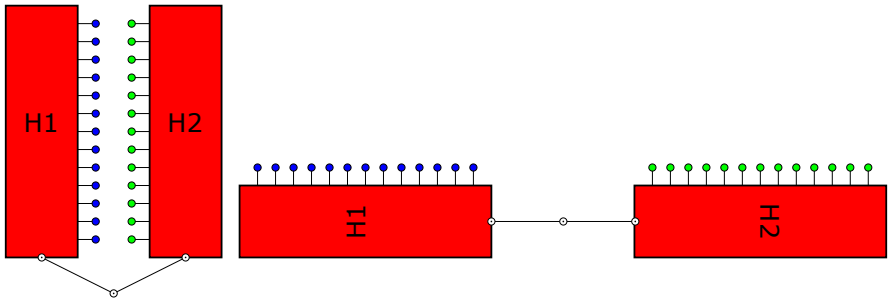


Figure 2.14: Before and after diagram of simulation using two simple helix models, both with the same length. 3) V-shaped hinge with the point of rotation at the base of the helix and the base of the V. Axes of rotation are marked with a small circle with a dot at its centre.

The force vs. angle curves generated from each simulation were plotted and are shown in Figure 2.15. All three curves start at the same unfolding force value indicating that the type of hinge and unfolding motion does not affect the unfolding force threshold. However, the hinges do alter the response of the tail end of the curve. This is highlighted by the blue curve (line a) in Figure 2.15, where the end of the curve can be seen slightly increasing in gradient after $\sim 70^{\circ}$. As the different links between the alpha-helices did not affect the unfolding force values,

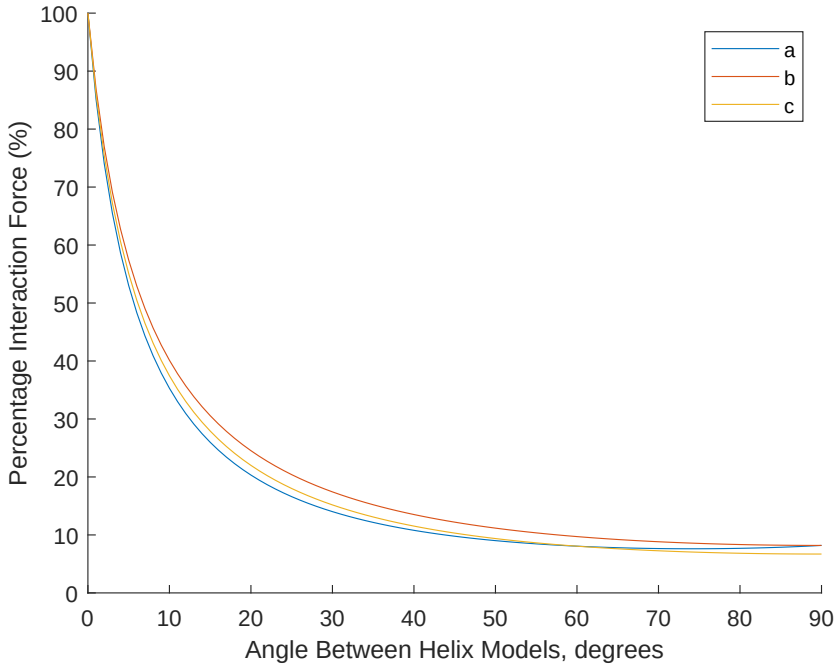


Figure 2.15: Percentage interaction force vs. angle, showing the unfolding response for the three hinge types shown above. Line a) corresponds to the setup in Figure 2.12. Line b) corresponds to the setup in Figure 2.13. Line c) corresponds to the setup in Figure 2.14.

they could be ignored in future simulations without impacting the calculation of unfolding force thresholds.

The lengths of the alpha-helices directly determine the number of sidechains expressed on the alpha-helices. The sidechains contain the atoms and molecules that produce the helices' chemical and physical properties, such as electrostatic charge and the ability to bond to other molecules. Additionally, physical properties of the sidechains such as electrostatic charge are what produce the intermolecular interactions between alpha-helices. The alpha-helix length and the number, position and type of sidechains need to be modelled together. To accurately model these alpha-helix sidechains and their properties, the structural data of the alpha-helices within the rod domain of talin needs to be processed and the relevant data extracted.

The simple simulations described in this section were each written to accomplish a predefined and specific task and did so using very simple models. However, they were time-consuming to write and error-prone as each model and simulation was bespoke. The next step was to develop simulations of four and five helices that used helix models with high levels of detail. This step is covered in Chapter 3. The complexity of these new simulations increased as more detailed helix models were used and more objects were simulated. Therefore, a generalised simulation

framework was implemented with repetitious and complex operations abstracted away. The framework was designed to minimise errors during the creation of simulations by managing the simulation and its common operations. It also minimised the time required to write them. The framework focused on three areas: the generation of models from raw data, the management and movement of model objects within the simulation space, and performing the interaction calculations on the simulation objects.

In summary, several key aspects were identified in this chapter that need to be incorporated in the alpha-helix models and simulation framework to determine the unfolding force thresholds of talin's rod subdomains. These aspects are as follows:

- The unfolding response of talin's rod subdomains can be modelled. However, the model will only be useful once the unfolding force thresholds are known and linked to each subdomain. They currently are not.

- Alpha-helix length is linked to the number of sidechains. The number and type of sidechains determines the interactions between helices. Therefore, protein structural data is required to accurately determine the sidechain's positions and properties in order to generate detailed alpha-helix models.

- The sidechain properties and the position and orientation of the alpha-helices within a rod subdomain determine the unfolding threshold. The linkage between the alpha-helices, the hinge mechanism and the unfolding motion do not affect the force threshold, but do affect the tail response of the force curve. Therefore, the connection between the alpha-helices is not a key factor in determining the unfolding force thresholds and does not need to be modelled in the simulation.

- The complexity of the simulation rapidly increases with model complexity and number of objects. Therefore, to minimise errors and the writing time of new simulations, a framework should be used to manage simulation objects and abstract away repetitious and complex operations.

3 Implementation of the Simulation Framework

The original contribution of this thesis lies in the simulation framework and its calculations operating directly within the force domain, the new model of the approximated alpha-helix structure, the use of coordinate frames to manage and manipulate the biological objects within the simulation framework, and the method by which the coordinate frames are deterministically attached to the models.

The previous chapter highlighted a number of requirements for the next simulation. Therefore, the new talin simulations needs to include: models of alpha-helices derived directly from protein structural data from the PDB, data on sidechain properties and positions, and a generalised simulation framework for managing simulation objects and calculating interaction forces.

In this chapter, the basic structure of amino-acids and their polymers (3.1.1) will be covered in order to explain the novel alpha-helix structural model developed for this simulation. An overview of the alpha-helix model and its components is given (3.1.2). Afterwards, the Protein Data Bank (PDB) and protein structural data (`pdb`) are discussed (3.1.3) so that the way in which the data from the `pdb` file was processed and used to generate the model can be explained (3.1.4). Next, the mathematical process that was created to transform the `pdb` data into the approximated alpha-helix model for the backbone (3.1.5) and the sidechain components (3.1.6) is covered. This is followed by a description of coordinate frames (3.2), their transforms, the definition of the local coordinate frames, and the method used to deterministically attach them to the alpha-helix model (3.2.1). Coordinate frames allow for efficient compartmentalisation of the simulation space and the objects contained within. Additionally, methodologies such as statics that have already been developed and tested in the field of robotics can be used. Finally, an example simulation is constructed (3.3) using all the concepts that have been discussed, and it is used to explain the implementation of the electrostatic calculation (3.3.4).

3.1 The Alpha-Helix Model

The alpha-helix model was designed to reduce the complexity of the structure of the alpha-helix while retaining the key data required to run the simulation. This process of reducing structural complexity for simulation models is similar to

the concept used when developing coarse-grained structural models for molecular dynamics simulations, as discussed in Section 1.3.3. A model with reduced complexity allows for longer simulation intervals and/or larger structures to be simulated using the same amount of computational resources. Two assumptions about alpha-helix structures were applied. First, that alpha-helices are straight and rigid structures. Second, that the size of the alpha-helix is constant along its length.

3.1.1 Amino-Acid and Alpha-Helix Structures

The structure is of an amino-acid monomer and can be broken into four distinct parts: the amino group (NH_2) in blue, the carboxyl group (COOH) in green, the carbon-alpha (Cα) in orange, and the sidechain group (R) in red. A reference amino-acid structure is shown in Figure 3.1. The sidechain group determines the type of amino-acid. The atoms that compose the sidechain can range from a single hydrogen atom to polycyclic aromatic rings (large, planar, rings of atoms that extend from the amino-acid).

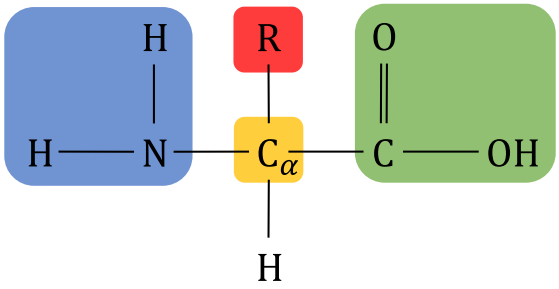


Figure 3.1: Structural formula of a reference amino-acid. The region highlighted in blue is the amino group. The green region is the carboxyl group. The yellow region is the carbon-alpha atom and the red region represents the sidechain.

Amino-acid monomers can be incorporated into a peptide chain; a chain of amino-acid monomers that form oligopeptides or polypeptides through condensation polymerisation. For this to occur, a hydrogen ion (H^+) from the amino group of one amino-acid and a hydroxyl ion (OH^-) from the carboxyl group of another amino-acid are released as a molecule of water, creating a peptide bond between the two amino-acids. The diagram in Figure 3.2 shows the result of two amino-acid monomers bonding, with the peptide bond highlighted in red. The amino-acid residue chain is the primary structure of a protein. An amino-acid residue is the remaining part of an amino-acid after it has lost parts of its amino and carboxyl groups during polymerisation. The (N — Cα — C) core of the

amino-acid residue is called the residue backbone chain.

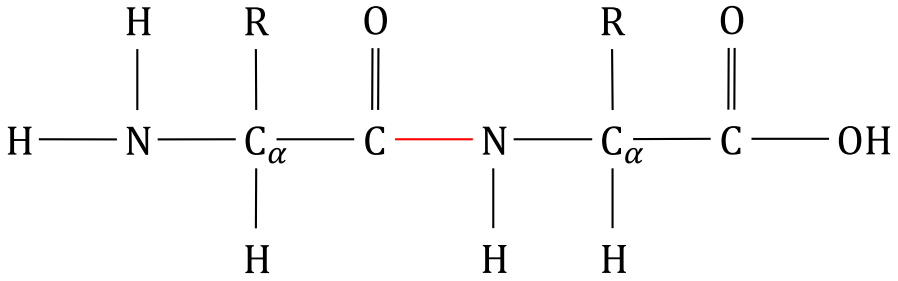


Figure 3.2: Structural formula of a reference dipeptide amino-acid sequence. The red line at the centre of the image indicates the peptide bond.

Alpha-helices are a type of secondary protein structure. In an alpha-helix, the residue backbone chain forms the alpha-helix backbone structure. Its spiral structure formed through every backbone (N — H) group hydrogen bonding with the (C = O) group of the amino-acid residue four positions previous in the peptide chain (Figure 3.3). This structure is a tight coil of the amino-acid residue backbone chain with the sidechains pointing outwards from the coil. The alpha-helix is often described as a 3.6-helix, as a complete 360° turn of the alpha-helix occurs every 3.6 amino-acid residues along the chain. This results in a repeating pattern where the sidechains extend from the alpha-helix at 100° rotational intervals.

3.1.2 Alpha-helix Model Overview

The remaining parts of this chapter contain content that has been published in a paper with Scientific Reports [191].

The alpha-helix model was constructed from two conceptual and structural parts: the alpha-helix backbone, and the collection of sidechains. In the model, both parts are approximated forms of the original data. However, as the sidechain part contains the key position and electrostatic data required for the interaction force calculation, the backbone is heavily approximated to make the model as simple as possible. The backbone component of the alpha-helix model was approximated as a primitive cylinder, which was described by: a start and end position that both defined the central axis through the helix backbone as well as the position and orientation of the helix; and a radius that approximated the width of the backbone. Local coordinate frames were used to create a local coordinate space. These local coordinate frames simplified the calculation and attachment of the relative positions of the collection of sidechains. They also simplified the manipulation of the models in the simulations.

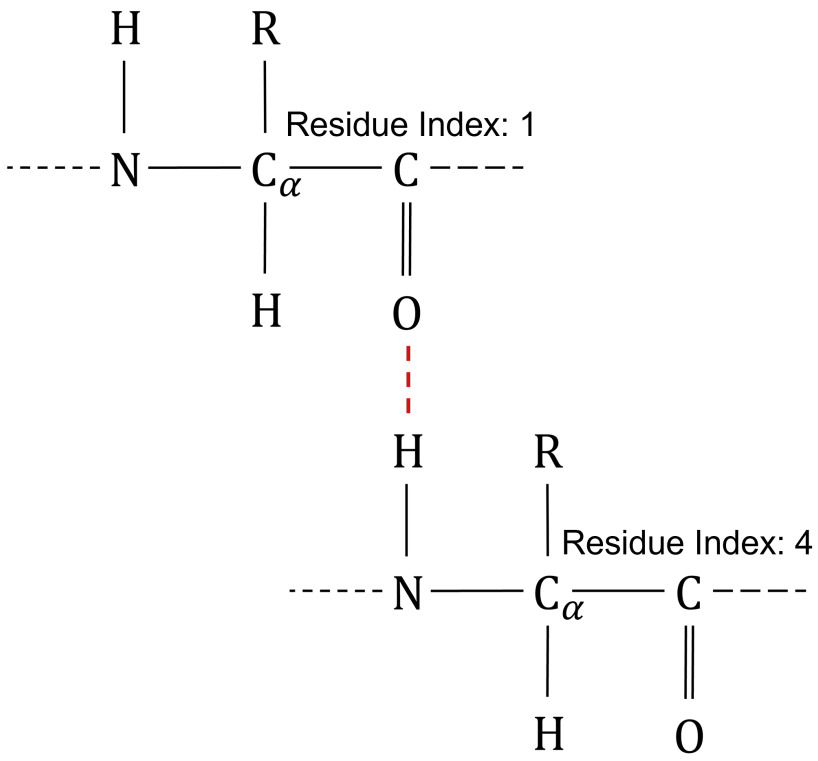


Figure 3.3: Structural diagram of the interlinking hydrogen bonds that stabilise alpha-helix structures. Each residue's amino group bonds to the carboxyl group of the residue four positions previous in the chain. The red dashed line represents the hydrogen bond between the two residues.

A Single Pseudo Atom (SPA) approximation was used. This implementation is similar to the Levitt-Warshel coarse-grained structural model. This approximation was chosen for two primary reasons: 1) for its level of simplicity it produced accurate results. Whereas the other models were much more complex, and 2) it enabled the sidechain model to have rotational symmetry. A more detailed description of the reasons for using this representation are discussed in Section 5.2 of the discussion chapter. The sidechain component of the model contained the position and electrostatic properties of each sidechain on the alpha-helix. Each pseudo-atom corresponded to the centroid position of each sidechain's atom and all electrostatic properties were calculated from this position. Additionally, a radius for each pseudo-atom was calculated to represent the sidechains' approximate size.

The resulting alpha-helix model has two components: the alpha-helix cylinder that corresponds to the position, orientation and relative size of the alpha-helix, and the collection of sidechain pseudo-atoms with their physical and chemical properties that are positioned relative to the helix cylinder. The next five subsec-

tions cover an explanation of protein structural data, its extraction, and processing, the process used to create the alpha-helix backbone approximation, and the sidechain approximation from the original structural data. This is followed by the attachment of a local coordinate frame to create the final alpha-helix model.

3.1.3 Protein Data-bank and Protein Structural Data

Before the components of the alpha-helix model ccould be created, the protein structural data needed to create the alpha-helix model had to be selected and processed as highlighted at the end of Chapter 2. This protein structural data is made available by the Protein Data Bank.

The Protein Data Bank (PDB) was first established at the Brookhaven National Laboratories at the start of the 1970s. It has grown to be the publicly available, worldwide archive of structural data for proteins and other biological macromolecules [189, 190]. Structural data is generated from experimental measurements using X-ray crystal structural determination, nuclear magnetic resonance imaging (NMR), cryoelectron microscopy, or from direct structural modelling.

The structural data are contained within `pdb` files. In these files, the position of each atom of the protein (excluding hydrogen) is stored as three-dimensional Cartesian coordinates. Hydrogen atoms are not included in the `pdb` structural data as the methods used to determine atomic positions have too low a resolution to accurately determine hydrogen atom positions [192]. Identifiers such as atom index, residue index, occupancy and temperature factor, are also included in the `pdb` data. An extract of a `pdb` file is given in Table 3.1 with abbreviated headings.

Table 3.1: Extract from a `pdb` file. Column Headings: Record Type (Type), Atom Serial Number (A_#), Atom Name (A_Name), Residue Name (Res_Abbrev), Residue Sequence Number (Res_#), X Orthogonal Angstrom Coordinate (X_Pos), Y Orthogonal Angstrom Coordinate (Y_Pos), Z Orthogonal Angstrom Coordinate (Z_Pos), Occupancy (Occ), Temperature Factor (Temp).

Type	A_#	A_Name	Res_Abbrev	Res_#	X_Pos	Y_Pos	Z_Pos	Occ	Temp
ATOM	13	CE1	HIS	2	17.349	-8.140	-14.399	1.00	10.15
ATOM	14	C	HIS	2	19.647	-10.762	-11.438	1.00	10.15
ATOM	15	O	HIS	2	20.462	-10.472	-12.312	1.00	10.15
ATOM	16	N	ALA	3	19.852	-10.479	-10.139	1.00	10.15

The PDB publishes `pdb` entries that contain large and small fragments of the talin-1 rod domains taken from both human and mouse samples. For example, a full segment of the rod domain R6 appears in entry 2l10.1.A. It was recorded

using NMR and has 1101 atom records. The `pdb` files do not contain secondary or tertiary structural data, which means that it does not indicate which atoms are parts of alpha-helices. Therefore, SWISS-MODEL's modelling servers were used to determine what sections of the `pdb` data were linked to which alpha-helix within each domain.

SWISS-MODEL operates a protein structure homology-modelling server [193, 194]. Using SWISS-MODEL's modelling service, each `pdb` file was assessed and the residues that made up each alpha-helix were determined. The starting and ending residue index values for each alpha-helix in the `pdb` file were stored and used in Section 3.1.4 to extract the alpha-helix data from the `pdb` files.

Using the SWISS-MODEL template matching server, each sequence was selected by comparing template quality and sequence matching scores to other `pdb` structures of the rod domains covering the same residue ranges. The templates that matched the sequences most closely and with the highest quality scores were selected [195]. The final template for each alpha-helix bundle was downloaded from the PDB as a `pdb` file (PDBID: 6R9T, 2L7A, 2LQG, 5FZT, 4F7G, 2KVP, 3DYJ, 2JSW) [126, 196–203].

3.1.4 PDB Data, Extraction, and Processing

MATLAB's Bioinformatics Toolbox [204] contains a function, `pdbread()`, that can parse `pdb` files into a usable data structure. The resulting data is a list of data structures that correspond to each atom entry in the original `pdb` file. The primary pieces of data for each atom entry were: position X, Y, and Z; atom name; atom index number; residue index number; and amino acid name abbreviation.

The developed alpha-helix model was built on top of two structural approximations, the approximation of the alpha-helix backbone and the approximation of the residue sidechains. Therefore, these alpha-helix backbone and sidechain structures first needed to be separated from each other for each alpha-helix in each rod domain. For each `pdb` file, the atom entries were separated into an individual list of each alpha-helix using the alpha-helix start and end residue indices obtained from SWISS-MODEL. Each alpha-helix list was then separated by residue index number so each alpha-helix had a list of amino-acid residues linked to a number of atom entries. Finally, the collection of atom entries for each amino-acid residue was separated into backbone and sidechain atoms. Backbone atoms were selected if their atom name was either "O", "N", "C"" or "Cα" (as these atom names correspond to the atoms identified as the amino-acid residue backbone in Section

3.1.1). All remaining atoms were grouped as sidechain atoms.

This new representation of the rod domain structure as a list of backbone positions and corresponding sidechains was generated as an image (Figure 3.4) and compared to the original graphical representation from the PDB generated using RCSB's NGL viewer (Figure 3.5).

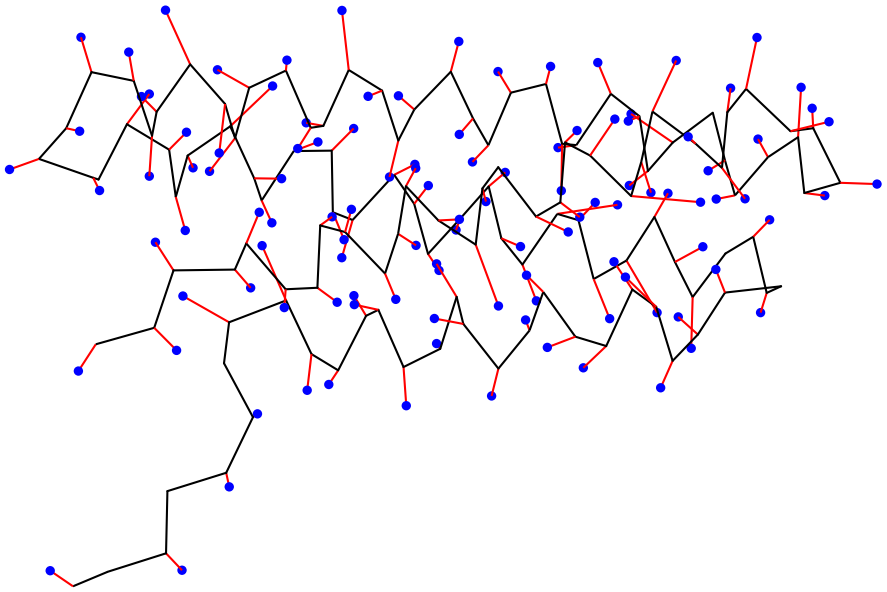


Figure 3.4: Figure generated from the processed backbone and sidechain positions, which shows the full talin-1 R3 domain containing all four alpha-helices. Backbone positions were connected with blue lines, sidechain positions are indicated with red dots and the link between sidechain and their corresponding backbone position was indicated with a green line. Residue index labels were applied at intervals of 5 residue indices.

3.1.5 Defining the Alpha-helix Backbone Approximation

After processing the raw structural data for each rod domain, the backbone and sidechain components for each alpha-helix model were generated. The steps for generating the backbone components are covered in this section.

Alpha-helices are not straight and have twists and curves throughout their structure. The cylinder approximation for the alpha-helix backbone does not need to match these curves, as the sidechain component of the model accounts for this. Further detail is given on this in Section 5.1. To define a cylinder, two positions (one for the centre of each opposing face of the cylinder) and a radius are needed. The two positions at each end of the cylinder create a central axis that passes through the cylinder. Here, this axis for the cylinder from the alpha-helix backbone structure is determined. The axis is used to generate the starting and

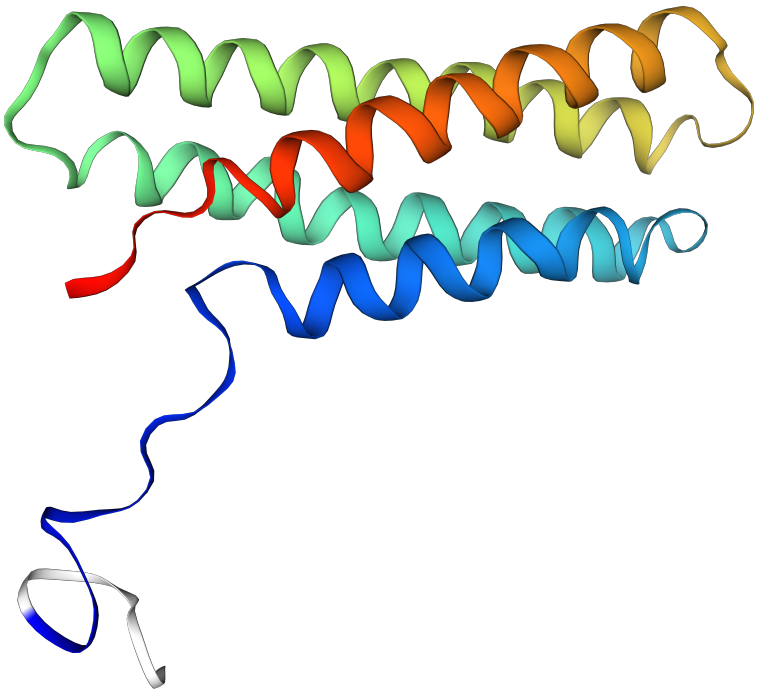


Figure 3.5: Figure generated from the original pdb data using RSCB's NGL viewer. It shows the full talin-1 R3 domain containing all four alpha-helices.

ending positions for the cylinder approximation. Afterwards, an estimation of the radius of the alpha-helix is calculated for the cylinder radius.

Previously, while processing the alpha-helix data from the `pdb` files, each alpha-helix's backbone atoms were approximated to a single point defined by that residue's Cα atom. From here onwards, this position is called the backbone position. The position of the Cα atom in the backbone structure was selected as the backbone position for two reasons. Firstly, it is placed centrally in the residue backbone structure and secondly, the residue's sidechain structure is directly connected to the backbone through this atom. A matrix $\mathbf{M}$ was defined for each alpha-helix where $\mathbf{M} = \left[\underline{b_1} \dots \underline{b_i} \dots \underline{b_n}\right]^\intercal$ and the vector $\underline{b_i}$ is a column vector that contains Cartesian coordinates of each backbone position of the alpha-helix in the original coordinate space of the `pdb` file. This matrix contains all the backbone positions for each residue in the alpha-helix. The symbol i in $\underline{b_i}$ denotes the i^{th} amino-acid residue by index, where $i = 1, \dots, n$ and n is the total number of amino-acid residues in the alpha-helix.

Next, the mean position of all backbone positions was determined in Equation 3.1. The position vector $\underline{b_c}$ corresponds to the centroid of backbone positions of the alpha-helix.

$$\underline{b}_c = \frac{1}{n}\sum_{i=1}^{n}\underline{b}_i \tag{3.1}$$

Principal component analysis (PCA) was used to calculate the axis for the cylinder approximation from the alpha-helix backbone structure. PCA calculates a line of best fit through an n-dimensional array of data. It is used here to determine the central axis through the backbone positions; the line along which the alpha-helix lies.

The PCA was calculated using single value decomposition (SVD) of the zero mean Cα backbone positions. The variance-covariance matrix $\mathbf{S}$ was formed as shown below in Equation 3.2.

$$\mathbf{S} = \frac{\mathbf{B}^\intercal\mathbf{B}}{n-1} \tag{3.2}$$

$$\begin{aligned} \text{Where } \mathbf{B} &= \left[\left(\underline{b}_1 - \underline{b}_c\right)\ldots\left(\underline{b}_i - \underline{b}_c\right)\ldots\left(\underline{b}_n - \underline{b}_c\right)\right] \\ \text{Which can be computed by } &= \mathbf{M} - \underline{r}\underline{b}_c \end{aligned} \tag{3.3}$$

Where $\mathbf{B}$ is a n-by-3 matrix of the alpha-helix backbone (Cα) positions with a zero-mean (Equation 3.3). Matrix $\mathbf{B}$ was created by element-wise subtraction of $\underline{b}_c$ from $\mathbf{M}$ (this can also be constructed by setting temporary vector $\underline{r}$ to a column vector containing 1 in all n rows). Thus, $\mathbf{S}$ is simply the variance-covariance matrix of $\mathbf{B}$.

Matrix $\mathbf{S}$ was then decomposed using SVD. This was done to obtain the principal axis of the input data from the left matrix of singular vectors that corresponds to the largest singular value. Golub and Kahan showed that it is possible to partition any matrix into three subunits such that $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal$. Where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices meaning that $\mathbf{U}^\intercal\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^\intercal\mathbf{V} = \mathbf{I}$, and $\mathbf{\Sigma}$ is a diagonal matrix where the diagonal values are called the singular values. These properties make SVD a useful tool for linear algebra.

$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal = \mathbf{S} \tag{3.4}$$

The vector approximating the direction of the alpha-helix backbone structure was estimated from the first left singular vector, which corresponds to the largest singular value. The vector $\hat{\underline{d}}$ is the first principle axis through the positions in $\mathbf{B}$ and is the unit vector specifying the direction from point $\underline{p}_s$ to point $\underline{p}_e$, in

other words, the direction of the central axis through the alpha-helix backbone. The value $u_{*,1}$ specifies the vector forming the first column of elements from **U**. Therefore, $\underline{\hat{d}}$ is the first singular vector of **S** and all other singular vectors were ignored as they correspond to secondary axes through that backbone structure which were not needed.

$$\underline{\hat{d}} = \underline{u}_{*,1} \tag{3.5}$$

The line through the alpha-helix backbone structure, nominally the backbone axis, is thus defined by noting that the two positions $\underline{p}_s$ and $\underline{p}_e$ are the *start* and *end* axis positions. These positions were calculated using the parametric form of the line equation as shown in Equation 3.6 below.

$$\begin{aligned} \underline{p}_s &= \underline{b}_\mathrm{c} + \underline{\hat{d}}\,|\underline{b}_1 - \underline{b}_\mathrm{c}| \\ \underline{p}_e &= \underline{b}_\mathrm{c} + \underline{\hat{d}}\,|\underline{b}_\mathrm{n} - \underline{b}_\mathrm{c}| \end{aligned} \tag{3.6}$$

where vectors $\underline{b}_1$ and $\underline{b}_\mathrm{n}$ are the first and last position vector elements of the alpha-helix backbone positions **B**. The positions $\underline{p}_s$, $\underline{b}_c$ and $\underline{p}_e$ are collinear, meaning they lie along the same line in space. The directionality of the axis through the start and end positions is checked after the sidechains have been approximated. This step must happen, as the vector calculated through PCA always points away from the origin of the coordinate space and does not take into account the original direction of the helix. Figure 3.6 shows the primary axis from the PCA calculation drawn through helix 2 of the R3 domain with $\underline{p}_s$ and $\underline{p}_e$ marked.

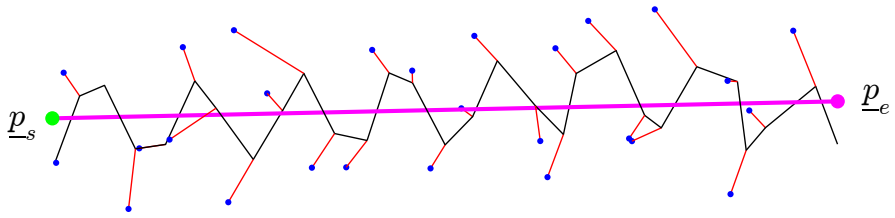


Figure 3.6: Helix 2 of the R3 subdomain. The primary axis determined by PCA is highlighted in magenta. Position $\underline{p}_s$ is marked by the green cirlce and position $\underline{p}_e$ is marked by the magenta circle.

The radius of the alpha-helix can now be estimated using Equation 3.7. The shortest distance between each backbone position and the backbone axis line is calculated using the position along the backbone axis line that results in a perpendicular line from the backbone axis line to the backbone position.

$$l_i = \left\| \underline{b}_i - \underline{p}_a \right\| = (\underline{b}_i - \underline{p}_a)^\intercal (\underline{b}_i - \underline{p}_a) \tag{3.7}$$

where $\underline{p}_a$ is the position along the backbone axis line which is the result of Equation (3.8), l_i is the Euclidean distance between $\underline{p}_a$ and the i^{th} backbone position $\underline{b}_i$. In Equation 3.8, $\underline{p}_a$ is the position calculated along the line defined by positions $\underline{p}_s$ and $\underline{p}_e$, t specifies the distance along this line and is constrained by $0 \leq t \leq 1$.

$$\begin{aligned} \underline{p}_a(t) &= \underline{p}_s + t(\underline{p}_e - \underline{p}_s) \\ t(\underline{p}_x) &= \underline{\hat{d}} \frac{(\underline{p}_s - \underline{p}_x)}{|\underline{p}_e - \underline{p}_s|} \\ \underline{\hat{d}} &= \frac{(\underline{p}_e - \underline{p}_s)}{|\underline{p}_e - \underline{p}_s|} \end{aligned} \tag{3.8}$$

3.1.6 Defining the Sidechain Approximation

To approximate each amino acid residue sidechain as a single position, the centre position of atoms was determined by taking the mean value of all positions of the atoms in that sidechain. The reasons why this single position sidechain representation was chosen, are discussed in Section 5.2. Additionally, only the relative position of the sidechain to the alpha-helix backbone was determined, rather than the connection point of the backbone to the sidechain, as discussed at the end of the previous chapter.

For each sidechain on the alpha-helix, $\underline{p}_{\text{SC},i,j}$ was set as the 3D position vector of a sidechain atom, where i specifies the sidechain in the alpha-helix where $i = 1, \ldots, n$, and j is the sidechain atom index where $j = 1, \ldots, m$, n is the total number of backbone and sidechain positions and m is the total number of sidechain atoms in the i^{th} sidechain. Equation 3.9 describes the calculation used to approximate the position of the sidechain, where $\underline{s}_i$ is the approximated position of the i^{th} sidechain. The i value links the i^{th} sidechain to the i^{th} backbone position. Figure 3.7 illustrates the sidechain approximation.

$$\underline{s}_i = \frac{1}{m} \sum_{j=1}^{m} \underline{p}_{\text{SC},i,j} \tag{3.9}$$

In order to maintain the size of the sidechain, the sidechains were approximated as spheres centred at the approximated sidechain position $\underline{s}_i$. The radius of these approximation spheres was estimated using the Euclidean distance between the

approximated sidechain position and each of the sidechain atoms' positions. The mean value of these distances is the radius of the sidechain sphere, $\underline{r}_i$, as shown in Equation 3.10.

Amino acid sidechains are generally planar in configuration. However, they possess at least one freedom of rotation around the covalent bond that links the sidechain to its backbone's Cα atom. A sphere approximation at the mean position of the sidechain atoms is appropriate, as it naturally accounts for all possible rotations of the sidechain structure around this covalent bond. $\underline{r}_i$ is the approximated size of the i^{th} sidechain. The amino-acid residue name corresponding to each sidechain is then used to attach the correct electrostatic charge.

$$\underline{r}_i = \frac{1}{m} \sum_{j=1}^{m} \left\| \underline{p}_{\text{SC}_{i,j}} - \underline{s}_i \right\| \tag{3.10}$$

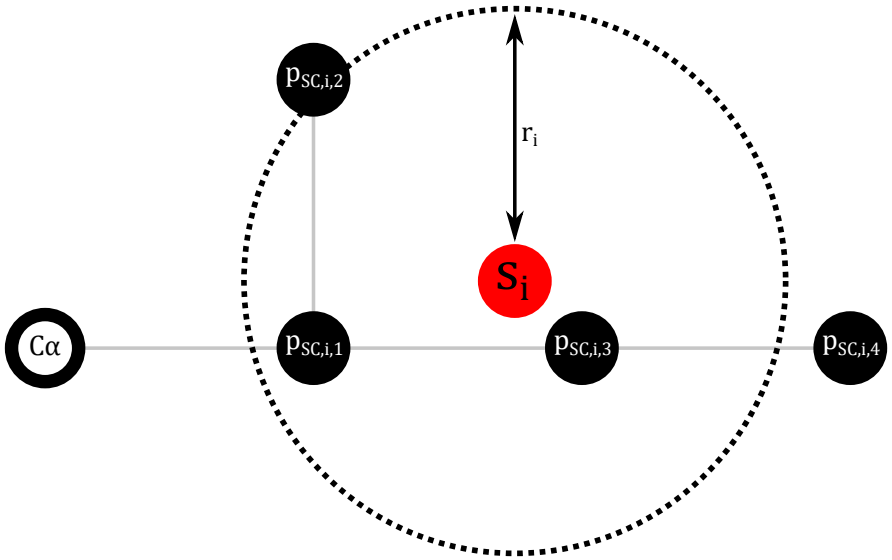


Figure 3.7: Diagram of the sidechain approximation. The sidechain atom positions are marked as solid black circles. The backbone carbon-alpha is marked as a hollow black circle. The resulting approximated sidechain position is marked as a solid red circle, and the approximated sidechain size is marked with a dotted black line.

After the sidechain positions are created, the orientation of the alpha-helix model is checked against the original alpha-helix amino-acid residue direction. This is done by comparing the start position of the helix model to that of the first and last sidechain positions. If the Euclidean distance between the start position and the first sidechain is shortest, the orientation of the helix is correct. If not, the start and end positions are swapped. This aligns the direction of the alpha-helix backbone axis to the direction of sidechain indexing.

3.2 Attaching Local Coordinate-frames

All positional coordinates so far have been within the original coordinate space from the `pdb` file. This coordinate space is separate from the local coordinate space created for alpha-helix model, and from the simulation coordinate space. A coordinate frame defines a coordinate space whose position and axial orientation is defined relative to another coordinate space. The reasons for using a coordinate frame system are covered in more detail in Section 5.4. A coordinate transform can be used to define this position and axial offset. It can also be used to convert points in space from one coordinate frame to another. This is also known as referencing a position from another coordinate frame. By attaching local coordinate frames to each alpha-helix and referencing their position relative to a 'simulation' coordinate space, each alpha-helix can be manipulated in space independently of the other alpha-helices. Importantly, this can be done by changing only the position and orientation of the alpha-helix coordinate frame, rather than having to update all the positions of the alpha-helix components such as the cylinder start and end position and each sidechain position. This is because all sidechain and cylinder positions are referenced via the local alpha-helix coordinate frame.

The homogeneous coordinate transform is used to define these coordinate frames and their respective transformations [205]. This method of coordinate transform definition follows the kinematics techniques from robotics. Similar techniques are also used in computer graphics to construct scene-graphs. This is expanded on in Section 5.4 of the discussion, but in short, this thesis follows the conventions used in robotics as this eases the calculations of force and torque. Statics enables calculations for propagating forces and torques through coordinate frames which are used to develop dynamic simulations.

Although there are alternative methods to specify the location and orientation of a coordinate frame, the homogeneous transform representation given in Equation 3.11 is the most common and is therefore adopted here [205]. The super- and sub-script notation of coordinate frames, rotations and transforms from Craig [205] is used.

$$ {}^{A}_{B}\mathbf{T} = ({}^{A}_{B}\mathbf{R}, {}^{A}\underline{P}_{B}) = \begin{bmatrix} & {}^{A}_{B}\mathbf{R} & & {}^{A}\underline{P}_{B} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \underline{\hat{x}} & \underline{\hat{y}} & \underline{\hat{z}} & {}^{A}\underline{P}_{B} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.11} $$

$$ {}^{A}\underline{p_0} = {}^{A}\underline{P}_B + {}^{A}_{B}\mathbf{R}\,{}^{B}\underline{p_0} = \begin{bmatrix} {}^{A}\underline{P}_{Bx} \\ {}^{A}\underline{P}_{By} \\ {}^{A}\underline{P}_{Bz} \end{bmatrix} + \begin{bmatrix} {}^{A}_{B}\mathbf{R}_{11} & {}^{A}_{B}\mathbf{R}_{12} & {}^{A}_{B}\mathbf{R}_{13} \\ {}^{A}_{B}\mathbf{R}_{21} & {}^{A}_{B}\mathbf{R}_{22} & {}^{A}_{B}\mathbf{R}_{23} \\ {}^{A}_{B}\mathbf{R}_{31} & {}^{A}_{B}\mathbf{R}_{32} & {}^{A}_{B}\mathbf{R}_{33} \end{bmatrix} \begin{bmatrix} {}^{B}\underline{p_0}_x \\ {}^{B}\underline{p_0}_y \\ {}^{B}\underline{p_0}_z \end{bmatrix} \tag{3.12} $$

$$ {}^{A}\underline{p_0} = {}^{A}_{B}\mathbf{T}\ {}^{B}\underline{p_0}^{*} \tag{3.13} $$

$$ \text{Where: } {}^{B}\underline{p_0}^{*} = \begin{bmatrix} {}^{B}\underline{p_0} \\ 0 \end{bmatrix} \tag{3.14} $$

Equation 3.12 shows the standard method of transforming a position vector p_0 in frame B so that it is referenced in frame A. The position vector is multiplied by the rotation matrix defining the difference in orientation of frame B from frame A before the result is translated by the difference in position of frame B from frame A. The homogeneous transform is shown in Equation 3.11. ${}^{A}_{B}\mathbf{T}$ defines rotation and translation that is required to reference a vector in frame B from frame A. Thus, a vector $\underline{\mathbf{V}}$ stated in frame B can be re-stated in frame A with ${}^{A}\mathbf{V} = {}^{A}_{B}\mathbf{T}({}^{B}\mathbf{V})$.

A homogeneous transform ${}^{A}_{B}\mathbf{T}$ defines the transformation of a position vector in frame B to a new position and orientation in frame A, but it also represents the position and orientation of coordinate frame B expressed in frame A. In this way, homogeneous transforms are operations of position vectors within a coordinate system and also the definition of coordinate frames themselves.

The local coordinate frame, B, defined by the homogeneous transform in Equation 3.11 exists at position $\underline{p_0}$ and has its axes aligned to unit vectors $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{z}}$ respectively, all referenced from the coordinate space of A. This transform is equivalent to having a duplicate set of axes at the origin of frame A that are rotated using rotation matrix $\underline{R}$ and then translated by the vector $\underline{p_0}$. The homogeneous transform is applied to a position vector as shown in Equation 3.13, where an addition zero is concatenated vertically to the bottom of the position vector.

To apply a local coordinate frame to each alpha-helix model, a deterministic mapping of the position and axes of the new local coordinate frame to the geometry of the model is required to ensure that the simulations are repeatable and generate the same output data for the same input data. This mapping is shown in Figure 3.8 and defined as follows:

1. The origin point of the coordinate frame is set to the starting position ($\underline{p_s}$) of the alpha-helix model.

2. The X-axis is aligned to the model's cylinder axis such that the positive direction of the axis points towards the end position (p_e) from the start position (p_s).

3. The Y-axis is aligned to the perpendicular line drawn from the alpha-helix cylinder axis to the first (by index value), non-glycine sidechain position.

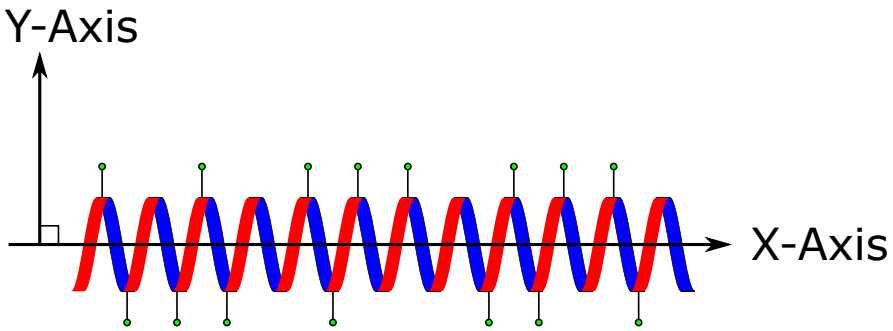


Figure 3.8: Diagram of the local coordinate frame axes overlaid on a representation of an alpha-helix. The X-axis is aligned along the axis of the helix. The Y-axis is aligned to the first non-glycine sidechain, which is indicated by the small green circle extending from the helix. The Z-axis is not shown but would be constructed using the right-hand rule and therefore extend out of the page towards the reader.

The Y-axis had to be defined specifically by the first 'non-glycine' sidechain position along the helix. Glycine is an amino-acid that has a sidechain consisting of a single hydrogen atom. As mentioned at the start of this chapter, `pdb` files store positions for all atoms, except for hydrogen. Thus, no sidechain atoms exist in glycine to create an appropriated sidechain position. Therefore, to work around the edge-case where the first sidechain on an alpha-helix was glycine, the first non-glycine sidechain position was selected. The vector defining this sidechain position is p_a.

3.2.1 Local Coordinate Frame Transformations

The $\hat{\underline{x}}$, $\hat{\underline{y}}$ and $\hat{\underline{z}}$ axes of the local coordinate frame were defined relative to the `pdb` coordinate space by the three unit vectors as follows:

$$\begin{aligned} \hat{\underline{x}} = \hat{\underline{d}} &= \left\| \underline{p}_e - \underline{p}_s \right\|_2 = \frac{(\underline{p}_e - \underline{p}_s)}{|\underline{p}_e - \underline{p}_s|} \\ \hat{\underline{y}} &= \left\| \underline{s}_1 - \underline{p}_a \right\|_2 = \frac{(\underline{s}_1 - \underline{p}_a)}{|\underline{s}_1 - \underline{p}_a|} \\ \hat{\underline{z}} &= \hat{\underline{x}} \times \hat{\underline{y}} \end{aligned} \tag{3.15}$$

where the X-axis direction $\hat{\underline{x}}$ is defined by the alpha-helix cylinder axis $\hat{\underline{d}}$ or calculated directly from the alpha-helix model's start ($\underline{p}_s$) and end ($\underline{p}_e$) positions. The Y-axis direction $\hat{\underline{y}}$ was defined using the first non-glycine sidechain position $\underline{s}_1$ and the point along the alpha-helix cylinder axis $\underline{p}_a$ that was used to draw a perpendicular line to $\underline{s}_1$. The position $\underline{p}_a$ was calculated using the parametric line equation previously shown in Equation 3.8, where $\underline{p}_x$ was the position $\underline{s}_1$. The Z-axis can then simply be determined using the cross product of $\hat{\underline{x}}$ and $\hat{\underline{y}}$. The rotational sub-matrix $\mathbf{R}$ and origin position vector $\underline{P}$ which construct the homogeneous transform and define the local coordinate frame within the `pdb` space are as follows in Equation 3.16a and Equation 3.16b.

$$\begin{aligned} {}^{\text{pdb}}_{\text{local}}\mathbf{R} &= \begin{bmatrix} \hat{\underline{x}} & \hat{\underline{y}} & \hat{\underline{z}} \end{bmatrix} \\ {}^{\text{pdb}}\underline{P} &= \underline{p}_s \end{aligned} \tag{3.16a}$$

$${}^{\text{pdb}}_{\text{local}}\mathbf{T} = ({}^{\text{pdb}}_{\text{local}}\mathbf{R}, {}^{\text{pdb}}\underline{P}) \tag{3.16b}$$

Now that the transform from pdb space to the alpha-helix's local coordinate ${}^{\text{pdb}}_{\text{local}}T$ has been defined, the positional data of the alpha-helix model is transformed from the original pdb space to the new local coordinate frame. After this, the position can be referenced through the local coordinate frame. A template notation for this data transform is given in Equation 3.17. Where position ${}^{\text{pdb}}\mathbf{P}$ stands for any positional data in the pdb space, ${}^{\text{pdb}}_{\text{local}}\mathbf{T}^{-1}$ is the inverse of the coordinate transform from pdb to local space and ${}^{\text{local}}\mathbf{P}$ is the resulting position in the local alpha-helix coordinate frame. Note, the inverse of the transform ${}^{\text{pdb}}_{\text{local}}\mathbf{T}$ is the transpose of its rotation matrix and the multiplication of its translation vector by

the negative transpose of the rotation matrix. This is shown in Equation 3.18. The inverse of the transform can be used for the transformation from pdb coordinate space into the local space. The normal transform converts a position in local coordinate space into its equivalent location in pdb coordinate space.

$$^{\text{local}}\mathbf{P} = {}^{\text{pdb}}_{\text{local}}\mathbf{T}^{-1}\ {}^{\text{pdb}}\mathbf{P} \tag{3.17}$$

$$\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{R}^{\intercal} & -\mathbf{R}^{\intercal}\mathbf{P} \\ 0 & 1 \end{bmatrix} \tag{3.18}$$

3.3 Simulation Framework

So far in this chapter, the selection and processing of pdb data have been covered. The alpha-helix model was described and the mathematical definitions of its construction and the attachment of its local coordinate frame detailed. In the remainder of this chapter, the structure of the simulation framework is explained, and an example simulation is set up using all the concepts covered above. The reasons for using a framework were previously discussed at the end of Section 2.2.2.

The simulation framework was written using MATLAB 2020b and holds all the data structures and code required to generate the alpha-helix models from raw `pdb` files. It also constructs and manipulates the coordinate frames which structure and control the simulation space, and generate graphical representations of the simulation. The framework is the basis of the simulations used to calculate the electrostatic interaction forces between the alpha-helix models.

The simulation framework is split into two components. First, the data structures for amino-acid structures, the alpha-helix model and coordinate frames. And second, the functions used by the framework to manipulate the data structures.

3.3.1 Implementation in MATLAB

In this section, the implementation details of the simulation framework's data structures and functions are described. The tables below show each of the data structures for the amino-acid structures (Table 3.2), alpha-helix models (Table 3.3) and coordinate frame (Table 3.4). In each table the core variables are stated and their variable type given. In the coordinate frame data structure, the homogeneous transform was seperated into a position vector and a rotation matrix.

This is one of the alternative forms. While the homogeneous transform is conceptually and notationally more efficient, it is more computational efficient to break it apart if it is known in advance that the homogeneous transform does not perform any scaling of perspective shift transformations. The transforms used in the coordinate framework only perform translation and rotation transforms and so the stages of the transform were computed separately.

The `Amino_Acid_Chain` class shown in Table 3.2, represents the approximated amino-acid structure after parsing and processing the `pdb` file. A range of index values are used to select sections of the structure to be turned into helix objects. It stores: the location of the original `pdb` data file, a list of backbone position vectors as a matrix, a list of sidechain position vectors as a matrix, and a list of sidechain sizes as a vector. The same index value is used to access all position and size data variables.

Table 3.2: Data table of the `Amino_Acid_Chain` class. The names, backbone positions, sidechain positions and sidechain sizes are length N where N is the total number of residues in the `pdb` file. This data object stores the parsed and processed `pdb` file data after the backbone and sidechain positions have been approximated.

Name	Type	Description
names	N *string* list	List of AA residue names
backbone_positions	$3 \times N$-matrix	List of approximated backbone positions
sidechain_positions	$3 \times N$-matrix	List of approximated sidechain positions
sidechain_sizes	$1 \times N$-vector	List of sidechain radii
pdb_file_path	*string*	Path to original `pdb` file

Once a range of indices from the amino-acid data is selected, it is converted into a `Helix_Object` class shown in Table 3.3. Each instance of this class stores the data for each alpha-helix in a simulation. It also stores data for the alpha-helix model, such as the start and end position of the helix cylinder and its radius. Additionally, it determines and stores the properties of each sidechain, such as its hydrophobicity and electrostatic charge value. Lastly, it stores the original `pdb` coordinate data so the helices can be transformed to their original conformation once in the simulation framework's coordinate system.

The last class is the `Coordinate_Frame` shown in Table 3.4. Each instance of this class is used to define a coordinate frame. The base variable is used to denote if this coordinate frame is the global coordinate frame. In the next section, this value will be used as part of a recursive algorithm that traverses the coordinate frame system. The parent and children variables are used to create a

Table 3.3: Data table of the `Helix_Object` class. The variables: position one, position two and the up direction are used to store the $\underline{\hat{x}}$ and $\underline{\hat{y}}$ axes values of the alpha-helix prior to the local coordinate frame being attached. The values 'original pos', 'original direction' and 'original up' are where the original pdb coordinate space values are stored after the cylinder and sidechain position have been transformed into the local coordinate space. The 'parent frame' is a reference pointer to the local coordinate frame once it has been attached and is used as part of the recursive algorithms discussed later in this section. The 'sidechain properties' hold any physical or chemical properties for each sidechain that may be needed for the simulation. In this case, it holds the sidechains electrostatic charge as either a 1, 0 or -1 value.

Name	Type	Description
position_one	1×3-vector	Cylinder Start Position
position_two	1×3-vector	Cylinder End Position
radius	*value*	Cylinder Radius
up_direction	1×3-vector	Y-Axis Direction
parent_frame	*reference*	Reference to Parent Coordinate Frame
original_pos	1×3-vector	Original Position in pdb space
original_direction	1×3-vector	Original Orientation in pdb space
original_up	1×3-vector	Original Y-Axis in pdb space
sidechain_names	N *string* list	List of Residue Sidechain Names
sidechain_positions	$3 \times N$-matrix	List of Sidechain Positions
sidechain_sizes	$1 \times N$-vector	List of Sidechain Radii
sidechain_properties	$1 \times N$-vector	List of Sidechain Electrostatic Charges

doubly-linked list between coordinate frames referenced within the current frame (child frames) and the coordinate frame that the current frame is referenced from (parent frame). The object reference variable stores a pointer to the simulation object that this coordinate frame is attached to, such as a helix object. Finally, the position start, position end, rotation from parent and translation from parent are used to define the coordinate transform itself. The rotation and translation variables are separate components of the homogeneous transform. The position start and end variables are used to help define object within the coordinate frame.

Each of the simulation framework's data structures discussed above were written as a MATLAB *handle* class. This allowed a single instance of each object to exist, and a reference passed to other objects and functions, rather than MATLAB creating a new copy of the object each time. Referencing and updating an existing object is much faster than recreating it anew whenever it is altered. The collection of functions of the simulation framework were also written together as a *handle* class. This class was also used to maintain the original instance of each of the amino-acid, helix object and coordinate frame data structures. In addition to the functions that create helix objects from raw `pdb` data, the most important

Table 3.4: Data table of the `Coordinate_Frame` class. The base value is set so that the recursive algorithms knows when it has reached the simulation coordinate frame. The parent variable links to the frame's parent coordinate frame. The children list links to all coordinate frames that are nested within this frame. The object value links to a helix object if one is attached. The rotation from parent and translation from parent together define the coordinate transform to move from the parent to this local coordinate frame.

Name	Type	Description
base	*boolean*	Set if this frame is the Simulation Frame
children	*reference* list	List of References to all Child Coordinate Frames
parent	*reference*	Reference to parent coordinate Frame
object	*reference*	Reference to Helix object
position_start	1×3-vector	Position of Local Origin
position_end	1×3-vector	End Position of Helix Object if present
rotation_from_parent	3×3-matrix	Rotation from Parent Frame to this Frame
translation_from_parent	1×3-vector	Translation from Parent Frame to this Frame

set of functions govern the coordinate transforms.

3.3.2 Traversing Nested Coordinate Frames

The most important operations in the simulation framework are the transformations of data between coordinate frames. In this section, the system of connected coordinate frames and their use in calculating sidechain positions are explained. This process is vital for the simulation framework, as it must be performed for all electrostatic interaction calculations between helix objects.

Each simulation starts with a *global* simulation coordinate frame, 'SIM' for short. This frame is sometimes known as the *world* or *root* frame in other coordinate frame system implementations. Any helix objects, their coordinate frames, and child frames are referenced from this initial SIM frame. An example set up is shown in Figure 3.9, where the numbered circles represent coordinate frames and the SIM frame is numbered zero and marked with an asterisk. Two helix objects are inserted into the simulation. Helix 1 in frame 1 is inserted directly into the SIM frame and the Helix 2 in frame 3 is inserted into the SIM frame via frame 2, an empty joint. All positions in a simulation calculation, such as the sidechain

positions in the electrostatic interaction calculation, are done from the reference of the SIM frame. In other words, when the electrostatic interaction is calculated, all sidechain positions are first transformed into the SIM frame's coordinate space before being used in the calculation. As the architecture of the coordinate frame system is tree-like, where the SIM frame is the root node and all other objects are children of that root node, a tree-walking algorithm can be used to move between the coordinate frames.

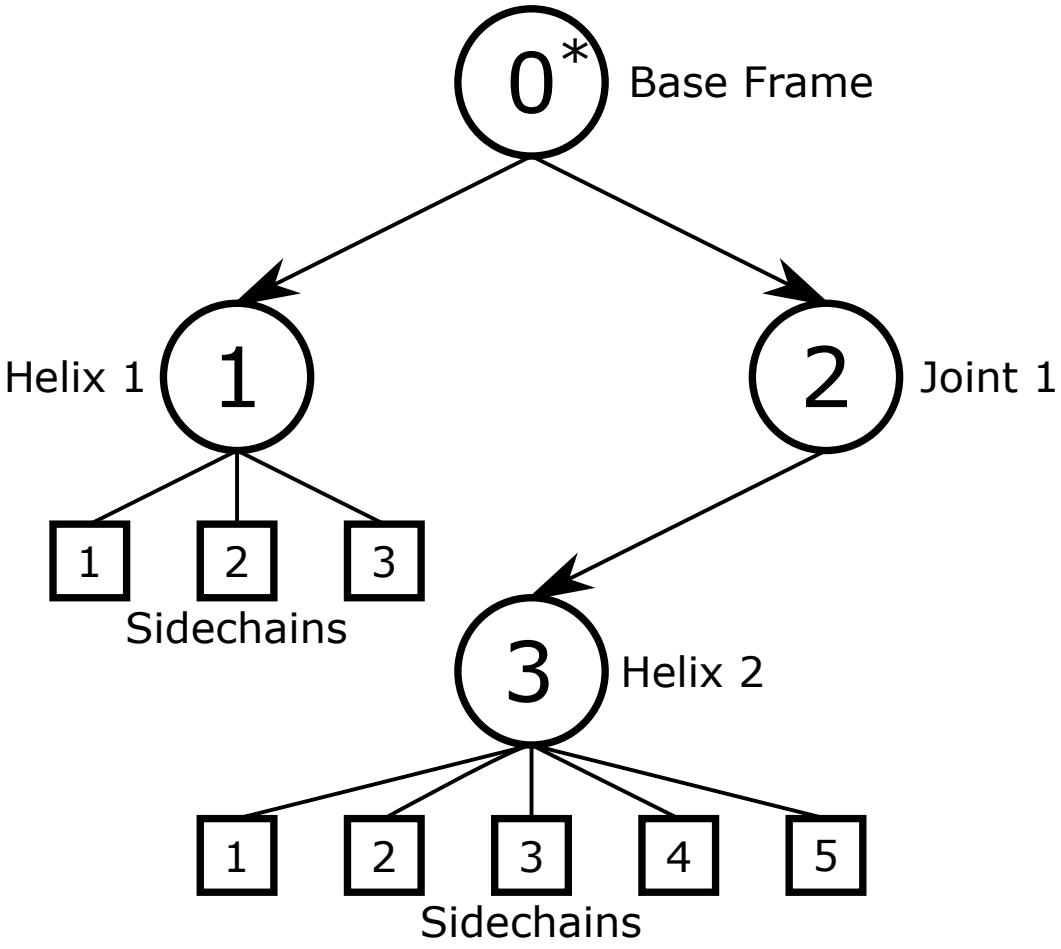


Figure 3.9: Diagram showing the connections between coordinate frames in an example simulation set up. Coordinate frames are shown as numbered circles. The global coordinate frame (SIM) is indicated by the circle containing 0^*. The arrows indicate the connections between frames. Sidechains are expressed as numbered squares.

Equation 3.19 shows the tree-walking algorithm which was designed as a recursive function $f_r(i, \underline{P_{out}})$. The function takes two input values: the index of the current frame, i, and the output position vector $\underline{P_{out}}$ passed between the levels of recursion. As matrix transformation operations need to be applied backwards from the input matrix or input vector, the recursive function is started at the

helix containing the sidechain positions that need to be transformed into the base frame.

$$f_r(i, \underline{P_{out}}) = \begin{cases} \underline{P_{out}} & i = 0 \\ f_r(i-1, {}^{i-1}_{i}T\ \underline{P_{out}}) & i > 0 \end{cases} \tag{3.19}$$

Using Figure 3.9 as an example, the recursive algorithm starts at frame 3 with $f_r(3, {}^3\underline{P_s})$. Where ${}^3\underline{P_s}$ is the starting helix position in frame 3, equating to $[0\ 0\ 0]^\intercal$. As the current frame is not the base frame, the bottom part of the piecewise function is used. The next frame index is set to $i-1=2$ and the second value is calculated by transforming ${}^3\underline{P_s}$ into coordinate frame 2's coordinate space by using the ${}^2_3T$. As discussed in Section 3.2.1, ${}^2_3T$ is equivalent to ${}^{i-1}P_i + {}^{i-1}_{i}R$, whose values are determined from the current frame's instance of the `Coordinate_Frame` class. The step above is repeated until the current frame is SIM frame 0 where $\underline{P_{out}}$ is given as the final result. This algorithm, therefore, walks up the tree-like structure of the simulation coordinate frame system until it encounters the SIM frame. At each coordinate frame, it applies the coordinate transform to the output position vector in order to move to the next frame. In the simulation framework, this function is called `Get_Position_World_Frame()`.

3.3.3 Example Simulation Setup

Using the explanation above, an example simulation was set up that matches the configuration shown in Figure 3.9 to illustrate the functionality of the simulation framework. The PDBID of the protein structure used was 2l7a.1.A and corresponded to the R3 rod subdomain of talin. The data was extracted from the `pdb` file using `pdbread()` and processed into an `Amino_Acid_Chain` instance. The first two alpha-helices were selected from the processed data using the residue indices $10-26$ and $30-57$. Each of these selections were then converted into `Helix_Object` instances.

Next, the SIM frame was set up with two new local coordinate frames for the helices and one coordinate frame for the joint, ${}^{\text{SIM}}_{\alpha H1}T$, ${}^{J1}_{\alpha H2}T$ and ${}^{J1}_{\alpha H2}T$. Newly constructed coordinate frames always have an origin position and an identity rotation matrix, $T(\underline{P}, \mathbf{R}) = T([0\ 0\ 0]^\intercal, I_3)$. For the simplicity of this explanation, the $J1$ joint frame does not alter the orientation or position of the $\alpha H2$ frame, meaning that ${}^{J1}_{\alpha H2}T = {}^{\text{SIM}}_{\alpha H2}T$ Each helix object is attached to their local coordinate frame. This also generates the transform that allows the local coordinate frame to be positioned in its helix's original conformation from the `pdb` data. These

transforms are applied as shown in equation 3.20. The translation that subtracts the first alpha-helix's start position from both coordinate frames is not required for the accuracy of the simulation, but moves the helices into the centre of the SIM coordinate space while maintaining their `pdb` conformation relative to each other. This makes visualising the simulation easier.

$$\begin{aligned} {}^{\text{SIM}}_{\alpha H1}T &= \left({}^{PDB}_{\alpha H1}T \; {}^{\text{SIM}}_{\alpha H1}T\right) - {}^{\alpha H1}\underline{P}_s \\ {}^{J1}_{\alpha H2}T &= \left({}^{PDB}_{\alpha H2}T \; {}^{J1}_{\alpha H1}T\right) - {}^{\alpha H2}\underline{P}_s \end{aligned} \tag{3.20}$$

A function of the simulation framework was then used to generate a visual representation of simulation. This is shown in Figure 3.10. The two helix objects are displayed as red cylinders with the blue spheres representing their sidechain's position and size. The first non-glycine sidechain of the first helix is shown with a magenta line connecting it to its helix.

3.3.4 Electrostatic Implementation

In order to produce force results from the simulation, a force calculation has to be implemented. The last section of this chapter covers the electrostatic interaction calculation used in the simulation, and its application to the example simulation.

The simulation framework uses the equation for Coulomb's Law, derived in Section 1.3.4, to estimate the magnitude of the electrostatic interactions between the alpha-helices within each rod subdomain. The calculation is implemented where the electrostatic interactions of each non-repeating pair of alpha-helices in each bundle are determined in turn. Each possible non-repeating pair of sidechains between two helices are combined to calculate an interaction force. The results of the calculations between each pair of helices are summed. Equation 3.21a defines the calculation between two alpha-helices a and b. Where a_n and b_m are the total number of sidechains on each alpha-helix, k_e is Coulomb's constant defined as $\frac{1}{4\pi\epsilon_0}$, ${}^{a}e_i$ and ${}^{b}e_j$ are the i^{th} and j^{th} sidechains' electrostatic charge from alpha-helix a and b respectively. Finally, d is the Euclidean distance between the i^{th} and j^{th} sidechains' electrostatic charge.

$$F_{pair}(a,b) = \sum_{i=0}^{{}^{a}n} \left(\sum_{j=0}^{{}^{b}m} \left(\frac{k_c \; {}^{a}e_i \; {}^{b}e_j}{d^2} \right) \right) \tag{3.21a}$$

To calculate the value d, the recursive tree-walking function described by Equation 3.19 in Section 3.3.2 was used. Taking the same example simulation from the

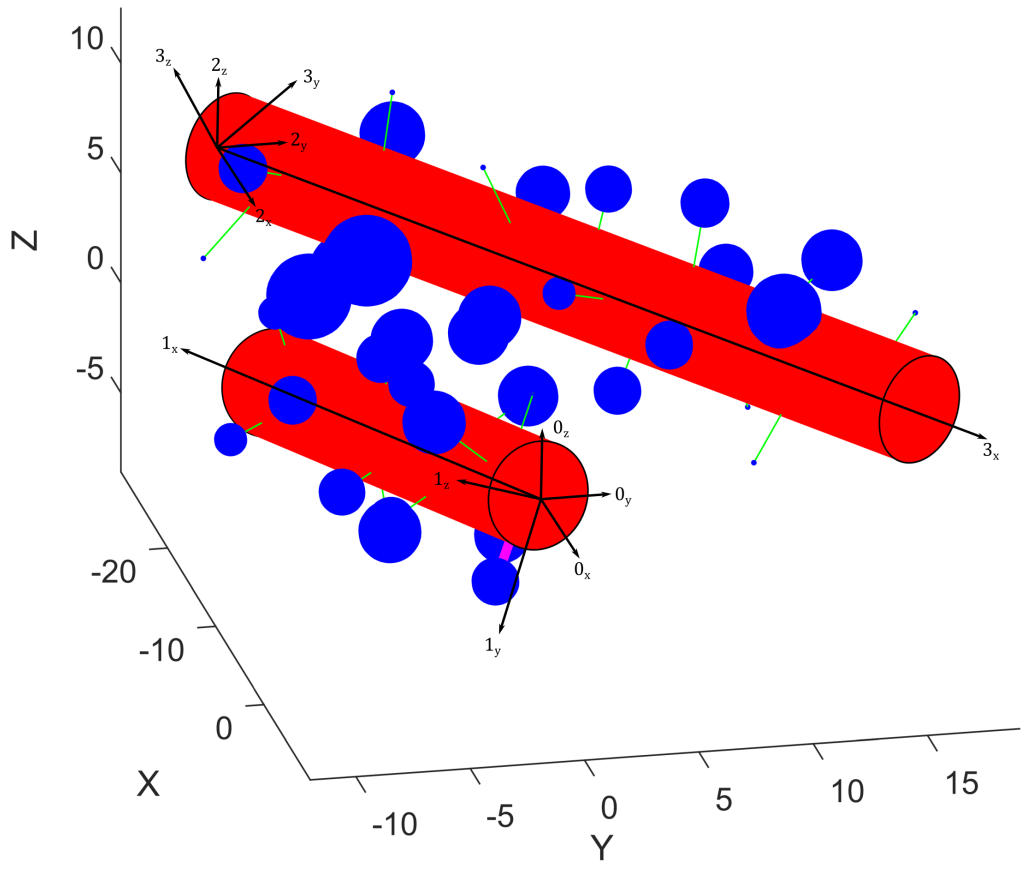


Figure 3.10: Image generated by the simulation framework of R3 helix one and helix two alpha-helix models. Red cylinder represents the backbone approximation and the blue spheres are the sidechain approximations. The green lines link the sidechain spheres to the backbone. The purple line indicates the first non-glycine sidechain link. The helix positions are shown after coordinate frame attachment and the transformation to the `pdb` conformation and translation back to simulation origin.

last section, ${}^{\alpha H1}P_i$ and ${}^{\alpha H2}P_j$ are the position of the i^{th} sidechain on alpha-helix one from the reference of frame $\alpha H1$ and the position of the j^{th} sidechain on alpha-helix two from the reference of frame $\alpha H2$. Both alpha-helices' sidechain positions are transformed into the SIM coordinate frame and the Euclidean distance between them is determined.

The positions of the two sidechains s_i and s_j were first converted from their alpha-helices' local coordinate frames to the SIM frame.

$$\begin{aligned} {}^{\text{SIM}}P_i &= {}^{\text{SIM}}_{\alpha H1}T\ {}^{\alpha H1}P_i \\ {}^{\text{SIM}}P_j &= {}^{\text{SIM}}_{\alpha H2}T\ {}^{\alpha H2}P_j \end{aligned} \tag{3.22a}$$

To determine the Euclidean distance between the two sidechain positions, the

2-norm is taken of the difference between the two positions.

$$d = \left\| {}^{SIM}P_i - {}^{SIM}P_j \right\|_2 \tag{3.22b}$$

This method was applied for each pair of sidechain interactions. The resulting force value $\sim$4.5 pN for the example simulation above with two alpha-helices gave results in the right order of magnitude for talin's rod domain. Additionally, an optimisation was made to increase the efficiency of the calculation. This was done computationally by vectorisation of the mathematical functions. This applied the coordinate transforms simultaneously to all sidechains positions of a helix by the recursive function, instead of executing the recursive function and transforming the sidechain pair positions individually.

In summary, to set up the simulation for a set of alpha-helices, the structural data for the alpha-helices was selected and pre-processed. Each alpha-helix was then approximated as a cylinder representing the backbone structure of the helix, and as a collection of spheres that represent the sidechain position. A local coordinate frame was attached to each alpha-helix model, which transformed all the model's positional data to be referenced through local coordinate frame. At the same time, a transform was created which positioned the local coordinate frame so its contained alpha-helix is returned to its original pdb position and orientation. This allowed for the original conformation of the alpha-helices to be restored after the local coordinate frames were attached. The single position and orientation of the local coordinate frame could then be manipulated instead of all the individual positions and orientations for each part of the alpha-helix model.

After the alpha-helix models and their coordinate frames were created, they were attached to the simulation coordinate frame. The simulation coordinate frame contained all other coordinate frames in the simulation. Their positions and orientations were referenced from this frame. Finally, in combination with a recursive algorithm, the electrostatic interactions between each alpha-helix were calculated and combined into the interaction force value for the alpha-helices.

4 Static Force Simulation Results & Force-based Structural Optimisation

The validation and results discussed in this chapter have been published in a paper with Scientific Reports [191]. A second paper on structural optimisation is currently under review.

In the previous chapter, the new alpha-helix model's backbone and sidechain approximations were described, along with the core structure of the simulation framework. In this chapter, the backbone and sidechain approximations and assumptions are validated. Next, the force results from the new simulation are detailed, followed by a discussion on further increasing the accuracy of the force results by improving the structural data. This leads on to the explanation of implementing an optimisation algorithm to search for more energetically stable conformations and the results of the subsequent simulation. Finally, the results of the two simulations are compared to experimental data in the form of force vs. extension curves.

4.1 Validation of Approximations

The alpha-helix model and the simulation framework are considered successful and meet the original specification when: the backbone and sidechain models are plausible and can approximate a distribution of alpha-helix and sidechain shapes and sizes; the electrostatic forces from the simulation are approximate to the physiologically relevant range of forces experienced by talin, and experimentally measured results, thereby improving upon the results of previous simulation methodologies; and finally, when the rod domain force values can be used to generate a force vs. extension curve that is similar to the experimental curves.

4.1.1 Backbone Approximation

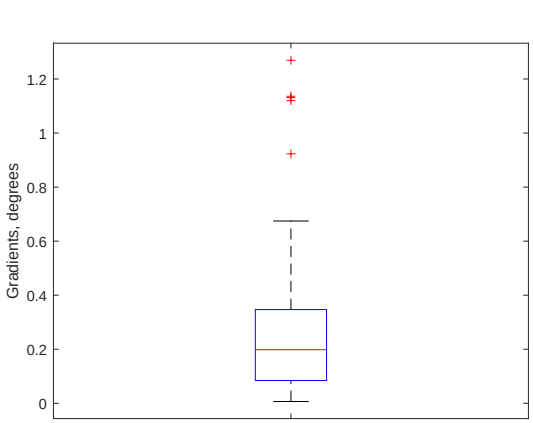
The approximation of the alpha-helix backbone used two assumptions to simplify the structure: 1) an alpha-helix is a straight, rigid structure, and 2) the size of the rigid alpha-helix is constant along its length. However, structures from X-ray crystallography show that alpha-helices in coiled-coil domains can be curved and twisted [206]. Even so, as long as the positions of the sidechain relative to each other and relative to the original position of the alpha-helix backbone are

maintained, the assumptions used to approximate the backbone of the alpha-helix did not alter the force calculations. These assumptions allowed the simplification of the alpha-helix amino-acid backbone positions by replacing them with a straight axis through the core of the alpha-helix. This axis represented the direction and length of the alpha-helix. The width of the alpha-helix was approximated as the mean distance between the axis and each backbone position.

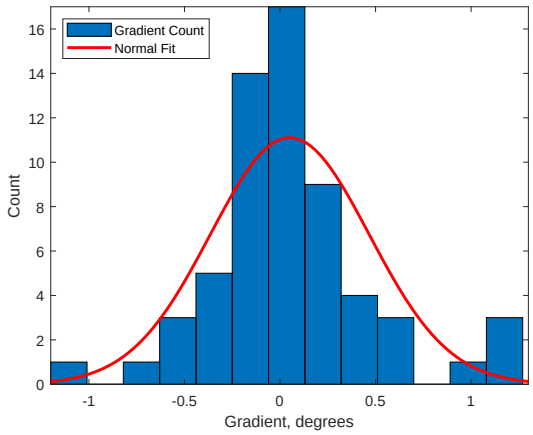
In order to test the alpha-helix backbone approximation and to validate the assumptions used to create the approximation, the original positions of the amino-acid backbone were compared to the approximated alpha-helix axis. This was accomplished by determining the closest Euclidean distance from the backbone position to the helix axis. A regression line through the distance data was calculated and its gradient determined. If the gradient was large, this meant that there was skew between the approximated helix axis and the backbone positions. This indicated that the curve in the helix was too large or the curve contained multiple changes in direction which prevented it from being approximated as a straight axis. A small or zero gradient meant that the axis approximated the backbone well. The gradient values from all helices were collected and plotted in a box plot, as a histogram with a normal distribution overlaid, and an ordered gradient distribution in Figure 4.1.

The values from the box plot in Figure 4.1a indicate the helix axes approximate the backbone position well, as the mean and median values are close to zero with a small standard deviation. The histogram in Figure 4.1b shows that the data was not biased. The plot in Figure 4.1c shows the regression line gradients ordered by magnitude. Almost two-thirds of the gradients are below the mean line and less than 10 of these gradients are outliers. Individual distance vs. helix index scatter plots were generated of two helices with good axis approximations. These are shown in Figure 4.2. For both, there is minimal difference between the green mean line and the red regression lines. However, as shown in the box plot, there were approximately six outliers. Two of these outliers were also plotted individually and are shown in Figure 4.3. Both regression lines have a larger difference compared to the mean line, but the histogram plot shows that the data is not biased.

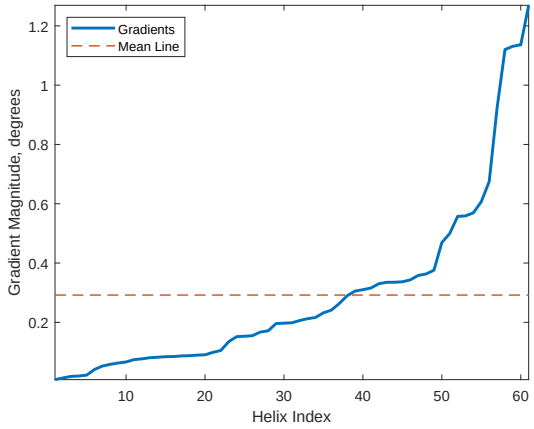
Possible improvements, such as splitting the backbone into individual segments that can be oriented separately from each other, or implementing a curved backbone axis are discussed in Section 5.1 of the discussion chapter. However, the majority of the backbone axes were approximated well by the helix axis, as their



(a) Box plot of the distribution of regression line angles.



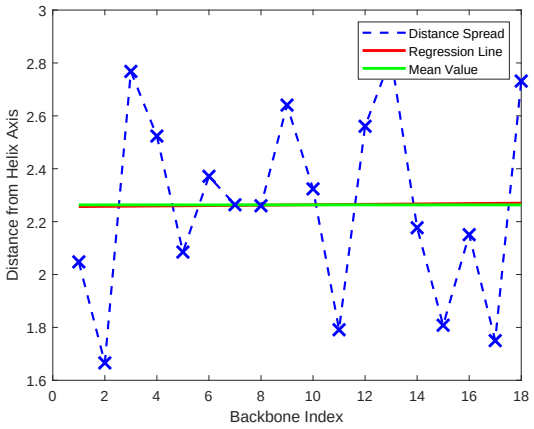
(b) Histogram of the distribution of regression line angles.



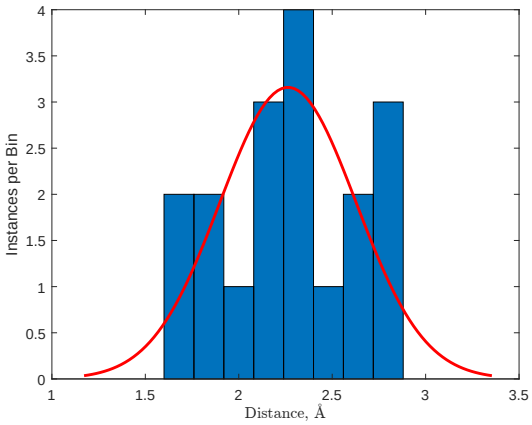
(c) Ordered gradient plot.

Figure 4.1: Different plots of combined gradient data of the regression lines calculated for each helix. (a) Box plot of absolute values: Median = 0.198, Mean = 0.292 and standard deviation = 0.299. (b) Histogram with normal distribution. (c) Ordered plot of absolute gradients with mean line indicated. (MATLAB, 2020b, https://www.mathworks.com)

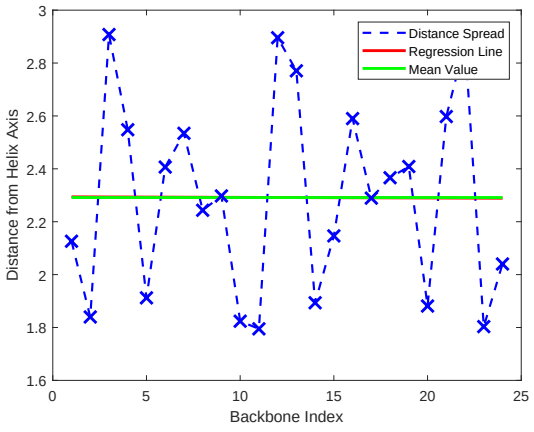
regression lines showed only a small difference in gradient (only six helices had regression lines with large gradients,$>$ 0.6, relative to the median value 0.198). Therefore, the backbone approximation achieved the requirements set, namely that it is plausible and that it can approximate a distribution of alpha-helices.



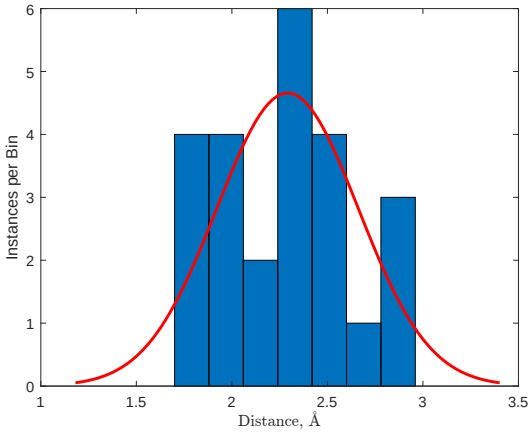
(a) Helix backbone radius and skew, subdomain R4, helix 3



(b) Box plot of the distribution of regression line angles.

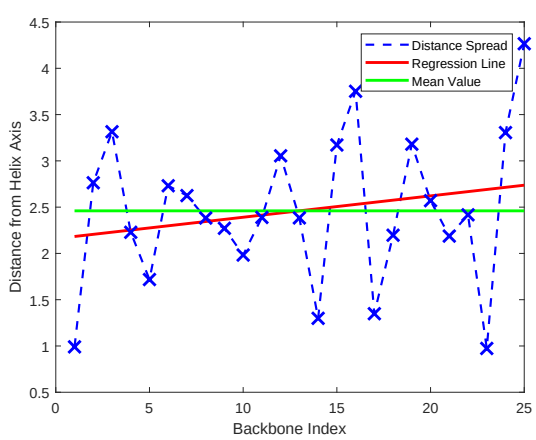


(c) Helix backbone radius and skew, subdomain R12, helix 2

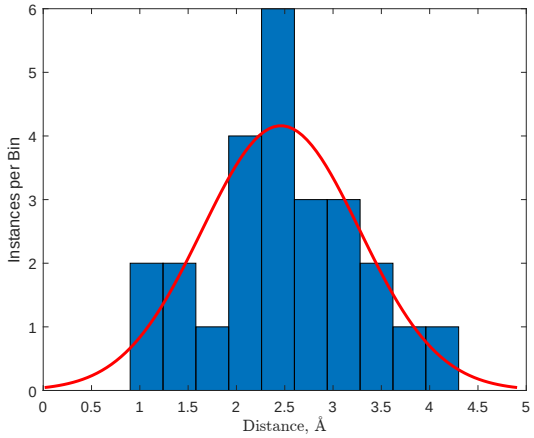


(d) Box plot of the distribution of regression line angles.

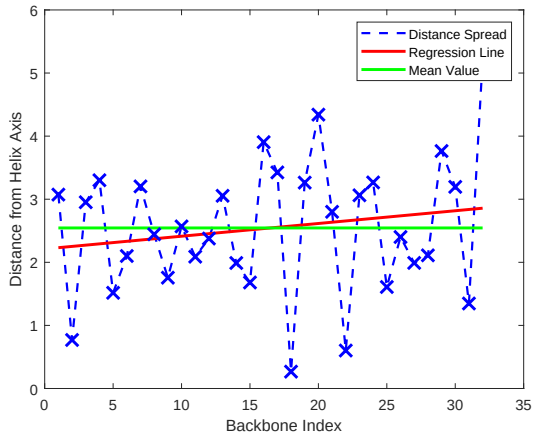
Figure 4.2: Individual Plots for R7H1 (a & b) and R12H2 (c & d) representing well approximated helices. (a & c) Distance and regression plots. Blue line and markers show the distances between the backbone positions and the helix axis. Green line is the mean distance line and the red line is the regression line. For these well approximated helix backbones, the difference between the mean line and the regression line is minimal at a gradient of $0.041°$ and $-0.012°$. (b & d) Histogram plots of the backbone distances. The graphs indicate that the distribution of distances are not biased, the approximated helix axis is appropriate, and the mean distance was a reasonable choice for the backbone cylinder radius. (MATLAB, 2020b, https://www.mathworks.com)



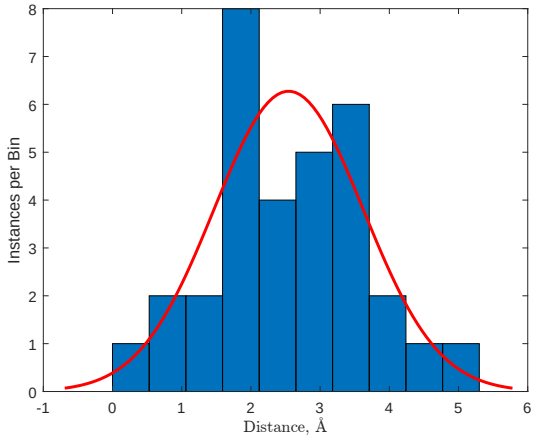
(a) Helix backbone radius and skew, subdomain R9, helix 2



(b) Box plot of the distribution of regression line angles.



(c) Helix backbone radius and skew, subdomain R1, helix 3



(d) Box plot of the distribution of regression line angles.

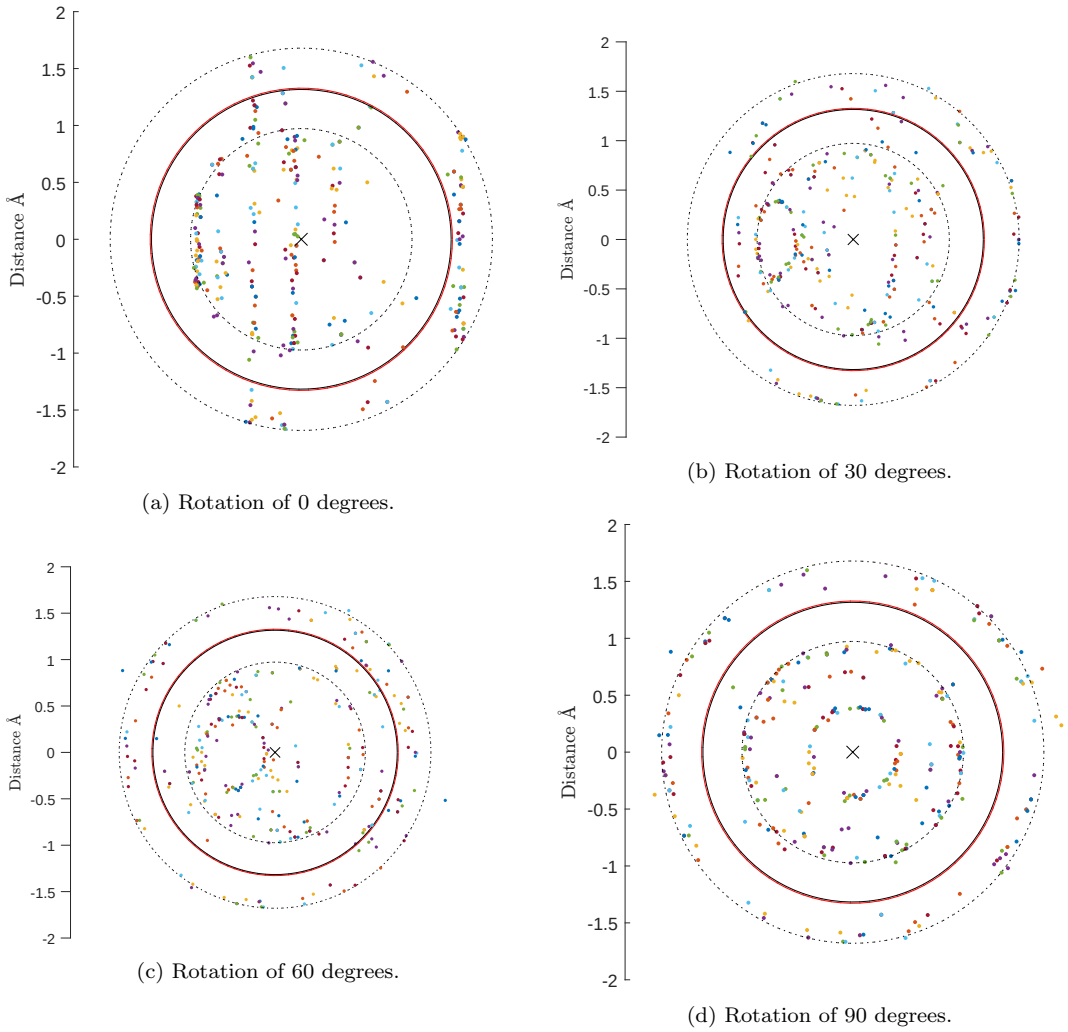
Figure 4.3: Individual Plots for R9H2 (a & b) and R1H3 (c & d) representing relatively poorly approximated helices. (a & c) Distance and regression plots. Blue line and markers show the distances between the backbone positions and the helix axis. Green line is the mean distance line and the red line is the regression line. For these poorly approximated helix backbones, the difference between the mean line and the regression line is larger at a gradient of 1.27° and 1.12°. (b & d) Histogram plots of the backbone distances. The graphs indicate that the distribution of distances are not biased. (MATLAB, 2020b, https://www.mathworks.com)

4.1.2 Sidechain Approximation

The sidechain approximation used the following assumptions to simplify the sidechains of the alpha-helices: 1) the position of the sidechain atoms can be represented by a single position at their centroid, 2) the mean distance of the sidechain atoms to their centroid can be used to approximate the size of the sidechain, 3) a sphere can be used to represent the shape of the sidechain, and 4) the properties of the sidechain, such as electrostatic charge, are equivalent to a point charge at the centroid position of the sidechain. Several coarse-grain structural approximations of amino-acid residues have successfully applied similar approximations to simplify their models [155, 157, 207], which gives precedent to this approach.

To further verify the approximations, the distances between each sidechain atom and the centroid of their position were calculated for every sidechain on all alpha-helices used in the simulations. Also, the position of the sidechains was calculated and zero-meaned with respect to their corresponding backbone position, enabling comparisons between them. Additionally, the orientation of the sidechains was aligned. This was done by rotating each collection of sidechain atoms so that the vector direction of the line between their carbon-alpha position and their centroid position was equal. As a result, the average shape of each sidechain could be visualised, irrespective of its orientation. This process was applied to every atom of every sidechain. Next, all sidechain atoms were grouped by sidechain type. To help visualise this data, the sidechain atom positions for all isoleucine sidechains were plotted from different perspectives shown in Figure 4.4. These plots include guide circles that define four regions in the data. The inner and outer dashed circle show the first and third quartile distance from the atoms to the centroid position at the origin. The solid black and red lines show the median and mean distance respectively.

Three other sidechains were plotted in the same manner and are shown alongside isoleucine in Figure 4.5. In plots 4.5a, 4.5b and 4.5c, the majority of the sidechain atoms can be seen within the red mean circle and the atoms position are evenly distributed around the plot. This shows that this approximation worked well for these sidechains across all helices. In plot 4.5d the sidechain atoms are less evenly distributed with a bias towards the left and right sides of the plot. However, the majority of the atoms in 4.5a are still within the mean line, because the black median line is also within the mean line. This approximation is valid for cases such as histidine, as it generalises the response to collisions without needing to consider the directions that collisions come from. Other methods of approxi-



(a) Rotation of 0 degrees.

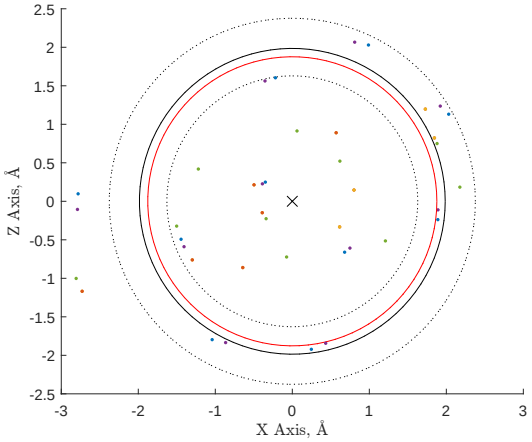
(b) Rotation of 30 degrees.

(c) Rotation of 60 degrees.

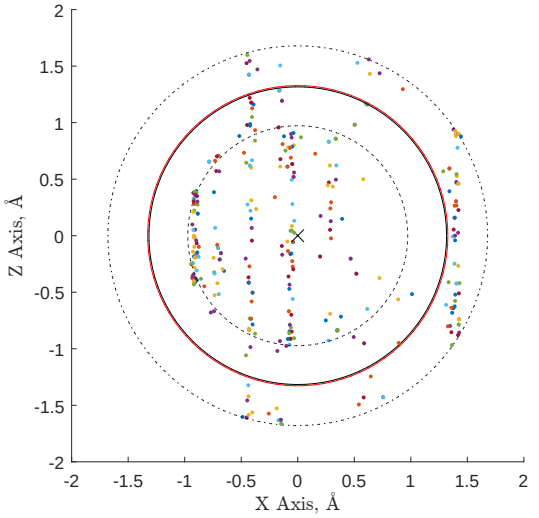
(d) Rotation of 90 degrees.

Figure 4.4: (a-d) The representation of sidechain atoms around their centroid position for all the isoleucine sidechains. Each panel in the figure shows the same data but from a different angle. (MATLAB, 2020b, https://www.mathworks.com/)

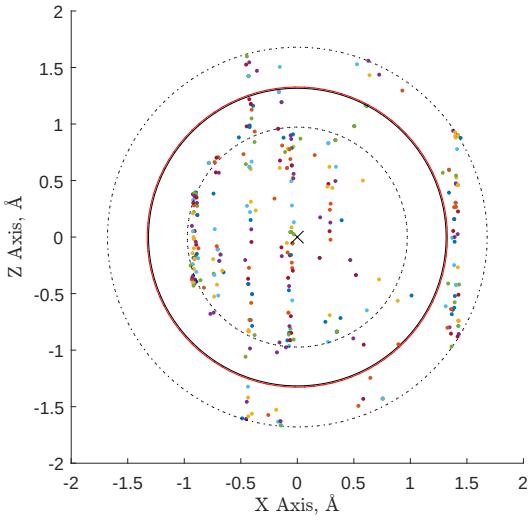
mating the sidechains such as CABS, UNRES and Dual Pseudo Atom (DPA) were also considered. However, the increase in the accuracy of these other methods was outweighed by the large increase in approximation complexity. These other methods are discussed in detail in Section 5.2 of the discussion chapter, along with the reasons for selecting the method used in this thesis and why it worked well in this context.



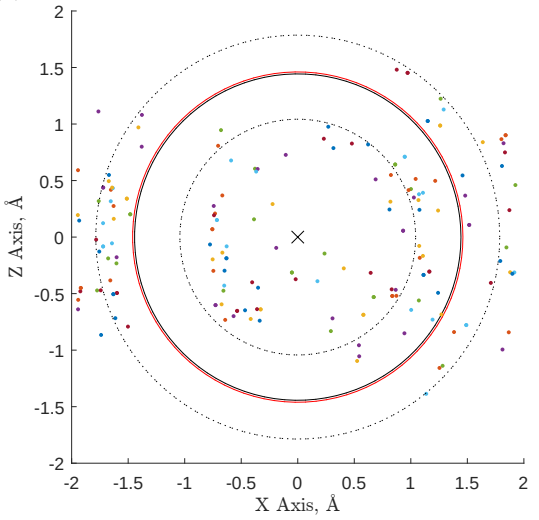
(a) Scatter plot of tryptophan sidechain atoms positions.



(b) Scatter plot of isoleucine sidechain atoms positions.



(c) Scatter plot of methionine sidechain atoms positions.



(d) Scatter plot of histidine sidechain atoms positions.

Figure 4.5: Sidechain atom position scatter plots of four different sidechains. (a) Tryptophan. (b) Isoleucine. (c) Methionine. (d) Histidine. Solid black circle indicates the median distance and solid red line indicates the mean distance. Inside dashed circle and outside dashed circle indicates the first and third quartile distances respectively. (MATLAB, 2020b, https://www.mathworks.com/)

4.2 Simulation Results

To generate force results for each rod subdomain of talin, a simulation was set up for each rod subdomain using the process described in Section 3.3.2. The selected structural data for each rod domains and its corresponding alpha-helix residue index ranges are shown in Table 4.1.

Table 4.1: The selected pdb structure shown as PDBID for each rod domain of talin. The index ranges for each alpha-helix are also listed. The H5 helix of R2, R3, R4 and R8 is blank as these rod domains have only four helices.

Domain	PDBID	H1	H2	H3	H4	H5
R1	6R9T	5-29	41-75	85-116	121-140	149-170
R2	6R9T	8-38	44-68	79-100	108-136	——
R3	2L7A	10-26	30-57	61-90	95-110	——
R4	2LQG	2-27	41-63	68-94	102-130	——
R5	6R9T	10-34	45-72	77-102	106-135	140-162
R6	2L10	2-18	30-55	59-84	90-119	124-148
R7	5FZT	2-19	31-58	61-91	239-264	269-295
R8	5FZT	107-123	130-156	161-188	194-217	——
R9	4F7G	5-30	42-66	71-98	102-129	137-167
R10	2KVP	8-21	31-58	63-87	91-120	125-149
R11	3DYJ	2-26	38-63	67-94	100-127	135-166
R12	3DYJ	12-31	43-66	69-94	101-130	134-160
R13	2JSW	2-25	42-74	81-117	122-143	152-177

After each rod subdomain's simulation was set up, the electrostatic interaction forces were determined using the method from Section 3.3.4. The electrostatic interaction force results for each rod domain bundle are shown below in Table 4.2.

In terms of absolute force values, the simulation force results range from 5.26 pN to 43.6 pN, encapsulating the ranges of unfolding force events recorded experimentally by Yao et al. [12, 161]. The force values were also at the order of magnitude, 10^{-12} to 10^{-13} N, matching the physiologically relevant range of force for talin. This is a significant improvement on the accuracy of the previous computational results, where results ranged from 10^{-9} to 10^{-12} N [9, 142]. This improved range of interaction force values shows that the new simulation method has indeed improved upon efforts to determine unfolding forces for talin's rod domain. The data in Table 4.2 were used to generate a force vs. extension curve that was plotted with the digitised version of Yao et al.'s data in Figure 4.6. Each unfolding event was set to extend the rod domain by 40 nm and the R7 and R8 event were combined with a double extension length (Unlike Yao et al.'s data

Table 4.2: The calculated electrostatic force values between the alpha-helices for each alpha-helix coiled-coiled subdomain within talin's rod domain. Negative force values represent attractive interaction and positive force values represent repulsive interactions. The three outliers were R2, R4 and R8 with repulsive interactions.

Bundle	Force, pN
R1	-30.9
R2	18.6
R3	-8.90
R4	22.0
R5	-40.3
R6	-20.5
R7	-7.28
R8	12.2
R9	-5.27
R10	-24.5
R11	-29.3
R12	-11.0
R13	-43.6

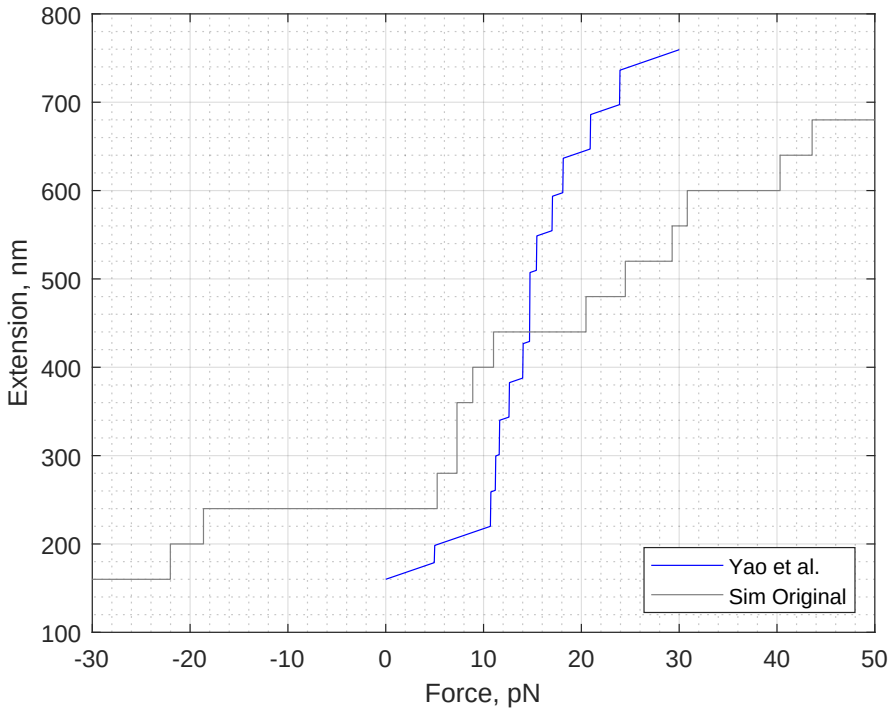


Figure 4.6: Estimated unfolding force response of talin rod domain using static force data from simulation. Digitised data from Yao et al. [12] shown for comparison.

this plot did not include the extension contribution from the stretching of linker domains and other unstructured sections of talin's rod domain).

This data and plot highlight three outlying values. The R2, R4 and R8 domain values were all positive at 18.6 pN, 22.0 pN and 12.2 pN respective and shown extending far to the left of the graph. Due to the polarity of Coulomb's law, interactions that result in negative force values are attractive interactions and interactions that result in a positive force value are repulsive. Therefore, the R2, R4 and R8 domains are overall repulsive in nature. The repulsive values are in contrast to the current evidence for the unfolding events of talin's rod domain, where all the domain are folded in their resting state. Interestingly, all three subdomains only contain four alpha-helices. There may be a correlation between the four helices and the repulsive outcome. Since R3 also has four helices but was not repulsive, the correlation was not strong. The repulsiveness was later resolved using an optimisation simulation. The results of this additional simulation showed that the R2 and R4 domain were attractive, and also explored the special case of the R8 domain.

4.3 Structural Optimisation

While investigating a method to explain the previously generated, repulsive forces of the R2, R4 and R8 subdomains, Figure 4.7 was created. This data came from a simulation arrangement where two helices from the same rod domain bundle were attached to the simulation frame and positioned in their original pdb conformation. Instead of calculating a single electrostatic interaction, the helices were rotated and at each rotation interval an interaction force was calculated. This resulted in a force value for each possible combination of rotations for each helix. These values were plotted as a surface plot in Figure 4.7 where the X and Y axes define the rotation of the first and second helix respectively and the Z-axis shows the resulting force value. Unexpectedly, this figure indicated that the helices were in their most unstable configuration.

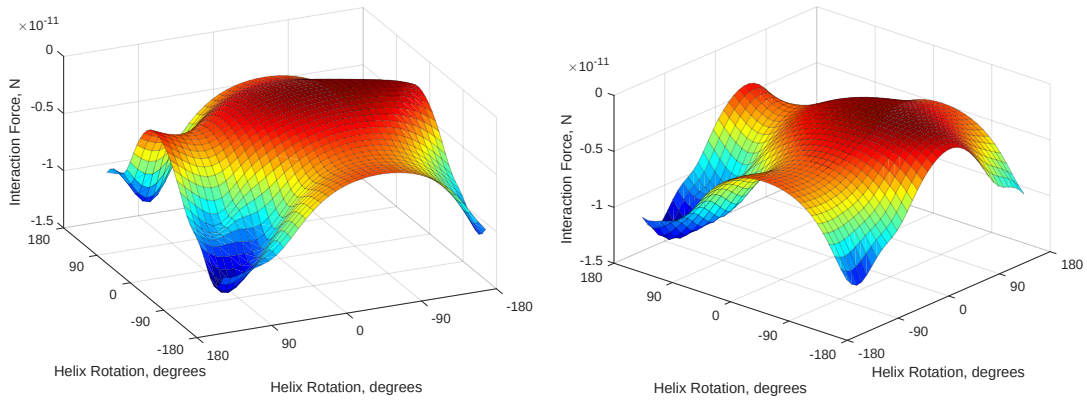


Figure 4.7: (a-b) Surface plot of interaction force values of helix one and two for the R6 rod domain. The two helices rotated through 360 degrees. Figure (b) contains the same data as (a) but with a perspective rotated by 90 degrees.

The position $(0, 0)$ in Figure 4.7 indicates the same rotation and orientation of the alpha-helix as in its original pdb conformation. This position's force value is almost at the highest point of the surface plot with the lowest repulsive force (red area), where the higher repulsive force values correspond to lower stability of the configuration. The 2^{nd} law of thermodynamics indicates that in the absence of external forces talin's protein structures in a living cell tends towards the lowest stable energy state. This means that talin's rod domains tend towards being as attractive as possible. However, the data in Figure 4.7 shows that the pdb conformation of these two alpha-helices is not in the most stable state as more attractive interactions are achievable with other rotational conformations. Using these more stable conformations in the simulation may result in an improved model for the unfolding of talin that more closely matches the experimental data. Obviously, the data shown in Figure 4.7 is not taking into account the other alpha-helices within the R6 domain. Therefore, the next step was to confirm that more stable conformations exist for the full R6 domain and the other rod subdomain.

However, each alpha-helix increases the dimensional complexity of the resulting data and two alpha-helices are already the limit for visual plots. Furthermore, the computation time increases dramatically with each additional alpha-helix, as the generation of the force data from the rotations is an $O(n^2)$ complexity problem with respect to time and grows exponentially. Therefore, an optimisation algorithm was used to decrease the computational time complexity by 'searching' the problem space for other lower energy states.

4.3.1 Optimisation Algorithm

The electrostatic interaction values in Table 4.2 (on page 87) were the starting point for the optimisation algorithm and the baseline for comparison of the results from said optimisation algorithm. In addition to rotating each alpha-helix, the new simulation also translates each helix along its axis in order to explore additional conformations.

The optimisation algorithm *Fmincon* (from MatLab's Optimisation Toolbox) was used and allows optimisation with imposed restraints (see Equation 4.1). The *Fmincon* algorithm minimises the output of a constrained non-linear multivariable function, in this case the force calculation. In other words, using the algorithm it is possible to identify new alpha-helix bundle conformations that have lower internal energy states by translating and rotating alpha-helices within a set of constraints. The *MultiStart* configuration was used to run multiple instances of the optimisation algorithm in parallel. Input variable x was the position and rotation of each helix within a bundle, lb was the vector of lower constraints of x and likewise, ub were the upper constraints of x.

$$\min_x f(x) \text{ such that } \{lb \leq x \leq ub\} \tag{4.1}$$

The initial results from the optimisation algorithm included attractive forces that were larger than the physiological relevant ranges of talin by several orders of magnitude. This was the result of charged sidechains intersecting each other during the simulation. A penalty function was designed and implemented into the cost function used by the optimisation algorithm. This prevented sidechains from colliding and intersecting each other. The penalty function is defined in Equation 4.2.

$$\begin{aligned} x &= d - (r_1 + r_2) \\ F &= -k_s x \end{aligned} \tag{4.2}$$

The penalty function was modelled on a simple spring interaction where F was the penalty force, k_s was a coefficient value set at 10^{-6} which prevented sidechain overlap while also preventing the optimisation algorithm becoming too stiff. As input values near their constraints, a stiff optimisation algorithm can become very slow. Value x was the Euclidean distance of intersection between the two sidechains. Where r_1 and r_2 were the radii of the two sidechains and d was the

Euclidean distance between the centres of the two sidechains. If the x value of the penalty function was negative, meaning that the distance between the centres of the two sidechains was less than the total radii of the sidechains, the calculated force was added to the cost function. This increased the value of the cost function, which steered the optimisation algorithm's search away from sidechain collisions.

The optimisation algorithm modifies the alpha-helices' axial rotation and translation along the alpha-helices' primary axis to determine the conformation with the highest internal attractive force between all the alpha-helices within the bundle simultaneously. The upper and lower bounds of the rotation and translation of the helices were set at ± 15 degrees and ± 10Å respectively compared to the original position defined in the pdb structures. These values were chosen to ensure that the resulting orientation and position of the alpha-helices did not diverge too far from the original pdb data, as the pdb conformations have clearly yielded usable results in the past.

4.3.2 Results of Optimisation

The optimisation of the thirteen rod domain bundles was conducted while constraining the bounds of the optimisation search so that confirmations wildly different from the PDB structure were excluded. The initial results before the implementation of the cost function based collision system are shown in Table 4.3. Three subdomains exhibited extremely high attractive forces that were orders of magnitude larger than the original force values and the physiologically relevant range of force for talin. These subdomains were: the R6 at 600 pN, the R10 at 130 pN, and the R12 at 1630 pN. Inspection of the resulting structures of these subdomains after optimisation clearly showed that some sidechain positions had overlapped. Therefore, the cost function based collision system was implemented to steer the optimisation process away from these incorrect conformations.

Following optimisation with the cost function based collision system, twelve of the thirteen bundles showed a change in conformation to a more energetically favourable one, resulting in a negative shift in their electrostatic interaction force values, corresponding to an increase in the alpha-helices' overall electrostatic attractiveness to each other within each bundle (Table 4.4 and Figure 4.8). The attractive interactions increased on average by 12.1 pN across the thirteen bundles.

Table 4.3: Comparison of original alpha-helix bundle forces with the forces values generated from the optimisation simulation before the implementation of the cost function based collision system. The difference between the original structure's force and the optimised structure's force is given as the force delta. Subdomains where collisions produced large force values: R3, R10 and R12.

Bundle	Original Force, pN	Optimised Force, pN	Force Delta, pN
R1	-30.8	-63.8	-33
R2	18.6	-0.958	-19.6
R3	-8.9	-28.8	-19.9
R4	22	-41.1	-63.1
R5	-40.3	-75.2	-34.9
R6	-20.5	-600	-580
R7	-7.28	-18	-10.7
R8	12.2	3.89	-8.33
R9	-5.26	-28	-22.7
R10	-24.5	-130	-106
R11	-29.3	-42.9	-13.6
R12	-11	-1630	-1610
R13	-43.6	-79	-35.4

Table 4.4: Comparison of original alpha-helix bundle forces with the forces values generated from the optimisation simulation after the implementation of the cost function based collision system. The difference between the original structure's force and the optimised structure's force is given as the force delta. The mean value of all original force is $-13\,\mathrm{pN}$. The mean force of the optimised force is $-25.1\,\mathrm{pN}$. Bundle R9 was the only bundle to have no change between its optimised structure and its original structure. Position and rotation delta boundaries were not hit for any bundles in the simulation.

Bundle	Original Force, pN	Optimised Force, with Collision pN	Force Delta, pN
R1	-30.8	-39	-8.2
R2	18.6	-0.958	-19.6
R3	-8.9	-28.8	-19.9
R4	22	-6.31	-28.3
R5	-40.3	-44.5	-4.18
R6	-20.5	-25.7	-5.21
R7	-7.28	-18.7	-11.4
R8	12.2	3.89	-8.33
R9	-5.26	-5.26	0
R10	-24.5	-32.6	-8.08
R11	-29.3	-42.2	-12.9
R12	-11	-34.7	-23.7
R13	-43.6	-44.4	-0.837

The change in position and orientation of each helix was also collected and shown in Table 4.5. There were no correlations between the change in magnitude of the position or rotation deltas and the change in magnitude of the force delta. Therefore, the position and rotation deltas are a poor indicator for changes in the force delta.

R9 was the only bundle that showed an unchanged conformation after the simulation. None of the simulated bundles' optimisation parameters hit the limit constraints after optimisation was complete. This highlights that the optimisation is successful and that the boundary conditions did not limit or alter the simulation results.

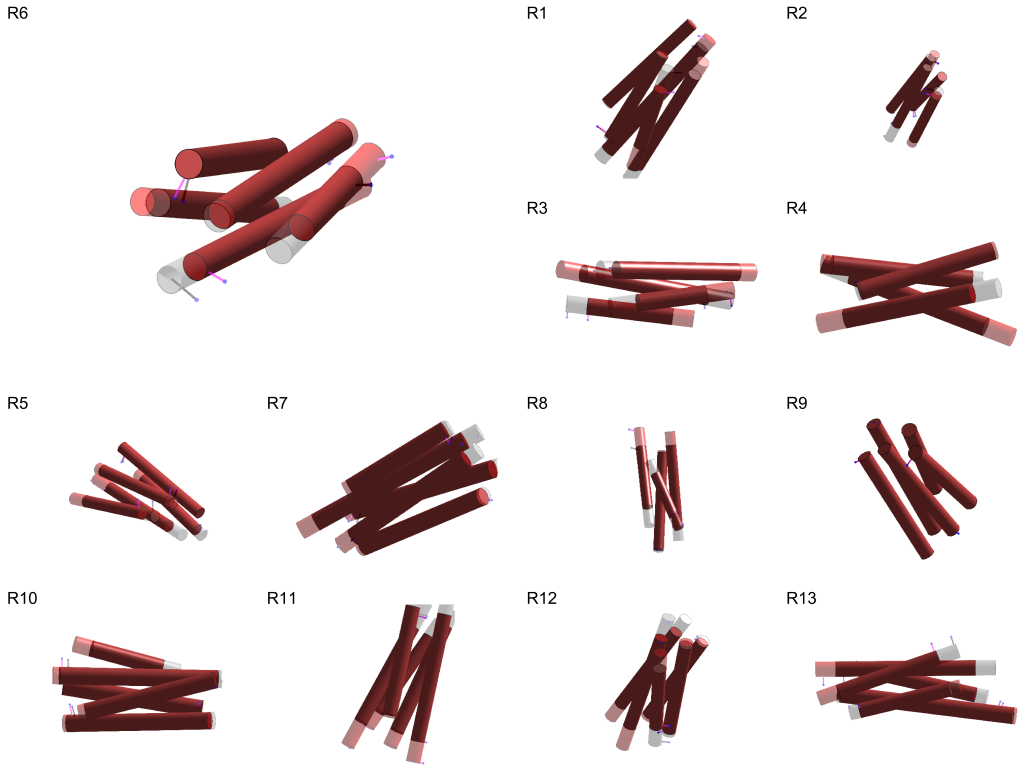


Figure 4.8: Original alpha-helix bundle R6 superimposed with the final conformation after optimal force interaction simulation. Original bundle coloured grey and the simulated bundle is coloured red. The first sidechains are given on each helix to visualise the difference in rotation. Original helices have the first sidechain coloured black, and the simulated helices are coloured magenta. A rotation limit of ± 15 degrees of pdb conformation and a translational limit of ± 10 Å were set as the bounds for the simulation. Resulting force was $-25.7\,\mathrm{pN}$. Axial rotations of each helix were $(10.5, 13.9, -5.67, -11.0, -11.8)$ degrees. The remaining 12 alpha-helix bundle images are tiled.

After the application of the optimisation algorithm to the complete bundles, R2 and R4 shifted to attractive electrostatic interactions. Only R8 remained repulsive at $3.89\,\mathrm{pN}$. The repulsive nature of R8 may be explained by previous work

Table 4.5: The resulting change in position and rotation of each helix within the structure is given by the position and rotation delta. The order of the position and rotation deltas were as follows: [helix one, helix two, helix three, helix four, helix five]. The four helix bundles: R2, R3, R4 and R8 only display four values as they contain only four helices.

Bundle	Position Delta, Å	Rotation Delta, Degrees	Mean Postion Delta, Å	Mean Rotation Delta, Degrees
R1	[0.401, 6.9, -4, 0.617, 9.14]	[-9.39, 0.397, 9.03, 12.6, -7.36]	2.61	1.06
R2	[9.37, 0.45, -0.41, 4.14]	[13.8, -12.7, -9.88, -2.03]	3.39	-2.69
R3	[-7.61, 4.97, 7.18, 5.6]	[-13.4, 12.9, 2.3, 2.52]	2.54	1.08
R4	[2.61, 0.769, 7.22, 7.65]	[-10.6, -10.3, -8.03, 13.5]	4.56	-3.86
R5	[7.76, -0.193, 2.62, -9.26, 3.59]	[-10.8, 13.3, -7.14, -9.29, -10.4]	0.905	-4.89
R6	[-7.88, -0.0381, -1.71, 7.34, 5.16]	[10.5, 13.9, -5.67, -11, -11.8]	0.574	-0.834
R7	[1.37, -2.14, 4.52, -4.07, 0.969]	[-9.55, -8.68, 0.911, -7.73, -9.55]	0.13	-6.92
R8	[-6.48, -6.06, 0.716, -9.07]	[11.7, -2.88, 12.4, 1.84]	-5.22	5.77
R9	[0, 0, 0, 0, 0]	[0, 0, 0, 0, 0]	0	0
R10	[4.79, -0.94, 0.0922, -1.76, 1.89]	[8.16, -9.81, 7.97, 9.63, 6.1]	0.812	4.41
R11	[-6.06, 0.716, -9.07, 5.38, -6.93]	[11.7, -2.88, 12.4, 1.84, -9.72]	-3.19	2.67
R12	[7.4, 1.21, 1.18, 8.86, -6.68]	[1.41, -7.11, 13.2, 0.649, -0.872]	2.39	1.45
R13	[3.67, -6.22, -0.879, 3.11, 6.47]	[-4.79, -4.87, 3.72, 5.94, -2.27]	1.23	-0.452

documented in literature that has indicated that R8 is unstable and allows vinculin to bind to its cryptic binding site in the absence of force [134]. Additionally, it has been proposed that this binding of vinculin to the R8 subdomain in the absence of force is required to promote focal adhesion maturation and to increase traction force [208]. The R7 subdomain likely stabilises the R8 subdomain in place with the R7 bundle unfolding at approximately $-15\,\text{pN}$ [12]. The non-optimised values for R7 and R8 were $-7.28\,\text{pN}$ and $12.2\,\text{pN}$, a net repulsion of $4.92\,\text{pN}$. The optimised structures, however, put the interaction value at $-18.7\,\text{pN}$ and $3.89\,\text{pN}$, resulting in a net attractive value of $-14.81\,\text{pN}$. This strong similarity between experimental and in silico results for the R7 subdomain, validates the electrostatic force simulation developed for this thesis.

To evaluate the results from both the optimisation simulation and the previous force-based simulation, the force values were compared to existing experimental results. To begin, rod domains were ranked in ascending order of unfolding force (Table 4.6). Previous studies have calculated that the average unfolding extension of talin's rod domain bundles is $40\,\text{nm}$ [8, 12]. Using the unfolding force values from the results in Table 4.2 and the unfolding extension per bundle of $40\,\text{nm}$, a force-extension graph was generated for the results of both the previous force-based simulation and the optimised structures simulation. These were compared against a digitised version of the force-extension results recorded experimentally by Yao et al. [12] (this comparison is visualised in Figure 4.9).

Experimental data from literature shows that the rod domains of talin unfold

Table 4.6: Estimated unfolding force thresholds for the optimised structures and the corresponding estimated unfolding order of the alpha-helix bundles.

Unfolding Order	1	2	3	4	5	6	7	8	9	10	11	12
Bundle	R2	R9	R4	R7(R8)	R6	R3	R10	R12	R1	R11	R13	R5
Force	-0.96	-5.26	-6.31	-14.9	-25.7	-28.8	-32.6	-34.7	-39.0	-42.2	-47.6	-48.3

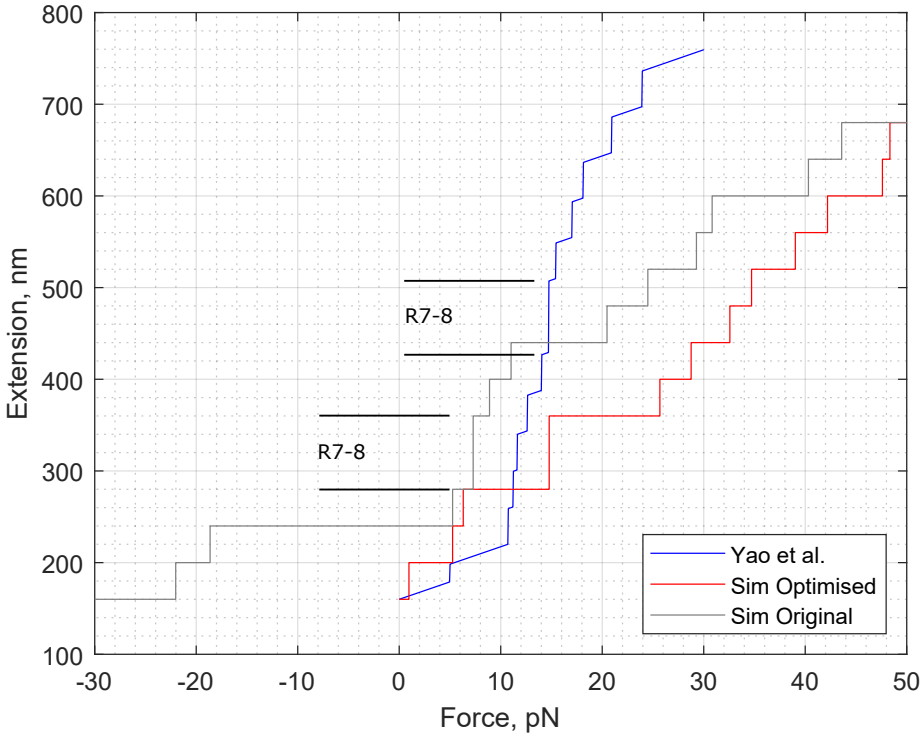


Figure 4.9: Estimated force vs extension response using original conformation unfolding force data, grey line. Estimated force vs extension response using optimised conformation unfolding force data, red line. Experimental data for comparison, blue line [12]. Double unfolding event from R7-8 bundles is highlighted. A scaling factor of around 3.5 is needed to align the previous force-based simulation data to the experimental data. A scaling factor of 2 is needed to align the optimised simulation data with the experimental data.

within the range of 5 pN to 25 pN [12]. Previous simulations have been unable to replicate this force range, producing peak unfolding forces of up to 1500 pN [9, 141–143]. Using PDB structures, the force-based simulation produced unfolding forces between −22 pN to 44 pN. This calculated force range covers 330% of the physiologically relevant range of force for talin, which is ∼5 pN to ∼25 pN. Following optimisation of bundle structures to their most energetically favourable conformation, the new simulation produced unfolding forces between ∼1 pN to ∼49 pN, covering ∼ 240% of the physiologically relevant range of force for talin (Figure 4.9). Furthermore, the large unfolding event at 15 pN seen in the ex-

perimental data which corresponded with the dual unfolding of R7 and R8 was replicated in our optimised simulation (Figure 4.9). Optimisation of structures, therefore, increased the alignment of the range of force values by 38%, as well as removing erroneous repulsive bundles. Previous results from molecular dynamics simulations covered $\sim$ 7500% of the range. Therefore, the results presented in this thesis show a more than $\sim$ 30 times improvement in matching the physically relevant range of unfolding forces for talin's rod domain compared to previous computation methods.

To further validate the simulation, the R7 domain S1589D mutation was explored using the simulation. The S1589D mutation is caused by the phosphorylation of a serine residue at the start of the fourth helix of R7. This mutation was created by Gough et al. to investigate the stability and interactions of the R7 and R8 domains [209]. The mutation was shown to reduce the stability of the two domains, resulting in their combined unfolding event occurring at a lower force threshold of $\sim$10 pN compared to $\sim$14 pN. A new pdb file for the mutant form was created from the original pdb file by swapping a serine residue of the R7R8 pdb at index 1589 for an aspartic acid residue. This resulted in an effective charge change of 0 to -1. Following this alteration, the unfolding force of the R7–S1589D domain was calculated using the simulation. The results are shown in Table 4.7.

Table 4.7: Comparison of force results for the S1589D mutation of the R7 domain.

Bundle	Original Force, pN	Optimised Force, with Collision pN	Force Delta, pN
R7	-7.28	-18.7	-11.4
R7–S1589D	-7.15	-15.9	-8.73

The result of the unfolding force simulation for the mutant R7-S1589D pdb was -15.9 pN compared to -18.9 pN for the wild-type R7 pdb. This shows that in the simulation, the R7-S1589D mutation reduced the stability of the subdomain and therefore its unfolding force threshold. Calculating the unfolding force threshold of the combined R7R8 subdomain with the S1589D mutation results in a value of -12.0 pN compared with -14.9 pN without the mutation. This change in force value also shows a reduction in stability. These new unfolding force values corroborate the experimental results by Gough et al. [209] and further validate the simulation methodology.

The optimisation of rod domain structures improved the range of electrostatic

interaction values for the rod domain bundles. This indicates that further validation and refinement of structures from the PDB could improve other simulations of talin's rod domains. It also validates the simulation framework and methodology described here as a useful tool for manipulating and simulating alpha-helix interactions. Future work involving the application of the framework to other research questions than static force estimation is covered in the discussion.

Results from both the previous simulation and the optimised structures simulation are much closer in range and absolute values to the experimental values and the known physiologically relevant range of forces, compared to force results from previous computational methodologies. Even though the force-based coordinate-frame simulation framework shown here currently computes only electrostatic interactions, these are very encouraging results. Possible areas for improvement in the models and the simulation framework have been identified, including: implementing additional non-bonded intermolecular interactions and enhancing the accuracy of the alpha-helix model by breaking it into multiple flexible segments. These possible improvements are being considered for future work and are covered in more detail in the next chapter.

5 Discussion

In this thesis, the following has been achieved: a novel simulation framework that operates directly within the force-domain was developed. Approximated models of alpha-helices were designed and used in conjunction with the simulation framework to simulate the unfolding forces of the rod subdomains of talin. This simulation produced unfolding forces in the range of ∼1 pN to ∼49 pN. The physiologically relevant range of forces for the unfolding of talin's rod domain is ∼5 pN to ∼25 pN. This is a significant improvement over previously used computational methods such as molecular dynamics, where simulations produced force ranges between ∼1 pN to ∼1500 pN. The improvement of these results is equivalent to a greater than 30 times reduction in the computed unfolding force range of talin, only 2.4 times larger than the physiologically relevant range for talin.

In this chapter, several topics that warrant further discussion are covered.

1) The helix backbone model was shown to be an appropriate approximation for the majority of helices in Section 4.1.1. However, there were Approximately six outlying helices that were not approximated well by this method. In Section 5.1, improvements that are considered in order to improve the backbone model in future simulation are discussed.

2) For the sidechain approximation, a simple model was chosen. A more detailed model of the residue sidechains may have improved the accuracy of the model and therefore the results of the unfolding force simulation. Section 5.2 describes other considered sidechain models and argues why a more detailed model was not the best option.

3) Hydrophobic interactions play an important role in protein folding similar to electrostatic forces, but these interactions were not included in the simulations. The reasons for not including these interactions are explored in Section 5.3, along with a description of the efforts made to approximate the hydrophobic effects as direct forces.

4) Section 5.4 covers the considerations for using a coordinate frame system as the method of spatial partitioning in the simulation framework, when other alternatives exist and are used in other fields.

In the final section, an overview is given of potential avenues for further research that have been highlighted earlier in this thesis and in this discussion chapter.

5.1 Helix Backbone Model

In section 4.1.1, the approximated model of the alpha-helix backbone structure was validated as an accurate representation. However, the validation data mentioned above and shown in the figures from Section 4.1.1 indicated that the helices with higher degrees of curvature were not approximated as well by the backbone model. For example, helix two of R9 and helix three of R1 were highlighted as cases of poor backbone approximation, see Figures 4.3a and 4.3c in Section 4.1.1. Both regression lines (which indicate the skew of the backbone position around the approximated axis) had a significantly higher gradient when compared to other helices. This was an average increase of 1.17° or an increase of a factor of ~ 44 from the best examples of backbone approximation (shown in Figures 4.2a and 4.2a).

For the current iteration of the simulation framework and the static force simulations, the accuracy of the backbone axis for each helix did not need to be very high. The two positions defining the backbone axis are currently used to form the local coordinate frame for that helix, for generating visualisations of the helix and finally to act as a reference point for determining the positions of the sidechains. The sidechain's positions are directly determined from the helices' structural data before being referenced from the helix model. Therefore, the sidechain positions are not affected by the backbone approximation. This means that the calculations based on interaction force are also not affected by the backbone positions. However, simulations developed in the future might require more accurate helix backbone models, for example because the helix structure could be used in collision calculations.

The approximated model of alpha-helix backbone structures, which assumed a straight helix with uniform width, was unable to approximate these helices accurately. Two possibilities were considered for causing the poor approximations: 1) the radius of the actual helix's backbone structure was increasing along the length of the helix or 2) the helix was curved, resulting in positions near the middle of the backbone being consistently further away from the predicted helix axis. The first possibility was discounted, as alpha-helices have a constant rotation and set number of amino-acids over each section of length. These are 100° rotation per amino-acid, 3.6 amino-acids per complete turn, and an extension of 1.5 Å per amino-acid. This means that alpha-helices have a very constrained structure, which results in a consistent width. Therefore, the second possibility was most likely. Visualisations of structural data for talin's rod subdomains show that most

helices have a small amount of curvature along their length (examples shown in Figure 5.1). It is possible that the curvature is due to flexible spring-like behaviour of the alpha-helix.

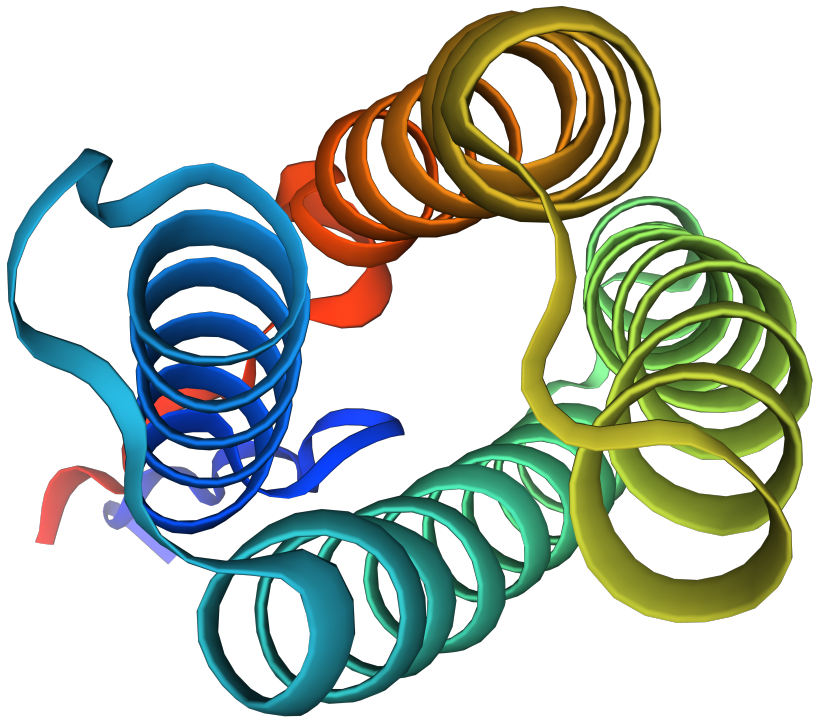


Figure 5.1: Figure generated from PDBID 5ic0.1 using RSCB's NGL viewer. This shows the curvature of the helices in the R8 domain.

Future simulations that need to make use of the model of the helix backbone, such as simulations propagating force through the rod subdomain structure or calculating collisions, might need an improved backbone approximation. Two approaches were considered as possible modifications to the helix backbone model: 1) the helix backbone is split into subsections that are joined and articulated using local coordinate frames, or 2) rather than the current straight line, a curve is fitted through the backbone data and the alpha-helix model is built around the curve as a replacement to the axis. Both options increase the alignment of the backbone approximation to the helix structural data. However, the first option also allows for simple manipulation of the helix subsections during simulations, allowing the helix to be treated as flexible.

5.2 Sidechain Model

The validation of the sidechain approximation (Section 4.1.2) showed that the shape and size of the sidechains were well approximated, with the majority of the sidechain atoms sitting within the mean distance line. The sidechain representation used for the approximation was selected from a number of different candidates. Four models were considered: a single pseudo-atom model (SPA) based on the Levitt-Warshel model [157], a dual pseudo-atom model (DPA) where an additional pseudo-atom was used to represent the position of the sidechain's electrostatic charge, the UNRES model [207], and the CABS model [155]. The different representations of these models are shown in Figure 5.2.

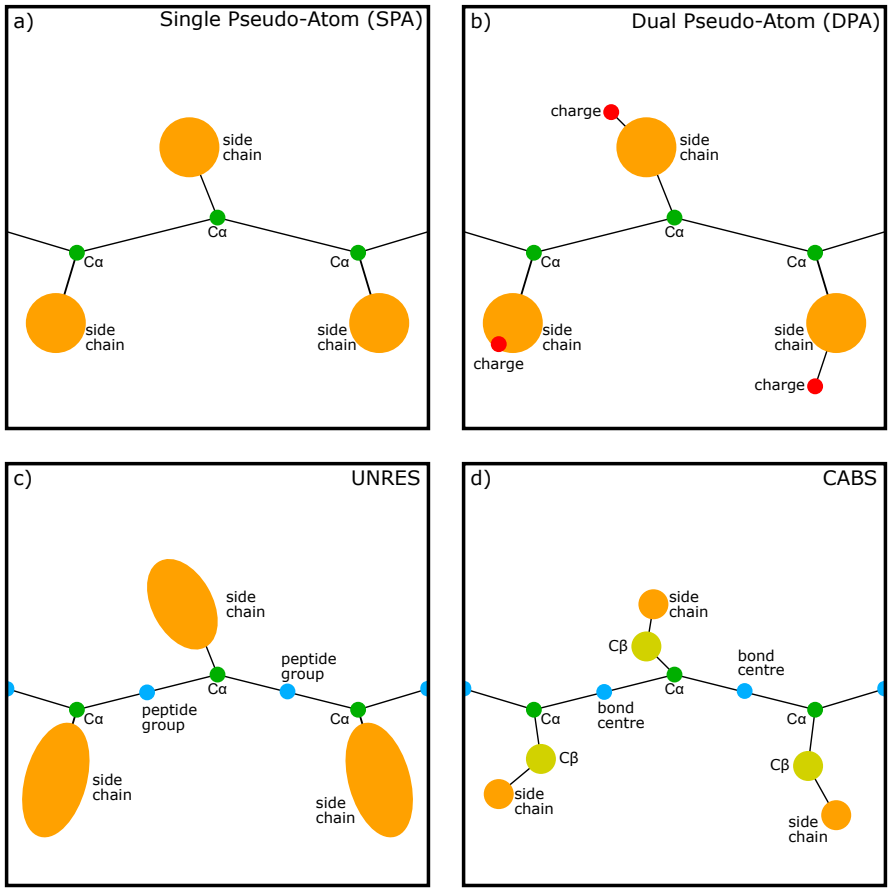


Figure 5.2: Diagrams of different possible sidechain representations. (a) Single Pseudo-Atom model based on the Levitt-Warshel model. This was the model used in the alpha-helix model. (b) Dual Pseudo-Atom model where an additional position that represented the electrostatic charge position was attached to the sidechain. (c) The UNRES model where the sidechains are modelled as a spheroid that encompasses the sidechain atoms. (d) The CABS model where the carbon-beta atom is also included in the representation as a separate pseudo-atom.

The models are similar in that the pseudo-atom the sidechains are connected to is the carbon-alpha atom of the amino-acid residue. Otherwise, each model

represents the sidechains differently. The SPA model replaces all sidechain atoms with a single pseudo-atom which is placed at the centroid of the sidechain atoms. The radius of the pseudo-atoms is defined as the mean distance between the sidechain atoms and the centroid position. All properties, such as electrostatic charge, originate from the centroid position. The DPA model also replaces all sidechain atoms with a pseudo-atom placed at the centroid. However, a second pseudo-atom (bound to the sidechain pseudo-atom) is placed at the position of the electrostatic charge, if the sidechain has one.

The UNRES model, originally developed by Liwo et al. in 1998 [207], adds a pseudo-atom called the peptide group position to the amino-acid residue chain. It also changes the shape of the sidechain pseudo-atom from a sphere to a prolate spheroid. The spheroid is positioned to encompass all sidechain atoms. As most sidechain structures are long chains of atoms, the prolate spheroid better approximates the overall shape of the sidechain as a 3D shape. The CABS model was developed by Kolinski et al. in 2004 [155]. Like the UNRES model, this model adds a pseudo-atom to the amino-acid residue backbone, but it also breaks the sidechain pseudo-atom in two. The first pseudo-atom of the sidechain is positioned at the carbon-beta atom. The second sidechain pseudo-atom is placed at the centre of mass of the remaining sidechain atoms. The long-range interactions are calculated from the pseudo-atom position at the centre of mass.

The SPA model was chosen for two primary reasons: firstly, the increased detail of the other models provided only a small increase in accuracy. The level of precision of positions within the pdb structures was between $\sim$2 Å to $\sim$6 Å. The mean potential change between the SPA and DPA models in the position of the electrostatic charge is $\sim$0.3 Å. The DPA model would have resulted in a more precise sidechain representation since the electrostatic charge position is completely maintained during the approximation. However, the difference in position is very small and falls within the margin of error of the resolution of the pdb structural data. Therefore, making this change would have a negligible impact on the interaction calculations.

Secondly, the increased detail of the other models dramatically increased the complexity of the sidechain representation. UNRES and CABS include additional pseudo-atoms and UNRES also defines the shape of the sidechain. Each additional pseudo-atom increases the degrees of freedom for each sidechain. For models like DPA and CABS, these additional pseudo-atoms also remove rotational symmetry, as illustrated in Figure 5.3. The SPA model has rotational symmetry around the

bond connecting the sidechain pseudo-atom to the backbone carbon-alpha. If the sidechain pseudo-atom is rotated, the structure does not change. However, if the sidechain pseudo-atom in the DPA model is rotated, the position of the charge pseudo-atom moves with it. Because of the motility of the charge pseudo-atom, its current position must be determined each time an interaction calculation occurs. Its position is influenced by all other charged pseudo-atoms in its vicinity. As a result, it would be more computationally expensive to calculate the positions of the charge pseudo-atoms. The CABS model would have similar complexities as the sidechain is at the end of the chain of pseudo-atoms. Therefore, the single pseudo-atom (SPA) model was selected for use and found to be an adequate representation in this research context.

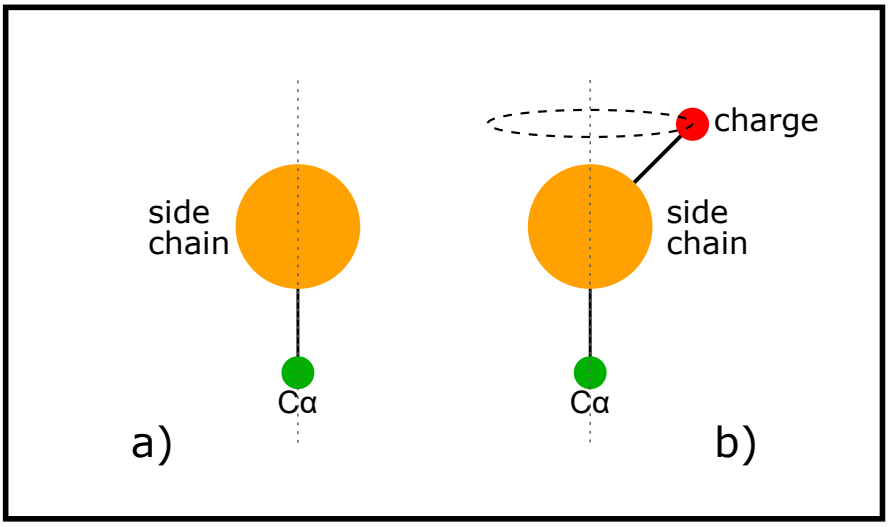


Figure 5.3: Diagram of the used sidechain model (SPA) and the DPA with vertical dashed lines showing the rotational symmetry and a circular dashed line showing the possible positions of the charge pseudo-atom.

5.3 Hydrophobic Interactions

In Section 2.2.2 it was mentioned that the hydrophobic interactions were not included in the simulation. The reasons for this will be discussed in detail here.

First, some concepts and terminology of hydrophobic interactions need to be explained. The hydrophobic interaction is a result of the entropic state of a closed system such as one containing a protein in aqueous solutions [210]. Entropy can be considered as the quantity representing the degree of disorder in a system. The second law of thermodynamics states that the entropy of a closed system always either remains the same or increases over time, but never decreases. This

means that it is energetically unfavourable for molecules to be constrained, as this increases the order of the system. Therefore, a molecular system will tend towards configurations with minimal constraints. Water molecules interact with the exposed sidechains on the surface of a protein. Normally, these interactions happen on the outside of a folded protein, as the water molecules are not able to make their way to the inside of the protein.

If the sidechains on the outside of a protein are non-polar in nature, the water molecules cannot form hydrogen bonds with them. Therefore, the water molecules have to turn towards the remainder of the water lattice in order to form bonds. Non-polar sidechains are therefore described as hydrophobic sidechains. The turning away of the water molecules constrains the degrees of freedom of the water molecules, as they are unable to make bonds in certain directions. This increases the order of the system, which is energetically unfavourable as describe by the 2nd law of thermodynamics [211].

On the other hand, the polar and electrostatically charged sidechains are able to form bonds with the water molecules without restricting the configuration of water molecules in the water lattice. These charged sidechains are described as hydrophilic sidechains. The unrestricted bonding of the water molecules to the hydrophilic sidechains decreases the ordered-ness of the water lattice, which is energetically favourable. Therefore, a protein's conformation tends towards one where the largest number of polar or electrostatically charged sidechains are exposed on the surface of the protein, and the non-polar sidechains facing away from the water molecules [212].

The results of the hydrophobic effect can be measured as forces and torques on the protein structure. However, the hydrophobic interaction itself is not a force, but is the result of the system's energy level tending towards its minimum level. It is often calculated as a resulting difference in the potential energy level of a system with a protein present in solution and without a protein in solution [144]. By having to take the extra step of calculating the hydrophobic effect as potential energy, it makes calculating a resulting force more complex.

In molecular dynamics force-fields, there are two ways of calculating the results of the hydrophobic effect: explicit and implicit solvation. In explicit solvation, each molecule of water, each bond formed between them and the protein's surface are individually simulated. The potential energy of the system is first calculated with the protein present in the solvent, and then with just the solvent. The change in potential energy contributed by the hydrophobic effect is calculated as

the difference in the potential energy between the two calculations.

In implicit solvation, an approximation of the interaction between water molecules and the surface of a protein is used. This is computationally less expensive as water molecules and their binds are not simulated individually. One such approximation is called the Solvent Accessible Surface Area (SASA) calculation [213]. Empirical evidence has shown a very strong proportional correlation between the surface area of a protein contributed by each atom A_i, an experimentally determined solvation parameter of each atom σ_i, and the change in free energy of solvation of the protein ΔG_{solv}.

$$\Delta G_{\text{solv}} = \sum_{\text{atoms } i} \Delta\sigma_i A_i \tag{5.1}$$

The formulation of SASA (shown in Equation 5.1) can then be used to determine the change in potential energy of the system and the effect the hydrophobic interactions have on this.

Both methods for calculating the impact of hydrophobic interactions operate directly within the energy domain. Molecular dynamics simulations can take this energy value and convert it to a force during the Euler integration phase of their calculations. However, there is no existing method for calculating the forces from hydrophobic interactions directly. Therefore, some time was spent on exploring the hydrophobic data, to work towards developing a method for approximating forces resulting from the hydrophobic effect. This was done with the hope of incorporating this method into the simulations described in this thesis. A series of calculations were developed that used the simulation framework to extract hydrophobicity data from each alpha-helix.

First, each sidechain was assigned a value of hydrophobicity. These values were taken from the Kyte and Doolittle scale of hydrophobicity [214] where positive values are hydrophobic and negative values are hydrophilic, ranging between 4.5 and -4.5. As hydrophobic sidechains hide within the folded core of a protein, it was important to determine whether each sidechain was on the surface of the rod subdomain or hidden within its core. Rod subdomains that have four alpha-helices have one core that can hide sidechains from solution. Those that have five alpha-helices have an additional core due to the placement of the fifth helix, as shown in Figure 5.4. Each of these cores must to be considered separately.

In order to determine the distance between the sidechains and these cores, an axis that runs parallel to the helices and through each core was calculated.

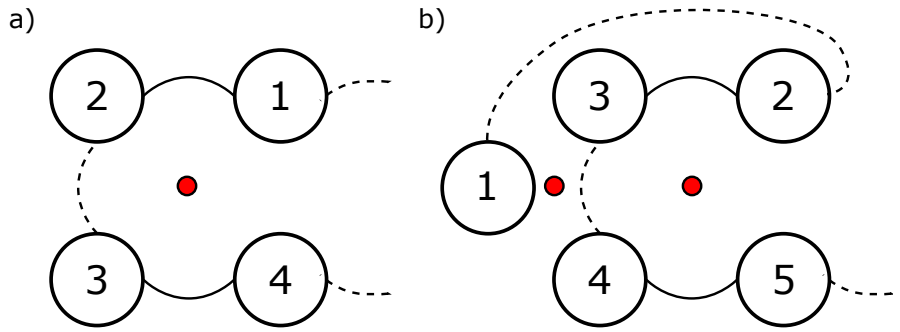


Figure 5.4: Diagram of the two rod subdomain layouts from a top-down perspective. Solid lines indicate the amino-acid chain linkers connected to the top of the helices and the dashed line indicates the linkers connected to the bottom of the helices. (a) The four alpha-helix layout. (b) The five alpha-helix layout. The new helix (helix one) is positioned next to the 3rd and 4th helices. The small red circles indicate the hydrophobic cores for each layout.

For the four helices, the positions of each set of ends of the helices were grouped and the mean position calculated. This resulted one position at each end of the rod subdomain which defined an axis through that core. The same was done for the core created by the three helices (1,3 and 4). For each core axis, the Euclidean distance between each sidechain and the closest point on the core axis was calculated. The core's sidechain distance could then be plotted as a scatter graph of distance vs. hydrophobicity.

As shown in Figures 5.5 and 5.6, the scatter plots were segmented into quadrants, showing the hydrophobic and hydrophilic sidechain within and outside the of the core respectively. To determine which sidechains were inside or outside the core, a distance from the axis to the mid-point between the closest sidechain and the furthest sidechain was chosen. To determine which sidechains were inside or outside the core, the distance from the axis to the mid-point between the closest sidechain and the furthest sidechain was chosen.

In the three hydropathy vs. distance plots, the upper left and lower right quadrants contained the majority of the sidechains. This showed that, as expected, the hydrophobic sidechains were positioned primarily in the core of the protein and the hydrophilic sidechains were exposed on the surface of the protein(this is similar in layout to lipid membrane structures). These results align with the expected hydrophobic configurations of the sidechains.

In order to make comparisons between the hydropathy data of the rod subdomains, the data discussed above was used to create comparative ratios between the quadrants. For example, the number of hydrophobic sidechains inside the core was compared to the number of hydrophilic sidechains in the core. This was

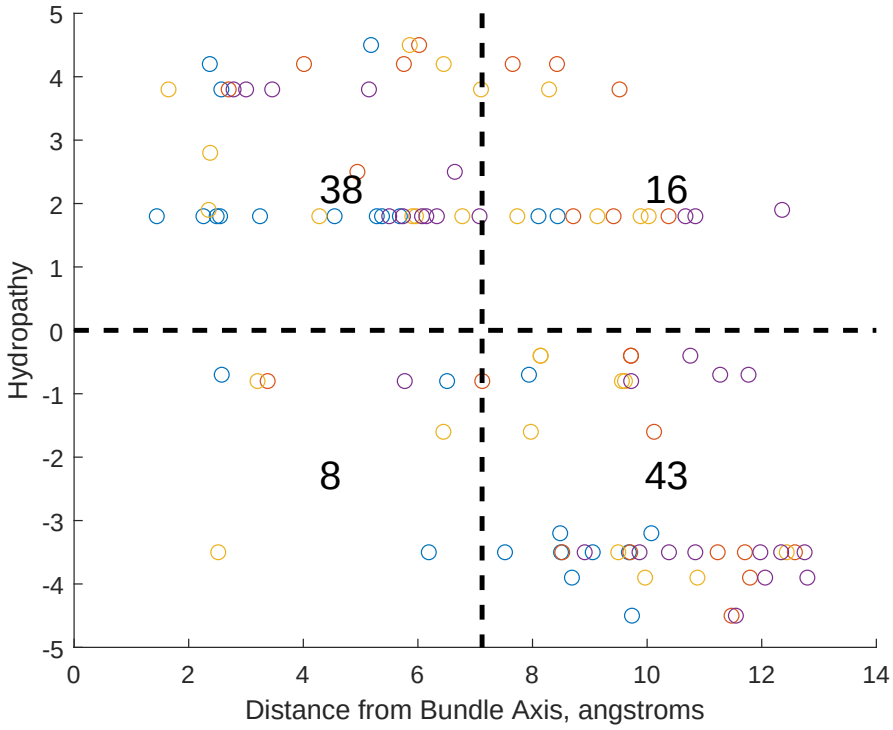
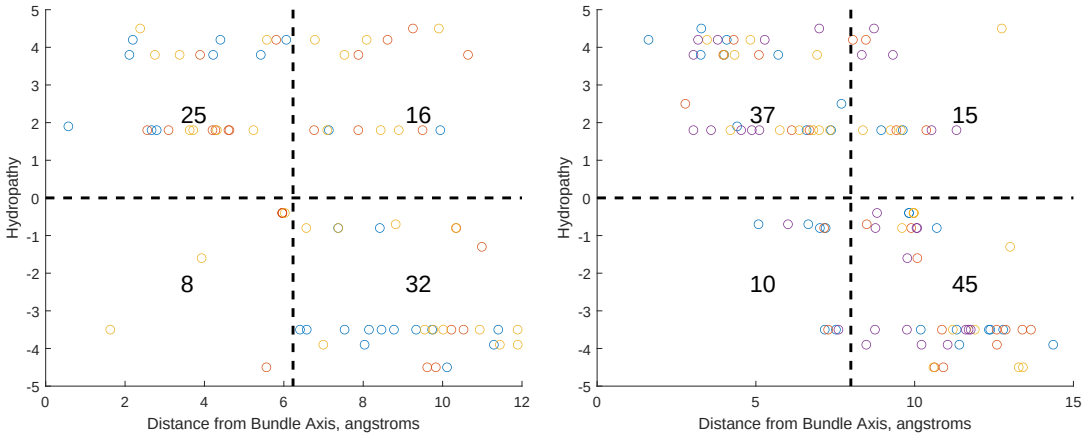


Figure 5.5: Subdomain R4: Scatter plot of sidechain hydropathy vs distance from rod subdomain core. Positive values of hydropathy indicate hydrophobic sidechains. The upper left quadrant contains the hydrophobic sidechains within the core, the upper right quadrant contains the hydrophobic sidechains outside the core. Likewise, the bottom left quadrant and the lower right quadrant contain the hydrophilic sidechains inside and outside the core respectively. Kyte and Doolittle hydropathy scale.



(a) (b)

Figure 5.6: Subdomain R5: Scatter plots of sidechain hydropathy vs distance from rod subdomain cores. (a) Shows the values from the core comprising the three helices (1st, 3 and 4). (b) Shows the value from the core comprising four helices (2, 3, 4 and 5). Positive values of hydropathy indicate hydrophobic sidechains. Kyte and Doolittle hydropathy scale.

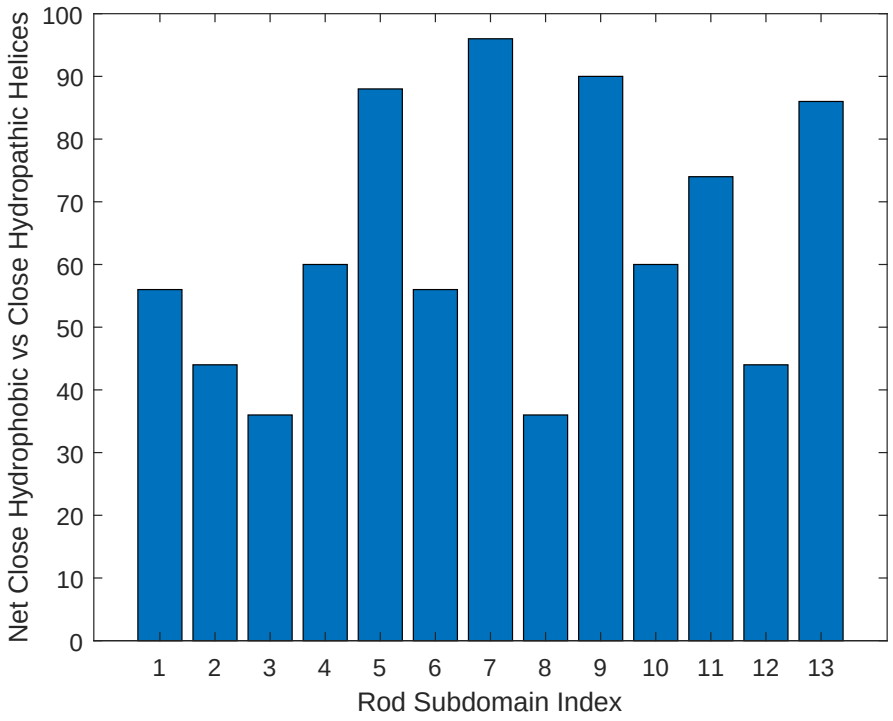


Figure 5.7: Plotted data of the number of close hydrophobic sidechains vs. the number of close hydropathic sidechains. Each subdomains number of inner hydropathic sidechains were subtracted from the number of inner hydrophobic sidechains.

done by subtracting the number of hydrophilic core sidechains from the number of hydrophobic resulting in the net number of sidechains trying to keep themselves hidden. These results are shown in Figure 5.7. However, there were no consistent, strong correlations. The graph shows the R3 and R8 subdomains as having the lowest values, indicating a lower hydrophobic tendency. This lines up with the low unfolding force thresholds of R3 and R8 ($\sim$5 pN). However, the R7 domain shows the second-highest value, indicating it has a very high hydrophobic tendency. While the R7 does have a higher unfolding force than the R3 and R8 subdomain, it does not have the strongest. Approximately five subdomains unfold at higher force thresholds than the R7 subdomain. Additionally, when compared to the estimated unfolding force data from Table 4.6, the data in Figure 5.7 does not align with the unfolding force values. Given the scope of this research project and the amount of time that would be required to draw meaningful conclusions from this type of approach, it was decided to leave this direction of inquiry for future work.

This brief excursion into mapping the hydrophobic effect has shown that much more work could be done. Calculating hydrophobic interactions as forces could be a very useful topic for future work. Being able to directly calculate the hydropho-

bic interaction as forces rather than resorting to molecular dynamics simulations would dramatically decrease the computation time for such values and allow for investigations into rod subdomains that are highly influenced by the hydrophobic effect.

5.4 Coordinate Frame Advantages

Coordinate frames are a core tool used in the creation of the simulation objects and the simulation framework. They allow for the definition of frames of reference that spatial coordinates can be defined in. At the same time, they act as an established mathematical transform that can be used to transition coordinates from one frame to another.

Other methods exist to achieve similar functionality such as scene graphs from the field of computer graphics [215]. While scene graphs do not necessarily define spatial relationships between objects, modern 3D computer graphics often integrate this functionality, for example with hardware accelerated ray tracing.

Coordinate frames have been used in the field of robotics for several decades, and a number of complimentary works and tools that can be used alongside them have been developed. One of these mathematical tools is statics. Statics calculates the propagation of forces and moments from one coordinate frame to the next [205]. Future simulations could use statics by building directly on top of what was designed and implemented here. This existing work may enable force propagation to be implemented into the simulation framework. This would be a component of a fully dynamic, time-dependent simulation of the alpha-helices of the talin rod domain. An example of such a simulation could be the stretching of a rod subdomain using constant, pulling velocity and tracking the domains unfolding as the pulling force is propagated through the alpha-helices. This may help highlight properties, such as intermediate unfolding states. This knowledge would improve our understanding of general protein folding and unfolding. Coordinate frame and statics research may be transferable to other structural protein fields.

5.5 Future Work

Improvements to the alpha-helix model were discussed, and the backbone approximation was shown to have promising options. Specifically, the segmentation of the alpha-helix backbone into separate, flexible sections would enable the curve of backbone and alpha-helix to be more accurately approximated. While this

would not improve the static force results presented in Section 4.2, it may enhance dynamic simulations by allowing for the modelling of a helix's resistance to tensile stress and its elastic behaviour. This could improve the accuracy of force and torque propagations, as the correct amount of force could be converted into elastic potential energy stored in bending helices.

Implementing statics calculations into the coordinate frame system of the simulation framework would allow for dynamic simulations to be built. These simulations would be able to explore the changes in the rod subdomain conformations over time as the domain unfolds. They could also be able to explore the unfolding pathways of the rod subdomain. This information may be used to analyse how different components in the alpha-helices and rod subdomains impact the unfolding forces and the trajectories of the unfolding alpha-helices.

During the initial investigation into talin in Section 2.1, talin was shown to have a highly complex signalling response. In that section, a categorisation was shown of talin's signalling response as a series of states and transitions between those states. The states and transitions were also visualised using a state diagram (Figure 2.3) that is commonly used to describe finite state machines (FSMs). FSMs are adept at computing a series of pre-defined operations that move between known states, depending on a set of expected inputs. They could be used to compute the changes in state of a rod subdomain as it reacts to forces, and therefore predict signalling responses. Therefore, in future work, the FSM representation may offer a computationally light method to simulate the individual subdomains of talin in order to observe its signalling responses. A simulation such as this would not only be computationally lightweight, but simple and efficient, making it appropriate for large scale simulations of focal adhesions involving large numbers of talin interacting with other signalling pathways.

Progress on developing a direct force approximation for the hydrophobic effect was covered in Section 5.3. Additional work is required to complete the development of this approximation. By developing a force calculation for the hydrophobic effect, two of the strongest intermolecular interactions would be used in the future simulations. This may increase the accuracy of the unfolding force results and would also allow for investigations into rod subdomains that have been shown to be highly influenced by the hydrophobic effect, such as the R3 domain [8].

Finally, there are additional applications for the simulation framework in its current form. Proteins such as vinculin and talin-2 contain alpha-helix coiled-coil domains and are therefore likely to have mechanosensitive properties. These

proteins could be examined to determine the unfolding force thresholds of these coiled-coil domains and therefore any possible signalling responses. As vinculin interacts with focal adhesions to strengthen and stabilise them, it is important to understand vinculin's full impact on the stabilisation and maturation of focal adhesions.

Other factors that impact focal adhesion formation could be explored, such as specific helix to helix interaction. Rap1/RIAM activates talin and localises it to the cell membrane in order to bind and activate integrins. The slip bond between talin and rap1 is the result of a single helix from rap1 joining the R2 or R3 subdomain [126]. R2 and R3 are subdomains with four helices and the addition of the single helix from rap1 creates a hybrid five helix complex. The simulation could be used to calculate the interaction forces between the Rap1 helix and the remainder of the other helices in order to determine the debonding force. This value may be used to estimate the conditions under which talin is localised to the cell membrane, and the debonding force values impact on early focal adhesion formation.

While an alpha-helix model was created in this research, it will be possible to create a beta-sheet model for use within the current simulation. By creating other models of secondary protein structures, more protein domains could be explored using the same simulation methodology. Scaling this up to models of whole domains, protein to protein surface binding interactions forces may also be explored.

Some mechanosensitive proteins do not have available pdb structures like talin, and therefore can not be investigated using simulations. Alpha Fold is an AI system developed by Deep Mind (Alphabet Inc.) using machine learning techniques that predict protein structures from amino-acid residue sequence data [216]. Alpha Fold could be used to generate protein structural data for these proteins which the simulation can then use to determine their unfolding force thresholds. Similarly, post-translational mutations could be explored using a combination of Alpha Fold and this simulation. By exploring these mutations, possible candidates for useful modifications that can alter subdomain unfolding force thresholds may be determined.

6 Conclusion

Talin's signalling response to force is determined by the unfolding of its rod subdomains. As talin is a core structural and signalling protein in platelet focal adhesions, it determines a platelet's response to both external and internal forces. If the force threshold at which each of talin's rod subdomains unfolds and signals can be determined, talin's signalling impact on platelets during haemostasis may be understood and predicted. Therefore, the aim of this thesis was to answer the research question: can the calculated magnitude and range of force values for the unfolding thresholds of talin's rod domain be improved upon by making simulations that operate directly within the force domain and are designed for the spatial and temporal ranges that talin operates in? The research question was answered by creating a novel model of talin's rod domain alpha-helices. A new simulation framework was developed in which these models were subsequently used to calculate intermolecular electrostatic interactions. This approach was chosen as other computational methods such as molecular dynamics simulations produce erroneously large unfolding force values of up to 1500 pN (The physiologically relevant range of forces for talin's rod domain is ∼5 pN to ∼25 pN).

The unique features of this method were: the new coarse-grain-like alpha-helix models derived from protein structural data, the use of a coordinate frame system to manage and manipulate the objects within the simulation, and the direct calculation of interaction forces. Additionally, it was noticed that the protein structural data was not in its most energetically stable conformation. Therefore, the optimisation algorithm *fmincon* was used to improve the data's conformation to maximise electrostatic interaction forces while constraining the maximum translation and rotation of the helices.

Through this method, the following results were obtained: a range of unfolding forces for talin's rod subdomains between ∼1 pN to ∼49 pN. As stated in Section 4.3.2, this range of unfolding forces covers ∼ 240% of the physiologically relevant range for talin's rod domain as based on experimental data. Previous results from molecular dynamics simulations covered ∼ 7500% of the range. Therefore, the results presented in this thesis show a greater than 30 times improvement in matching the physiologically relevant range of unfolding forces for talin's rod domain compared to previous computational methods. These results prove that indeed an improvement was made to the range of unfolding force values by using a force-based simulation framework.

Points of improvement were also identified. These are: updating the alpha-helix model so that its approximated backbone accurately follows the curvature of alpha-helices, evaluating additional intermolecular interactions for the simulation framework and developing force-based approximations such as for hydrophobic or van der Waal interactions, and finally, implementing dynamics simulations so that time-dependent unfolding simulations can be designed. By working on these in the future, the results may be improved so that the range of unfolding forces matches even more closely the range of experimental forces. Thus, research questions like "What is each rod subdomain's unfolding force threshold?" or "What motions do the alpha-helices in a subdomain go through during an unfolding event?" may be answered.

Most cells or proteins exist in a high force environment, meaning that they must be able to generate, withstand or transmit high levels of forces. The signalling response of talin, for example, is the product of its rod domains unfolding and either ejecting bound ligands or allowing new ligands to bind. Modelling force effects at a molecular level is therefore helpful in understanding when talin signals other platelet processes, such as focal adhesion maturation and platelet contraction. This, in turn, may impact the understanding of diseases related to platelet binding and aggregation, as well as thrombus formation, contraction, and lysis. In the future, it may be possible to alter thrombogenicity by altering the forces that determine platelet adhesion strength.

This research has shown that there are other viable alternative molecular simulation methodologies than molecular dynamics. Areas of interest that molecular dynamics was not designed for, such as the calculation of inter-domain forces, can benefit from these types of novel simulation approaches. Importantly, this work shows that the improvements in results are significant enough to warrant the effort in developing these new simulation approaches in the first place. Presumably, this is the case not only for talin but for other molecules as well. For talin specifically, this research has shown that the PDB data may require some evaluation as more optimal configurations appear to be available, which would in turn increase the accuracy of research such as on ligand docking.

In summary, the research question was indeed answered: the calculated magnitude and range of force values for the unfolding thresholds of talin's rod domain can be significantly improved upon by making simulations that operate directly within the force domain and are designed for the spatial and temporal ranges that talin operates in.

References

[1] A. P. Bye, A. J. Unsworth, and J. M. Gibbins, "Platelet signalling: A complex interplay between inhibitory and activatory networks," *Journal of Thrombosis and Haemostasis*, vol. 14, no. 5, pp. 918–930, 2016.

[2] C. I. Jones, N. E. Barrett, L. A. Moraes, J. M. Gibbins, and D. E. Jackson, "Endogenous inhibitory mechanisms and the regulation of platelet function," in *Platelets and Megakaryocytes: Volume 3, Additional Protocols and Perspectives*, J. M. Gibbins and M. P. Mahaut-Smith, Eds. New York, NY: Springer New York, 2012, pp. 341–366.

[3] R. Rahikainen, T. Ohman, P. Turkki, M. Varjosalo, and V. P. Hytönen, "Talin-mediated force transmission and talin rod domain unfolding independently regulate adhesion signalling," *Journal of Cell Sciences*, vol. 132, jcs226514, 2019.

[4] D. R. Critchley and A. R. Gingras, "Talin at a glance," *Journal of Cell Science*, vol. 121, pp. 1345–1347, 2008.

[5] J. D. Humphries, P. Wang, C. Streuli, B. Geiger, M. J. Humphries, and C. Ballestrem, "Vinculin controls focal adhesion formation by direct interactions with talin and actin," *Journal of Cell Biology*, vol. 179, pp. 1043–1057, Dec. 3, 2007.

[6] V. P. Hytönen and V. Vogel, "How force might activate talin's vinculin binding sites: Smd reveals a structural mechanism," *PLOS Computational Biology*, vol. 4, no. 2, e24, Feb. 15, 2008.

[7] B. T. Goult, J. Yan, and M. A. Schwatz, "Talin as a mechanosensitive signalling hub," *Journal of Cell Biology*, vol. 217, no. 11, pp. 3776–3784, Sep. 25, 2018.

[8] M. Yao, B. T. Goult, H. Chen, P. Cong, and J. Sheetz M. P. Yan, "Mechanical activation of vinculin binding to talin locks talin in an unfolded conformation," *Scientific Reports*, vol. 4, Apr. 9, 2014.

[9] A. W. M. Haining, M. von Essen, S. J. Attwood, V. P. Hytönen, and A. del Rio Hernandez, "All subdomains of the talin rod are mechanically vulnerable and may contribute to cellular mechanosensing," *ACS Nano*, vol. 10, pp. 6648–6658, Jul. 5, 2016.

[10] M. E. Carr and S. L. Zekert, "Measurement of platelet-mediated force development during plasma clot formation," *The American Journal of the Medical Science*, vol. 302, no. 1, pp. 13–18, Jul. 1991.

[11] D. R. Myyers, Y. Qiu, M. E. Fay, M. Tennenbaum, D. Chester, J. Cuadrado, Y. Sakurai, J. Baek, R. Tran, J. Ciciliano, B. Ahn, R. Mannino, S. Bunting, C. Bennett, M. Briones, A. Fernadnez-Nieves, M. L. Smith, A. C. Brown, T. Sulchek, and W. A. Lam, "Single-platelet nanomechanics measured by high-throughput cytometry," *Nature Materials*, vol. 16, no. 2, pp. 230–235, Apr. 10, 2017.

[12] M. Yao, B. T. Goult, B. Klapholz, X. Hu, C. P. Toseland, Y. Guo, P. Cong, M. P. Sheetz, and J. Yan, "The mechanical response of talin," *Nature Communications*, vol. 7, Jul. 7, 2016.

[13] S. Sorrentino, J.-D. Studt, O. Medalia, and K. Tanuj Sapra, "Roll, adhere, spread and contract: Structural mechanics of platelet function," *European Journal of Cell Biology*, vol. 94, no. 3, pp. 129–138, Mar. 2015.

[14] C. L. Wagner, M. A. Mascelli, D. S. Neblock, H. F. Weisman, B. S. Coller, and R. E. Jordan, "Analysis of gpiib/iiia receptor number by quantification of 7e3 binding to human platelets," *Blood*, vol. 88, no. 3, pp. 907–914, Aug. 1, 1996.

[15] J. M. Burkhart, M. Vaudel, S. Gambaryan, S. Radau, U. Walter, L. Martens, J. Geiger, A. Sickmann, and R. P. Zahedi, "The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways," *Blood*, vol. 120, no. 15, e73–e82, Oct. 11, 2012.

[16] J. S. Bennett, "Platelet-fibrinogen interactions," *Annals of the New York Academy of Sciences*, vol. 936, no. 1, pp. 340–354, Jan. 25, 2005.

[17] C. T. Esmon, "The regulation of natural anticoagulant pathways," *Science*, vol. 235, no. 4794, pp. 1348–1352, Mar. 12, 1987.

[18] A. Tripodi, C. Legnani, V. Chantarangkul, B. Cosmi, G. Palareti, and P. M. Mannucci, "High thrombin generation measured in the presence of thrombomodulin is associated with an increased risk of recurrent venous thromboembolism," *Journal of Thrombosis and Haemostasis*, vol. 6, no. 8, pp. 1327–1333, Jul. 17, 2008.

[19] J. Mehta and P. Mehta, "Platelet function in hypertension and effect of therapy," *American Journal of Cardiology*, vol. 47, no. 2, pp. 331–334, Feb. 1981.

[20] A. Para, D. Bark, A. Lin, and D. Ku, "Rapid platelet accumulation leading to thrombotic occlusion," *Annals of Biomedical Engineering*, vol. 39, pp. 1961–1971, Mar. 22, 2011.

[21] H. Heijboer, D. P. Brandjes, H. R. Büller, A. Sturk, and J. W. ten Cate, "Deficiencies of coagulation-inhibiting and fibrinolytic proteins in outpatients with deep-vein thrombosis," *New England Journal of Medicine*, vol. 323, no. 22, pp. 1512–1516, Nov. 29, 1990.

[22] A. T. Nurden and P. Nurden, "Congenital platelet disorders and understanding of platelet function," *British Journal of Haematology*, vol. 165, no. 2, pp. 165–178, Nov. 29, 2013.

[23] M. J. Davies and A. Thomas, "Thrombosis and acute coronary-artery lesions in sudden cardiac ischemic death," *New England Journal of Medicine*, vol. 310, no. 18, pp. 1137–1140, May 3, 1984.

[24] H. K. Nieuwenhuis, K. S. Sakariassen, W. P. M. Houdijk, P. F. E. M. Nievelstein, and J. J. Sixma, "Deficiency of platelet membrane glycoprotein ia associated with a decreased platelet adhesion to subendothelium: A defect in platelet spreading," *Blood*, vol. 68, no. 3, pp. 692–695, Sep. 1986.

[25] K. Kato, T. Kanaji, S. Russel, T. J. Kunicki, K. Furihata, S. Kanaji, P. Marchese, A. Reininger, Z. M. Ruggeri, and J. Ware, "The contribution of glycoprotein vi to stable platelet adhesion and thrombus formation illustrated by targeted gene deletion," *Blood*, vol. 102, no. 5, pp. 1701–1707, 2003.

[26] M. Cattaneo, "Inherited platelet-based bleeding disorders," *Journal of Thrombosis and Haemostasis*, vol. 1, no. 7, pp. 1628–1636, Jun. 20, 2003.

[27] G. E. Raskob, P. Angchaisuksiri, A. N. Blanco, H. Buller, A. Gallus, B. J. Hunt, E. M. Hylek, A. Kakkar, S. V. Konstantinides, M. McCumber, Y. Ozaki, A. Wendelboe, and J. I. Weitz, "Thrombosis: A major contributor to global disease burden," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 34, no. 11, pp. 2363–2371, Nov. 2014.

[28] D. Woufle, J. Yang, and L. Brass, "Adp and platelets: The end of the beginning," *The Journal of Clinical Investigation*, vol. 107, no. 12, pp. 1503–1505, Jun. 15, 2001.

[29] S.-H. Yun, E.-H. Sim, R.-Y. Goh, J.-I. Park, and J.-Y. Han, "Platelet activation: The mechanisms and potential biomarkers," *BioMed Research International*, vol. 2016, p. 9 060 143, Jun. 15, 2016.

[30] M. Hamilos, S. Petousis, and F. Parthenakis, "Interaction between platelets and endothelium: From pathophysiology to new therapeutic options," *Cardiovascular Diagnosis and Therapy*, vol. 8, no. 5, pp. 568–580, Oct. 2018.

[31] J. Sai, D. Raman, Y. Liu, J. Wikswo, and A. Richmond, "Parallel phosphatidylinositol 3-kinase (pi3k)-dependent and src-dependent pathways lead to cxcl8-mediated rac2 activation and chemotaxis*," *Mechanisms of Signal Transduction*, vol. 283, no. 39, pp. 26 538–26 547, Sep. 2008.

[32] S. P. Watson, J. M. Auger, O. J. T. McCarty, and A. C. Pearce, "Gpvi and integrin alphaiib beta3 signalling in platelets," *Journal of Thrombosis and Haemostasis*, vol. 3, no. 8, pp. 1752–1762, Aug. 8, 2005.

[33] K. Gilio, I. C. A. Munnix, P. Mangin, J. M. E. M. Cosemans, M. A. H. Feijge, P. E. J. van der Meijden, S. Olieslagers, M. B. Chrzanowska-Wodnicka, L. Rivka, S. Schoenwaelder, S. Koyasu, S. O. Sage, S. P. Jackson, and J. W. M. Heemskerk, "Non-redundant roles of phosphoinositide 3-kinase isoforms α and β in glycoprotein vi-induced platelet signalling and thrombus formation," *Journal of Biological Chemistry*, vol. 284, no. 49, pp. 33 750–33 762, Oct. 8, 2009.

[34] D. Varga-Szabo, A. Braun, and B. Nieswandt, "Calcium signaling in platelets," *Journal of Thrombosis and Haemostasis*, vol. 7, no. 7, pp. 1057–1066, Jun. 26, 2009.

[35] J. R. Crittenden, W. Bergmeier, Y. Zhang, C. L. Piffath, Y. Liang, D. D. Wagner, D. E. Housman, and A. M. Graybiel, "Caldag-gefi integrates signalling for platelet aggregation and thrombus formation," *Nature Medicine*, vol. 10, no. 9, pp. 982–286, Aug. 29, 2004.

[36] F. Lagarrigue, C. Kim, and M. H. Ginsberg, "The rap1-riam-talin axis of integrin activation and blood cell function," *Blood*, vol. 128, no. 4, pp. 479–487, May 20, 2016.

[37] S. Stritt, K. Wolf, V. Lorenz, T. Vögtle, S. Gupta, M. R. Bösl, and Nieswandt, "Rap1-gtp–interacting adaptor molecule (riam) is dispensable for platelet integrin activation and function in mice," *Blood*, vol. 125, no. 2, pp. 219–222, Jan. 8, 2015.

[38] Z. Sun, M. Costell, and R. Fässler, "Integrin activation by talin, kindlin and mechanical forces," *Nature Cell Biology*, vol. 21, no. 1, pp. 25–31, Jan. 2, 2019.

[39] B. S. Coller and S. J. Shattil, "The GPIIb/IIIa (integrin IIb 3) odyssey: A technology-driven saga of a receptor with twists, turns, and even a bend," *Blood*, vol. 112, no. 8, pp. 3011–3025, Oct. 15, 2008.

[40] H. R. Gralnick, S. B. Williams, and B. S. Coller, "Fibrinogen competes with von willebrand factor for binding to the glycoprotein IIb/IIIa complex when platelets are stimulated with thrombin," *Blood*, vol. 64, no. 4, pp. 797–800, Oct. 1984.

[41] C. Buensuceso, M. de Virgilio, and S. J. Shattil, "Detection of integrin $\alpha_{IIb}\beta_3$ clustering in living cells," *Journal of Biological Chemistry*, vol. 278, no. 17, pp. 15 217–15 224, Apr. 25, 2003.

[42] H. Gong, B. Shen, P. Flevaris, C. Chow, S. C.-T. Lam, T. A. Voyno-Yasenetskaya, T. Kozasa, and X. Du, "G protein subunit gα13 binds to integrin αiibβ3 and mediates integrin "outside-in" signalling," *Science*, vol. 327, no. 5963, pp. 340–343, Jan. 15, 2010.

[43] W. Bergmeier and L. Stefanini, "Novel molecules in calcium signaling in platelets," *Journal of Thrombosis and Haemostasis*, vol. 7, no. 1, pp. 187–190, Jul. 13, 2009.

[44] Q. Ren, S. Ye, and S. W. Whiteheart, "The platelet release reaction: Just when you thought platelet secretion was simple," *Current Opinion in Hematology*, vol. 15, no. 5, pp. 537–541, Sep. 2008.

[45] G. L. Reed, M. L. Fitzgerald, and J. Polgár, "Molecular mechanisms of platelet exocytosis: Insights into the "secrete" life of thrombocytes," *Blood*, vol. 96, no. 10, pp. 3334–3342, Nov. 15, 2000.

[46] S. Offermanns, "Activation of platelet function through g protein-coupled receptors," *Circulation Research*, vol. 99, no. 12, pp. 1293–1304, Dec. 8, 2006.

[47] P. Ohlmann, K. L. Laugwitz, B. Nürnberg, K. Spicher, G. Schultz, J. P. Cazenave, and C. Gachet, "The human platelet adp receptor activates gi2 proteins," *Biochemical Journal*, vol. Pt 3, no. Pt 3, pp. 775–779, Dec. 5, 1995.

[48] S. Kim, C. Foster, A. Lecchi, T. M. Quiton, D. M. Prossor, J. Jin, M. Cattaneo, and S. P. Kunapuli, "Protease-activated receptors 1 and 4 do not stimulate g(i) signalling pathways in the absence of secreted adp and cause human platelet aggregation independently of g(i) signalling," *Blood*, vol. 99, no. 10, pp. 3629–3636, May 15, 2002.

[49] I. Knezevic, C. Borg, and G. C. Le Breton, "Identification of gq as one of the g-proteins which copurify with human platelet thromboxane a2/prostaglandin h2 receptors," *Journal of Biological Chemistry*, vol. 268, no. 34, pp. 26 011–26 017, Dec. 5, 1993.

[50] Y. Djellas, J. M. Manganello, K. Antonakis, and G. C. Le Breton, "Identification of galpha13 as one of the g-proteins that couple to human platelet thromboxane a2 receptors," *Journal of Biological Chemistry*, vol. 274, no. 20, pp. 14 325–14 330, May 14, 1999.

[51] M. J. Cho, I. Pestina Tamara, S. A. Steward, A. Lowell Clifford, C. W. Jackson, and T. K. Gartner, "Role of the src family kinase lyn in txa2 production, adenosine diphosphate secretion, akt phosphorylation, and irreversible aggregation in platelets stimulated with gamma-thrombin," *Blood*, vol. 99, no. 7, pp. 2442–2447, Apr. 1, 2002.

[52] Z. Li, G. Zhang, J. Liu, A. Stojanovic, C. Ruan, C. A. Lowell, and X. Du, "An important role of the src family kinase lyn in stimulating platelet granule secretion," *Journal of Biological Chemistry*, vol. 285, no. 17, pp. 12 559–12 570, Apr. 23, 2010.

[53] Z. Li, G. Zhang, G. C. Le Breton, X. Gao, A. B. Malik, and X. Du, "Two waves of platelet secretion induced by thromboxane a2 receptor and a critical role for phosphoinositide 3-kinases," *Journal of Biological Chemistry*, vol. 278, no. 33, pp. 30 725–30 731, Aug. 15, 2003.

[54] T. J. Kovacsovics, C. Bachelot, A. Toker, C. J. Vlahos, B. Duckworth, L. C. Cantley, and J. H. Hartwig, "Phosphoinositide 3-kinase inhibition spares actin assembly in activating platelets but reverses platelet aggregation,"

Journal of Biological Chemistry, vol. 270, no. 19, pp. 11 358–11 366, May 12, 1995.

[55] D. Woufle, H. Jiang, A. Morgans, R. Monks, M. Birnbaum, and L. F. Brass, "Defects in secretion, aggregation, and thrombus formation in platelets from mice lacking akt2," *Journal of Clinical Investigations*, vol. 113, no. 3, pp. 441–450, Feb. 2004.

[56] J. Chen, S. De, D. S. Damron, W. S. Chen, N. Hay, and T. V. Byzova, "Impaired platelet responses to thrombin and collagen in akt-1-deficient mice," *Blood*, vol. 104, no. 6, pp. 1703–1710, Sep. 15, 2004.

[57] A. Stojanovic, J. A. Marjanovic, V. M. Brovkovych, X. Peng, N. Hay, R. A. Skidgel, and X. Du, "A phosphoinositide 3-kinase-akt-nitric oxide-cgmp signalling pathway in stimulating platelet secretion and aggregation," *Journal of Biological Chemistry*, vol. 281, no. 24, pp. 16 333–16 339, Jun. 16, 2006.

[58] Z. Li, G. Zhang, J. A. Marjanovic, C. Ruan, and X. Du, "A platelet secretion pathway mediated by cgmp-dependent protein kinase," *Journal of Biological Chemistry*, vol. 279, no. 41, pp. 42 469–42 475, Oct. 8, 2004.

[59] V. Randriamboavonjy, J. Schrader, R. Busse, and I. Fleming, "Insulin induces the release of vasodilator compounds from platelets by a nitric oxide-g kinase-vamp-3-dependent pathway," *Journal of Experimental Medicine*, vol. 199, no. 3, pp. 347–356, Feb. 2, 2004.

[60] T. Kozasa, X. Jiang, M. J. Hart, P. M. Sternweis, W. D. Singer, A. G. Gilman, G. Bollag, and P. C. Sternweis, "P115 rhogef, a gtpase activating protein for galpha12 and galpha13," *Science*, vol. 280, no. 5372, pp. 2109–2111, Jun. 26, 1998.

[61] B. Klages, U. Brandt, M. I. Simon, G. Schultz, and S. Offermanns, "Activation of g12/g13 results in shape change and rho/rho-kinase-mediated myosin light chain phosphorylation in mouse platelets," *Journal of Cell Biology*, vol. 144, no. 4, pp. 745–754, Feb. 22, 1999.

[62] S. Offermanns, C. F. Toombs, Y. H. Hu, and M. I. Simon, "Defective platelet activation in g alpha(q)-deficient mice," *Nature*, vol. 389, no. 6647, pp. 183–186, Sep. 11, 1997.

[63] S. Murugappan, F. Tuluc, R. T. Dorsam, H. Shankar, and S. P. Kunapuli, "Differential role of protein kinase c delta isoform in agonist-induced dense

granule secretion in human platelets," *Journal of Biological Chemistry*, vol. 279, no. 4, pp. 2360–2367, Jan. 23, 2004.

[64] O. Konopatskaya, K. Gilio, M. T. Harper, Y. Zhoa, J. M. E. M. Cosemans, Z. A. Karmin, S. W. Whiteheart, J. D. Molkentin, P. Verkade, S. P. Watson, J. W. M. Heemskerk, and A. W. Poole, "Pkcalpha regulates platelet granule secretion and thrombus formation in mice," *Journal of Clinical Investigation*, vol. 119, no. 2, pp. 399–407, Feb. 2009.

[65] J. Boyles, J. E. Fox, D. R. Phillips, and P. E. Stenberg, "Organization of the cytoskeleton in resting, discoid platelets: Preservation of actin filaments by a modified fixation that prevents osmium damage," *Journal of Cell Biology*, vol. 101, no. 4, pp. 1463–1472, Oct. 1985.

[66] J. E. Fox, D. E. Goll, C. C. Reynolds, and D. R. Phillips, "Identification of two proteins (actin-binding protein and p235) that are hydrolyzed by endogenous ca2+-dependent protease during platelet aggregation," *Journal of Biological Chemistry*, vol. 260, no. 2, pp. 1060–1066, Jan. 25, 1985.

[67] J. H. Hartwig and M. DeSisto, "The cytoskeleton of the resting human blood platelet: Structure of the membrane skeleton and its attachment to actin filaments," *Journal of Cell Biology*, vol. 112, no. 3, pp. 407–425, Feb. 1991.

[68] J. G. White, "The substructure of human platelet microtubules," *Blood*, vol. 32, no. 4, pp. 638–648, Oct. 1, 1968.

[69] J. G. White and G. H. Rao, "Microtubule coils versus the surface membrane cytoskeleton in maintenance and restoration of platelet discoid shape," *The American Journal of Pathology*, vol. 152, no. 2, pp. 597–609, Feb. 1998.

[70] R. D. Allen, L. R. Zacharski, S. T. Widirstky, R. Rosenstein, L. M. Zaitlin, and D. R. Burgess, "Transformation and motility of human platelets: Details of the shape change and release reaction observed by optical and electron microscopy," *Journal of Cell Biology*, vol. 83, no. 1, pp. 126–142, Oct. 1, 1979.

[71] T. D. Pollard and G. G. Borisy, "Cellular motility driven by assembly and disassembly of actin filaments," *Cell*, vol. 112, no. 4, pp. 453–465, Feb. 21, 2003.

[72] B. Diagouraga, A. Grichine, A. Fertin, J. Wang, S. Khochbin, and K. Sadoul, "Motor-driven marginal band coiling promotes cell shape change

during platelet activation," *Journal of Cell Biology*, vol. 204, no. 2, pp. 177–185, Jan. 20, 2014.

[73] E.-K. Shin, H. Park, J.-Y. Noh, K.-M. Lim, and J.-H. Chung, "Platelet shape changes and cytoskeleton dynamics as novel therapeutic targets for anti-thrombotic drugs," *Biomolecules & Therapeutics*, vol. 25, no. 3, pp. 223–230, May 1, 2017.

[74] J. Jin, Y. Mao, D. Thomas, S. Kim, J. L. Daniel, and S. P. Kunapuli, "Rhoa downstream of gq and g12/13 pathways regulates protease-activated receptor-mediated dense granule release in platelets," *Biochemical Pharmacology*, vol. 77, no. 5, pp. 835–844, Mar. 1, 2009.

[75] W. T. Arthur, "Integrin engagement suppresses rhoa activity via a c-src-dependent mechanism," *Current Biology*, vol. 10, no. 12, pp. 719–722, Jun. 15, 2000.

[76] P. Flevaris, A. Stojanovic, H. Gong, A. Chishti, E. Welch, and X. Du, "A molecular switch that controls cell spreading and retraction," *Journal of Cell Biology*, vol. 179, no. 3, pp. 553–565, Nov. 5, 2007.

[77] J. H. Hartwig, "Mechanisms of actin rearrangements mediating platelet activation," *Journal of Cell Biology*, vol. 118, no. 6, pp. 1421–1442, Sep. 1992.

[78] J. E. Aslan and O. J. T. McCarty, "Rho gtpases in platelet function," *Journal of Thrombosis and Haemostasis*, vol. 11, no. 1, pp. 35–46, Jan. 2013.

[79] S. D. J. Calaminus, J. M. Auger, O. J. T. McCarty, M. J. O. Wakelam, L. M. Machesky, and S. P. Watson, "Myosiniia contractility is required for maintenance of platelet structure during spreading on collagen and contributes to thrombus stability," *Journal of Thrombosis and Haemostasis*, vol. 5, no. 10, pp. 2136–2145, Oct. 2007.

[80] J. V. Small, T. Stradal, E. Vignal, and K. Rotter, "The lamellipodium: Where motility begins," *Trends in Cell Biology*, vol. 12, no. 3, pp. 112–120, Mar. 2002.

[81] C. Valet, G. Chicanne, C. Severac, C. Chaussade, M. A. Whitehead, C. Cabou, M.-P. Gratacap, F. Gaits-Iacovoni, B. Vanhaesebroeck, B. Payrastre, and S. Severin, "Essential role of class ii pi3k-c2α in platelet membrane morphology," *Blood*, vol. 126, no. 9, pp. 1128–1137, Aug. 27, 2015.

[82] S. H. Zigmond, "Formin-induced nucleation of actin filaments," *Current Opinion in Cell Biology*, vol. 16, no. 1, pp. 99–105, Feb. 2004.

[83] M. Vinzenz, M. Nemethova, F. Schur, J. Mueller, A. Narita, E. Urban, C. Winkler, C. Schmeiser, S. A. Koestler, K. Rottner, G. P. Resch, Y. Maeda, and J. V. Small, "Actin branching in the initiation and maintenance of lamellipodia," *Journal of Cell Science*, vol. 125, no. 11, pp. 2775–2785, Jun. 1, 2012.

[84] G. Dimchev, A. Steffen, F. Kage, V. Dimchev, J. Pernier, M.-F. Carlier, and K. Rottner, "Efficiency of lamellipodia protrusion is determined by the extent of cytosolic actin assembly," *Molecular Biology of the Cell*, vol. 28, no. 10, pp. 1311–1325, May 15, 2017.

[85] G. L. Ryan, H. M. Petroccia, N. Watanabe, and D. Vavylonis, "Excitable actin dynamics in lamellipodial protrusion and retraction," *Biophysical Journal*, vol. 102, no. 7, pp. 1493–1502, Apr. 2012.

[86] F. Kage, M. Winterhoff, V. Dimchev, J. Mueller, T. Thalheim, A. Freise, S. Brühmann, J. Kollasser, J. Block, G. Dimchev, M. Geyer, H.-J. Schnittler, C. Brakebusch, T. E. B. Stradal, M.-F. Carlier, M. Sixt, J. Käs, J. Faix, and K. Rottner, "FMNL formins boost lamellipodial force generation," *Nature Communications*, vol. 8, p. 14 832, Mar. 22, 2017.

[87] A. Bender Markus andEckly, J. H. Hartwig, M. Elvers, I. Pleines, S. Gupta, G. Krone, E. Jeanclos, A. Gohla, C. Gurniak, C. Gachet, W. Witke, and B. Nieswandt, "Adf/n-cofilin-dependent actin turnover determines platelet formation and sizing," *Blood*, vol. 116, no. 10, pp. 1767–1775, Sep. 9, 2010.

[88] G. G. Borisy and T. M. Svitkina, "Actin machinery: Pushing the envelope," *Current Opinion in Cell Biology*, vol. 12, no. 1, pp. 104–112, Feb. 2000.

[89] A. J. Ridley and A. Hall, "The small gtp-binding protein rho regulates the assembly of focal adhesions and actin stress fibres in response to growth factors," *Cell*, vol. 70, no. 3, pp. 389–399, Aug. 7, 1992.

[90] J. L. Guan, "Role of focal adhesion kinase in integrin signaling," *The International Journal of Biochemistry and Cell Biology*, vol. 29, no. 8-9, pp. 1085–1096, 1997.

[91] L. Hemmings, D. J. Rees, V. Ohanian, S. J. Bolton, A. P. Gilmore, B. Patel, H. Priddle, J. E. Trevithick, R. O. Hynes, and D. R. Critchley, "Talin contains three actin-binding sites each of which is adjacent to a

vinculin-binding site," *Journal of Cell Science*, vol. 109, no. 11, pp. 2715–2726, Nov. 1996.

[92] H. Chen, D. M. Choudhury, and S. W. Craig, "Coincidence of actin filaments and talin is required to activate vinculin," *Journal of Biological Chemistry*, vol. 281, no. 52, pp. 40 389–40 398, Dec. 29, 2006.

[93] A. Carisey, R. Tsang, A. M. Greiner, N. Nijenhuis, N. Heath, A. Nazgiewicz, R. Kemkemer, B. Derby, J. Spatz, and C. Ballestrem, "Vinculin regulates the recruitment and release of core focal adhesion proteins in a force-dependent manner," *Current Biology*, vol. 23, no. 4, pp. 271–281, Feb. 18, 2013.

[94] B. Nieswandt, M. Moser, I. Pleines, D. Varga-Szabo, S. Monkley, D. Critchley, and R. Fässler, "Loss of talin1 in platelets abrogates integrin activation, platelet aggregation, and thrombus formation in vitro and in vivo," *Journal of Experimental Medicine*, vol. 204, no. 13, pp. 3113–3118, Dec. 17, 2007.

[95] T. A. Doggett, G. Girdhar, A. Lawshé, D. W. Schmidtke, I. J. Laurenzi, S. L. Diamond, and T. G. Diacovo, "Selectin-like kinetics and biomechanics promote rapid platelet adhesion in flow: The gpib(alpha)-vwf tether bond," *Biophysical Journal*, vol. 83, no. 1, pp. 194–205, Jul. 2002.

[96] N. A. Mody and M. R. King, "Platelet adhesive dynamics. part ii: High shear-induced transient aggregation via gpibalpha-vwf-gpibalpha bridging," *Biophysical Journal*, vol. 95, no. 5, pp. 2556–2574, Sep. 2008.

[97] Y. Li, P. Bhimalapuram, and A. R. Dinner, "Model for how retrograde actin flow regulates adhesion traction stresses," *Journal of Physics, Condensed Matter*, vol. 22, no. 19, p. 194 113, May 19, 2010.

[98] P. W. Oakes, Y. Beckham, J. Stricker, and M. L. Gardel, "Tension is required but not sufficient for focal adhesion maturation without a stress fibre template," *Journal of Cell Biology*, vol. 196, no. 3, pp. 363–374, Feb. 6, 2012.

[99] M. Dembo and Y. L. Wang, "Stresses at the cell-to-substrate interface during locomotion of fibroblasts," *Biophysical Journal*, vol. 76, no. 4, pp. 2307–2316, Apr. 1999.

[100] N. Q. Balaban, U. S. Schwarz, D. Riveline, P. Goichberg, G. Tzur, I. Sabanay, D. Mahalu, S. Safran, A. Bershadsky, L. Addadi, and B. Geiger,

"Force and focal adhesion assembly: A close relationship studied using elastic micropatterned substrates," *Nature Cell Biology*, vol. 3, no. 5, pp. 466–472, May 2001.

[101] A. del Rio, R. Perez-Jimenez, R. Liu, P. Roca-Cusachs, J. Fernandez M., and M. P. Sheetz, "Stretching single talin rod molecules activates vinculin binding," *Science*, vol. 323, no. 5914, pp. 638–641, Jan. 30, 2009.

[102] D. P. Felsenfeld, D. Choquet, and M. P. Sheetz, "Ligand binding regulates the directed movement of beta1 integrins on fibroblasts," *Nature*, vol. 383, no. 6599, pp. 438–440, Oct. 3, 1996.

[103] M. Chrzanowska-Wodnicka and K. Burridge, "Rho-stimulated contractility drives the formation of stress fibers and focal adhesions," *Journal of Cell Biology*, vol. 133, no. 6, pp. 1403–1415, Jun. 1996.

[104] M. D. Welch, A. Mallavarapu, J. Rosenblatt, and T. J. Mitchison, "Actin dynamics in vivo," *Current Opinion in Cell Biology*, vol. 9, no. 1, pp. 54–61, 1997.

[105] D. Raucher and M. P. Sheetz, "Cell spreading and lamellipodial extension rate is regulated by membrane tension," *The Journal of Cell Biology*, vol. 148, no. 1, pp. 127–136, Jan. 10, 2000.

[106] G. J. Johnson, L. A. Leis, M. D. Krumwiede, and J. G. White, "The critical role of myosin IIA in platelet internal contraction," *Journal of Thrombosis and Haemostasis*, vol. 5, no. 7, pp. 1516–1529, Jul. 2007.

[107] J. V. Small and G. P. Resch, "The comings and goings of actin: Coupling protrusion and retraction in cell motility," *Current Opinions in Cell Biology*, vol. 17, no. 5, pp. 517–523, Oct. 2005.

[108] S. Pellegrin and H. Mellor, "Actin stress fibres," *Journal of Cell Science*, vol. 120, no. 20, pp. 3491–3499, Oct. 15, 2007.

[109] A. M. Swenson, D. V. Trivedi, A. A. Rauscher, Y. Wang, Y. Takagi, B. M. Palmer, A. Málnási-Csizmadia, E. P. Debold, and C. M. Yengo, "Magnesium modulates actin binding and adp release in myosin motors," *Journal of Biological Chemistry*, vol. 289, no. 34, pp. 23 977–23 991, Aug. 22, 2014.

[110] O. V. Kim, R. I. Litvinov, M. S. Alber, and J. W. Weisel, "Quantitative structural mechanobiology of platelet-driven blood clot contraction," *Nature Communications*, vol. 8, no. 1, p. 1274, Nov. 2, 2017.

[111] Y. Cai, N. Biais, G. Giannone, M. Tanase, G. Jiang, J. M. Hofman, C. H. Wiggins, P. Silberzan, A. Buguin, B. Ladoux, and M. P. Sheetz, "Non-muscle myosin IIA-dependent force inhibits cell spreading and drives f-actin flow," *Biophysical Journal*, vol. 91, no. 10, pp. 3907–3920, Nov. 2006.

[112] C. J. Jen and L. V. McIntire, "The structural properties and contractile force of a clot," *Cell Motility*, vol. 2, no. 5, pp. 445–455, 1982.

[113] X. M. Liang, S. J. Han, J.-A. Reems, D. Gao, and N. J. Sniadecki, "Platelet retraction force measurements using flexible post force sensors," *Lab on a Chip*, vol. 10, pp. 991–998, Jan. 20, 2010.

[114] W. A. Lam, O. Chaudhuri, A. Crow, K. D. Webster, T.-D. Li, A. Kita, J. Huang, and D. A. Fletcher, "Mechanics and contraction dynamics of single platelets and implications for clot stiffening," *Nature Materials*, vol. 10, no. 1, pp. 61–66, Jan. 2011.

[115] Y. Qiu, A. C. Brown, D. R. Myers, Y. Sakurai, R. G. Mannino, R. Tran, B. Ahn, E. T. Hardy, M. F. Kee, S. Kumar, G. Bao, T. H. Barker, and W. A. Lam, "Platelet mechanosensing of substrate stiffness during clot formation mediates adhesion, spreading, and activation," *PNAS*, vol. 111, no. 40, pp. 14 430–14 435, Oct. 7, 2014.

[116] S. Schwarz Henriques, R. Sandmann, A. Strate, and K. Sarah, "Force field evolution during human blood platelet activation," *Journal of Cell Science*, vol. 125, no. 16, pp. 3912–3920, Aug. 15, 2012.

[117] X. Zhang, G. Jiang, Y. Cai, S. J. Monkely, D. R. Critchley, and M. P. Sheetz, "Talin depletion reveals independence of initial cell spreading from integrin activation and traction," *Nature Cell Biology*, vol. 10, pp. 1062–1068, Aug. 17, 2008.

[118] A. Elosegui-Artola, R. Oria, Y. Chen, A. Kosmalska, C. Pérez-González, N. Castro, C. Zhu, X. Trepat, and P. Roca-Cusachs, "Mechanical regulation of a molecular clutch defines force transmission and transduction in response to matrix rigidity," *Nature Cell Biology*, vol. 18, pp. 540–548, Apr. 11, 2016.

[119] D. Critchley, "Cytoskeletal proteins talin and vinculin in integrin-mediated adhesion," *Biochemical Society Transactions*, vol. 32, no. 5, pp. 831–836, Nov. 1, 2004.

[120] M. Moser, K. R. Legate, R. Zent, and R. Fassler, "The tail of integrins, talin, and kindlins," *Science*, vol. 324, no. 5929, pp. 895–899, May 15, 2009.

[121] E. Goksoy, Y.-Q. Ma, X. Wang, X. Kong, D. Perera, E. F. Plow, and J. Qin, "Structural basis for the autoinhibition of talin in regulating integrin activation," *Molecular Cell*, vol. 31, no. 1, pp. 124–133, Jul. 11, 2008.

[122] N. Watanabe, L. Bodin, M. Pandey, M. Krause, S. Coughlin, V. A. Boussiotis, M. H. Ginsberg, and S. J. Shattil, "Mechanisms and consequences of agonist-induced talin recruitment to platelet integrin αiibβ3," *Journal of Cell Biology*, vol. 181, no. 7, pp. 1211–1222, Jun. 23, 2008.

[123] H.-s. Lee, C. J. Lim, W. Puzon-McLaughlin, S. J. Shattil, and M. H. Ginsberg, "Riam activates integrins by linking talin to ras gtpase membrane-targeting sequences," *Journal of Biological Chemistry*, vol. 284, no. 8, pp. 5229–5127, Feb. 2009.

[124] J. Yang, L. Zhu, H. Zhang, J. Hirbawi, K. Fukuda, P. Dwivedi, J. Liu, T. Byzova, E. F. Plow, J. Wu, and J. Qin, "Conformational activation of talin by riam triggers integrin-mediated cell adhesion," *Nature Communications*, vol. 5, p. 5880, Dec. 18, 2014.

[125] F. Margadant, L. L. Chew, X. Hu, H. Yu, N. Bate, X. Zhang, and M. Sheetz, "Mechanotransduction in vivo by repeated talin stretch-relaxation events depends upon vinculin," *PLOS Biology*, vol. 9, no. 12, e1001223, Dec. 20, 2011.

[126] B. T. Goult, T. Zacharchenko, N. Bate, R. Tsang, F. Hey, A. R. Gringas, P. R. Elliot, G. C. K. Roberts, C. B1strem, D. R. Critchley, and I. L. Barsukov, "Riam and vinculin binding to talin are mutually exclusive and regulate adhesion assembly and turnover," *The Journal of Biological Chemistry*, vol. 288, no. 12, pp. 8238–8249, Mar. 22, 2013.

[127] M. L. Gardel, B. Sabass, L. Ji, G. Danuser, U. S. Schwarz, and C. M. Waterman, "Traction stress in focal adhesions correlates biphasically with actin retrograde flow speed," *Journal of Cell Biology*, vol. 183, no. 6, pp. 999–1005, Dec. 15, 2008.

[128] V. Swaminathan and C. M. Waterman, "The molecular clutch model for mechanotransduction evolves," *Nature Cell Biology*, vol. 18, pp. 459–461, Apr. 27, 2016.

[129] P. Atherton, B. Stutchbury, D.-Y. Wang, D. Jethwa, R. Tsang, E. Meiler-Rodriguez, P. Wang, N. Bate, R. Zent, I. L. Barsukov, B. T. Goult, D. R. Critchley, and C. Ballestrem, "Vinculin controls talin engagement with the actomyosin machinery," *Nature Communications*, vol. 6, p. 10 038, Dec. 4, 2015.

[130] R. M. Saunders, M. R. Holt, L. Jennings, H. Sutton Deborah, I. L. Barsukov, A. Bobkov, R. C. Liddington, E. A. Adamson, G. A. Dunn, and D. R. Critchley, "Role of vinculin in regulating focal adhesion turnover," *European Journal of Cell Biology*, vol. 85, no. 6, pp. 487–500, Jun. 14, 2006.

[131] T. R. Strick, J. .-. Allemand, D. Bensimon, and V. Croquette, "Behavior of supercoiled dna," *Biophysical Journal*, vol. 74, no. 4, pp. 2016–2028, Apr. 1998.

[132] H. Chen, H. Fu, X. Zhu, P. Cong, F. Nakamura, and J. Yan, "Improved high-force magnetic tweezers for stretching and refolding of proteins and short dna," *Biophysical Journal*, vol. 100, no. 2, pp. 517–523, Jan. 19, 2011.

[133] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy," *Nature Methods*, vol. 5, no. 6, pp. 491–505, Jun. 2008.

[134] A. R. Gingras, N. Bate, B. T. Goult, I. L. Barsukov, G. C. K. Roberts, and D. R. Critchley, "Central region of talin has a unique fold that binds vinculin and actin," *Journal of Biological Chemistry*, vol. 285, no. 38, pp. 29 577–29 587, Sep. 2010.

[135] G. Jiang, G. Giannone, D. R. Critchley, E. Fukumoto, and M. P. Sheetz, "Two-piconewton slip bond between fibronectin and the cytoskeleton depends on talin," *Nature*, vol. 424, pp. 334–337, Jul. 17, 2003.

[136] R. Rahikainen, M. von Essen, M. Schaefer, L. Qi, L. Azizi, K. Conor, T. O. Ihalainen, B. Wehrle-Haller, M. Bastmeyer, C. Huang, and V. P. Hytönen, "Mechanical stability of talin rod controls cell migration and substrate sensing," *Scientific Reports*, vol. 7, p. 3571, Jun. 15, 2017.

[137] X. Hu, C. Jing, X. Xu, N. Nakazawa, V. W. Carnish, F. M. Margadant, and M. P. Sheetz, "Cooperative vinculin binding to talin mapped by time-resolved super resolution microscopy," *Nano Letters*, vol. 16, no. 7, pp. 4062–4068, May 23, 2016.

[138] J. A. McCammon, B. G. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, pp. 585–590, Jun. 16, 1977.

[139] A. Li and V. Daggett, "Investigation of the solution structure of chymotrypsin inhibitor 2 using molecular dynamics: Comparison to x-ray crystallographic and nmr data," *Protein Engineering*, vol. 8, no. 11, pp. 1117–1128, Nov. 1, 1995.

[140] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulation of protein structure and function," *Current Opinion in Structural Biology*, vol. 19, no. 2, pp. 120–127, Apr. 8, 2009.

[141] A. W. M. Haining, R. Rahikainen, E. Cortes, D. Lachowski, A. Rice, M. von Essen, V. P. Hytönen, and A. del Rio Hernandez, "Mechanotransduction in talin through the interaction of the r8 domain with dlc1," *PLOS Biology*, vol. 16, no. 7, e2005599, Jul. 20, 2018.

[142] V. V. Mykuliak, A. W. M. Haining, M. von Essen, A. del Rio Hernandez, and V. P. Hytönen, "Mechanical unfolding reveals stable 3-helix intermediates in talin an α-catenin," *PLOS Computational Biology*, vol. 14, no. 4, e1006126, Apr. 26, 2018.

[143] V. V. Mykuliak, M. Sikora, J. J. Booth, M. Cieplak, D. V. Shalashilin, and V. P. Hytönen, "Mechanical unfolding of proteins — a comparative nonequilibrium molecular dynamics study," *Biophysical Journal*, vol. 199, pp. 939–949, Sep. 1, 2020.

[144] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, "Coarse-grained protein models and their applications," *Chemical Reviews*, vol. 116, pp. 7898–7936, Jun. 22, 2016.

[145] C. D. A., H. M. Aktulga, K. Belfon, I. Y. Ben-shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, H. Gilson M. K. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, C. Jin, C. Lin, T. Liu J. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M. Merz, Y. miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Anufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Shimmerling, N. R. Skynnikov, J. Smith, J. Swails, R. C. Walker,

J. Wang, H. Wei, R. M. Wolf, X. Wu, Y. Xue, D. M. York, S. Zhao, and P. A. Kollman, *Amber 2021*. San Francisco: University Of California, 2021.

[146] B. R. Brooks, C. L. Brooks, III, A. D. MacKerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, S. Bartels C. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. W. York, and M. Karplus, "Charmm: The biomolecular simulation program," *Journal of Computational Biology*, vol. 30, no. 10, pp. 1545–1614, Jul. 30, 2010.

[147] P. Bauer, B. Hess, and E. Lindahl, *Gromacs 2022 manual*, Feb. 2022.

[148] J. C. Phillips, J. Hardy David, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot, and E. Tajkhorshid, "Scalable molecular dynamics on cpu and gpu architectures with namd," *Journal of Chemical Physics*, vol. 153, no. 4, p. 044 130, Jul. 28, 2020.

[149] C.-e. A. Chang, Y.-m. M. Huang, L. J. Mueller, and W. You, "Investigation of structural dynamics of enzymes and protonation states of substrates using computational tools," *Catalysts*, vol. 6, no. 6, p. 82, May 31, 2016.

[150] P. Maragakis, K. Lindorff-Larsen, M. P. Eastwood, R. O. Dror, J. L. Klepeis, I. T. Arkin, M. Ø. Jensen, H. Xu, N. Trbovic, R. A. Friesner, A. G. Palmer, and D. E. Shaw, "Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins," *Journal of Physical Chemistry B*, vol. 112, no. 19, pp. 6155–6158, 2008.

[151] R. Sun, O. Sode, J. F. Dama, and G. A. Voth, "Simulating protein mediated hydrolysis of atp and other nucleoside triphosphates by combining qm/mm molecular dynamics with advances in metadynamics," *Journal of Chemical Theory and Computation*, vol. 13, no. 5, pp. 2332–2341, Mar. 27, 2017.

[152] H. J. Dyson, P. E. Wright, and H. A. Scheraga, "The role of hydrophobic interactions in initiation and propagation of protein folding," *PNAS*, vol. 103, no. 35, pp. 13 057–13 061, Aug. 29, 2006.

[153] A. CCumberworth, J. M. Bui, and J. Gsponer, "Free energies of solvation in the context of protein folding: Implications for implicit and explicit solvent models," *Journal of Computational Chemistry*, vol. 37, no. 7, pp. 629–640, Nov. 12, 2015.

[154] N. Pokala and T. M. Handel, "Energy functions for protein design i: Efficient and accurate continuum electrostatics and solvation," *Protein Science*, vol. 13, no. 4, pp. 925–936, Jan. 1, 2009.

[155] A. Kolinski, "Protein modelling and structure prediction with a reduced representation," *Acta Biochimica Polonica*, vol. 51, no. 2, pp. 349–371, May 27, 2004.

[156] A. Liwo, M. Baranowski, C. Czaplewski, E. Golas, Y. He, D. Jagiela, P. Krupa, M. Maciejczyk, M. Makowski, M. A. Mozolewska, A. Niadzvedtski, S. Oldziej, H. A. Scheraga, A. K. Sieradzan, R. Slusarz, T. Wirecki, Y. Yin, and B. Zaborowski, "A unified coarse-grained model of biological macromolecules based on mean-field multipole–multipole interactions," *Journal of Molecular Modelling*, vol. 20, p. 2306, Jul. 15, 2014.

[157] M. Levitt and A. Warshel, "Computer simulation of protein folding," *Nature*, vol. 253, no. 1, pp. 694–698, Feb. 1, 1975.

[158] M. Zachariak, "Protein-protein docking with a reduced protein model accounting for side-chain flexibility," *Protein Science*, vol. 12, no. 6, pp. 1271–1282, Jan. 1, 2009.

[159] B. G. Petrich, P. Marchese, Z. M. Ruggeri, S. Spiess, R. A. M. Weichert, F. Ye, R. Tiedt, R. C. Skoda, S. J. Monkley, D. R. Critchley, and M. H. Ginsberg, "Talin is required for integrin-mediated platelet function in hemostasis and thrombosis," *Journal of Experimental Medicine*, vol. 204, no. 13, pp. 3103–3111, Oct. 17, 2007.

[160] F. Lagarrigue, D. S. Paul, A. R. Gingras, A. J. Valadez, H. Sun, J. Lin, M. N. Cuevas, J. N. Ablack, M. A. Lopez-Ramirez, W. Bergmeier, and M. H. Ginsberg, "Talin-1 is the principal platelet rap1 effector of integrin activation," *Blood*, vol. 136, no. 10, pp. 1180–1190, Sep. 3, 2020.

[161] J. Yan, M. Yao, B. T. Goult, and M. P. Sheetz, "Talin dependent mechanosensitivity of cell focal adhesions," *Cellular and Molecular Bioengineering*, vol. 8, no. 1, pp. 151–159, Mar. 1, 2015.

[162] S. Le, R. Liu, C. Teck Lim, and J. Yan, "Uncovering mechanosensing mechanisms at the single protein level using magnetic tweezers," *Methods*, vol. 94, pp. 13–18, Feb. 1, 2016.

[163] D. Provasi, A. Negri, B. S. Coller, and M. Filizola, "Talin-driven inside-out activation mechanism of platelet αiibβ3 integrin probed by multimicrosecond, all-atom molecular dynamics simulations," *Proteins*, vol. 82, pp. 3231–3240, Feb. 20, 2014.

[164] C. W. Hopkins, S. Le Grand, R. C. Walker, and A. E. Roitberg, "Long-time-step molecular dynamics through hydrogen mass repartitioning," *Journal of Chemical Theory and Computation*, vol. 11, no. 4, pp. 1864–1874, Mar. 16, 2015.

[165] D. S. Kudryashov and E. Reisler, "Atp and adp actin states," *Biopolymers*, vol. 99, no. 4, pp. 245–256, Apr. 2013.

[166] Y. Tsuda, H. Yasutake, A. Ishijima, and T. Yanagida, "Torsional rigidity of single actin filaments and actin-actin bond breaking force under torsion measured directly by in vitro micromanipulation," *PNAS*, vol. 93, no. 23, pp. 12 937–12 942, Nov. 12, 1996.

[167] C. S. Peskin, G. M. Odell, and G. F. Oster, "Cellular motions and thermal fluctuations: The brownian ratchet," *Biophysical Journal*, vol. 65, pp. 316–324, Jul. 1993.

[168] M. M. Kozlov and A. D. Bershadsky, "Processive capping by formin suggests a force-driven mechanism of actin polymerization," *Journal of Cell Biology*, vol. 167, no. 6, pp. 1011–1017, Dec. 3, 2004.

[169] A. Mogilner and G. Oster, "Cell motility driven by actin polymerization," *Biophysical Journal*, vol. 71, no. 6, pp. 3030–3045, Dec. 1996.

[170] G. W. Greene, T. H. Anderson, H. Zeng, B. Zappone, and J. N. Israelachvili, "Force amplification response of actin filaments under confined compression," *PNAS*, vol. 106, no. 2, pp. 445–449, Jan. 13, 2009.

[171] M. Vicente-Manzanares, X. Ma, and A. R. Horwitz, "Non-muscle myosin ii takes centre stage in cell adhesion and migration," *Nature Reviews Molecular Cell Biology*, vol. 10, pp. 778–790, Nov. 2009.

[172] B. Brenner, "The stroke size of myosins: A reevaluation," *Journal of Muscle Research & Cell Motility*, vol. 27, p. 173, Feb. 10, 2006.

[173] J. T. Finer, R. M. Simmons, and J. A. Spudich, "Single myosin molecule mechanics: Piconewton forces and nanometre steps," *Nature*, vol. 268, pp. 113–119, Mar. 10, 1994.

[174] A. Ishijima, Y. Harada, H. Kojima, T. Funatsu, H. Higuchi, and T. Yanagida, "Single-molecule analysis of the actomyosin motor using nano-manipulation," *Biochemical and Biophysical Research Communications*, vol. 199, no. 2, pp. 1057–1063, Mar. 15, 1994.

[175] N. Hundt, W. Steffen, S. Pathan-Chhatbar, M. H. Taft, and D. J. Manstein, "Load-dependent modulation of non-muscle myosin-2a function by tropomyosin 4.2," *Scientific Reports*, vol. 6, p. 20 554, Feb. 5, 2016.

[176] J. H. Hartwig, "Chapter 4 — the platelet cytoskeleton," in *Platelets (Second Edition)*, A. D. Michelson, Ed., Second Edition, Burlington: Academic Press, 2007, pp. 75–97.

[177] G. Totsukawa, Y. Yamakita, S. Yamashiro, D. J. Hartshorne, Y. Sasaki, and F. Matsumura, "Distinct roles of rock (rho-kinase) and mlck in spatial regulation of mlc phosphorylation for assembly of stress fibres and focal adhesions in 3t3 fibroblasts," *Journal of Cell Biology*, vol. 150, no. 4, pp. 797–806, Aug. 21, 2000.

[178] G. Totsukawa, Y. Wu, Y. Sasaki, D. J. Hartshorne, Y. Yamakita, S. Yamashiro, and F. Matsumura, "Distinct roles of mlck and rock in the regulation of membrane protrusions and focal adhesion dynamics during cell migration of fibroblasts," *Journal of Cell Biology*, vol. 164, no. 3, pp. 427–439, Feb. 2, 2004.

[179] M. H. Kroll, J. D. Hellums, L. V. McIntire, A. I. Schafer, and J. L. Moake, "Platelets and shear stress," *Blood*, vol. 88, no. 5, pp. 1525–1541, Sep. 1, 1996.

[180] J. M. Brockman, A. T. Blanchard, V. P.-Y. Ma, W. D. Derricotte, Y. Zhang, M. E. Fay, W. A. Lam, F. A. Evangelista, A. L. Matthyses, and K. Salaita, "Mapping the 3d orientation of piconewton integrin traction forces," *Nature Methods*, vol. 15, pp. 115–118, Dec. 11, 2017.

[181] S. Chakraborty, S. Banerjee, M. Raina, and S. Haldar, "Force-directed "mechanointeractome" of talin-integrin," *Biochemistry*, vol. 58, no. 47, pp. 4677–4695, Aug. 8, 2019.

[182] A. Agnihotri, P. Soman, and C. A. Siedlecki, "Afm measurements of interactions between the platelet integrin receptor gpiibiiia and fibrinogen," *Colloids and Surfaces B: Biointerfaces*, vol. 71, no. 1, pp. 138–147, Jun. 1, 2009.

[183] A. R. Gingras, W. H. Ziegler, R. Frank, I. L. Barsukov, G. C. K. Roberts, D. R. Critchley, and J. Emsley, "Mapping and consensus sequence identification for multiple vinculin binding sites within the talin rod," *Journal of Biological Chemistry*, vol. 280, no. 11, pp. 37 217–37 223, Nov. 2005.

[184] Y. Wang, J. Yan, and B. T. Goult, "Force-dependent binding constants," *Biochemistry*, vol. 58, no. 47, pp. 4696–4709, Jul. 18, 2019.

[185] R. Flaumenhaft and D. S. Sim, "The platelet as a model for chemical genetics," *Cell Chemical Biology*, vol. 10, no. 6, pp. 481–486, Jun. 1, 2003.

[186] D. Bishop-Bailey, "The platelet as a model system for the acute actions of nuclear receptors," *Steroids*, vol. 75, no. 8-9, pp. 570–575, Sep. 22, 2010.

[187] J. R. Meyers, "Zebrafish: Development of a vertebrate model organism," *Current Protocals*, vol. 16, e19, May 4, 2018.

[188] K. A. Dill, "Dominant force in protein folding," *Biochemistry*, vol. 29, no. 31, pp. 7133–7155, Aug. 7, 1990.

[189] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Sindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, Jan. 1, 2000.

[190] H. Berman and H. Nakamura, "Announcing the worldwide protein data bank," *Nature Structural & Molecular Biology*, vol. 10, p. 980, Dec. 1, 2003.

[191] G. G. Butcher, W. S. Harwin, and C. I. Jones, "An efficient alpha-helix model and simulation framework for non-dynamic electrostatic interaction force estimation.," *Scientific Reports*, vol. 11, pp. 1–11, Apr. 12, 2021.

[192] V. Sobolev, A. Sorokien, J. Prilusky, E. Abola, and M. Edelman, "Automated analysis of interatomic contacts in proteins," *Bioinformatics*, vol. 15, no. 4, pp. 327–332, Apr. 1999.

[193] S. Bienert, A. Waterhouse, T. A. P. de Beer, G. Tauriello, G. Studer, L. Bordoli, and T. Schwede, "The swiss-model repository—new features and

functionality," *Nucleic Acids Research*, vol. 45, no. D1, pp. D313–D319, Nov. 2016.

[194] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, and T. Schwede, "Swiss-model: Homology modelling of protein structures and complexes," *Nucleic Acids Research*, vol. 46, no. W1, W296–W303, May 2018.

[195] P. Benkert, M. Biasini, and T. Schwede, "Toward the estimation of the absolute quality of individual protein structure models," *Bioinformatics*, vol. 27, no. 3, pp. 343–350, Dec. 2010.

[196] D. Dedden, S. Schumacher, C. F. Kelley, M. Zacharias, C. Biertimpfel, R. Fässler, and N. Mizuno, "The architecture of talin1 and autoinhibition mechanism," *Cell*, vol. 179, no. 1, 120–131.e13, Sep. 19, 2019.

[197] T. Zacharchenko, X. Qian, B. T. Goult, D. Jethwa, T. B. Almeida, C. Ballestrem, D. R. Critchley, D. R. Lowy, and I. L. Barsukov, "Ld motif recognition by talin: Structure of the talin-dlc1 complex," *Structure*, vol. 24, no. 7, pp. 1130–1141, Jul. 6, 2016.

[198] X. Song, J. Yang, J. Hirbawi, S. Ye, H. D. Perera, E. Goskoy, P. Dwivedi, E. F. Plow, R. Zhang, and J. Qin, "A novel membrane-dependent on/off switch mechanism of talin ferm domain at sites of cell adhesion," *Cell Research*, vol. 22, pp. 1533–1545, Jun. 19, 2012.

[199] B. T. Goult, A. R. Gingras, N. Bate, I. L. Barsukov, D. R. Critchley, and G. C. K. Roberts, "The domain structure of talin: Residues 1815–1973 form a five-helix bundle containing a cryptic vinculin-binding site," *FEBS Letters*, vol. 584, no. 11, pp. 2237–2241, Apr. 20, 2010.

[200] A. R. Gingras, W. H. Ziegler, A. A. Bobkov, M. G. Joyce, D. Fasci, M. Himmel, S. Rothemund, A. Ritter, J. G. Grossmann, B. Patel, N. Bate, B. T. Goult, J. Emsley, I. L. Barsukov, G. C. K. Roberts, R. C. Liddington, M. H. Ginsberg, and D. R. Critchley, "Structural determinants of integrin binding to the talin rod," *Journal of Biological Chemistry*, vol. 284, no. 13, pp. 8866–8876, Mar. 27, 2009.

[201] A. R. Gingras, N. Bate, B. T. Goult, L. Hazelwood, I. Canestrelli, J. G. Grossmann, H. J. Liu, N. S. M. Putz, G. C. K. Roberts, N. Volkmann, D. Hanein, L. Barsukov Igor, and D. R. Critchley, "The structure of the

c-terminal actin-binding domain of talin," *EMBO*, vol. 27, no. 2, pp. 458–469, Dec. 20, 2007.

[202] "Introduction to protein data banks format." (2007), [Online]. Available: `https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/pdbintro.html` (visited on 03/27/2022).

[203] "Atomic coordinate entry format version 3.3." (Nov. 21, 1012), [Online]. Available: `http://www.wwpdb.org/documentation/file-format-content/format33/sect9.html` (visited on 03/27/2022).

[204] *Matlab version 9.9.0*, The MathWorks, Inc., Natick, Massachusetts, 2020.

[205] J. J. Craig, *Introduction to Robotics: Mechanics and Control*. Prentice Hall, 1989.

[206] D. J. Barlow and J. M. Thornton, "Helix geometry in proteins," *Journal of Molecular Biology*, vol. 201, no. 3, pp. 601–619, Jun. 5, 1988.

[207] A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, "A united-residue force field for off-lattice protein-structure simulations. i. functional forms and parameters of long-range side-chain interaction potentials from protein crystal data.," *Journal of Computational Chemistry*, vol. 18, no. 7, pp. 849–873, Dec. 7, 1998.

[208] m. S. J. Han, E. V. Azarova, A. J. Whitewood, A. Bachir, E. Guttierez, A. Groisman, A. R. Horwitz, B. T. Goult, and K. M. Dean, "Pre-complexation of talin and vinculin without tension is required for efficient nascent adhesion maturation," *eLife*, vol. 10, e66151, Mar. 30, 2021.

[209] R. E. Gough, M. C. Jones, T. Zacharchenko, L. Shimin, M. Yu, G. Jacquemet, S. P. Muench, J. Yan, J. D. Humphries, C. Jørgensen, M. J. Humphries, and B. T. Goult, "Talin mechanosensitivity is modulated by a direct interaction with cyclin-dependent kinase-1," *Journal of Biological Chemistry*, vol. 297, no. 1, p. 100 837, Jul. 2021.

[210] L. Wang, R. A. Friesner, and B. J. Berne, "Hydrophobic interactions in model enclosures from small to large length scales: Non-additivity in explicit and implicit solvent models," *Faraday Discussions*, vol. 146, pp. 247–401, Mar. 10, 2011.

[211] R. M. Lynden-Bell, "From hydrophobic to hydrophilic behaviour: A simulation study of solvation entropy and free energy of simple solutes," *Journal of Chemical Physics*, vol. 107, pp. 1981–1991, Jun. 4, 1998.

[212] M. Munson, S. Balasubramanian, K. G. Fleming, A. D. Nagi, R. O'Brien, J. M. Sturtevant, and L. Regan, "What makes a protein a protein? hydrophobic core designs that specify stability and structural properties.," *Protein Science*, vol. 5, pp. 1584–1593, Aug. 1996.

[213] D. Eisenberg and A. D. McLachlan, "Solvation energy in protein folding and binding," *Nature*, vol. 319, pp. 199–203, Jan. 16, 1986.

[214] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, May 5, 1982.

[215] E. Axelsson, J. Costa, C. Silva, C. Emmart, A. Bock, and A. Ynnerman, "Dynamic scene graph: Enabling scaling, positioning, and navigation in the universe," *Computer Graphics Forum*, vol. 36, no. 3, pp. 459–468, Jul. 4, 2017.

[216] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, pp. 583–589, Jul. 15, 2021.