Unravelling the Mechanobiology Landscape of Force Sensing Paradigms through Molecular Simulations and Enhanced Sampling Methodologies

by

Yuvraj Singh

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Chemistry
New York University
May, 2024

Glen M. Hocky

© Yuvraj Singh

All rights reserved, 2024

DEDICATION

To my grandfather

ACKNOWLEDGMENTS

This is going to be the part of my thesis where I'm going to be speaking directly from the heart as much as I can. That being said I have to give appreciation to many people I have met at NYU, my friends from college, and my family.

Starting off with my mentors at NYU. First and foremost many thanks go to my advisor Prof. Glen Hocky. I must confess, even though Glen and I weren't always in agreement, much of which had to do with my own bold and unreserved nature, there is no doubt Glen and I can still work together to create great science. At the end of the day, I am grateful for you giving me a chance and taking me under your mentorship. The skills and confidence that have grown in me would not have been possible without you. No matter what sector I go into, I'll always be proud of the work you and I have accomplished, whether it was with collaborators, our own research projects, or even departmental duties. I promise that everything I learned while in the Hocky lab will be put to great use. If there's anything I can help with in the future, one message or email and I'll be there. I also have to give credit to the rest of my thesis committee for their guidance throughout; Prof. Mark Tuckerman, Prof. Paramjit Arora, Prof. Alexej Jerschow, and Prof. Tania Lupoli. Mark, thank you for being my committee chair and my initial inspiration for coming to NYU. Bobby, I love your knowledge of the science and that you speak your mind when things make sense or not, something I respect and relate to. Alexej, I've always appreciated your calm demeanor and your stimulating insights on Theory/Simulation. Tania, thank you again for being my reader; I look forward to you learning about the work I've done and am excited to know your thoughts. Other faculty members I've enjoyed meeting and having conversations with are Prof. Zlatko Bacic, Prof. Marc Walters, and Prof. Enrique Rojas from the Bio department. I cannot forget my collaborators Prof. Brad Nolen from the University of Oregon and Dr. Aaron Blanchard from Duke University.

I have to give much credit to my mentors and recommenders from Rutgers-Camden. If it weren't for them I wouldn't be writing this to begin with- Dr. Luca Larini, Dr. George Kumi, Dr. David Salas De la Cruz, RJ Portella, Ana Rodriguez, BQ, and Lesley Bryant. RJ, BQ, and Ana were mentors from my first generation college student support group and were with me at every step of college, especially RJ during my preparation for Grad school. Larini was the one who motivated me and put me on the path to grad school when I was totally clueless, even though I was far from the best student. If you ever read this one day just know I've never forgotten your patience and all that you did for me, especially helping me pick out schools and going over my applications. I hope we cross paths someday to show you how far I've come.

Although I was distant at the very beginning, the truth is, there are many people at NYU I've grown fond of over the years, be it administrators, staff, grad students, facilities managers, or security personnel alike. Within the Hocky lab, I have to give many thanks to Gaurav, Nico, Triasha, Abby, and Michael our first postdoc. I would also like to mention the wonderful REU and GSTEM students I was fortunate enough to guide and mentor- Alex, Matilda, Deborah, and Lauren. Gaurav, you're without a doubt my closest colleague here. I've always enjoyed your company during those late nights in lab and in general, our conversations ranging from research to everything under the sun without any reservations, and your sense of humor. I know you've been through a lot over the years yet you continue to get stronger and better. You'll always have my respect for that and I hope we stay in touch. Nico, you've really learnt a lot in such a short span of time. Keep up the great work and never stop getting better. Triasha, even in your second year of grad school I've seen you work in lab late nights like Gaurav and I, something even I didn't do. Keep your head up and stay motivated. I know you'll accomplish great things. Michael, thank you for being one of my first mentors here at NYU. You might have been here for a brief period but you did show me you can do research, volunteer, network, and have an awesome personality all at once. Other fellow graduate students who started the same year as me are Amiel and Richard. Other people from the theory suite I have to give a shout-out to are Stephanie, Eric, Tommy, and Tuckerman lab alumni Leslie and Joey. Michael, Muhammad, Iren, Markus, Tom, Phillip, and Federica; even though you guys haven't been here for a full year I've enjoyed being around you all. Michael and Muhammad, I not only value you both as scientists for your enthusiasm, multitasking skills, and eagerness to help others but even getting to know you as friends has been a joy. I also have to give much appreciation to my fellow GSO members from all over the department and previous members who served with me while I was a general member and the "point of contact", especially Tommy and Veronika for their passion, leadership, and trying their hardest to bring people closer. In fact, no matter what the division, there are several other grad students and postdocs I've had good conversations with and learnt something from, especially from the MDI laboratories. If anything, there are some people I wish I could've been closer with to some extent or at least be on better terms. Regardless, I'm thankful to everyone I've met and had close contact with here.

I cannot forget the hardworking staff and administrators within and outside the department. Starting with Julie, thank you for having faith in me when it came to departmental service and in general and teaching me a thing or two about keeping a positive attitude even in the most difficult situations. It doesn't matter that you might not be a scientist, you're definitely someone worth learning something from, I know I did. Kathleen, you have to be one of the kindest people I've ever met. The Simons Center should be lucky to have someone who is always ready to help and take initiative when needed as you. Hafsah, even though you mainly worked in the MDI and MDI related events, it was always a pleasure seeing and talking to you at the Simons events or whenever we crossed paths for that matter. I could always rely Shenglong and the rest of the NYU HPC's team technical expertise whenever I faced an issue while using the cluster.

Now I give thanks to my friends and fraternity brothers from Rutgers. I know I was a bit aloof at the beginning of grad school but I'm so grateful to have each and every one of you on my journey even after so many years. Starting off with Taylor, Eric, Parth, Bill, and Arpreet with whom I had my first trip to Costa Rica. Taylor, it's been great getting to know you over the years. Fraternity brother or not, I'm so glad to call you a friend and have you on this list. Eric, you and I haven't spent too much time together but it's been so great meeting and hanging out with you, I couldn't be happier calling you a brother. Parth and Bill, both of you were my mentees back in college but I think it's safe to say you both have come a long way since then. Parth, keep doing great things and being a great person. Bill, you've really grown up a lot since college. I'm so proud of you and glad we got to reconnect on this trip. Also, you're one of the funniest people I know. Arpreet was not only with me in Costa Rica but also during my first trip to Hawaii. Even though we haven't met or spoken so long before Hawaii, I love how we can just continue along and have fun like nothing's changed. You really are and have been a great brother. Some of my closest thanks have to go to my "big" Nick. It's not only because of Nick I became a groomsman for the first time in my life at his wedding, but I know Nick is one of the most solid and loyal individuals I have in my circle. Nick if you're reading this, I'm not going to be the smartest man ever but maybe a step closer. Some of my other favorite people in this world who have stood by me during difficult times are Ricky, Amirah, Adam, Bryce, Jake, Jason and PT.

I also want to give much thanks to two of my closest fraternity brothers from college- Andres and Fink. The three of us not only travelled together to Hawaii and Costa Rica but also a third trip to Colombia which I hold very near and dear. Fink, you're one of the few people who reached out to me while I was in grad school, that too while you were literally saving lives at the beginning of COVID. It felt great to get back in touch after all this time and travel together. Andres and I both agree you can be a little "intense" every now and then, but no one can ever question your passion or caring nature. That's why you're at the top of this list. Andres, I feel like you and I along with Fink have been on a long journey these past 10 years and it's still going- from living in the same freshman dorms to joining the same fraternity, living in the same fraternity house, to graduating college at the same time and even starting our careers in New York at the same time. Like Nick and Fink, you're one of the most loyal people I have in my circle. There were many moments in graduate school when I was low and your reaching out to hang out easily lifted my spirits. Even some of your messages in the common group chats we are in were enough at times. Andres, you've truly been a great friend with a big heart, and there's no questioning that. Also, I could never forget the rest of the Del Gallego household- your brother Alex, your parents Edna and Raul, and of course your dogs Coco and Princess. The generosity and warmth of Edna and Raul especially while we were in Colombia will always hold a special place in my heart.

Finally, my closest regards go to my brother, Mom, Dad, grandmother, and the rest of my family back in India. My Mom and Dad are without a doubt my biggest fans and supporters for any milestone I achieve or even when I fail, always supporting me the best they can. Just the fact that I come home and know that they are there is itself a blessing I never take for granted. As far as my brother goes, he's my oldest and best friend. He's probably the only person in this world who knows me better than I know myself at times. In fact, this entire journey began when he and his best friend Hua were the first people who got me settled in when I landed from India 10 years ago. Whether it was pizza nights at your place, random Walmart runs with Joey, spontaneous nights in Philly with Sedrick, hours-long gym sessions with Tony, Eddie, or JR, or just lengthy thought-provoking conversations. There was no way I could even complete this section without acknowledging you, Hua, Sedrick, Costa, Joey, Eddie, and Tony. All these experiences really gave me a solid foundation my first year in the States and taught me how to navigate through life problems and make meaningful friendships and connections that I mentioned before. If there is anyone in this world I'm writing this for it will have to be you, Mom, and Dad.

ABSTRACT

This thesis aims to unravel the intricacies of Mechanobiology surrounding Molecular Machines within cells called Force Sensors, using the latest approaches in Computational Chemistry and Statistical Mechanics. Mechanical forces are present throughout cells and play a pivotal role in processes such as motility, transport, and growth. Force Sensors respond to these forces, undergoing conformational shifts to facilitate essential cellular functions. To overcome the limitations of experimental techniques, computational methodologies are employed to probe the effect of forces at the molecular level, providing a deeper understanding of Force Sensor functionality. This thesis focuses on two systems where response to force is integral to function: Actin Cytoskeleton and so-called Molecular Springs. For Actin Cytoskeleton, we address a significant challenge by presenting simulation data supporting a multiple-step activation pathway of Arp2/3 complex, a known actin filament nucleator unique for its ability to form branched actin filaments. Shifting to Molecular Springs, we introduce a simulation technique designed to predict the molecular-level effects of mechanical forces. A subsequent study enhances this technique by combining it with another effective enhanced sampling tool. The overarching goal is to predict the behavior of various Molecular Springs, contributing to the engineering of piconewton magnitude cellular and subcellular force sensors.

Contents

Dedication **iii**

Acknowledgments **iv**

Abstract **ix**

List of Figures **xv**

List of Tables **xxxiii**

1 Introduction **1**

1.1 Forces at the Molecular level 1

1.1.1 Thermal Fluctuations: How particles move on their own 1

1.1.2 Effect of force on Thermodynamics and Kinetics 2

1.1.3 The Linear Response Regime 4

1.2 Biomolecular Force-Sensing Paradigms 6

1.2.1 Mechanical Allostery 8

1.2.2 Force-sensitive Binding Kinetics 8

1.2.3 Actin Cytoskeleton and Mechanotransductions 9

1.2.4 Molecular Springs 10

1.3 Experimental Approaches to Measure Forces during Biological Processes 12

1.3.1 Single molecule force spectroscopy . . . 13

1.3.2 Tension Sensor Modules . . . 15

1.4 Computational techniques to study Mechanobiology . . . 18

1.5 Dissertation Outline . . . 21

2 Molecular dynamics simulations support a multi-step Arp2/3 complex activation pathway 23

2.1 Introduction . . . 25

2.2 Results . . . 29

2.2.1 Description of system setup . . . 30

2.2.2 The splayed and short pitch conformations are maintained in microsecond unbiased MD simulations . . . 32

2.2.3 Flattening and adoption of the short pitch conformation are not tightly linked . . . 34

2.2.4 Contacts with the mother filament stabilize flattened Arp3 . . . 37

2.2.5 Twisting of Arp2 closes the barbed end groove and weakens its interactions with the daughter filament . . . 39

2.2.6 Splayed Arp2/3 complex maintains approximately the same interface area with the mother filament as short pitch Arp2/3 complex . . . 41

2.2.7 Flexible segments in Arp2/3 complex maintain conformation-insensitive contacts with the mother filament of actin . . . 44

2.3 Discussion . . . 46

2.4 Supporting Information . . . 50

2.4.1 Simulation Setup . . . 50

2.4.1.1 Construction . . . 50

2.4.1.2 Equilibration . . . 51

2.4.1.3 Production 52
2.4.2 Data Analysis 53
2.4.3 Data Availibility 54
2.4.4 Supplemental Data 54

3 Infinite switch simulated tempering in force (FISST) 61

3.1 Introduction 63
3.2 Theory and Methods 65
3.2.1 Simulations under constant force 65
3.2.2 Simulated Tempering in Force 66
3.2.3 The Infinite Switch Limit 67
3.2.4 Effective Potential 71
3.2.5 Algorithm for learning weights 71
3.3 Results 73
3.3.1 Analytical Potential 73
3.3.2 Beaded Helix 76
3.3.3 Alanine-10 79
3.4 Discussion and conclusions 83
3.5 Simulation details 85
3.5.1 Code and Data Availability 85
3.5.2 Langevin dynamics on test potentials 85
3.5.3 Beaded helix simulations 86
3.5.4 Atomistic molecular dynamics on peptides 86
3.6 Supporting Information 88
3.6.1 V-Shape Potential 88
3.6.2 Alanine-10 89

4 Improved prediction of molecular response to pulling by combining force tempering with replica exchange methods 94

4.1 Introduction . . . 96

4.2 Theory . . . 99

4.2.1 FISST overview . . . 99

4.2.2 Overview of replica exchange methods . . . 101

4.3 Combining FISST with Replica Exchange . . . 102

4.4 Methods . . . 103

4.4.1 System details . . . 104

4.4.2 Production runs . . . 105

4.4.2.1 Overview . . . 105

4.4.2.2 FISST details . . . 105

4.4.2.3 REST3 Simulations . . . 105

4.4.3 Data Analysis . . . 106

4.5 Results and Discussion . . . 106

4.5.1 Simulations of polyalanine validate implementation of hybrid sampling approach . . . 106

4.5.2 Simulations of Aib_9 show improved performance from hybrid FISST+RE sampling . . . 110

4.5.3 Villin (NLE/NLE) mutant simulations allow us to assess performance on a TSM-like molecule . . . 114

4.6 Conclusions . . . 117

4.7 Data Availability . . . 118

4.8 Supporting information . . . 118

4.8.1 Simulation Details . . . 118

4.8.1.1 System Construction . . . 118

4.8.1.2 Minimization 119
4.8.1.3 Equilibration and Production run inputs 119
4.8.1.4 Production runs 120
4.8.2 Alanine Decamer 123
4.8.3 Aib$_9$ 126
4.8.4 Villin (NLE/NLE) Mutant 128

5 Influence of Linker Sequence on Tension Sensor Module Behaviour: Force Induced Structural Insights 129
5.1 Introduction 130
5.2 Preliminary Results 132
5.2.1 Results from CAMPARI simulations 133
5.2.2 Results from GROMACS simulations 136
5.3 Discussion and ongoing work 140
5.4 Supporting Information 142
5.4.1 Simulation setup and data collection 142
5.4.2 Trajectory analysis and visualization 143

6 Concluding Remarks and Future Studies 144
6.1 Future directions in Arp2/3 activation 145
6.1.1 Inactive and Active Arp2/3 free energy barrier 146
6.1.2 Importance of Arp subunit nucleotide state during activation 148
6.2 Future directions for Tension Sensor prediction 149

Bibliography 152

List of Figures

1.1 Energy landscape of a two-state model as a function of an arbitrary 1D coordinate Q with force (red solid line) and without force (black solid line). Q_A, $Q^\dagger$, and Q_B defines state A, the transition state, and state B respectively. The energy barrier between the two states with and without force are denoted by $E_F^\dagger$ and $E^\dagger$ respectively. ΔE_{BA} represents the energy difference between states A and B. (Schematic adapted from Ref. [4]) 2

1.2 **Overview of some Biomolecular Force-sensing Paradigms**. (A) Illustration of Flow activation of GPCR GPR68. Structures shown are derived from homology models taken from inactive and active μ-opioid receptor (PDB: 4DKL [8] and PDB: 5C1M [9] respectively). (B) "Catch" bond behavior seen in Bacterial adhesion FimH with glycolipid sugars on the surface of endothelial cells [10] and "Slip" bond behavior in T-Cell Receptors (TCR) and Major Histocompatibility Complex (MHC) on the surface of antigens [11, 12]. (C) Actin nucleation by Formin dimer [13]. (D) Different Molecular springs and their force-sensitivities. These Molecular Springs can be engineered into Tension Sensor Modules (TSMs) for cellular and sub-cellular force measurement [6, 14–16]. 7

1.3 Overview of actin filament formation denoting the rate-limiting nucleation step, elongation, and stationary phase. (Schematic adapted from Ref. [26]). 10

1.4 **Properties of a Tension Sensor molecule**. (A) Force-extension curve of Wild-type Villin-headpiece domain (PDB: 1YRF [30]), a well-documented Tension Sensor molecule that begins to unfold between 6-8 pN of force [31]. (B) Comparison of unfolding force (pN) at different loading rates (pN/s) for a Tension Sensor molecule and Non Tension Sensor molecule (Schematic adapted from Ref. [15]). (C) Unfolding-folding cycles demonstrating lack of hysteresis in a Tension Sensor molecule as opposed to a Non-Tension Sensor molecule (Schematic adapted from Ref. [15]). 11

1.5 **Overview of Single Molecule Force Spectroscopy Techniques**. Schematics for (A) Optical Tweezers, (B) Magnetic Tweezers, and (C) Atomic Force Microscopy. (Schematics adapted from Ref. [32]). 12

1.6 Calibrating a FRET based Tension Sensor Module (TSM) using optical tweezers. Λ_{exc} (orange) denotes the light at the excitation wavelength specific to the Donor fluorophore (yellow). Force-distance data acquired through optical tweezers, combined with simultaneous FRET responses at corresponding distances, establishes a direct correlation between FRET response and applied force (Schematic adapted from Ref. [15]). (Inset) Jablonski diagrams illustrating the ground and excited states of the Donor (yellow) and Acceptor (cyan) fluorophores. Solid arrows indicate radiative processes (Absorption, Fluorescence), while dashed arrows represent non-radiative processes (Vibrational Relaxation, FRET) (Schematic adapted from Ref. [44]). 17

2.1 Simplified schematics of the conformational pathway to activation of Arp2/3 complex in a concerted versus multistep model of activation. Previous data indicate that while WASP triggers the splayed to short pitch conformational change, both states can exist with or without WASP bound to the complex [93, 94]. For clarity, neither these conformational states nor their reversibility is depicted here. For a more detailed diagram of the two proposed mechanisms that includes these states, see Fig. 2.16. Text boxes to the right of each scheme list the key features of each mechanism. The splayed/flattened state (marked with red question mark) may not be adopted because of steric clash (see Discussion). 29

2.2 Starting structures used for each of the four unbiased all-atom MD simulations described here. The PDB file used to build each structure is indicated in the lower right corner. Arp2/3 complex, actin-related protein 2/3 complex; MD, molecular dynamics. 30

2.3 (A) Plot of the distance of the center of geometry (COG) of subdomains 3 and 4 of Arp2 to the COG of subdomain 3 and 4 of Arp3 as a function of simulation time. Dashed and dotted lines show the corresponding distance in the branch junction structure, 7TPT, and the inactive Arp2/3 complex structure, 4JD2, respectively. Data for simulations are shown as average smoothed over 5 ns (50 frames) for this and all other plots from the unbiased simulations. The shaded area shows the standard deviation over the smoothing window. (B) Surface representation of Arp3 from inactive (4JD2) or active structure (7TPT) showing average contact scores over the entire trajectory for Arp3 residues that contact Arp2. Contact scores for 4JD2 and 7TPT are shown on the left for reference. Contact scores were calculated using PyContact, as described in the Supporting Information. . . 32

2.4 (left) Backbone trace of Arp2 showing measurement of the twisting/flattening angle. Subdomains of Arp2 are labeled 1 to 4. (right) Plot of Arp2 and Arp3 twisting/flattening angle (φ) versus simulation time for the free Arp2/3 complex from branch junction simulation. Arp2 and Arp3 twisting/flattening angles from inactive (4JD2) and the active Arp2/3 complex structure (7TPT) are shown as dashed or dotted lines, as indicated. 34

2.5 (A) Plot of the Arp3 subunit twisting/flattening dihedral versus the Arp2-Arp3 COG, which measures movement into the short pitch conformation. Enclosed regions indicate the most probable conformations in the simulation, as defined by conformations that are within a radius of one free energy unit from the lowest energy conformation. Circles show the corresponding measurements for selected active and inactive cryo-EM or X-ray crystal structures. (B) Plot as described in (A), except the Arp2 twisting/flattening dihedral angle is plotted on the x-axis. . . 34

2.6 Ribbon diagram of the free Arp2/3 complex from branch junction simulation output at 1 μs showing that the Arp2 D-loop maintains contact with ARPC3 even when the complex moves into a short-pitch, twisted conformation. The D-loop of Arp2 is highlighted in yellow. The distance between the globular domain of ARPC3 and subdomains 1 and 2 of Arp2 is indicated with a black line. Inset shows a zoomed in view of the interaction. 35

2.7 (A) Plot showing the distance between Arp2 D-loop and ARPC3 for Arp2/3 complex (active) and (inactive) simulations versus simulation time, with 7TPT and 4JD2 plotted for reference. (B) plot showing the distance between the COGs of $\text{Arp2}_{(\text{Sub1\&2})}$ and ARPC3 versus simulation time in the Arp2/3 complex active simulation. Arp2/3 complex, actin-related protein 2/3 complex; cryo-EM, cryo-electron microscopy; COG, center of geometry. 35

2.8 (A) Diagram showing the twisting/flattening angle (φ) of Arp3. The four subdomains of Arp3 are labeled 1 to 4. (B) plot of Arp3 twisting/flattening dihedral angle versus simulation time for all unbiased simulations that started in the active conformation. Arp3 twisting/flattening angles from inactive (4JD2) and active Arp2/3 complex structures (7TPT) are shown in dotted or dashed lines, as indicated. . . . 37

2.9 (A) Plot of the surface area of subdomain 4 of Arp3 buried on the mother filament versus simulation time. Buried surface area of subdomain 4 in the branch junction structure (7TPT) and a model of 4JD2 on an actin filament are shown as dotted or dashed lines. (B) Ribbon and surface representation of the last frame of the branch junction without daughter filament simulation showing residues within actin filament that interact with Arp2/3 complex upon subunit flattening in Arp3. Actin filament residues that interact with ARPC3 or Arp3^{Sub4} in the MF-bound Arp2/3 complex simulation (average contact score > 1, colored green) are mapped onto the surface of the actin filament. Subdomain 4 of Arp3 is labeled. The Arp3 dihedral angle that flattens Arp3 is shown as angle φ 38

2.10 (A) Ribbon diagram showing the interaction of the D-loop of actin D2 with the barbed end groove of Arp2. Arp2 and actin D2 subdomains are labeled 1 to 4. PE: pointed end, BE: barbed end. Right panel shows closeup of the interaction with distances measured in (B). Actin D2 from 0.67 μs (transparent light blue ribbon) in the branch junction simulation was placed by overlaying Arp2 from the 0.67 μs frame in the trajectory with Arp2 from 7TPT. BEG: Barbed-end groove. Start: position of actin D-loop at the beginning of the simulation. (B) Plot of twisting/flattening angle of Arp2 (φ) in the branch junction and the branch junction without daughter filament (no daughter) simulations. Arp2 twisting/flattening angles from inactive (4JD2) and active Arp2/3 complex structures (7TPT) are shown in dotted or dashed lines, as indicated. (C) surface representation of branch junction model (7TPT) showing the interface between Arp2 and Arp3 and the pointed end of the nucleated daughter filament. 39

2.11 (A) Plot of the W-loop opening (x_1) and D-loop to Arp2 distance (x_2) in the branch junction simulation. See Fig. 2.10A for definition of x_1 and x_2. Distances x_1 and x_2 in the branch junction structure are plotted for reference. (B) Plot of W-loop distance (x_1) in Arp2 subunits for all unbiased simulations. Distance x_1 in the branch junction (7TPT) and inactive Arp2/3 complex structure (4JD2) is plotted as dashed or dotted lines, as indicated. (C) Identical to (B), except x_1 for Arp3 from each simulation is plotted. 40

2.12 (A) Ribbon diagram of inactive Arp2/3 complex (4JD2) showing the center of geometry (COG) of each subunit. For the steered MD simulations, a bias was applied to move the COGs from their positions in the active complex to their positions in the inactive (4JD2) complex. (B) Plot of the distance between the Arp2 and Arp3 subdomains 3 and 4 COGs as a function of simulation time. The same distances for the inactive (4JD2) and branch junction (7TPT) structures are plotted as dashed and dotted lines, as indicated. (C) Plot of the total interaction area of Arp2/3 complex with the mother filament in all three steered simulations. Data for simulations are shown as an average smoothed over 1 ns (10 frames) for this and all other plots for the steered simulations. The standard deviation over the smoothing window is shaded. Interaction surface area in the branch junction (7TPT) is plotted as a dashed line for reference. 42

2.13 (A) (Left panel) surface representation of Arp2/3 complex rendered using the starting coordinates of the 100 ns pulling simulation. The four rigid blocks that move independently when the complex undergoes subunit flattening (see Video S1) are outlined with red dashes. The “top” and “bottom” rigid blocks that move independently when the clamp twists are indicated with gray boxes behind the complex. Residues of Arp2/3 complex that contact the mother filament at the start of the simulation (PyContact calculated contact score >1) are colored gray. (Right panel) Same as left panel except surface representation is rendered from the final frame of the simulation and residues that have an average contact score > 1 over the last 1 ns of the simulation are colored gray. Arp2/3 complex is in the splayed conformation at the end of the simulation. (B) Comparison of mother filament binding contacts of activated (flattened, short pitch) Arp2/3 complex (7TPT) to those of the splayed Arp2/3 complex (final frame of MF-bound 150 ns pulling simulation—rendered in gray ribbon or transparent gray surface). The splayed Arp2/3 complex was modeled onto the mother filament by superposing block 1 onto block 1 in the branch junction model. Yellow arrow shows movement of blocks 2 and 4 stimulated by clamp twisting. Block 3 is omitted for clarity. BE: Barbed end of mother filament. PE: pointed end of mother filament. 43

2.14 (A) Plot of the interaction area of ARPC1 residues 287 to 326 with the mother filament versus simulation time for all pulling simulations. Dashed line shows the corresponding interaction areas in the branch junction structure 7TPT. (B) Surface and cartoon representation of Arp2/3 complex bound to the mother filament in the last frame of the 100 ns pulling simulation. The two flexible segments from the bottom half of the complex (ARPC1 287–326, green, and ARPC2 281–300, cyan) are shown in thicker cartoon representation. (C) Sequence alignment of the ARPC1 insert sequence from a diverse range of species showing conserved hydrophobic (green) and acidic (red) residues. The average contact score over the course of all three pulling simulations is plotted above the sequence for each residue. Error bars: standard deviation. 45

2.15 (A) Plot of the interaction area of ARPC2 residues 281 to 300 with the mother filament versus the simulation time for all pulling simulations. (B) Sequence alignment of ARPC2 C-terminal extension from a diverse range of species showing conserved hydrophobic (green), basic (red), or hydrophilic (cyan) residues. The average contact score over the course of the three pulling simulations is plotted above the sequence for each residue. Error bars: standard deviation. (F) Plot of the root mean squared fluctuation (RMSF) from the initial conformation for backbone atoms of the ARPC1 insert (green) or the ARPC2 C terminus (cyan) over the entire simulation, plotted separately for each pulling simulation. Black dashed rectangle highlights RMSF values for residues in ARPC2 C-terminal extension with the closest contacts to the mother filament. Yellow rectangle highlights RMSF values for residues in the ARPC1 insert with the closest contacts to the mother filament. 45

2.16 These schematics show the relationship between conformation and binding state in more detail than in Fig. 2.1. Key additions relative to Fig. 2.1 are: the short pitch conformation can be weakly adopted even in the absence of WASP [77, 78], Arp2/3 complex can adopt the short pitch or splayed conformation when bound to WASP [77, 78], and Arp2/3 complex from some species shows NPF-independent activity [79]. It is unknown (red question mark, concerted model) whether the NPF-independent activity of Arp2/3 complex requires preformed filaments, though filaments activate short-pitch crosslinked S. cerevisiae Arp2/3 complex [80]. Neither schematic shows individual WASP binding steps. Biochemical data indicate WASP binds more tightly to the Arp2 site [81], and that there may be differences in the contribution of WASP binding at each site to stimulating the short pitch conformation [78]. The splayed/flattened state (red question mark, multi-state model) may not be adopted because of steric clash (see Discussion). 57

2.17 Ribbon representation of Arp3 (orange) and Arp2 (red) from the branch junction structure (7TPT). Arp2 from the MF-bound Arp2/3 complex simulation at 0.76 μs is shown in pink and is positioned by overlaying the Arp3 backbone from the simulation onto Arp3 from the branch junction structure. The centers of geometry of subdomains 3 and 4 of Arp3 and Arp2 from each of the structures are shown as spheres and connected with blue (branch junction) or cyan (simulation) dashed lines. The yellow region shows where actin D1 would clash with Arp2 in the tilted conformation observed in the simulation. 58

2.18 Plot of the area of subdomain 4 of Arp3 and ARPC3 buried upon interaction with the mother filament versus simulation time. Buried surface area of the same regions in the branch junction structure (7TPT) and a model of inactive Arp2/3 complex (4JD2) on an actin filament are shown as dashed or dotted lines, respectively. 58

2.19 (A) Ribbon diagram of Arp2/3 complex from the branch junction structure and from the 60 ns pulling simulation showing the distances between subdomains 3 and 4 of Arp3 with subdomains 3 and 4 of Arp2. This distance measures movement of the complex into the short pitch position. The approximate position of the mother filament is shown in shaded grey. MF: Mother Filament. (B) Ribbon representation of Arp2 and Arp3 from the branch junction structure with the clamp subunits shown as semi-transparent surface. The dihedral angle (α) used to measure clamp twisting is shown in yellow. Residues used to define the centers of geometry marked a, b, c, and d are listed in the Supplementary Materials. (C) Plot of the clamp twisting angle (α) versus simulation time for all three pulling simulations. The clamp twisting angle for the branch junction structure (7TPT) and for an inactive Arp2/3 complex structure (4JD2) are shown as dashed lines. (D) (Left panel) Surface representation of Arp2/3 complex rendered using the starting coordinates of the MF-bound Arp2/3 complex 100 ns pulling simulation. The four rigid body blocks that move independently when the complex undergoes subunit flattening and clamp twisting (see Video S1) are indicated with red dashes. The "top" and "bottom" rigid body blocks that move independently when the clamp twists are indicated with grey boxes behind the complex. Residues of Arp2/3 complex that contact the mother filament at the start of the simulation (PyContact calculated contact score >1) are colored grey. (Right panel) Same as left panel except surface representation is rendered from the final frame of the simulations and residues that have an average contact score > 1 over the last 1 ns of the simulations are colored grey. Note that the complex is in the splayed conformation at the end of the simulations. 59

2.20 (A) Cartoon representation of Arp2/3 complex modeled into the short pitch, twisted conformation. Arp3 was constructed by superposing the backbone atoms of (well-ordered) residues in subdomains 1 and 2 of Arp3 from the inactive structure (4JD2, Arp3:6-32,78-153,375-408,33-37,60-77) with the same atoms in the branch junction structure (7TPT). Subdomains 3 and 4 from the superposed inactive structure along with ARPC3 were then used to replace the corresponding residues in the branch junction structure. Arp2 was constructed by superposing the backbone atoms of well-ordered residues in subdomains 3 and 4 of Arp2 in the inactive structure (4JD2, Arp2: 186-265, 151-185, 266-326,339-351) with the same atoms in the branch junction structure (7TPT). Subdomains 1 and 2 from the superposed inactive structure were then used to replace the corresponding residues in the branch junction structure. The dihedral angles that measure subunit twisting/flattening in the Arps are shown in cyan. The distance between the center of geometry of subdomains 3 and 4 of Arp3 and subdomains 3 and 4 of Arp2 is shown in yellow. (B) Cartoon representation of Arp2/3 complex modeled into the splayed and flattened conformation. Arp3 in this model was constructed by superposing the backbone atoms of (well-ordered) residues in subdomains 1 and 2 of Arp3 from the branch junction structure with the same atoms in the inactive structure (4JD2). Arp2 in this model was constructed by superposing the backbone atoms of well-ordered residues in subdomains 3 and 4 of Arp2 with the same atoms in the branch junction structure (7TPT). (C) Table showing the number of clashes in the experimentally determined versus modeled structures. Clashes were identified with the clash command in ChimeraX [82]. 60

3.1 (a) Contour plot of the potential (Eq. 3.31) used for Langevin dynamics. (b) Estimated effective potential (U_{eff}, Sec. 3.2.4) sampled by FISST using a force range of [-15:15] applied to $Q = y$. (c) Failure of running a simulation at a single force of $F = -15$, where the position is stuck in the right minimum for the entirety of the simulation. (d) Reweighted FISST data predicts free energy surfaces at -15, -7.5, 0, 7.5, and 15. In all cases (a)-(d), the minimum value of the potential is shifted to be zero. 74

3.2 (a) Left and right handed helices used as reference structures to quantify helicity. (b) Estimated free-energy surface for end-to-end distance and helicity sampled from single force simulations (top) and reweighted from a single FISST simulation (bottom), for forces of F = 0, 3, 4.5, and 8 ($k_B T/\sigma$). In this case, probability histograms were computed by a two-dimensional Gaussian kernel-density-estimate as implemented in `scipy` [182]. For all cases, the lowest free energy is shifted to zero. 76

3.3 (a) Probability density function of end-to-end distances at $F = -5$, 0, and 5 pN calculated from replica exchange, single force, and FISST simulations. (b) Jensen-Shannon distances as a function of force for single force and FISST simulations. The reference density is determined from 5 replica exchange simulations at -10, -5, 0, 5, and 10 pN and interpolated to desired force using EMUS [184]. 80

3.4 Ramachandran plots of alanine-10 peptide at 0, 33, 67, and 100 pN applied force. All plot were generated by reweighting data from a single 160 ns FISST simulation using a force range of [0:100] pN. The regions with the most significant change in free-energy upon upon pulling are the α/δ area at (-50,0), γ at (75,150), PPII at (-75,150) and β at (-140,150), using the nomenclature from Ref. [185]. 81

3.5 (a) Error of FISST [-15:15] and individual single force simulations run relative to the exact density (Eq. 3.4), using fractions of total work. In the 1/20 data set the error is very high compared to FISST despite the number of total steps across all forces being equal to that of the full FISST trajectory. Simulations on the order of the same length of FISST are required at each single force to replicate the same level of accuracy. (b) Average error relative to exact density over 5 replicates for FISST [-15:15] and single force simulations run at 20 different forces, each with 1/20th the number of steps. 88

3.6 Probability density of end-to-end distances at F = -10, -5, 0, 5, and 10 pN calculated from replica exchange, single force, and FISST simulations. 90

3.7 Relative error of single force simulations compared to FISST vs amount of computational work, as explained in the text above. The dashed line shows accuracy equal to FISST. 92

3.8 Ramachandran plots of alanine-10 peptide at 0, 33, 67, and 100 pN applied force, each simulated at a single applied force. 92

3.9 Ramachandran plots of alanine-10 peptide at 0 and 100 pN applied force, each calculated from Replica Exchange simulations. 93

3.10 Distribution of weights at various stages of a 160 ns deca-alanine simulation. (a) At short times, the weight distribution changes significantly. (b) At longer times, the weight distribution is essentially constant for the remainder of the simulation. 93

3.11 Error in FISST for a range of learning rates. A 160 ns FISST simulation was run with the weights updated every 100, 200, 300, 400, and 500 steps. We calculate the end-to-end distance probability densities at each force and calculate the error against results from replica exchange simulations. In the range considered, there is no obvious dependence on learning rate. An update period of 200 was used for all data discussed in the main text. 93

4.1 Systems probed in this study, shown without solvent for clarity. (A) Solvated alanine decamer starting in the extended state. (B) Solvated Aib$_9$ molecule starting from the left-handed helical state. Each residue is colored according to the residue ID number. (C) Solvated villin (NLE/NLE) mutant starting in the folded state. Locations of residue mutations are colored in ochre. In all cases, pulling forces are applied to the terminal C$_\alpha$ atoms. 99

4.2 (A) End-end distance distributions for Ala$_{10}$ at $F = 0$ for TRE (black solid line), FISST (blue spheres), FISST+TRE without freezing weights (red spheres), and FISST+TRE with freezing (gray spheres). (B) End-end distributions for F =-10, -5, 5, and 10 pN. (C) Comparison of free energies (see Fig. 4.7) computed from end-end distribution functions, comparing FISST+TRE with and without freezing to corresponding reference TRE data at -10, -5, 0, 5, 10 pN forces. 107

4.3 (A) (left) TRE and FISST+TRE Ramachandran plot at zero force and (right) scatter plot comparing FISST+TRE and TRE free energies at zero force. (B) Ramachandran plots at -10, -5, 5, and 10 pN forces for FISST+TRE (top row) and FISST+REST3 (bottom row). 109

4.4 (A) $F(\zeta')$ at zero force calculated from REST3 (black solid line), FISST (red spheres), and FISST+REST3 (gray spheres). (B) RMSE of $F(\zeta')$ for values below 6 $k_{\mathrm{B}}T$ using data points from different simulation time windows. 111

4.5 (A) Theoretical Force vs average End-to-end distance $\langle d_{End} \rangle$ curve for Aib$_9$ calculated from FISST (red dashed line) and FISST+REST3 (gray dashed line). End-to-end distance values calculated from REST3 simulations at 0, 10, and 20 pN forces (black), and unbiased MD (blue) are embedded for comparison. (B) Jenson-Shannon distance of $P(d_{End})$ calculated for FISST (red) and FISST+REST3 (gray) for 0, 10, 20 pN forces using REST3 simulation data as the reference. 113

4.6 (A) Villin Mutant Force vs average End-to-end distance $\langle d_{End} \rangle$ calculated for FISST + REST3 without freezing the weights and freezing the weights at 298K. $\langle d_{End} \rangle$ values calculated from REST3 simulations at $F = 10$ and 20 pN. Error bars represent 1/3 of the standard deviation in length at that force. (B). Similar data as (A) for simulations where the lowest replica is at $T = 360K$, which is close to the melting temperature. Also shown is data from a 310 μs trajectory at $T = 360K$ from Ref. [229]. 114

4.7 Alanine Decamer Free energy profiles for TRE (black solid line), FISST+TRE without freezing weights (red solid line), and FISST+TRE with freezing weights (gray solid line) calculated from the End-to-end distance probability distributions at -10, -5, 0, 5, and 10 pN forces shown in Fig. 4.2A and Fig. 4.2B. 123

4.8 (A) Alanine decamer End-to-end distance probability distribution functions calculated for TRE (black solid line), FISST (blue spheres), REST3 (orange spheres), FISST+REST3 without freezing the weights (red spheres), and FISST+REST3 with freezing the weights (gray spheres) at -10, 0, and 10 pN forces. (B) (left) Corresponding free energy profiles. (right) Free energy scatter plots comparing FISST+TRE and TRE data without and with freezing of the weights. 124

4.9 (left to right) Ramachandran plots calculated from REST3, FISST+REST3 without freezing the weights, FISST+REST3 with freezing the weights and corresponding free energy scatter plots comparing FISST+REST3 without and with freezing of the weights for (A) -10, (B) 0, and (C) 10 pN forces. 125

4.10 (left) Time series plot for ζ' coordinate generated from unbiased MD simulation of Aib$_9$. (right) Corresponding free energy profile of ζ'. (inset) Left and right-handed Aib$_9$ helices are marked in their respective basins. 126

4.11 Snapshots from Aib$_9$ FISST+REST3 trajectory depicting helical compositions at different forces and solute temperature replicas. 126

4.12 RMSE of $F(\zeta')$ at zero force calculated from full free energy profiles from REST3 (black), FISST (red), and FISST+REST3 (gray) using data points from different simulation time windows. 127

4.13 Villin Mutant End-to-end distance probability distribution reweighted to forces in the range [-10 pN:20 pN] from FISST+REST3 simulations at (A) Solute temperature of 298K and (B) Solute temperature of 360K. The color bar distinguishes the forces to which the distributions are reweighted. 128

5.1 Linker peptide repeats probed for MC and MD studies, plotted according to their percentage of Serine and Proline to illustrate their effect on linker stiffness. The color scheme for amino acid residues are Alanine: Blue, Glycine: White, Proline: Brown, Serine: Yellow. 132

5.2 End-to-end Distance probability distribution functions for each Linker CAMPARI simulation at zero force along with an embedded snapshot of 100 configurations across all frames. For each subplot, the blue histograms are calculated from roughly 25000 frames for each simulation, the orange solid line denotes the Worm Like Chain model fit described by the model in Ref. [255], and the black dashed line represents the linker contour length calculated by multiplying the approximate length of each amino acid with the total number of residues in each peptide. For each embedded snapshot, Proline residues are colored in blue, serine residues in yellow, and the rest of the peptide in cyan. 135

5.3 Linker Ramachandran plots generated from CAMPARI simulation for zero force at 298K (top row) and 310K (bottom row). 136

5.4 Mean z/L_c calculated for each linker trajectory, with error bars representing the standard deviation. The linker contour length L_c was calculated by multiplying the approximate length of each amino acid with the total number of residues in each peptide. Simulation data is from GROMACS MD runs at 0, 5, 10, 15, and 20 pN forces for 298K (blue) and 310K (red). The black dashed line represents Worm Like Chain (WLC) Model fitting by iteratively solving Marko-Sigio's equation [257] by defining an initial Persistence Length L_p for each linker. 137

5.5 Linker contour plots for ϕ and ψ dihedrals calculated from GROMACS MD runs for 0 (red solid line), 5 (blue solid line), 10 (green solid line), 15 (cyan solid line), and 20 pN (magenta solid line) forces at 298K and 310K. Enclosed regions represent conformations within one free energy unit of the lowest energy configuration at each force. 138

5.6 Mean PolyProline II content calculated for each linker trajectory as a percentage, with error bars representing the standard deviation. Simulation data is from GROMACS MD runs at 0, 5, 10, 15, and 20 pN forces for 298K (blue) and 310K (red). 139

6.1 2D and 1D free energy plots calculated from umbrella simulations for Active Arp2/3 Mother filament bound 150 ns pulling trajectory (top) and Free Active Arp2/3 150 ns pulling trajectory (bottom). Morph structures (black spheres) were generated from PyMOL [262] using PDB: 7TPT and PDB: 4JD2 as the initial and final structures respectively. 147

6.2 Contour plots for Arp2-Arp3 COG distance vs φ_{Arp2} and Arp2-Arp3 COG distance vs φ_{Arp3} from Active Arp2/3 Mother filament bound (top) and Free Active Arp2/3 Complex (bottom) simulations. Enclosed regions represent configurations within one free energy unit of the lowest energy conformation for each nucleotide state: ADP (blue solid liner) and ATP (gray solid line). 149

List of Tables

2.1 Definition of subdomains 1-4 in Bos taurus Arp3 and Arp2. Backbone atoms in the listed residues were used for center of geometry calculations. 56

2.2 Definition of centers of geometry in Bos taurus Arp3, ARPC2 and ARPC4 used for calculating the clamp twist angle. Backbone atoms in the listed residues were used for center of geometry calculations. 56

4.1 Overivew of system construction details for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35). We have also included system setup details for Villin (NLE/NLE) mutant described in Ref. [229], which we use as a reference for our results in Sec. 4.5.3. 118

4.2 System minimization details provided for Alanine decamer (Ala_{10}) and Villin (NLE/NLE) mutant (HP35). No entries for Left-handed AIB9 helix (Aib_9) have been provided as the equilibrated inputs were directly obtained from the authors of Ref. [224]. . 119

4.3 Equilibration and Production run parameters used for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35). 119

4.4 Summary of all single-process simulations with FISST. Force range and total simulation time collected provided for Alanine decamer (Ala_{10}) and Left-handed AIB9 helix (Aib_9). For Aib_9, weights were updated every 500 steps (1 ps) and both observable and restart data were also saved every 500 steps. For Ala_{10}, the weights were updated every 1000 steps (2 ps), and the observable data and restart data were also saved for the same number of steps. 120

4.5 Summary of all REST3 simulations without FISST implemented. Forces used, Replica Exchange setup, and total simulation time collected provided for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35). 120

4.6 Summary of all FISST+REST3 simulations without FISST implemented. Force ranges, Replica Exchange setup, and total simulation time collected provided for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35). We also specified if we ran additional simulations with the updated PLUMED source code, which freezes the FISST weights. For Aib_9, weights were updated every 500 steps (1 ps) and both observable and restart data were also saved every 500 steps. For Ala_{10} and HP35, the weights were updated every 1000 steps (2 ps), and the observable data and restart data were also saved for the same number of steps. 121

4.7 Summary of Temperature Replica Exchange (TRE) performed for Alanine decamer (Ala_{10}). Forces used, Replica Exchange setup, and total simulation time collected are provided. 121

4.8 Summary of FISST+TRE runs performed for Alanine decamer (Ala_{10}). Forces used, Replica Exchange setup, and total simulation time collected are provided. For Ala_{10}, the weights were updated every 1000 steps (2 ps), and the observable data and restart data were also saved for the same number of steps. 122

4.9 Summary of total simulation and wall times of all Aib$_9$ runs. 127

5.1 CAMPARI Monte Carlo Moveset parameters for implicit solvent simulations with ABSINTH forcefield. 143

CHAPTER 1

INTRODUCTION

1.1 FORCES AT THE MOLECULAR LEVEL

1.1.1 THERMAL FLUCTUATIONS: HOW PARTICLES MOVE ON THEIR OWN

In Chemistry, particles are in a state of motion that brings about different interactions and hence the formation of new compounds with a unique set of physical and chemical properties. Consider the simple reaction of adding a strip of Magnesium metal to a beaker of Hydrochloric acid

$$\mathrm{Mg\,(s) + 2\,HCl\,(aq) \longrightarrow MgCl_2(aq) + H_2(g)\uparrow}$$

According to *Collision Theory* [1], these particles must collide frequently with sufficient energy to break old chemical bonds and form new ones for a successful chemical reaction to occur. Even in the absence of any external factors, how are these particles able to move? The answer lies in understanding the *Thermal Fluctuations* particles are experiencing at the molecular level. Statistical Mechanics allows us to utilize concepts from statistics and mathematics to predict how an enormous number of particles at this molecular scale gives rise to the visible phenomena and properties scientists are keen on studying and also quantify the thermal fluctuations these particles experience [2]. While not the typical forces we are familiar with, we may convert these thermal fluctuations from molecular-scale energies in units of $k_{\mathrm{B}}T$ (k_{B} is the Boltzmann constant and T is the temperature in Kelvin) to units of force times distance: $k_{\mathrm{B}}T \sim 4.1$ pN · nm at room temperature (298K) and hence draw connections with the forces experienced and displacements associated with larger molecular machines, such as those inside cells (see below) [3, 4].

1.1.2 Effect of force on Thermodynamics and Kinetics

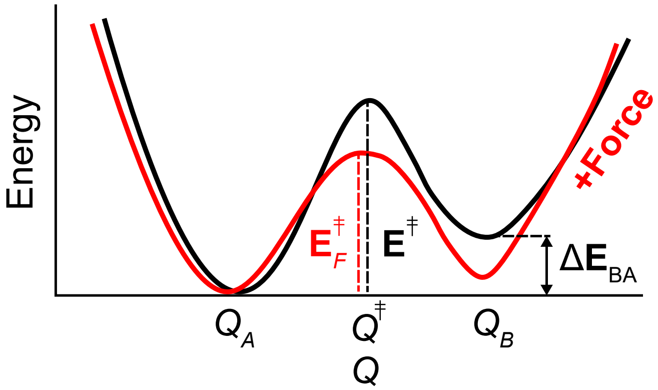


Figure 1.1: Energy landscape of a two-state model as a function of an arbitrary 1D coordinate Q with force (red solid line) and without force (black solid line). Q_A, $Q^{\dagger}$, and Q_B defines state A, the transition state, and state B respectively. The energy barrier between the two states with and without force are denoted by $E_F^{\dagger}$ and $E^{\dagger}$ respectively. ΔE_{BA} represents the energy difference between states A and B. (Schematic adapted from Ref. [4])

To better understand the consequences of external forces at the molecular level; we begin with the canonical Boltzmann distribution [2, 5]

$$P(\boldsymbol{X}) = \frac{e^{-\beta U(\boldsymbol{X})}}{\int d\boldsymbol{X}\, e^{-\beta U(\boldsymbol{X})}} \tag{1.1}$$

Where $\beta = 1/k_{\mathrm{B}}T$ is the inverse temperature and $U(\boldsymbol{X})$ is the potential energy of configuration $\boldsymbol{X}$ as described by the complete set of atomic coordinates.

Now consider the following expression of the Boltzmann distribution [3, 4]

$$P(\boldsymbol{X}, \boldsymbol{F}) = \frac{e^{-\beta U(\boldsymbol{X}) + \beta \boldsymbol{F}.\boldsymbol{Q}(\boldsymbol{X})}}{\int d\boldsymbol{X} \; e^{-\beta U(\boldsymbol{X}) + \beta \boldsymbol{F} \cdot \boldsymbol{Q}(\boldsymbol{q})}} \tag{1.2}$$

In Eq. 1.2 we have introduced an additional term in the exponent $-\boldsymbol{F} \cdot \boldsymbol{Q}(\boldsymbol{X})$; which represents the change in energy corresponding to mechanical work. Here, $\boldsymbol{F}$ is a vector for the constant applied force and $\boldsymbol{Q}(\boldsymbol{X})$ is the Collective Variable, which is a vector-valued function described by the set of the atomic coordinates to which $\boldsymbol{F}$ is being applied.

For simplicity, we will consider the case for a one-dimensional free energy landscape consisting of two different states A and B, with and without force (Fig. 1.1). We will also assume that the applied force is not sufficient to alter the shape or position of the free energy landscape. For a 1D Free Energy landscape, it can shown that up to an additive constant this external force "tilts" the free energy landscape as follows [3, 4]

$$A(X) = -k_{\mathrm{B}} T \ln P(X, F) = U(X) - FQ \tag{1.3}$$

The "tilting" from the applied force has a few consequences on the thermodynamics and kinetics of the system.

THERMODYNAMIC EFFECT. The equilibrium constant K_{eq} favors the direction of the applied force. We can express K_{eq} without force as a ratio of the probabilities of being in state B to state A as follows

$$K_{eq}(F = 0) = \frac{P(B)}{P(A)} = e^{-\beta \Delta E_{BA}} \tag{1.4}$$

Under an applied force, it can be shown that K_{eq} scales exponentially [3, 4]

$$K_{eq}(F) = K_{eq}(F = 0)e^{\beta F(Q_B - Q_A)} \tag{1.5}$$

From Eq. 1.5 we see that for $Q_B > Q_A$ an increase in F shifts equilibrium to state B.

Kinetic effect. The transition rate $k_{A\to B}$ favors the direction of the applied force. From Arrhenius kinetics, we write

$$k_{A\to B}(F = 0) = Ae^{-\beta E^\dagger} = Ae^{-\beta(U(Q^\dagger) - U(Q_A))} \tag{1.6}$$

Similar to Eq. 1.5 we can also show that $k_{A\to B}$ scales exponentially under applied force, assuming the positions of Q_A and $Q^\dagger$ do not change [3, 4]

$$k_{A\to B}(F) = k_{A\to B}(F = 0)e^{\beta F(Q^\dagger - Q_A)} \tag{1.7}$$

1.1.3 The Linear Response Regime

From Eq. 1.2, we can also predict an observable of interest $\boldsymbol{O}(\boldsymbol{X})$, such as the backbone dihedrals of a protein under the effect of force by performing the following ensemble average

$$\langle \boldsymbol{O} \rangle_F = \frac{\int d\boldsymbol{X}\, \boldsymbol{O}(\boldsymbol{X}) e^{-\beta U(\boldsymbol{X}) + \beta \boldsymbol{F}\cdot \boldsymbol{Q}(\boldsymbol{X})}}{\int d\boldsymbol{X}\, e^{-\beta U(\boldsymbol{X}) + \beta \boldsymbol{F}\cdot \boldsymbol{Q}(\boldsymbol{X})}} \tag{1.8}$$

Where $\langle ... \rangle_F$ represents an ensemble average at that force. The fluctuation $\boldsymbol{O}(\boldsymbol{X})$ experiences from the applied force is given by $\langle \delta \boldsymbol{O} \rangle_F = \langle \boldsymbol{O} \rangle_F - \langle \boldsymbol{O} \rangle_0$ [3, 4].

We have asserted that forces at the molecular level are on the magnitude of piconewtons, but how do piconewton forces affect molecular configurations? Consider Eq. 1.2. We can modify the exponential term by introducing $\delta \boldsymbol{Q}(\boldsymbol{X}) = \boldsymbol{Q}(\boldsymbol{X}) - \boldsymbol{Q}_0(\boldsymbol{X})$, which is the fluctuation of $Q(\boldsymbol{X})$ as a result of the applied force and compute the ensemble average of $\delta \boldsymbol{Q}(\boldsymbol{X})$ using Eq. 1.8

$$\langle \delta \boldsymbol{Q} \rangle_F = \frac{\int d\boldsymbol{X}\ \delta \boldsymbol{Q}(\boldsymbol{X}) e^{-\beta U(\boldsymbol{X}) + \beta \boldsymbol{F} \cdot \delta \boldsymbol{Q}(\boldsymbol{X})}}{\int d\boldsymbol{X}\ e^{-\beta U(\boldsymbol{X}) + \beta \boldsymbol{F} \cdot \delta \boldsymbol{Q}(\boldsymbol{X})}} \tag{1.9}$$

Expanding Eq. 1.9 to 2nd order terms simplifies to [3, 4]

$$\langle \delta \boldsymbol{Q} \rangle_F = \beta \boldsymbol{F} \sigma_Q^2 \tag{1.10}$$

Where $\sigma_Q^2 = \langle \delta \boldsymbol{Q} \delta \boldsymbol{Q} \rangle_0$ is the variance of $\boldsymbol{Q}(\boldsymbol{X})$ at zero force. Eq. 1.10 tells us that $\boldsymbol{Q}(\boldsymbol{X})$ varies linearly with $\boldsymbol{F}$ and hence termed *the linear response regime*. We can determine which forces satisfy this regime by asserting a margin error (say <10%) with respect to the equilibrium standard deviation as follows [3, 4]

$$\begin{aligned} \frac{\langle \delta \boldsymbol{Q}(\boldsymbol{X}) \rangle_F}{\sigma_Q} &= \beta F \sigma_Q < 0.1 \\ \implies F &< \frac{0.41\ \text{pN} \cdot \text{nm}}{\sigma_Q} \end{aligned} \tag{1.11}$$

Where in Eq. 1.11 we used $k_B T \sim 4.1$ pN · nm. We can use molecular simulations to test whether this is violated and also predict the results of larger forces.

1.2 Biomolecular Force-Sensing Paradigms

After establishing the existence of piconewton forces at the molecular level and using Statistical Mechanics to predict their effect, the inquiry now turns to their applicability on larger length scales, specifically within cellular and sub-cellular domains. Mechanical forces are prevalent throughout a multitude of cellular and subcellular functions including material transport, growth, and motility [4, 6]. Molecular machines known as *Force Sensors* are able to orchestrate these crucial functions through complex force transduction mechanisms and structural displacements. The field of *Mechanobiology* encompasses the study of characterizing the behavior of these Force Sensors and accurately quantifying and locating cellular forces that bring about these intricate force signaling and response cascades [7]. A few examples of such “Mechanoresponsive” molecular machines are shown in Fig. 1.2, each with their own implications within or outside the cellular environment in response to mechanical forces on the order of piconewton magnitudes. We will summarize how each of these paradigms responds to mechanical force and their biophysical implications to emphasize the significance of mechanical forces at the cellular level and the need to study these Mechanoresponsive systems.

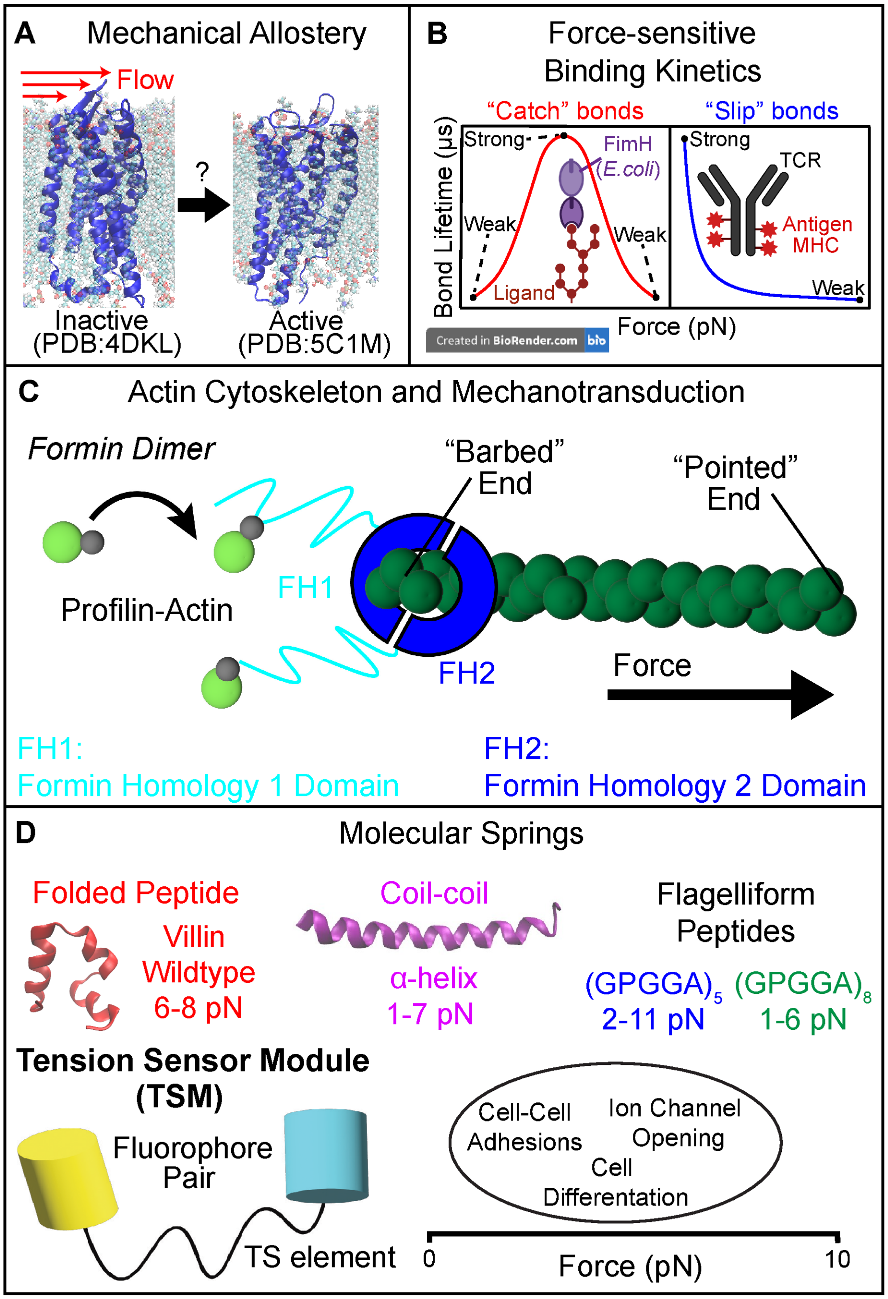

A
Mechanical Allostery
Flow
?
Inactive
(PDB:4DKL)
Active
(PDB:5C1M)
B
Force-sensitive
Binding Kinetics
"Catch" bonds
"Slip" bonds
Bond Lifetime (μs)
Strong
FimH
(E.coli)
Weak
Weak
Ligand
Strong
TCR
Antigen
MHC
Weak
Force (pN)
Created in BioRender.com
C
Actin Cytoskeleton and Mechanotransduction
Formin Dimer
"Barbed"
End
"Pointed"
End
Profilin-Actin
FH1
FH2
Force
FH1:
Formin Homology 1 Domain
FH2:
Formin Homology 2 Domain
D
Molecular Springs
Folded Peptide
Villin
Wildtype
6-8 pN
Coil-coil
α-helix
1-7 pN
Flagelliform
Peptides
(GPGGA)5
2-11 pN
(GPGGA)8
1-6 pN
Tension Sensor Module
(TSM)
Fluorophore
Pair
TS element
Cell-Cell
Adhesions
Ion Channel
Opening
Cell
Differentiation
0
Force (pN)
10

Figure 1.2: Overview of some Biomolecular Force-sensing Paradigms. (A) Illustration of Flow activation of GPCR GPR68. Structures shown are derived from homology models taken from inactive and active μ-opioid receptor (PDB: 4DKL [8] and PDB: 5C1M [9] respectively). (B) "Catch" bond behavior seen in Bacterial adhesion FimH with glycolipid sugars on the surface of endothelial cells [10] and "Slip" bond behavior in T-Cell Receptors (TCR) and Major Histocompatibility Complex (MHC) on the surface of antigens [11, 12]. (C) Actin nucleation by Formin dimer [13]. (D) Different Molecular springs and their force-sensitivities. These Molecular Springs can be engineered into Tension Sensor Modules (TSMs) for cellular and sub-cellular force measurement [6, 14–16].

1.2.1 MECHANICAL ALLOSTERY

The concept of Allostery refers to how the binding of a substrate or ligand at one particular site of biomacromolecule affects the functionality of another site. A pertinent example in Biochemistry is the action of non-competitive inhibitors binding to a different region of an enzyme which in turn alters the enzyme's kinetics at the active site and hence the overall catalytic activity [17]. *Mechanical Allostery* is the idea that mechanical force at one location brings about structural displacement at another location. A probable example of Mechanical Allostery is the flow activation of GPCR protein GPR68 [18–20]. The authors of Ref. [20] identified GPR68 in regulation Flow-Mediated Dilation (FMD) of blood vessels as a result of shear stress (or flow). The precise mechanism of GPR68 for sensing and responding to shear stress is widely unknown, more specifically the structural perturbations that arise from this flow activation (Fig. 1.2A).

1.2.2 FORCE-SENSITIVE BINDING KINETICS

We have discussed how force affects the kinetics of a simple two-state model (Sec. 1.1.2). The consequences mechanical forces have on kinetics are more prominent when studying binding interactions among larger biophysical entities. One such case is the "Catch" bond behavior of bacterial adhesion protein FimH binding to the Glycolipid sugars on the surface of Endothelial cells along the urinary tract [4, 21–23]. At low forces, the binding interactions are very weak. However, with increasing force as a result of flow, the binding interactions become stronger, causing the pathogenic bacteria to remain bound to the urinary tract leading to severe infections. The binding strength saturates up to a certain force and eventually weakens with increasing force (Fig. 1.2B). In Sec. 1.4, we will examine current computational approaches to predict the effect of mechanical forces on "Catch" bond kinetics. Another type of force-dependent non-covalent interaction is the "Slip" bond [11, 12, 21].

Unlike “Catch” bonds, “Slip” bonds are strongest at low forces and weaken with higher mechanical force as depicted in Fig. 1.2B. Slip bond behavior has been reported in the T-Cell receptor (TCR) binding of Major Histocompatibility Complex (MHC) of antigens during immune response [11, 12]. Depending on the type of antigen and T-cell, hybrid “Catch-Slip” bond behavior has also been observed [12].

1.2.3 Actin Cytoskeleton and Mechanotransductions

The cytoskeleton comprises the architectural framework of cells assembled from actin filaments. Actin filament polymerization is a proponent of numerous cellular processes including but not limited to the formation of cellular protrusions for movement, spindle fiber formation during meiosis, and maintenance of inter-cellular junctions (see Chapter 2) [24, 25]. The first step and rate limiting step of Actin filament formation is Nucleation; which involves the formation of an actin "nucleus", consisting of an aggregate of three actin monomers. After overcoming the energetically unfavorable nucleation step, the nucleus undergoes elongation, wherein actin monomers are added to both ends of the nucleus. Actin monomers are recruited at a faster rate (five or ten times faster) at one end termed the "Barbed" (plus, +) or growing end, compared to the "Pointed" (minus, -) end. As filament elongation continues and the freely available actin monomer concentration in the cell decreases until a critical actin monomer concentration is reached, a stationary phase is achieved where the rate of growth at the plus end is equal to the rate of dis-association at the minus end [24, 26]. Throughout the filament formation process, mechanical forces are propagated along these filaments which regulate assembly and disassembly [27]. The location and magnitude of mechanical forces along actin cytoskeleton networks have been identified through AFM-microscopy and FRET probe experiments, however, a mechanism of force propagation leading to assembly and disassembly has not been fully resolved [28, 29].

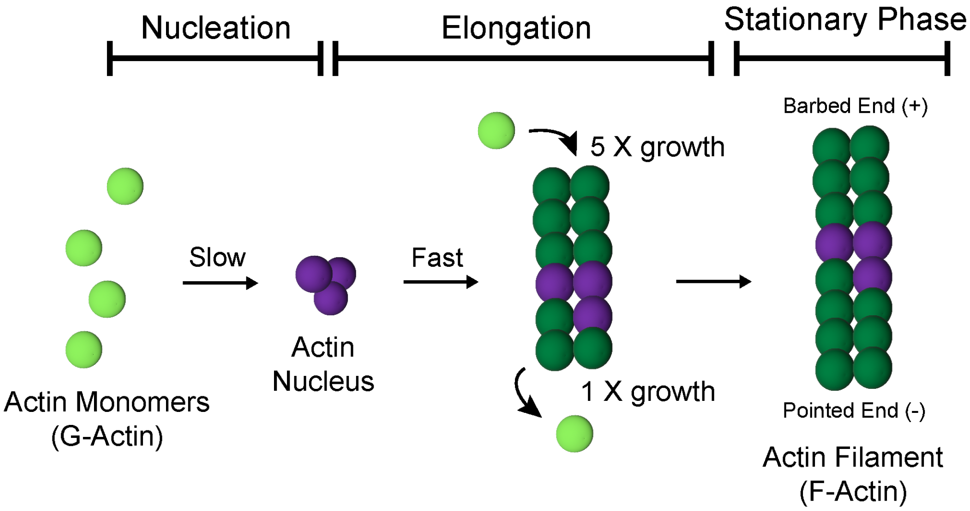

Nucleation
Elongation
Stationary Phase
Slow
Fast
5 X growth
1 X growth
Barbed End (+)
Pointed End (-)
Actin Monomers
(G-Actin)
Actin
Nucleus
Actin Filament
(F-Actin)

Figure 1.3: Overview of actin filament formation denoting the rate-limiting nucleation step, elongation, and stationary phase. (Schematic adapted from Ref. [26]).

1.2.4 MOLECULAR SPRINGS

At the cellular level certain amino acid sequences or folded peptides shown in Fig. 1.2D experience significant structural changes even at forces around or less than 10 pN. These "Molecular Springs" which we will term Tension Sensors have been indispensable in the field of Mechanobiology for the engineering of Tension Sensor Modules (TSMs) to quantify and locate forces associated with cellular and subcellular processes in Fig. 1.2D (see below) [6, 14–16]. Aside from low piconewton force sensitivity, there are a few other key characteristics of Tension Sensors summarized in Fig. 1.4. *Hysteresis* refers to how the measurement of some property of a system depends on the system's history (Fig. 1.4C), a characteristic good Tension Sensors should lack. In other words, performing repeated cycles of folding and unfolding by pulling on a Tension Sensor allows for a one-to-one mapping of the force applied and the length recorded at that force, which allows for unambiguous calibration when designing TSMs [15].

The molecular basis for which an amino acid sequence is able to possess the properties of a Tension sensor or a protocol to predict *a priori* whether an amino acid sequence is a Tension Sensor or not are still open-ended research problems in the field of Mechanobiology.

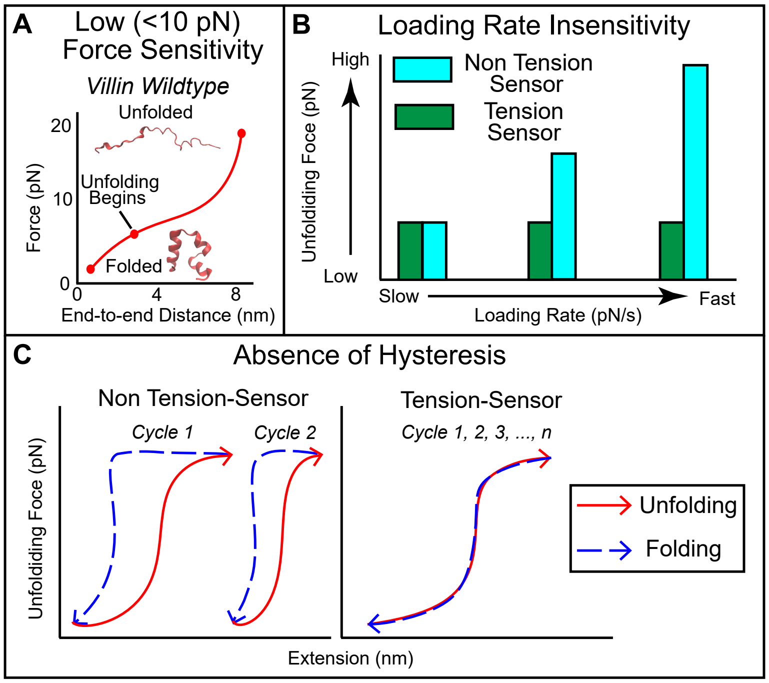


Figure 1.4: Properties of a Tension Sensor molecule. (A) Force-extension curve of Wildtype Villin-headpiece domain (PDB: 1YRF [30]), a well-documented Tension Sensor molecule that begins to unfold between 6-8 pN of force [31]. (B) Comparison of unfolding force (pN) at different loading rates (pN/s) for a Tension Sensor molecule and Non Tension Sensor molecule (Schematic adapted from Ref. [15]). (C) Unfolding-folding cycles demonstrating lack of hysteresis in a Tension Sensor molecule as opposed to a Non-Tension Sensor molecule (Schematic adapted from Ref. [15]).

1.3 Experimental Approaches to Measure Forces during Biological Processes

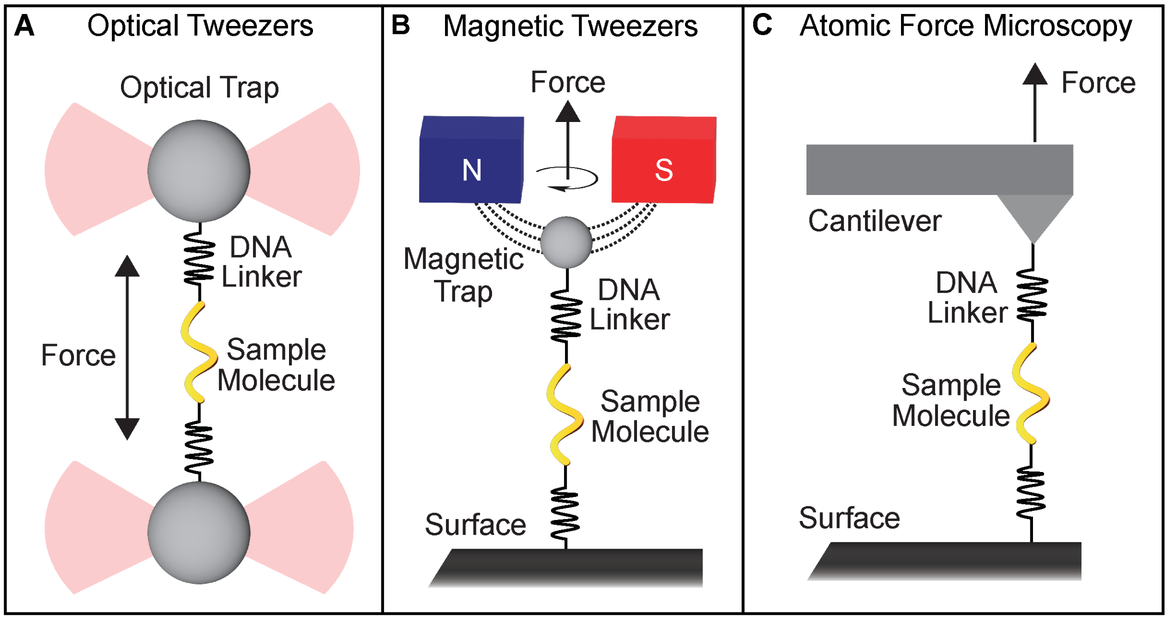

A
Optical Tweezers
Optical Trap
DNA Linker
Force
Sample Molecule
B
Magnetic Tweezers
Force
N
S
Magnetic Trap
DNA Linker
Sample Molecule
Surface
C
Atomic Force Microscopy
Force
Cantilever
DNA Linker
Sample Molecule
Surface

Figure 1.5: Overview of Single Molecule Force Spectroscopy Techniques. Schematics for (A) Optical Tweezers, (B) Magnetic Tweezers, and (C) Atomic Force Microscopy. (Schematics adapted from Ref. [32]).

In Sec. 1.2, we briefly mentioned a few experimental techniques for measuring cellular forces. Sec. 1.4, will delve into theoretical approaches to study Force-Sensing Paradigms in greater detail. Prior to delving into Theory and Simulation, a comprehensive understanding of the current functionalities and progress of available technologies is imperative. This includes outlining strengths and weaknesses. Additionally, identifying deficiencies in our current understanding is essential, as it will aid in formulating pertinent research questions for investigation using Theory and Simulation.

1.3.1 Single molecule force spectroscopy

Although the mechanics appear straightforward, *Single molecule force spectroscopy* techniques have proven to be powerful tools in Mechanobiology by offering illuminating insights into the forces associated with protein folding, ligand binding, and enzyme kinetics [33, 34]. The three most common techniques under this methodology we will be outlining are Optical Tweezers, Magnetic Tweezers, and Atomic Force-Microscopy (AFM) (Fig 1.5). The three methods vary in technical specifications such as the force range and resolution, but share the same principle in practice; a pulling force being applied to a molecule of interest and recording the extension [33].

A few other factors must be considered before performing any Single molecule force spectroscopy technique [33]. First, is the attachment of the molecule to the *measuring probe* and surface, which is implemented using open-ended DNA molecules or through ligand-receptor pairs such as biotin-avidin or antibody-antigen pairs to ensure tight binding to the probe during measurement. A second consideration is the accurate measurement of the measuring probe from its equilibrium position, as the mechanical force being applied is calculated directly from the probe's position. Therefore, to ensure highly precise measurements, these Single molecule force spectroscopy experiments have to be performed in environments devoid of any fluctuations in temperature or sound to minimize any sources of systemic error. Each of these techniques possesses distinct technical characteristics, applications, advantages, and limitations that must be carefully considered before putting them into practice.

Optical Tweezers. The usual forces and sizes of molecules vary between 0.1 to 100 piconewtons and from nanometers to micrometers, respectively (Fig. 1.5A). The measuring probe used is an "optical trap" consisting of a laser generated from a very powerful microscopic lens [33, 35]. Dielectric particles within the vicinity of the optical trap undergo polarization resulting in induced dipole-dipole interactions which produce a three-dimensional force along the direction of the laser.

The use of this high-focus laser beam offers high-precision force measurement for diverse biochemical assays such as RNA polymerase transcription of DNA [36]. However, a few drawbacks lie in sample purity, as samples used need to be highly pure to avoid any optical perturbations that could affect the focus of the laser. Moreover, the laser itself may cause local heating leading to unwanted temperature fluctuations during measurements [33].

Magnetic Tweezers. In contrast to optical tweezers, which create a three-dimensional force through dipole-dipole interactions, Magnetic Tweezers generate force by utilizing a permanent magnetic field that induces a magnetic dipole aligned with the field's direction (Fig. 1.5B). This magnetic force is then applied to either fold or unfold the targeted molecule. Magnetic Tweezers allow for larger force ranges up to nanonewton magnitudes and are more applicable in studying processes involving DNA unwinding by DNA topoisomerases [37] or ATP production by ATP synthase F_0F_1 ATPase [38]. Similar to Optical Tweezers, Magnetic Tweezers possess their own drawbacks. The one-dimensional force poses a restriction when studying the effect of forces on a molecule in different directions as opposed to the three-dimensional force produced in Optical Tweezers. Additionally, the use of a permanent magnetic field complicates the simultaneous integration of other techniques, such as the incorporation of fluorophores in Tension Sensor Module calibration (see Sec. 1.3.2) [15, 33].

Atomic Force Microscopy (AFM). Among the discussed Force Spectroscopy techniques, Atomic Force Microscopy (AFM) stands out as arguably the easiest to set up (Fig. 1.5C). It operates within the highest force range, spanning from 10 to 10000 pN, making it well-suited for investigating the unfolding and folding processes of proteins and nucleic acids. AFM is particularly effective in probing large biomolecules like actin-cross-linking protein, filamin [39]. In this technique, a sharp AFM tip, carrying the sample molecule, is attached to a flexible cantilever. Unlike other methods using lasers or magnetic fields, the force in AFM is directly generated as the cantilever moves vertically.

The displacement of a built-in laser caused by the cantilever's movement is detected by a position-sensitive detector, providing a direct correlation to the applied force [33, 40]. While AFM provides a more straightforward setup, it's important to note that there could be cases where the sample molecule is connected to a location other than the endpoint. Additionally, while the high forces generated are well-suited for examining the mechanics of protein folding and unfolding, they may not be optimal for investigating force distributions related to cells or cellular processes [33].

1.3.2 Tension Sensor Modules

While single-molecule force spectroscopy techniques have provided valuable insights into various biological processes across a spectrum of forces, the current challenge lies in accurately quantifying force distributions at the cellular and subcellular levels, as depicted in Fig. 1.2D [14, 15, 41, 42]. A significant breakthrough in addressing this challenge involves the development of Tension Sensor Modules (TSMs). The fundamental configuration of a TSM comprises a Molecular Spring or Tension Sensor molecule (refer to Sec. 1.2.4) tethered to two fluorophore dyes via DNA strands or terminal cysteine residues shown in Fig. 1.6. Various TSMs can be tailored based on the Molecular Spring type, fluorophore pairs, and their orientations [42, 43]. An example of a widely used Tension Sensor module is the FRET-based Tension Sensor Module. FRET (Fig. 1.6), which stands for Fluorescence Resonance Energy Transfer, results from an instantaneous dipole-induced dipole coupling between an excited-state fluorophore (Donor) and a ground-state fluorophore (Acceptor) [44, 45]. Utilizing a known single-molecule force spectroscopy technique such as optical tweezers to pull on the TSM ends and concurrently measuring the extension and FRET response enables the direct correlation of FRET response with applied force. The absence of hysteresis in Tension Sensor molecules ensures a clear one-to-one correspondence between FRET response and applied force. FRET's high sensitivity to distance ($E \sim 1/r^6$) allows for the calibration of extremely sensitive TSMs, capable of discerning minute piconewton forces within cells.

The development of numerous FRET-based tension sensor modules, featuring diverse Molecular Springs and thus varying force sensitivities, has provided insights into the spatial distribution and magnitude of forces during critical cellular processes. For instance, TSMs such as genetically encoded Flagellum silk Peptide and Wildtype Villin Headpiece Domain TSMs have been instrumental in confirming the existence of forces in the piconewton range (<10 pN) within cell-matrix networks known as Focal Adhesions [31, 46]. Another example includes a genetically encoded $\alpha-$helix tension sensor module specifically designed for the *in vivo* monitoring of cytoskeletal forces resulting from mechanical stress [47]. Various genetically encoded tension sensor modules have also unveiled forces associated with cell-cell junctions and mitosis [15]. Despite considerable advancements, fundamental questions persist regarding the actual pathways of force transduction. This prompts inquiries into the magnitude and duration of forces experienced by individual molecules. To comprehend the molecular-level impact of forces, computational tools such as Molecular Dynamics (MD) are essential [15].

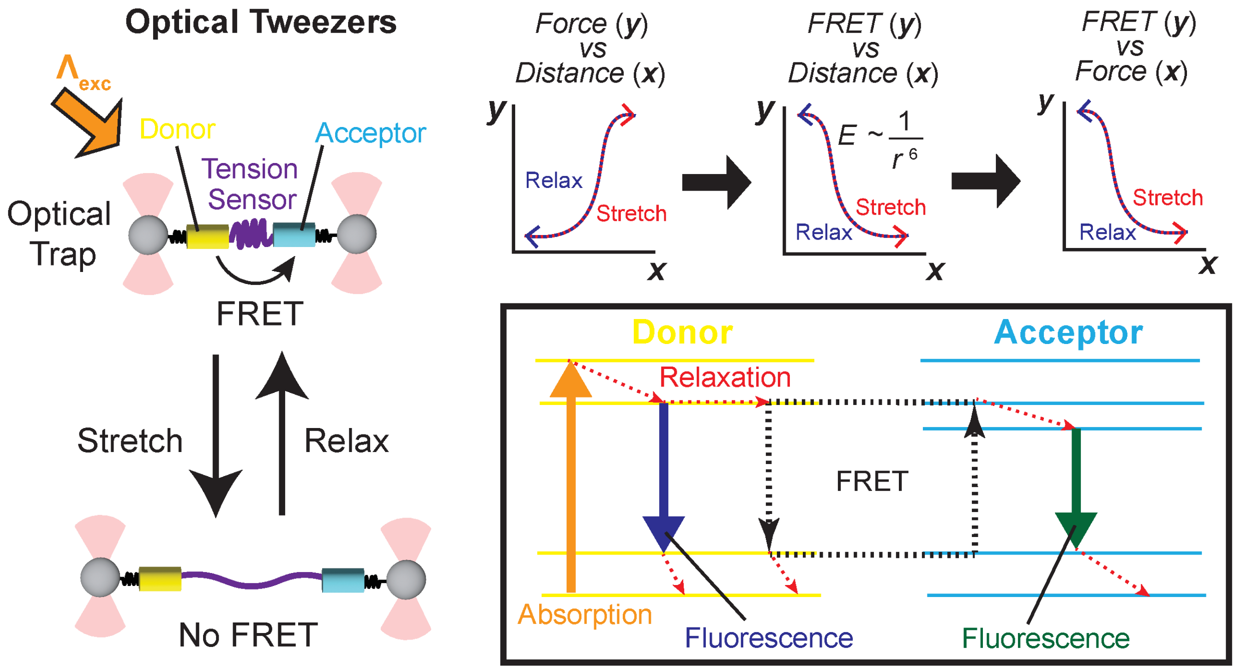

Optical Tweezers
Λexc
Donor
Acceptor
Tension Sensor
Optical Trap
FRET
Stretch
Relax
No FRET
Force (y) vs Distance (x)
y
x
Relax
Stretch
FRET (y) vs Distance (x)
E ~ 1/r6
Relax
Stretch
FRET (y) vs Force (x)
Relax
Stretch
Donor
Acceptor
Relaxation
FRET
Absorption
Fluorescence
Fluorescence

Figure 1.6: Calibrating a FRET based Tension Sensor Module (TSM) using optical tweezers. Λ_{exc} (orange) denotes the light at the excitation wavelength specific to the Donor fluorophore (yellow). Force-distance data acquired through optical tweezers, combined with simultaneous FRET responses at corresponding distances, establishes a direct correlation between FRET response and applied force (Schematic adapted from Ref. [15]). (Inset) Jablonski diagrams illustrating the ground and excited states of the Donor (yellow) and Acceptor (cyan) fluorophores. Solid arrows indicate radiative processes (Absorption, Fluorescence), while dashed arrows represent non-radiative processes (Vibrational Relaxation, FRET) (Schematic adapted from Ref. [44]).

1.4 COMPUTATIONAL TECHNIQUES TO STUDY MECHANOBIOLOGY

After exploring various experimental techniques for probing cellular force sensing and identifying some inherent limitations, we will now delve into several computational approaches aimed at enhancing our understanding of the molecular-level effects of forces. **Molecular Dynamics (MD)** simulations stand out as particularly promising tools for elucidating the molecular changes associated with force sensing within cells [48–50]. Depending on the system under investigation, be it a sequence of amino acids, a folded protein, or a nucleic acid, the general protocol for an MD simulation involves constructing a computational model of the system and solving Newton's equations of motion across multiple time steps to generate an ensemble of microscopic configurations or MD trajectories. These trajectories often yield valuable insights into crucial structural changes and significant molecular interactions, which can be correlated with experimental data.

The information obtained from MD simulations is invaluable, but the demanding computational resources and time required for data collection pose persistent challenges, especially with the increasing system size and complexity. Consequently, it becomes paramount to employ various **Enhanced Sampling** methods to expedite conformational exploration or sampling at a reduced computational cost. This often involves utilizing a biasing potential along a critical reaction coordinate (or **Collective Variable, CV**) [2]. One example of bias-based methods is *Metadynamics* [51–53]. The fundamental principle of this algorithm involves introducing a time-dependent bias potential in the form of Gaussian hills [54] centered at the position of the the biased CV. This facilitates overcoming energy barriers and reconstructing a Free Energy profile from the biased coordinates. While metadynamics enables the computation of static observables like Free Energy Surfaces, *Infrequent Metadynamics* (InfrMetaD) takes a step further to extract information about dynamic properties [55]. InfrMetaD extends the metadynamics framework by intermittently depositing the bias potential, allowing the system to freely explore specific barriers along the energy landscape.

Leveraging concepts from Transition State theory, rates between different states can be calculated from an InfrMetaD simulation [23, 55]. Although the method has demonstrated near-perfect results on test potentials and simple models, predicting more complex protein-ligand binding kinetics—such as the interaction between the protein streptavidin and its ligand biotin under mechanical force—remains an ongoing challenge in both simulation and experiment [23, 56].

An alternative approach employing a time-dependent bias potential is *Steered Molecular Dynamics* (Steered MD). In contrast to Metadynamics and its variations, which utilize a Gaussian hill bias potential, Steered MD simulations implement a time-dependent harmonic potential, as a means to effectively mimic Atomic Force Microscopy (AFM) experiments on proteins within computer simulations [4, 57]. Some of the earliest applications of Steered MD focused on investigating the unbinding kinetics of avidin-biotin under forces akin to those generated during AFM experiments [58]. In a notable study, Stirnemann et al. [59] conducted Steered MD simulations on the protein Ubiquitin (PDB: 1UBQ), successfully replicating Worm-Like Chain model behavior for mechanical properties, including persistence length, which aligned with experimental data. In Chapter 2, we present an MD study involving the application of Steered Molecular Dynamics (Steered MD) to investigate crucial structural motifs of Arp2/3 complex, a known molecular machine, offering valuable perspectives that contribute to our understanding of the system's dynamics and activation mechanism. It is also worth mentioning that the forces generated during a Steered MD simulation are often unphysical [60], so we also discuss another formulation to implement forces in our simulation from Chapter 3 onwards.

To further advance enhanced sampling methodologies, conducting parallel interacting MD runs through Replica Exchange Molecular Dynamics (REMD), presents a more sophisticated approach [2]. Rooted in a *Markov Chain Monte Carlo* [61] (MCMC) framework, REMD follows a *Metropolis-Hastings Criterion*, derived using principles from *Detailed Balance* [2, 62].

This criterion gives the probability of two consecutive replicas from a REMD run exchanging configurations, facilitating a “random walk” between different temperature spaces and enabling a more extensive exploration of the conformational landscape. In practical terms, Replica Exchange has proven highly effective for sampling small folded proteins [63] and larger Intrinsically Disordered Proteins (IDPs) associated with aggregation in Neurodegenerative diseases [64, 65], making it a valuable tool for understanding a system’s thermodynamic properties and structural changes during folding and unfolding. However, it is important to note that the computational expense of REMD scales considerably with increasing system size, the number of replicas, and the temperature range. To address the computational demands associated with scaling the temperature of the entire system, variants of Replica Exchange have emerged, focusing on scaling specific interactions, which in turn would require fewer replicas during setup [66–68]. In Chapter 4, we implement **Replica Exchange with Solute Tempering 3 (REST3)** [68], a critical component in improving a force tempering method presented in Chapter 3.

While Replica Exchange stands out as a potent sampling tool, and we have touched on a few REMD variants to address scalability concerns, the perennial challenge of the “*Curse of Dimensionality*” [69] remains. An alternative and effective strategy is the application of *Coarse Graining* [70–72]. This approach aims to mitigate the multitude of conformational degrees of freedom inherent in atomistic systems, alleviating the computational load by focusing on a few essential degrees of freedom and hence capturing relevant system behavior with enhanced computational efficiency. The versatility of simplifying systems to a select few crucial degrees of freedom makes Coarse Graining widely applicable across diverse realms of Theoretical Chemistry and Computational Biology, including *Ab Initio* modeling, intricate protein dynamics, and membrane modeling [71]. In the field of Mechanobiology, a recent study by Mitra et al. [73] exemplifies the implementation of Coarse Graining on *mobile binders* to replicate dynamic behaviors of biomimetic emulsion droplets, offering insights into the study of adhesion proteins such as cadherins.

In addition to these techniques, advanced computational technologies like Molecular Docking have produced groundbreaking insights into protein-ligand binding for drug discovery [74, 75]. However, the scope of this thesis will examine computational methodologies specifically designed for implementing forces or probing Mechanobology in Force Sensing Paradigms, providing a more focused exploration of the role of forces in cellular processes.

1.5 Dissertation Outline

At this juncture, we have identified molecular-level mechanical forces through Statistical Mechanics, highlighting their presence during cellular and subcellular processes. While postulating potential mechanisms, the examination of Force Sensing Paradigms has underscored the challenge of precisely uncovering how these forces induce critical conformational changes for cellular functions.

Recognizing the significance of mechanical forces at the cellular and subcellular levels, we examined experimental techniques developed to quantify and pinpoint these forces within cells. After identifying challenges and open-ended questions associated with experimental approaches, we shifted our focus to computational methodologies. These include Enhanced Sampling Molecular Dynamics, designed to accelerate conformational sampling in our simulations to further investigate structural changes of Force Sensing Paradigms under force. Building upon this foundational knowledge, we now outline the structure of this dissertation.

Chapter 2 focuses solely on the Actin Cytoskeleton and Mechanotransduction paradigm in Fig. 1.2C, specifically exploring the unique actin filament nucleator, Arp2/3 complex. The investigation aims to identify key structural changes during Arp2/3 complex activation and argues for a "Multi-Step" or unconcerted activation pathway with implications as a regulator in actin filament nucleation.

From Chapter 3 onwards we transition to the Molecular Springs paradigm (Fig. 1.2D). Chapter 3 introduces a Force Tempering Method known as **Infinite Switch Simulated Tempering in Force (FISST)**, a technique developed to implement a range of forces as an ensemble average in an MD simulation. The demonstration shows that this ensemble averaging approach facilitates crossing energy barriers and predicting the effect of multiple forces on any observable from a single MD simulation.

Chapter 4 presents a hybrid enhanced sampling method, combining the FISST method presented in Chapter 3 with Replica Exchange Molecular Dynamics. This approach effectively extracts benefits from both force tempering and conformational sampling, with results presented for different systems, including one resembling a known Tension Sensor molecule. The discussion argues that, with further finetuning, the hybrid technique can replicate Tension Sensor force response matching experiments and predict Tension Sensor behavior *a priori*.

Chapter 5, still within the Molecular Springs paradigm, focuses on predicting and verifying mechanical properties of linker peptides used in the design of Tension Sensor Modules. The preliminary data for this chapter offers insights into the secondary structure character of different linkers with varying amino acid compositions in implicit solvent Monte Carlo (MC) simulations and explicit solvent Molecular Dynamics (MD) simulations. We will discuss in detail that identifying and understanding the stability of different secondary structure motifs will help establish a more rational framework for linker design in not only Tension Sensor Modules but also Fusion proteins, which combine the functionalities of two or more bioactive protein domains.

Chapter 6 will summarize key findings and conclusions for all the studies presented in this thesis. We will present preliminary data that directly follow from some of the work presented here and also elaborate further on future computational studies for ongoing research that will expand the field of Mechanobiology.

Chapter 2

Molecular dynamics simulations support a multi-step Arp2/3 complex activation pathway

This chapter is adapted from Ref. [76]

Abstract

Actin-related protein 2/3 complex (Arp2/3 complex) catalyzes the nucleation of branched actin filaments that push against membranes in processes like cellular motility and endocytosis. During activation by WASP proteins, the complex must bind WASP and engage the side of a pre-existing (mother) filament before a branched filament is nucleated. Recent high-resolution structures of activated Arp2/3 complex revealed two major sets of activating conformational changes. How these activating conformational changes are triggered by interactions of Arp2/3 complex with actin filaments and WASP remains unclear. Here we use a recent high-resolution structure of Arp2/3 complex at a branch junction to design all-atom molecular dynamics simulations that elucidate the pathway between the active and inactive states. We ran a total of ~4.6 microseconds of both unbiased and steered all-atom molecular dynamics simulations starting from three different binding states, including Arp2/3 complex within a branch junction, bound only to a mother filament, and alone in solution. These simulations indicate that the contacts with the mother filament are mostly insensitive to the massive rigid body motion that moves Arp2 and Arp3 into a short pitch helical (filament-like) arrangement, suggesting actin filaments alone do not stimulate the short pitch conformational change. In contrast, contacts with the mother filament stabilize subunit flattening in Arp3, an intrasubunit change that converts Arp3 from a conformation that mimics an actin monomer to one that mimics a filamentous actin subunit. Our results support a multistep activation pathway that has important implications for understanding how WASP-mediated activation allows Arp2/3 complex to assemble force-producing actin networks.

2.1 INTRODUCTION

Filament nucleation is a critical step in the regulation of the actin cytoskeleton, as it controls when and where actin filament networks assemble in cells [77]. Spontaneous nucleation of actin filaments is energetically unfavorable, but once nucleated, new actin monomers can elongate both the fastgrowing barbed end and the slow-growing pointed end of filaments [78]. Cells rely on multiple classes of actin filament nucleators to catalyze nucleation and direct the assembly of new actin filaments, including tandem WH2 domain containing proteins, formins, and actin-related protein 2/3 complex (Arp2/3 complex) [78, 79]. Among these nucleators, Arp2/3 complex is the only one that can nucleate branched actin filaments [80]. Branched actin networks assembled by Arp2/3 complex play important roles in endocytosis, cellular migration, maintenance of cell–cell junctions, meiosis, DNA repair, and vesicle trafficking [81, 82].

To properly orchestrate complex cellular functions, the activity of Arp2/3 complex must be regulated so that nucleation occurs at the right time and location within the cell. On its own, the complex is inactive. Activation requires binding to a nucleation-promoting factor protein [78, 83]. WASP family proteins form a class of nucleation-promoting factors that activate Arp2/3 complex to create branches, but WASP is insufficient for activation; activation also requires that WASP recruit actin monomers to the complex and that the complex bind to a pre-existing "mother" filament of actin [83–85]. Upon activation by WASP, Arp2/3 complex nucleates a new actin filament with a free barbed end and its pointed end anchored to the complex at the newly formed branched actin filament junction [86]. How the nucleation activity of Arp2/3 complex is triggered by binding of WASP, WASP-recruited actin monomers, and actin filaments is unclear, despite its implications for understanding a wide range of cellular processes.

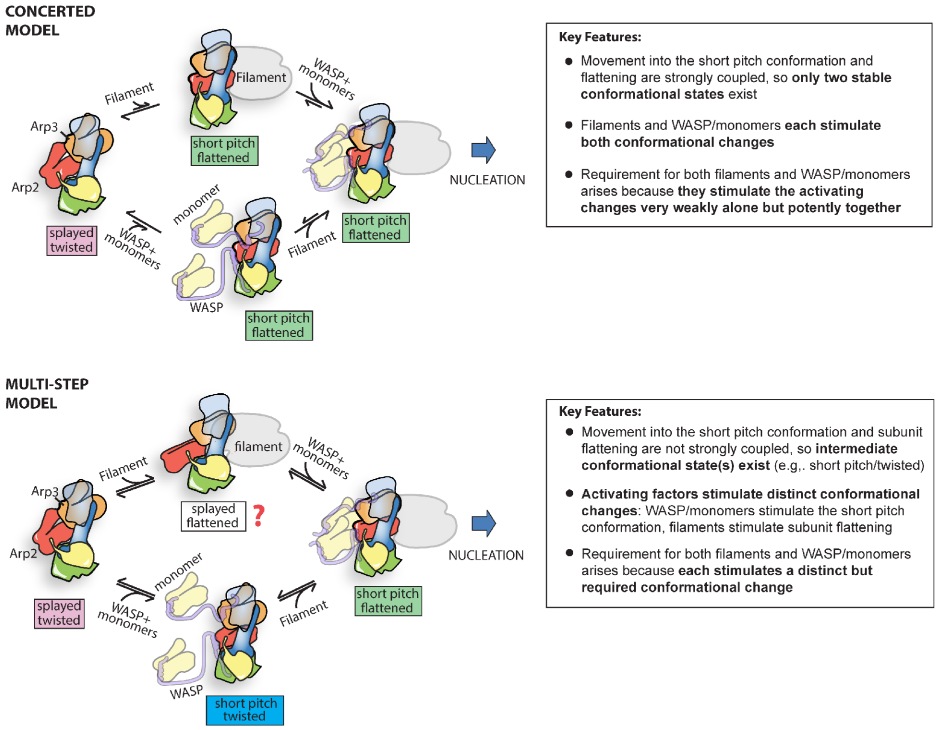
High-resolution structures of Arp2/3 complex in the inactive state have been available since 2001 [87], but high resolution structures of the activated complex have only recently become available because of advances in cryo-electron microscopy (cryo-EM) methods [88]. Among the recently solved cryo-EM structures are (a) a 9.0 Å structure of activated human Arp2/3 complex reconstructed from branch junctions imaged in cells [89], (b) a 3.9 Å structure of activated Bos taurus Arp2/3 complex at a branch junction [90], and (c) 3.5 Å and 3.9 Å structures of activated Schizosaccharomyces pombe Arp2/3 complex [91, 92]. These structures, along with biochemical data [93, 94], confirmed the long-standing hypothesis that during nucleation, the two actin-related subunits in the complex—Arp2 and Arp3—mimic a filamentous actin dimer to template the growth of a new filament [87]. Along with the previously solved inactive structures, the new structures revealed two major structural changes that bring Arp2 and Arp3 into a filamentous dimer-like conformation [90–92]. First, twisting of clamp subunits ARPC2 and ARPC4 rotates the bottom half of the complex (subunits ARPC1, ARPC5, Arp2, and the globular portion of ARPC2) to move Arp2 and Arp3 from an end-to-end ("splayed") conformation into an arrangement that mimics the positioning of two consecutive actin subunits along the short pitch helical axis of a filament (Video S1). Second, Arp2 and Arp3 each transition from a twisted state to a flattened state, an intramolecular change in which the four subdomains of each Arp move into approximately the same plane (Video S1). Flattening also occurs in actin when it transforms from a monomeric to a filamentous state [95–97], indicating this change is a key step in allowing the Arps to mimic a filamentous actin dimer. Flattening of the Arps is thought to trigger opening of grooves on their barbed ends for interactions with the first actin monomers in the newly nucleated (daughter) filament [90–92].

While the new structures revealed the key conformational changes required for activation, how binding of WASP, WASP-recruited actin monomers, and actin filaments stimulate these structural changes remain unclear. One model postulates that the structural changes are concerted (or strongly coupled) and that the activating factors bind cooperatively to Arp2/3 complex to stimulate them in a single step (Fig. 2.1 and Fig. 2.16) [91]. Two main observations support the concerted model. First, under some conditions, actin filaments increase the binding affinity of WASP for Arp2/3 complex, indicating cooperativity between WASP and filaments. Second, a large surface area is buried when Arp2/3 complex binds actin filaments, suggesting the availability of a large amount of binding energy to stimulate major conformational changes [85, 91, 98, 99]. A second model proposes that activation occurs via multiple steps (Fig. 2.1 and Fig. 2.16). In one step, WASP and WASP-recruited actin monomers stimulate movement of the complex into the short pitch conformation. In another step, bound actin filaments stimulate subunit flattening. The multistep model is supported by engineered crosslinking assays that show WASP and WASP recruited actin monomers stimulate movement into the short pitch conformation [93, 100] but actin filaments do not [100]. In addition, structures of activated Arp2/3 complex at a branch junction revealed interactions with actin filaments that can be made in the flattened but not the twisted conformation of Arp3, suggesting actin filaments stimulate subunit flattening [90]. Distinguishing between activation mechanisms is critical for understanding how the complex serves as a "coincidence detector" that triggers nucleation only when WASP, WASP-recruited actin monomers, and actin filaments are bound. The stringent requirement for each activator is thought to be critical for the ability of Arp2/3 complex to assemble functional actin networks in cells. For instance, the requirement for actin filaments ensures that when activated by WASP, Arp2/3 complex nucleates only branched actin filaments, which are optimal for pushing against broad flat surfaces like the plasma membrane at the leading edge of lamellipodia within motile cells [101, 102]. The requirement for WASP-recruited actin monomers helps control the density of branches nucleated within Arp2/3 complex assembled actin networks [103, 104]. The requirement for WASP connects Arp2/3 complex to cellular signaling pathways and targets the branching nucleation activity of Arp2/3 complex to the proper cellular location [81, 105]. Therefore, understanding how each of these factors contributes to the activating conformational changes in Arp2/3 complex is critical for understanding how the complex assembles force-producing actin networks in cells.

Computational simulations provide a powerful tool to understand the dynamics of biomolecular structures [49]. In the case of Arp2/3 complex, previous simulations have yielded important insights into multiple aspects of activation and nucleation. For instance, atomistic and coarse-grained molecular dynamics (MD) simulations have uncovered details about the binding of WASP and actin monomers to the complex and changes in its conformation caused by nucleotide binding and hydrolysis [106–108]. Unbiased all-atom MD simulations investigated the role of adenine nucleotides in controlling the conformation of Arp3 and Arp2, the branch angle and stability, and details of the interface of the complex with the mother filament [106, 109, 110]. All-atom steered molecular dynamics (SMD) showed that movement of Arp2/3 complex into the short pitch conformation occurs via twisting of the clamp subunits, ARPC2 and ARPC4 [111]. While each of these approaches has its strengths and weaknesses, all-atom MD simulations can be particularly useful because they provide atomistic details of the structural rearrangements that occur when a macromolecule transitions between states. However, such simulations are more informative if high-resolution structures are available to define each of the endpoints of a conformational pathway, thereby tethering simulation trajectories to the empirical data at multiple points along the pathway [112]. Therefore, the recent availability of high-resolution structures of activated Arp2/3 complex marked an important increase in the potential of all-atom simulations to yield insights into the pathway to nucleation by Arp2/3 complex.

Here we take advantage of the recently solved structure of activated B. taurus Arp2/3 complex bound at a branch junction—along with a high-resolution X-ray crystal structure of the inactive B. taurus Arp2/3 complex—to investigate the pathway between active and inactive states of Arp2/3 complex [90, 113]. We ran microsecond unbiased MD simulations starting from the active conformation of Arp2/3 complex anchored at a branch junction, bound to the side of an actin filament, or free in solution, along with a microsecond simulation of the complex free in solution starting from an inactive state. We also used SMD to pull Arp2/3 complex from the active to an inactive state when it was bound to the side of a filament. These data support a multistep activation pathway of Arp2/3 complex in which WASP and actin monomers stimulate the short pitch conformation and actin filaments stimulate subunit flattening. Specifically, the simulations provide evidence that actin filaments trigger only one of the two conformational changes required for activation: subunit flattening. They also reveal an intermediate conformation in which the complex is in the short pitch conformation, but Arp2 and Arp3 are twisted into the monomer-like conformation, a result that argues against a concerted mechanism. By providing information that allows us to distinguish between activation mechanisms, these simulations provide important insights into the conformational pathway to Arp2/3 complex activation by WASP.

2.2 RESULTS


CONCERTED MODEL
Arp3
Arp2
splayed twisted
Filament
Filament
short pitch flattened
WASP+ monomers
NUCLEATION
short pitch flattened
Filament
monomer
WASP+ monomers
WASP
short pitch flattened
MULTI-STEP MODEL
Arp3
Arp2
splayed twisted
Filament
filament
splayed flattened
?
WASP+ monomers
NUCLEATION
short pitch flattened
Filament
monomer
WASP+ monomers
WASP
short pitch twisted

Key Features:

- Movement into the short pitch conformation and flattening are strongly coupled, so **only two stable conformational states** exist
- Filaments and WASP/monomers **each stimulate both conformational changes**
- Requirement for both filaments and WASP/monomers arises because **they stimulate the activating changes very weakly alone but potently together**

Key Features:

- Movement into the short pitch conformation and subunit flattening are not strongly coupled, so **intermediate conformational state(s) exist** (e.g,. short pitch/twisted)
- **Activating factors stimulate distinct conformational changes**: WASP/monomers stimulate the short pitch conformation, filaments stimulate subunit flattening
- Requirement for both filaments and WASP/monomers arises because **each stimulates a distinct but required conformational change**

Figure 2.1: Simplified schematics of the conformational pathway to activation of Arp2/3 complex in a concerted versus multistep model of activation. Previous data indicate that while WASP triggers the splayed to short pitch conformational change, both states can exist with or without WASP bound to the complex [93, 94]. For clarity, neither these conformational states nor their reversibility is depicted here. For a more detailed diagram of the two proposed mechanisms that includes these states, see Fig. 2.16. Text boxes to the right of each scheme list the key features of each mechanism. The splayed/flattened state (marked with red question mark) may not be adopted because of steric clash (see Discussion).

2.2.1 Description of system setup

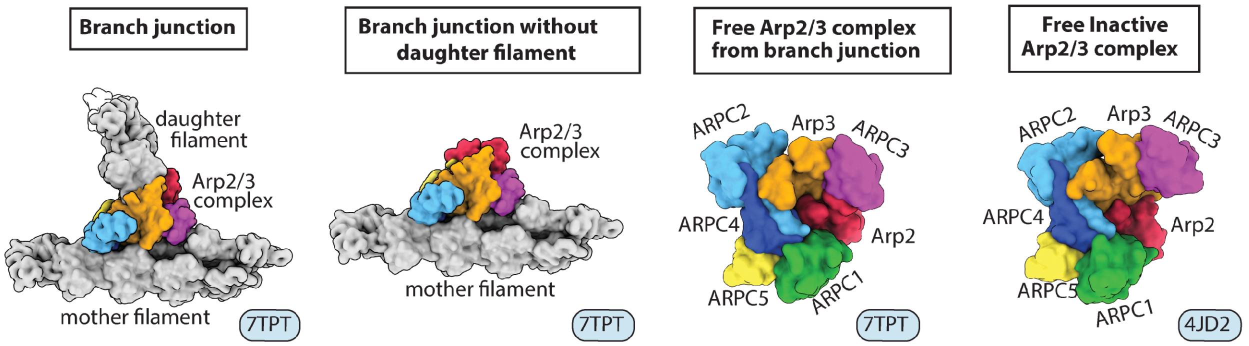

Branch junction
daughter filament
Arp2/3 complex
mother filament
7TPT
Branch junction without daughter filament
Arp2/3 complex
mother filament
7TPT
Free Arp2/3 complex from branch junction
ARPC2
Arp3
ARPC3
ARPC4
Arp2
ARPC5
ARPC1
7TPT
Free Inactive Arp2/3 complex
ARPC2
Arp3
ARPC3
ARPC4
Arp2
ARPC5
ARPC1
4JD2

Figure 2.2: Starting structures used for each of the four unbiased all-atom MD simulations described here. The PDB file used to build each structure is indicated in the lower right corner. Arp2/3 complex, actin-related protein 2/3 complex; MD, molecular dynamics.

Four different systems were set up for MD simulations (Fig. 2.2). The branch junction simulation was built from the recent cryo-EM structure of activated B. taurus Arp2/3 complex at a branch junction (PDB ID 7TPT, [90]). In this simulation, Arp2/3 complex was bound to the side of a mother filament consisting of 10 actin subunits and to the pointed end of the nucleated daughter filament containing four actin subunits, as in the cryo-EM reconstruction [90]. The branch junction without daughter filament simulation is identical to the branch junction simulation, except that the daughter filament was removed. In the simulation of free Arp2/3 complex from the branch junction, both the mother filament and daughter filament were removed from 7TPT. These three simulations allowed us to take advantage of the newly solved structures of Arp2/3 complex in the activated state, but it is important to note that any movement of these structures toward the inactive state represents the reverse direction of conformational changes that would be observed during the activation process. Lastly, the free inactive Arp2/3 complex simulation was built from the crystal structure of inactive B. taurus Arp2/3 complex bound to the inhibitor GMFγ (PDB ID 4JD2), as this structure provides the most complete model of inactive Arp2/3 complex [113].

GMFγ was removed from the coordinate file for the free inactive Arp2/3 complex simulation. In all but the branch junction simulation, Arp2 and Arp3 had bound ATP in their nucleotide clefts, consistent with a preactivation state. In the simulation of the branch junction, we modeled ADP into the clefts of Arp2 and Arp3, because both Arp2 and Arp3 hydrolyze ATP after branch formation [89, 114–116]. Similarly, ATP is hydrolyzed by actin subunits upon polymerization [117], so ADP is present in the nucleotide clefts of mother and daughter actin in the cryo-EM reconstruction and in the simulations here [90]. All systems were set up using the CHARMM22+CMAP forcefield with explicit TIP3P water and 50 mM neutral salt concentration, using K^+ and Cl^- ions [118]. Further details on system construction, minimization, heating, and production can be found in the Supporting Information. System sizes ranged from four hundred thousand to 1.4 million atoms. Due to the relatively large size of these systems, we limited the simulation times to 1 μs, a duration that—as described below—may not be long enough to reach the most stable ground state conformations for some of the systems. The SMD simulations were run exclusively on the branch junction without daughter filament, with the goal of pulling Arp2/3 complex from the short pitch to the splayed conformation while it remained bound to the side of a filament. A harmonic biasing restraint was applied by defining the center of geometry (COG) of each of the subunits and applying a spring constant of 10000 kJ/mol/Å between each COG. The relaxed spring length for each subunit pair was the distance between the centers of mass of each subunit when the complex was in the splayed conformation (as defined by the structure of the inactive Arp2/3 complex, 4JD2). We ran three different simulations in which the bias was applied over 60, 100, and 150 ns, after which the simulation was continued without the restraint to a total simulation time of ~200 ns.

2.2.2 The splayed and short pitch conformations are maintained in microsecond unbiased MD simulations

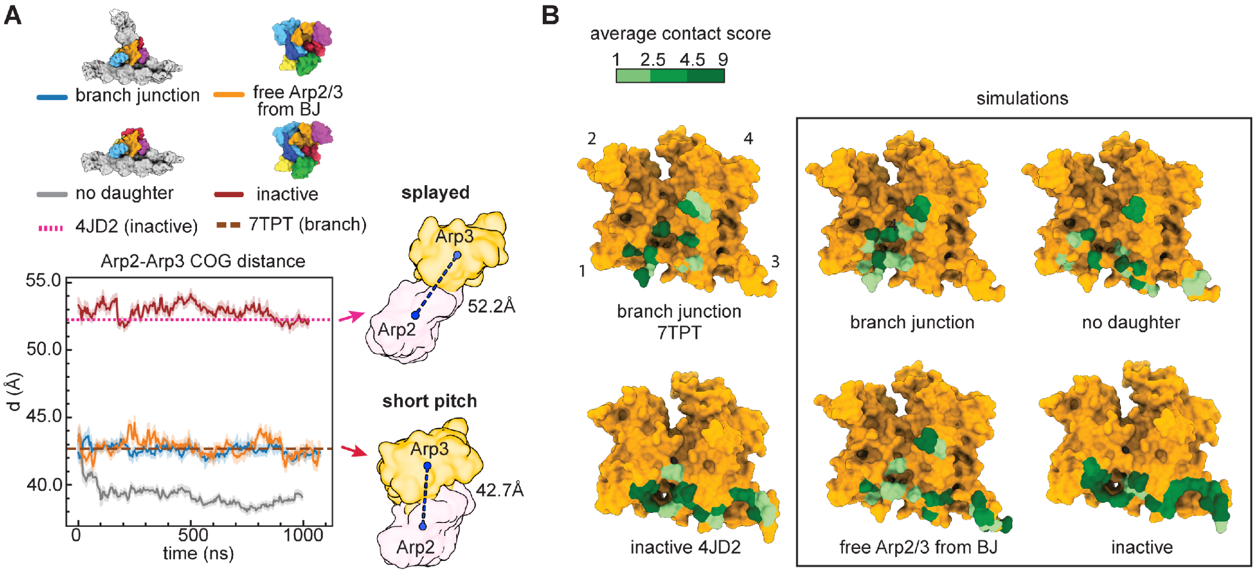


Figure 2.3: (A) Plot of the distance of the center of geometry (COG) of subdomains 3 and 4 of Arp2 to the COG of subdomain 3 and 4 of Arp3 as a function of simulation time. Dashed and dotted lines show the corresponding distance in the branch junction structure, 7TPT, and the inactive Arp2/3 complex structure, 4JD2, respectively. Data for simulations are shown as average smoothed over 5 ns (50 frames) for this and all other plots from the unbiased simulations. The shaded area shows the standard deviation over the smoothing window. (B) Surface representation of Arp3 from inactive (4JD2) or active structure (7TPT) showing average contact scores over the entire trajectory for Arp3 residues that contact Arp2. Contact scores for 4JD2 and 7TPT are shown on the left for reference. Contact scores were calculated using PyContact, as described in the Supporting Information.

We first asked whether Arp2/3 complex switched between the splayed and short pitch conformations in each of the simulations. We tracked this conformational change by measuring the distances between the COG of the inner domain (subdomains 3 and 4) of Arp2 with the inner domain (subdomains 3 and 4) of Arp3. High-resolution structures show that this distance decreases from 52.2 to 42.7 Å during the short pitch conformational switch [90, 113]. In the simulation of free inactive Arp2/3 complex, Arp3 and Arp2 remain splayed, with residues that stabilize the splayed interface remaining in close contact throughout the simulation (Fig. 2.3A, and Fig. 2.3B).

This result is consistent with previous biochemical and structural data showing that splayed conformation is strongly favored in the absence of activating factors [93, 94, 119]. In contrast, analysis of cryo-EM structures showed that the short pitch conformation is stabilized in the context of the branch junction [90, 91]. Consistent with these data, the short pitch conformation is maintained throughout the entire 1 μs branch junction simulation, with an average distance of 42.6 Å, very close to the distance in the branch junction cryo-EM structure (Fig. 2.3A, and Fig. 2.3B). Despite the loss of stabilizing interactions, the short pitch conformation was also maintained over the entire microsecond trajectory in the simulations of the branch junction without daughter filament and the free Arp2/3 complex from the branch junction (Fig. 2.3A). In both simulations, the Arp2 and Arp3 maintained contacts characteristic of the short pitch arrangement (Fig. 2.3B). Assuming the simulations represent a near average behavior, they point to an energy barrier of at least several $k_\mathrm{B}T$ between the splayed and short pitch conformations [120]. The stability of the short pitch state over the entire 1 μs simulation is consistent with experimental and theoretical studies that estimate rigid body motions in proteins to be on the order of micro- to milliseconds [120], although we would expect that the complex would relax to the splayed conformation in longer simulations of the branch junction or the branch junction without the daughter filament.

During the simulation of the branch without daughter filament, the COG of the inner domains of Arp2 and Arp3 moved together by ~4.0 Å (Fig. 2.3A). This is a result of Arp2 tilting toward the barbed end of Arp3, making additional contacts to Arp3, but adopting a conformation that would clash with actin subunit DA1 from the daughter filament (Fig. 2.17). Therefore, this state, which is stable throughout the trajectory, would not be expected to be nucleation competent, even though it has a conformation close to the short pitch arrangement seen in the branch junction structure (Fig. 2.17).

2.2.3 Flattening and adoption of the short pitch conformation are not tightly linked

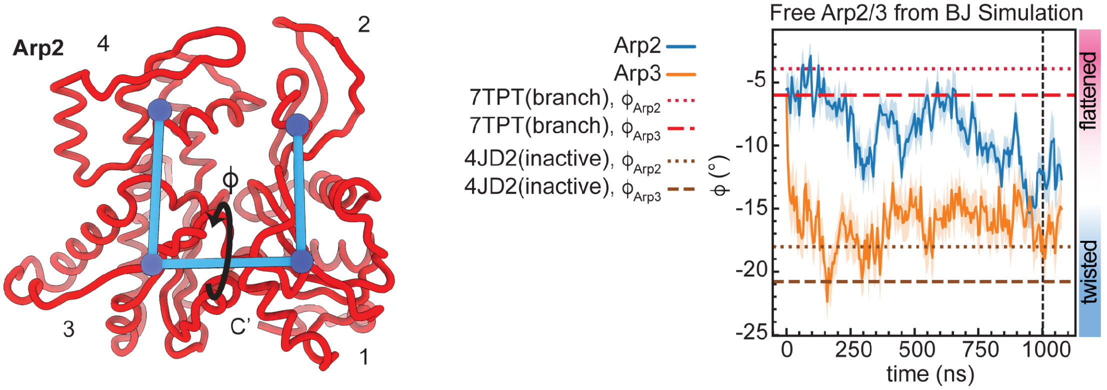


Figure 2.4: (left) Backbone trace of Arp2 showing measurement of the twisting/flattening angle. Subdomains of Arp2 are labeled 1 to 4. (right) Plot of Arp2 and Arp3 twisting/flattening angle (φ) versus simulation time for the free Arp2/3 complex from branch junction simulation. Arp2 and Arp3 twisting/flattening angles from inactive (4JD2) and the active Arp2/3 complex structure (7TPT) are shown as dashed or dotted lines, as indicated.

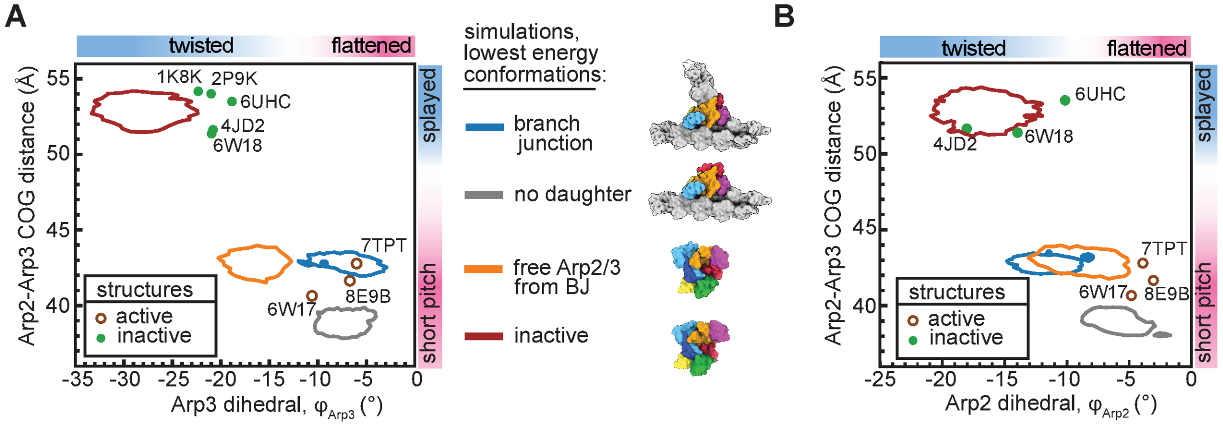


Figure 2.5: (A) Plot of the Arp3 subunit twisting/flattening dihedral versus the Arp2-Arp3 COG, which measures movement into the short pitch conformation. Enclosed regions indicate the most probable conformations in the simulation, as defined by conformations that are within a radius of one free energy unit from the lowest energy conformation. Circles show the corresponding measurements for selected active and inactive cryo-EM or X-ray crystal structures. (B) Plot as described in (A), except the Arp2 twisting/flattening dihedral angle is plotted on the x-axis.

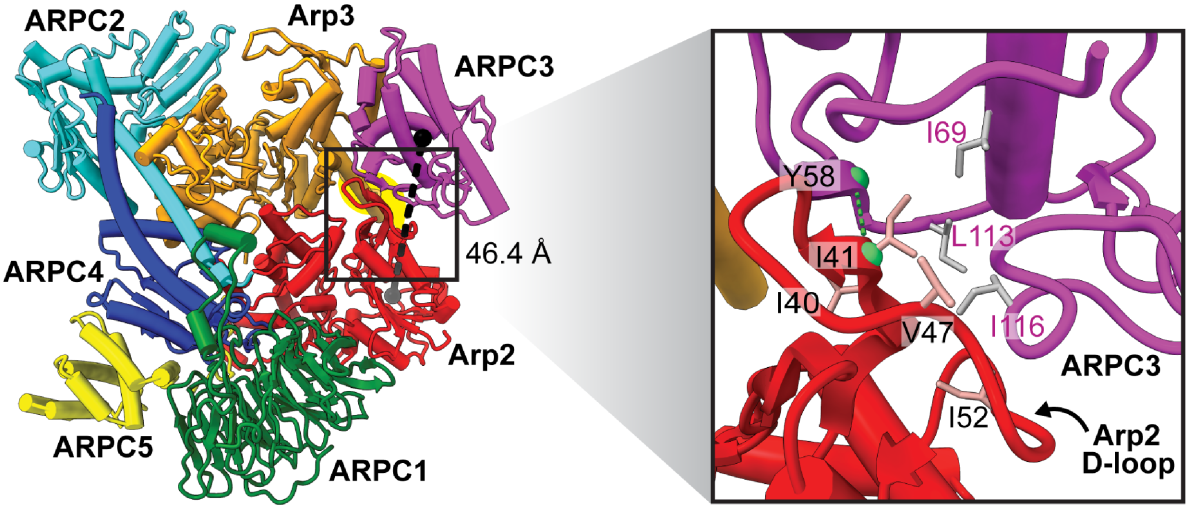


Figure 2.6: Ribbon diagram of the free Arp2/3 complex from branch junction simulation output at 1 μs showing that the Arp2 D-loop maintains contact with ARPC3 even when the complex moves into a short-pitch, twisted conformation. The D-loop of Arp2 is highlighted in yellow. The distance between the globular domain of ARPC3 and subdomains 1 and 2 of Arp2 is indicated with a black line. Inset shows a zoomed in view of the interaction.

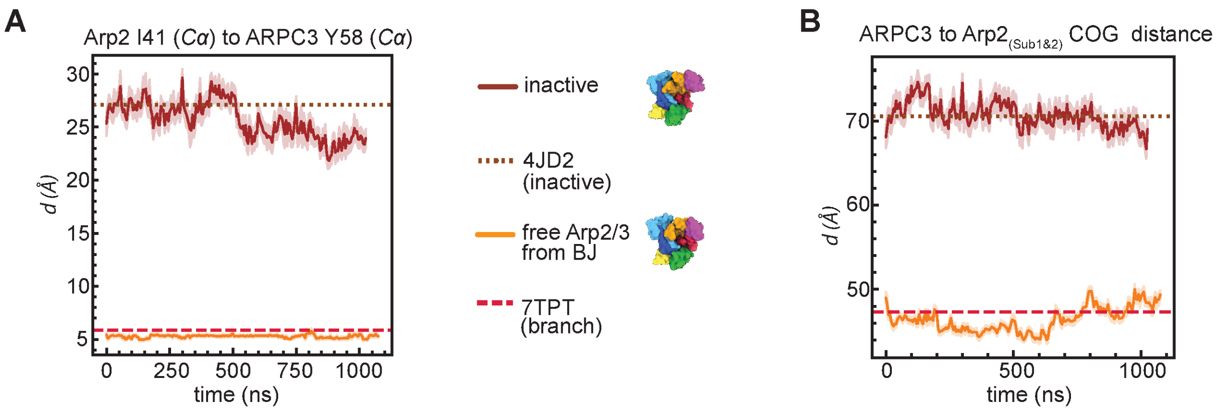


Figure 2.7: (A) Plot showing the distance between Arp2 D-loop and ARPC3 for Arp2/3 complex (active) and (inactive) simulations versus simulation time, with 7TPT and 4JD2 plotted for reference. (B) plot showing the distance between the COGs of Arp2$_{(\mathrm{Sub1\&2})}$ and ARPC3 versus simulation time in the Arp2/3 complex active simulation. Arp2/3 complex, actin-related protein 2/3 complex; cryo-EM, cryo-electron microscopy; COG, center of geometry.

A key feature of concerted models is that ligand-induced conformational changes are strongly linked so that the system exhibits high cooperativity and switch-like activation when all ligands are bound [121]. In a fully concerted Arp2/3 complex activation mechanism, movement into the short pitch conformation and flattening would be perfectly coupled. In contrast, in the multistep model for activation, these two conformational switches could occur independently and give switch-like activation [90, 100]. To investigate potential links between the activating conformational changes, we asked if Arp2 or Arp3 could switch from the flattened to twisted states in the simulations of activated Arp2/3 complex, even though the complex stays in the short pitch conformation (Fig. 2.3A). We used the dihedral angle between the COG of the four subdomains of the Arps as a metric for subunit flattening, as previously described ([92], Fig. 2.4). During the 1 μs simulation of free Arp2/3 complex from the branch junction, the dihedral angle of Arp3 rapidly moved toward the twisted state and stayed in a short-pitch, twisted conformation for much of the simulation (Fig. 2.4 and Fig. 2.5A). The short-pitch, twisted conformation was also adopted in Arp2 in the simulation of the branch junction and moderately populated in Arp2 in the simulation of free Arp2/3 complex from the branch junction (Fig. 2.5B). These data indicate that Arp2 and Arp3 can move from a flattened state into or close to the twisted conformation even when they are arranged in the short pitch conformation. This suggests that the two major activating conformational changes in Arp2/3 complex—adoption of the short pitch conformation and subunit flattening—are not tightly linked. Furthermore, the existence of a stable intermediate conformation of Arp2/3 complex, in which one activating conformational change (short pitch adoption) but not the other (flattening) has occurred argues against a concerted mechanism for activation (Fig. 2.1).

The observation that the Arps undergo subunit twisting when the complex is in the short pitch conformation was unexpected because flattening is thought to increase the buried surface area (BSA) between Arp2 and Arp3 in the short pitch complex [92]. However, in addition to contacts between Arp2 and Arp3, other intracomplex interactions appear to stabilize the short pitch conformation, including contacts between the D-loop of Arp2 and a hydrophobic pocket in ARPC3 (Fig. 2.6). This interaction—which was shown to be important for nucleation activity of the complex [90]—is maintained when the Arps twist (Fig. 2.6 and Fig. 2.7A). In contrast, the Arp2 D-loop/ARPC3 interaction is not observed in structures or in simulations of inactive Arp2/3 complex, in which the two actin-related subunits are in the splayed configuration, because movement into the short pitch conformation moves the centers of mass of Arp2 and ARPC3 apart by ~20 Å (Fig. 2.7A and Fig. 2.7B). These data provide an explanation for the apparent stability of the short-pitch, twisted intermediate state; because of the flexibility of the Arp2 D-loop, the ARPC3–Arp2 interaction can stabilize the short pitch conformation regardless of whether the Arps are flattened or twisted. We note that our observations differ from a recent analysis of the activated S. pombe Arp2/3 complex which suggested that the twisted conformation is incompatible with the short pitch conformation because of steric clash (see Discussion) [91].

2.2.4 Contacts with the mother filament stabilize flattened Arp3

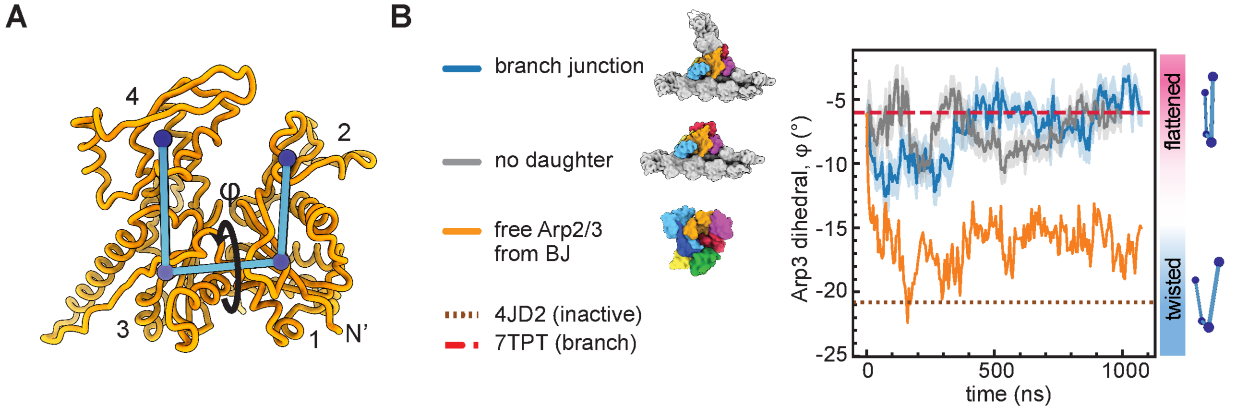


Figure 2.8: (A) Diagram showing the twisting/flattening angle (φ) of Arp3. The four subdomains of Arp3 are labeled 1 to 4. (B) plot of Arp3 twisting/flattening dihedral angle versus simulation time for all unbiased simulations that started in the active conformation. Arp3 twisting/flattening angles from inactive (4JD2) and active Arp2/3 complex structures (7TPT) are shown in dotted or dashed lines, as indicated.

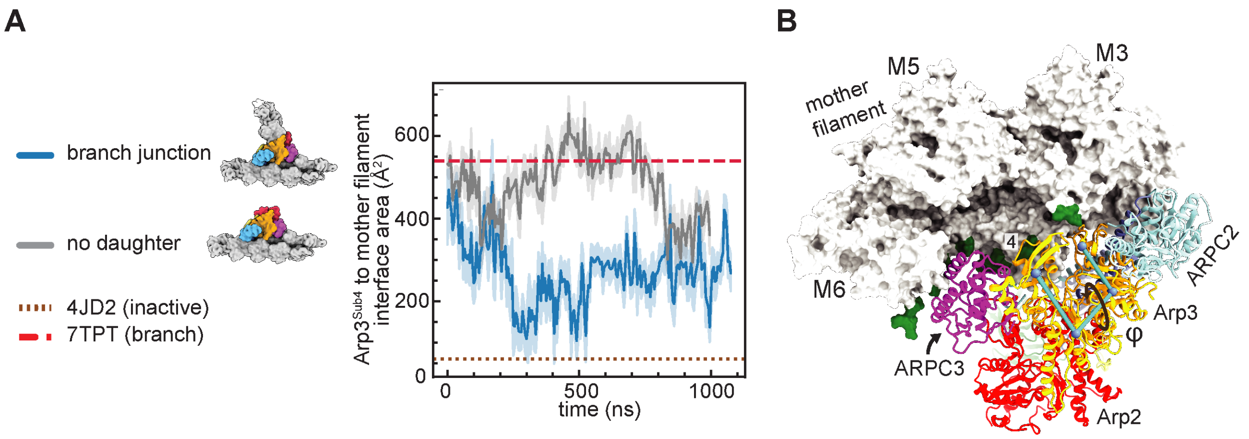


Figure 2.9: (A) Plot of the surface area of subdomain 4 of Arp3 buried on the mother filament versus simulation time. Buried surface area of subdomain 4 in the branch junction structure (7TPT) and a model of 4JD2 on an actin filament are shown as dotted or dashed lines. (B) Ribbon and surface representation of the last frame of the branch junction without daughter filament simulation showing residues within actin filament that interact with Arp2/3 complex upon subunit flattening in Arp3. Actin filament residues that interact with ARPC3 or $\text{Arp3}^{\text{Sub4}}$ in the MF-bound Arp2/3 complex simulation (average contact score > 1, colored green) are mapped onto the surface of the actin filament. Subdomain 4 of Arp3 is labeled. The Arp3 dihedral angle that flattens Arp3 is shown as angle φ

We next analyzed the simulations to determine whether interactions of Arp2/3 complex with the mother filament favor subunit flattening in Arp3, as predicted based on the branch junction structure of B. taurus Arp2/3 complex [90]. We found that Arp3 stays in or close to the flattened conformation in both the simulation of the branch junction and the simulation of the branch junction without daughter filament (Fig. 2.8A and Fig. 2.8B). This contrasts the simulation of the free Arp2/3 complex from the branch junction, in which Arp3 adopts twisted states (Fig. 2.4 and 2.8B). The flattened conformation is stabilized by contacts of subdomain 4 of Arp3 with the mother filament, which are maintained throughout the simulations (Fig. 2.9A and 2.9B). ARPC3 also interacts with the mother filament during these simulations (Fig. 2.18). This subunit is bound to subdomains 3 and 4 of Arp3, so flattening of Arp3 brings it into contact with the mother filament. Together, Arp3 subdomain 4 and ARPC3 bury an average of 1600 Å^2 and 700 Å^2 with the mother filament in the last 750 ns of the simulations of the branch junction and the branch junction without daughter filament, respectively (Fig. 2.18). These observations support a model in which binding to the mother filament stimulates flattening in the Arp3 subunit. We note that in steered MD simulations in which Arp2/3 complex is pulled from the short pitch to splayed conformation (see below), Arp3 twists but ARPC3 and subdomain 4 maintain contact with the mother filament. In these simulations, clamp subunits ARPC2 and ARPC4 bend to move Arp3 relative to the filament, allowing it to maintain all filament contacts even when twisted (Video S2). Bending of the clamp has not been observed in X-ray crystal structures or cryo-EM structures of the complex, so will be important to determine whether this conformation of the complex is biologically relevant.

2.2.5 Twisting of Arp2 closes the barbed end groove and weakens its interactions with the daughter filament

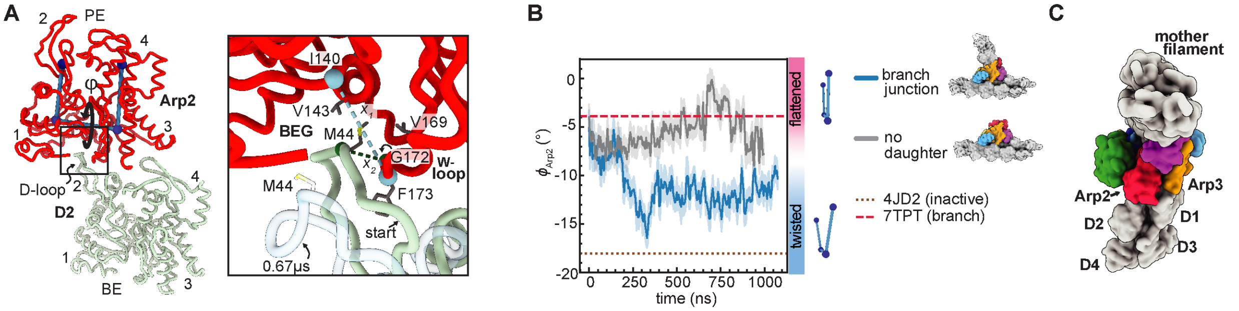


Figure 2.10: (A) Ribbon diagram showing the interaction of the D-loop of actin D2 with the barbed end groove of Arp2. Arp2 and actin D2 subdomains are labeled 1 to 4. PE: pointed end, BE: barbed end. Right panel shows closeup of the interaction with distances measured in (B). Actin D2 from 0.67 μs (transparent light blue ribbon) in the branch junction simulation was placed by overlaying Arp2 from the 0.67 μs frame in the trajectory with Arp2 from 7TPT. BEG: Barbed-end groove. Start: position of actin D-loop at the beginning of the simulation. (B) Plot of twisting/flattening angle of Arp2 (φ) in the branch junction and the branch junction without daughter filament (no daughter) simulations. Arp2 twisting/flattening angles from inactive (4JD2) and active Arp2/3 complex structures (7TPT) are shown in dotted or dashed lines, as indicated. (C) surface representation of branch junction model (7TPT) showing the interface between Arp2 and Arp3 and the pointed end of the nucleated daughter filament.

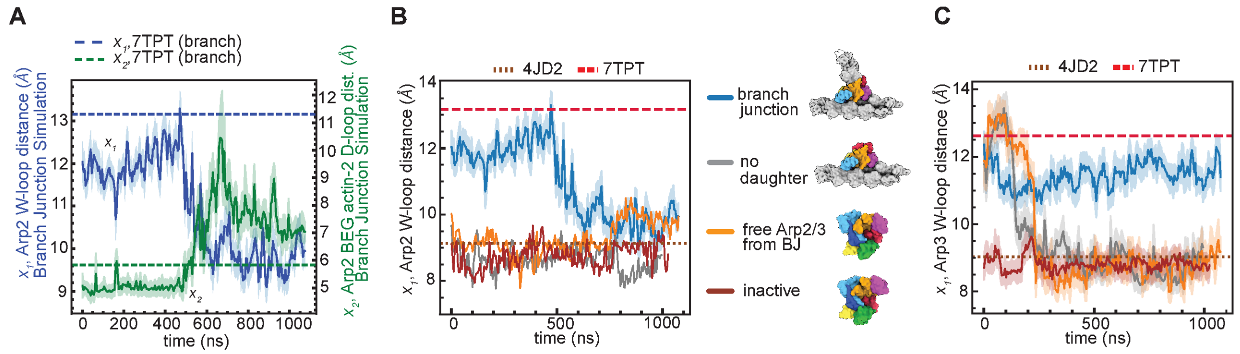


Figure 2.11: (A) Plot of the W-loop opening (x_1) and D-loop to Arp2 distance (x_2) in the branch junction simulation. See Fig. 2.10A for definition of x_1 and x_2. Distances x_1 and x_2 in the branch junction structure are plotted for reference. (B) Plot of W-loop distance (x_1) in Arp2 subunits for all unbiased simulations. Distance x_1 in the branch junction (7TPT) and inactive Arp2/3 complex structure (4JD2) is plotted as dashed or dotted lines, as indicated. (C) Identical to (B), except x_1 for Arp3 from each simulation is plotted.

While Arp3 stays near or in a flattened conformation in the simulation of the branch junction, Arp2 moves closer to the twisted state, populating mostly intermediate but sometimes nearly fully twisted states (Fig. 2.10A and 2.10B). To better understand how twisting influences the interface of Arp2 with the daughter filament pointed end (Fig. 2.10C), we measured the width of the barbed end groove in the simulation of the branch junction. We found that Arp2 twisting was accompanied by a closure of the Arp2 barbed end groove, as measured by the distance between F173 and I140 in Arp2 (Fig. 2.10A and Fig. 2.11A). This distance measures curling of the W-loop, a short loop that lines the barbed end groove. When the W-loop curls, it closes a pocket in the side of the barbed end groove. Curling of the W-loop in the simulation of the branch junction was accompanied by ejection of Met44 from the actin D2 D-loop from the pocket (Fig. 2.10A, Fig. 2.11A, and video S3). Met44 has previously been shown to be a target of MICAL enzymes, which oxidizes it to stimulate actin filament disassembly [122]. Together, our observations support a model in which flattening/twisting of the Arps allosterically influences the interaction of their barbed ends with actin subunits in the long pitch position. This allostery could play an important role in activation of Arp2/3 complex because it provides a mechanism by which actin filaments could promote nucleation by stimulating subunit flattening. To further probe the link between twisting/flattening of the Arps and the width of their barbed end grooves, we measured the W-loop position in Arp2 and Arp3 in all four simulations. In simulations lacking the daughter filament, the W-loop of both Arps remained or become curled even if the Arp subunit remained flat, suggesting that insertion of the D-loop of actin into the barbed end groove helps stabilize an open barbed end groove (Fig. 2.11B and 2.11C). The W-loop of Arp3 tended to remain uncurled through a greater portion of the simulations than in the Arp2 subunit, suggesting that the open state of the barbed end groove may be more favorable in Arp3 compared to Arp2. This finding may have implications for understanding how WASP-mediated recruitment of actin monomers to each Arp influences the activating conformational changes.

2.2.6 Splayed Arp2/3 complex maintains approximately the same interface area with the mother filament as short pitch Arp2/3 complex

Our MD simulations suggest actin filaments stimulate subunit flattening in Arp3, one of the major activating structural changes in Arp2/3 complex. It is also important to determine whether actin filaments can stimulate the other major activating conformational change: movement into the short pitch conformation. Filaments will stimulate the short pitch arrangement if the short pitch conformation of the complex interacts more favorably with the filament than the splayed conformation. However, there are no structures of Arp2/3 complex in the splayed state bound to the filament, so information about the inactive interface is limited [90, 91]. Therefore, to better understand the differences between short pitch and splayed mother filament contacts, we ran SMD simulations using the branch junction without daughter filament system (Fig 2.2). We used a harmonic biasing restraint to pull Arp2/3 complex from the short pitch to the splayed conformation while it remained bound to the side of the filament, as described above (Fig. 2.12A and Video S4). The pulling forces were applied over three different time intervals. In all three pulling inter-

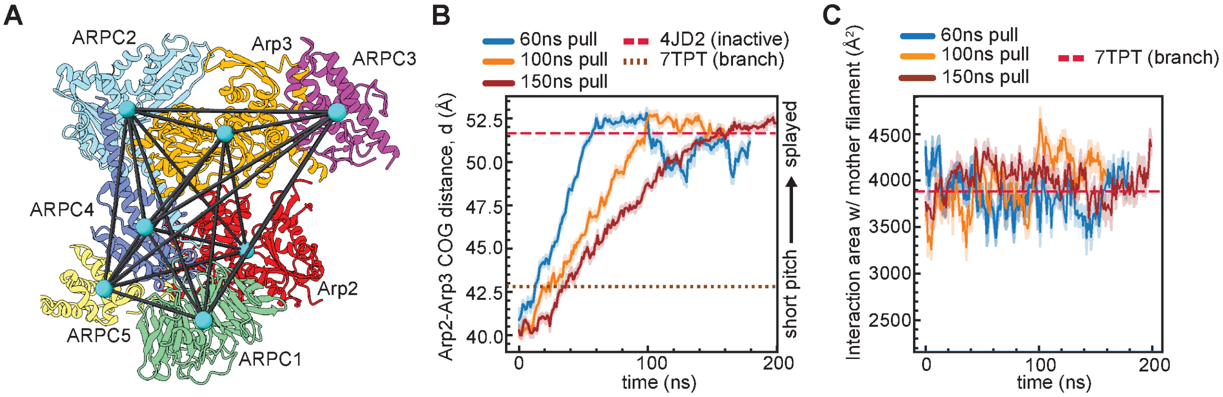


Figure 2.12: (A) Ribbon diagram of inactive Arp2/3 complex (4JD2) showing the center of geometry (COG) of each subunit. For the steered MD simulations, a bias was applied to move the COGs from their positions in the active complex to their positions in the inactive (4JD2) complex. (B) Plot of the distance between the Arp2 and Arp3 subdomains 3 and 4 COGs as a function of simulation time. The same distances for the inactive (4JD2) and branch junction (7TPT) structures are plotted as dashed and dotted lines, as indicated. (C) Plot of the total interaction area of Arp2/3 complex with the mother filament in all three steered simulations. Data for simulations are shown as an average smoothed over 1 ns (10 frames) for this and all other plots for the steered simulations. The standard deviation over the smoothing window is shaded. Interaction surface area in the branch junction (7TPT) is plotted as a dashed line for reference.

vals, the Arp2 and Arp3 subunits move into the splayed arrangement (Fig. 2.12B). The pathway to the splayed conformation is consistent with previous empirical and computational studies; it proceeds via a twisting of the clamp subunits (ARPC2 and ARPC4) that rotates the bottom half of Arp2/3 complex—including Arp2, ARPC1, ARPC5 and the globular portion of ARPC4—into the splayed position [90, 91, 111, 113] (Fig. 2.19A-C). These observations show that the bias applied in the steered MD simulations produces a splayed conformation of filament-bound Arp2/3 complex via a pathway consistent with the available empirical data. We note that the steered simulations change the conformation of Arp2/ 3 complex in a direction opposite to that occurring during branching nucleation. While it is unclear whether the conformations can interchange on the side of the filament, we expect this set of conformational changes to occur sometime after branch disassembly to reset the complex to the inactive conformation.

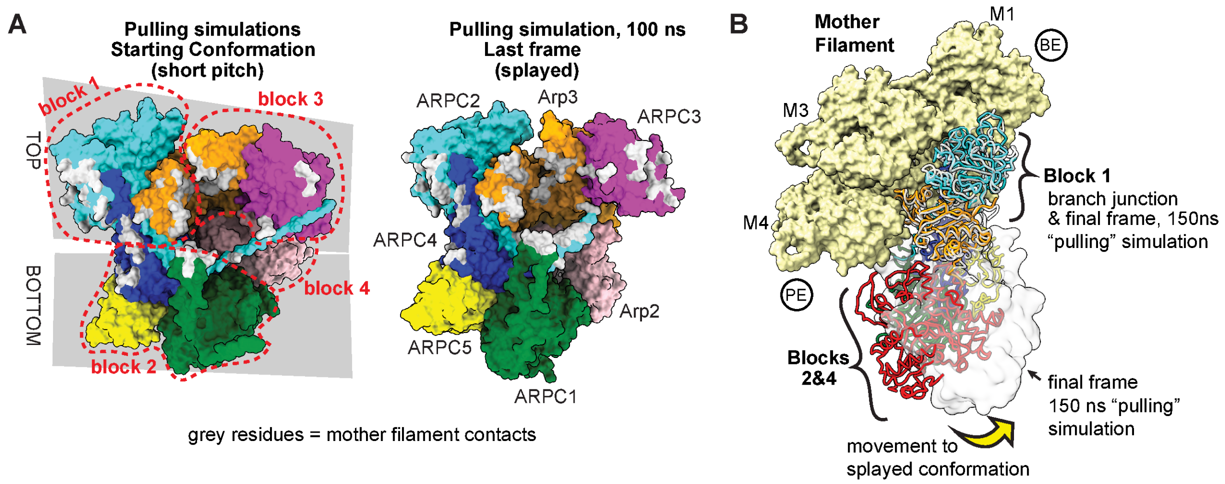

A
Pulling simulations
Starting Conformation
(short pitch)
block 1
block 3
TOP
BOTTOM
block 4
block 2
Pulling simulation, 100 ns
Last frame
(splayed)
ARPC2
Arp3
ARPC3
ARPC4
Arp2
ARPC5
ARPC1
grey residues = mother filament contacts
B
Mother
Filament
M1
BE
M3
M4
Block 1
branch junction
& final frame, 150ns
"pulling" simulation
PE
Blocks
2&4
final frame
150 ns "pulling"
simulation
movement to
splayed conformation

Figure 2.13: (A) (Left panel) surface representation of Arp2/3 complex rendered using the starting coordinates of the 100 ns pulling simulation. The four rigid blocks that move independently when the complex undergoes subunit flattening (see Video S1) are outlined with red dashes. The "top" and "bottom" rigid blocks that move independently when the clamp twists are indicated with gray boxes behind the complex. Residues of Arp2/3 complex that contact the mother filament at the start of the simulation (PyContact calculated contact score >1) are colored gray. (Right panel) Same as left panel except surface representation is rendered from the final frame of the simulation and residues that have an average contact score > 1 over the last 1 ns of the simulation are colored gray. Arp2/3 complex is in the splayed conformation at the end of the simulation. (B) Comparison of mother filament binding contacts of activated (flattened, short pitch) Arp2/3 complex (7TPT) to those of the splayed Arp2/3 complex (final frame of MF-bound 150 ns pulling simulation—rendered in gray ribbon or transparent gray surface). The splayed Arp2/3 complex was modeled onto the mother filament by superposing block 1 onto block 1 in the branch junction model. Yellow arrow shows movement of blocks 2 and 4 stimulated by clamp twisting. Block 3 is omitted for clarity. BE: Barbed end of mother filament. PE: pointed end of mother filament.

To assess how the conformation of Arp2/3 complex influences its contacts with the mother filament, we analyzed the surface area buried with the mother filament during the pulling simulations. Because BSA is roughly correlated to the energy of the binding interface [123], we reasoned that this analysis could provide insight into how binding to actin filaments might influence the short pitch conformational switch.

The total interaction area with the mother filament is relatively constant throughout the simulations, with an average interface area of 3880 Å^2 , 3971 Å^2 , and 3995 Å^2 during the 60, 100, and 150 ns simulations, compared to 3884 Å^2 for the branch junction structure [90] (Fig. 2.12C). Furthermore, the residues on Arp2/3 complex that contact the mother filament are nearly identical at the beginning of the simulations, when the complex is in the short pitch conformation, as they are at the end of the simulation, after the complex has been pulled into the splayed conformation (Fig. 2.13A and 2.19D). Examination of the trajectory revealed how contacts to the mother filament are maintained in the splayed conformation. As the clamp twists, the top portion of the complex (Arp3, ARPC3, and the globular domain of ARPC2) stays anchored to the mother filament and maintains the same or similar contacts to those it makes in an activated conformation (Fig. 2.13B and Video S5). The two rigid blocks shown in Fig. 2.13B that make up the bottom of the complex (2 and 4) rotate away from the filament as the clamp twists into the splayed conformation, but these blocks contact the filament almost exclusively through flexible segments that can engage the filament regardless of the conformation of Arp2/3 complex (see below). Therefore, these simulations suggest that clamp rotation and movement of Arp2/3 complex into the splayed conformation may not significantly influence its interface with the mother filament, in agreement with previous modeling experiments [90].

2.2.7 Flexible segments in Arp2/3 complex maintain conformation-insensitive contacts with the mother filament of actin

In addition to rigid block 1 and 3 in the top of the complex, flexible segments on ARPC1 and ARPC2 from the bottom of Arp2/3 complex remain attached to the mother filament as the complex moves from the short pitch to the splayed conformation. The ARPC1 insert forms a short $\alpha-$helix that binds to a hydrophobic groove in actin filament subunit M4 [89–91]. We showed previously that because the ARPC1 insert is flexible, it can stay bound to the filament throughout the pulling simulation even as the globular portion of ARPC1 moves away from the mother filament [90]. The ARPC1 insert helix has several conserved residues and buries an average of 434, 349, and 415 Å^2 throughout the 60, 100, and 150 ns pulling simulations, respectively (Fig. 2.14A-C and Video

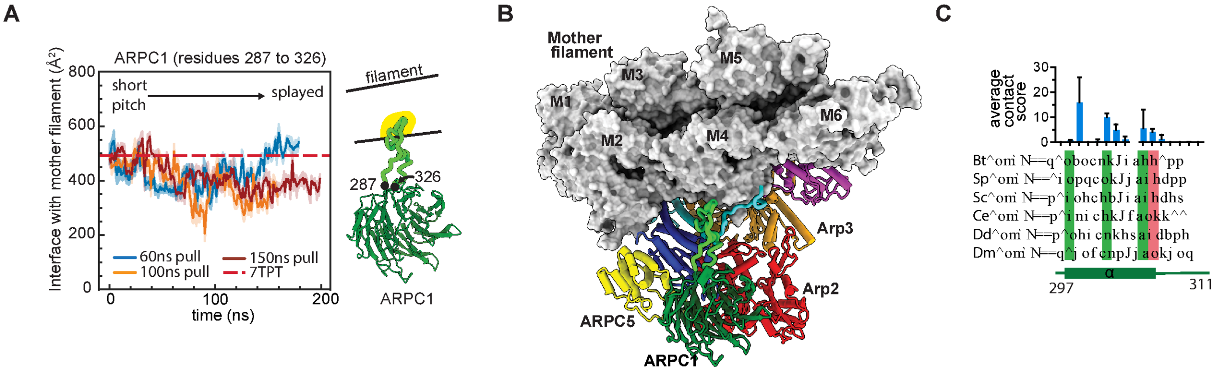


Figure 2.14: (A) Plot of the interaction area of ARPC1 residues 287 to 326 with the mother filament versus simulation time for all pulling simulations. Dashed line shows the corresponding interaction areas in the branch junction structure 7TPT. (B) Surface and cartoon representation of Arp2/3 complex bound to the mother filament in the last frame of the 100 ns pulling simulation. The two flexible segments from the bottom half of the complex (ARPC1 287–326, green, and ARPC2 281–300, cyan) are shown in thicker cartoon representation. (C) Sequence alignment of the ARPC1 insert sequence from a diverse range of species showing conserved hydrophobic (green) and acidic (red) residues. The average contact score over the course of all three pulling simulations is plotted above the sequence for each residue. Error bars: standard deviation.

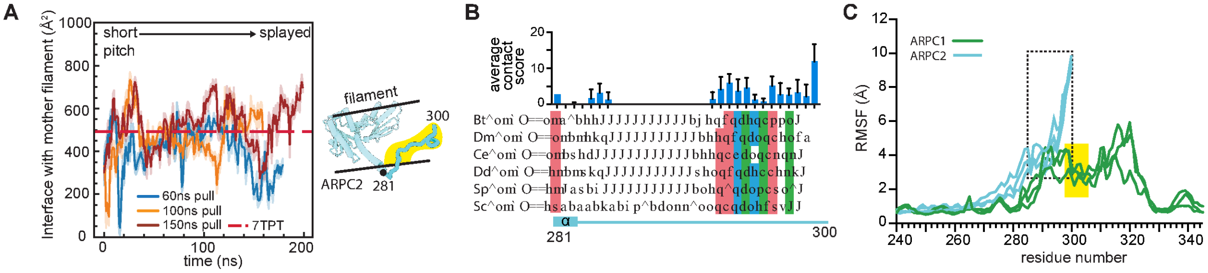


Figure 2.15: (A) Plot of the interaction area of ARPC2 residues 281 to 300 with the mother filament versus the simulation time for all pulling simulations. (B) Sequence alignment of ARPC2 C-terminal extension from a diverse range of species showing conserved hydrophobic (green), basic (red), or hydrophilic (cyan) residues. The average contact score over the course of the three pulling simulations is plotted above the sequence for each residue. Error bars: standard deviation. (F) Plot of the root mean squared fluctuation (RMSF) from the initial conformation for backbone atoms of the ARPC1 insert (green) or the ARPC2 C terminus (cyan) over the entire simulation, plotted separately for each pulling simulation. Black dashed rectangle highlights RMSF values for residues in ARPC2 C-terminal extension with the closest contacts to the mother filament. Yellow rectangle highlights RMSF values for residues in the ARPC1 insert with the closest contacts to the mother filament.

S6).

The C-terminal extension in ARPC2 also remains bound to the mother filament throughout most of the simulation. This segment is disordered in all structures except the recent cryoEM structure of S. pombe Arp2/3 complex at the branch junction [91]. Recent coarse-grained simulations suggested it may play a role in stabilizing interactions with the mother filament [107]. Because it extends from the C terminus of subunit ARPC2—which moves very little during the transition from the splayed to short pitch conformation—the extension stays positioned at the mother filament interface during activation, burying an average of 404, 471, and 525 Å^2 throughout the entire 60, 100, and 150 ns pulling simulations, respectively. Several residues in the C-terminal extension of ARPC2 are broadly conserved (Fig. 2.15B), suggesting the potential importance of this interaction. However, while the extension maintains contact with the mother filament throughout the trajectory, it remains relatively flexible, binding several different surfaces during the simulation (Fig. 2.15C and Video S6). This observation is consistent with previous structural data showing this segment is either completely disordered or has weak electron density, even when bound to the mother filament [90, 91].

2.3 Discussion

Here we use steered and unbiased all-atom MD simulations to investigate the conformational pathway of WASP-mediated Arp2/3 complex activation. WASP stimulates movement of the Arps into the short pitch conformation [93, 94, 124], but WASP was not included in any of our simulations, which instead focused on the role of actin filaments in activation. Using simulations to investigate the influence of WASP on activating structural changes will be an important next step, especially considering that while some experiments indicate that the primary function of WASP and WASP recruited actin monomers is to stimulate the short pitch conformation [93, 125], other experiments point to an additional WASP-mediated function in the activation process [94]. In addition, our studies included only one nucleotide state in the Arps, so it will be important to use simulations to investigate how other nucleotide states influence activating conformational changes. These studies will provide insights into why the nucleotide is required for activation and how it controls the stability of branches [116, 126–128].

A key conclusion from this work is that the two main activating conformational changes in Arp2/3 complex, movement into the short pitch conformation and subunit flattening, are unlikely to be concerted. Specifically, we show that the complex can adopt twisted states of Arp2 and Arp3 when the Arps are arranged into the short pitch conformation. These observations support a multistep model for activation, in which flattening and movement into the short pitch conformation can occur in separate steps and can be stimulated by different activating factors. An activation pathway that can be separated into multiple distinct steps has important implications for regulating the complex. First, it provides a mechanism to program triggered WASP release into the reaction mechanism. Previous experiments have shown that WASP is released before or concurrently with nucleation [129]. This feature of the activation mechanism is thought to prevent unproductive connections between growing actin networks and membrane-bound WASP, which decrease pushing forces of the network against the membrane [129]. In the multistep activation pathway, WASP (and actin monomer binding) could stimulate movement into the short pitch conformation, and subunit flattening stimulated by actin filaments could stimulate WASP release.

Another important advantage of a multistep activation model is that it would permit switch-like behavior of Arp2/3 complex at a wide range of concentrations of actin filaments, WASP and WASP-recruited actin monomers [121]. In contrast, concerted models of multisignal activation pathways have switch-like behavior only at low concentrations of activators; moderate to high concentrations cause rheostat-like activation, where each activating factor can trigger some activity on its own [121]. Therefore, a multistate model would allow for tighter regulation of Arp2/3 complex in cellular contexts where the local concentrations of activated WASP, actin monomers, and preformed actin filaments are high. We note that while our data support a multistep model, it is possible that there is some conformational coupling between the activation steps, so that the mechanism has some aspects of both a multistate and a cooperative model. This coupling could explain the reported cooperativity between WASP and actin filaments in binding to the complex [98, 99].

Our data points to the existence of stable intermediate conformations of Arp2/3 complex during the activation process. Specifically, we show that Arp2 and Arp3 can move into or part way toward the twisted conformation even with the complex in the short pitch state. Because of the limits on computation time, we were not able to assess whether the short pitch twisted states persist over periods longer than a microsecond. We note that during the simulations, Arp2/3 complex never adopted a state in which both Arp2 and Arp3 were fully twisted and short pitch (Fig. 2.5A and Fig. 2.5B). However, we were able to create a model of a fully short-pitch/twisted conformation by superposing the half of Arp2 and Arp3 bound to the clamp subunits from the active structures with twisted Arp2 and Arp3 with only minimal steric clash, which could be relieved through side chain minimization (Fig. 2.20). Therefore, we anticipate that with longer simulation times, this fully twisted/short-pitch conformation would be adopted. We note that we cannot eliminate the possibility that the short pitch twisted state we observe here is merely a transiently populated, high energy state.

In contrast, when we attempted to model a completely splayed/flattened conformation of the complex using the same procedure, we found clashes that could not be relieved by sidechain minimization or backbone remodeling of known flexible regions (Fig. 2.20). This may indicate that the flattened conformation is not stable when Arp2/3 complex is in the splayed conformation. Therefore, while there is currently no evidence for an obligatory sequence of binding events by activating factors, steric effects may require that Arp2/3 complex moves out of the splayed conformation before Arp2 or Arp3 flatten. Given that flattening of the Arps aligns residues in the nucleotide binding cleft for ATP hydrolysis [90, 92, 95, 130], stimulating this step only after the subunits are already aligned into a filament-like short pitch arrangement may help prevent unproductive hydrolysis of the nucleotide.

Our simulations showed that short pitch Arp2/3 complex is stable for at least a microsecond even without interactions with actin filaments or WASP. This was not surprising given the typical rates of rigid body motions of large groups of atoms (μs-ms) [120]. We expect that with longer simulation times, the complex would relax into a splayed conformation. In contrast, movement of the Arps from flattened to twisted conformations occurred within the microsecond simulations. Although this transition was context-dependent (e.g., it did not occur in Arp3 when the complex was filament-bound), it suggests a lower energy barrier separates movement from flat to twisted states than short pitch to splayed conformational states. Because both subunit twisting and movement into the splayed conformation would be expected to favor branch disassembly, this observation may have important implications for understanding how proteins like GMF and Coro7 stimulate branch disassembly upon binding to the complex [131, 132].

Using steered all-atom simulations, we showed that contacts between Arp2/3 complex and mother filaments remain largely unchanged as Arp2/3 complex transitions from the splayed to the short pitch conformation, suggesting contacts with actin filaments do not trigger the short pitch conformational change. These observations are consistent with crosslinking experiments using dual-cysteine engineered S. cerevisiae Arp2/3 complex, which showed that actin filaments do not stimulate the short pitch conformation [100]. However, these results differ from Förster resonance energy transfer measurements on S. pombe Arp2/3 complex, which showed that probes on the C termini of Arp2 and Arp3 have greater Förster resonance energy transfer efficiency when the complex binds actin filaments [133]. Empirical methods and the simulations presented here indicate the C termini of the Arps exhibit flexibility that could influence interpretation of these measurements [90, 95, 134]. Analysis of the recent cryo-EM structure of S. pombe Arp2/3 complex at branch junction suggested that movement into the splayed conformation decreases the interaction interface with the mother filament by ~35%, indicating an increased binding energy for the activated state that could be used to stabilize the short pitch conformation [91]. A key difference in the S. pombe Arp2/3 complex analysis is that it was assumed that the ARPC1 insert and the ARPC2 extension—which are disordered in the structure of inactive S. pombe Arp2/3 complex [92]—do not contact the mother filament when the complex binds filaments in an inactive state. Our data here suggest that both of these segments make conformation-insensitive contacts that provide over ~800 Å^2 of BSA with the mother filament. Nonetheless, interaction surface areas are only roughly correlated with binding energies [123], and small differences in the contacts could cause significant energetic differences. Therefore, additional biochemical/biophysical methods to probe the relationship between the conformation and binding states of Arp2/3 complex will be an important next step.

2.4 Supporting Information

2.4.1 Simulation Setup

2.4.1.1 Construction

Four systems were created for MD simulation studies. The “branch junction” consists of Arp2/3 complex, the mother filament (10 actin subunits), and daughter filament (4 actin subunits) with initial configurations derived from the Ding et al. structure, 7TPT [90]. The “branch junction without daughter filament” is the same but only contains Arp2/3 complex and the mother actin subunits. “Free Arp2/3 complex from branch junction” contains only Arp2/3 complex in the active configuration from 7TPT. “Free inactive Arp2/3 complex” complex is constructed from the X-ray crystal structure of Arp2/3 complex in an inactive state and bound to the inhibitor protein GMFγ (4JD2) [113]. The nucleotide states for each system are described in the main text. A Mg^{2+} ion was modeled in the nucleotide binding clefts based on its position in the structures. All residues of each actin or Arp2/3 complex subunit were included in the simulations. The following residues were missing from 7TPT: Arp3, 1 to 2, 417 to 418; Arp2, 1 to 3, 389 to 394;

ARPC1, 1, 365 to 372; ARPC2, 284 to 300; ARPC3, 1, 175 to 178; ARPC5, 1 to 9, 27 to 34. The following residues were missing from 4JD2: Arp3, 1 to 2, 40 to 51, 356 to 359, 417 to 418; Arp2, 1 to 3, 43 to 50, 389 to 394; ARPC1, 289 to 318; ARPC2, 284 to 300; ARPC3, 1, 36, 102, 121, 175 to 178; ARPC4, 1 to 2; ARPC5, 1 to 8, 29 to 30, 34 to 35; Missing residues were modeled by hand using Coot or PyMol or automatically with Modeller [135, 136]. The C terminus of ARPC2 was modeled as an extended $\beta-$strand that protrudes into the solvent. These systems were constructed and equilibrated as previously described [90, 137]. Briefly, all systems were built using VMD 1.9.3 and parameterized using a CHARMM22+CMAP forcefield with TIP3P water [118]. Bound nucleotides and surrounding water were modeled in the nucleotide binding cleft of each actin or Arp2 and Arp3 protein as previously described [138, 139]. Each system was solvated in water such that there is at least 1 nm of water in each direction surrounding the protein. Potassium (K^+) and Chloride (Cl^-) ions were added so that each system was neutralized, and the resulting concentration of salt was 50 mM.

2.4.1.2 Equilibration

Systems were equilibrated in several steps using NAMD 2.14 [140].

(1) Minimization: Energy minimization was carried out in four stages, each for 1000 time steps with 10 kcal/mol/Å^2 restraints on different groups of atoms, as previously described [138]. For the first stage, the protein, nucleotide, nucleotide-bound magnesium ions, and nucleotide waters (within 5 Å of the magnesium ion) were restrained. In the second stage, the protein backbone, nucleotide, nucleotide-bound magnesium ions, and nucleotide-proximal waters were restrained. Third, the nucleotide, nucleotide-bound magnesium ions, nucleotide-proximal waters, and finally, only nucleotide-bound magnesium ions and nucleotide-proximal waters were restrained.

(2) Heating/Equilibration: 10 kcal/mol/Å^2 restraints were applied to the protein backbone, nucleotide, nucleotide-bound magnesium ions, and nucleotide-proximal waters (As in the second stage of minimization). The system was heated from 0 K to 310 K over 100 ps, using a Langevin

Thermostat in the NVT ensemble. Immediately after heating, the system was run in the NPT ensemble with a target pressure of 1 atm using the Langevin piston Nose-Hoover method in NAMD 2.14. Equilibration with the same restraints as the heating stage was carried out over five cycles of 200 ps, with each stage reducing the restraint coupling by half. After the first five cycles, an additional equilibration of 200 ps was run with constraints set to 0.1 kcal/mol/Å^2 . A final cycle of equilibration was run for 400 ps without any constraints on any groups of atoms.

2.4.1.3 Production

Unbiased MD For the production runs, a python script (`psf2itp`) created by the CHARMM-GUI developers was used to convert the equilibrated system topologies and coordinates from NAMD format to GROMACS compatible files with the same forcefield and MD parameters as in the equilibration step [141]. Each system was relaxed for an additional 10 ns using GROMACS 2018 [142]. To prevent rotation of the simulation box; 10 kcal/mol/Å^2 constraints were applied to backbone atoms of the four terminal mother actin subunits (chains L, M, T, and U in 7TPT).

Biased MD PLUMED plugin was used to implement steered MD in our simulations [143]. The MOVINGRESTRAINT function was used to apply a time-dependent harmonic potential starting from the Arp2/3 complex short-pitch configuration to the splayed configuration for the two systems without daughter actin. The collective variables used for the biasing were the COG distances between each of Arp2/3 complex domains, as defined by their Cα positions. The target values of the collective variables were calculated from the crystal structure of 4JD2 [113]. The biasing simulations used a force constant of 10,000 kJ/mol/A2 for different pulling durations- 60 ns, 100 ns, and 150 ns. In addition, the pulling simulations were continued by biasing the DRMSD, which is the distance RMSD between the Cα of the domains of each frame in the trajectory to the Cα of the domains of the 4JD2.

2.4.2 DATA ANALYSIS

Simulation analysis was performed using the `mdtraj` library (version 1.9.4) in python [144]. Additional contact analysis and scoring was carried out using GUI-based interface PyContact [145]. Trajectory and structure files were also visualized in VMD 1.9.3, PyMOL, and ChimeraX [146].

Subdomains (1–4) of Arp2 and Arp3 were defined as shown in Table 2.1, with only backbone atoms used in the COG calculations. Each of the subdomains were mapped into single coarse-grained beads using the custom mapping script written in python and then used for our calculations (Available on the manuscript Github). The subunit twisting/flattening was measured by computing the dihedral angle defined by the four subdomains (subdomain 2, 1, 3, and 4). Subdomains 3 and 4 were used to calculate the distance between Arp3 and Arp2 so that the simulation data could easily be compared to X-ray crystal structures, in which subdomains 1 and 2 of Arp2 are often partially or completely disordered [87, 113].

To define clamp twisting, we used the centers of geometry of four small sets of backbone atoms that were close to atoms used in Shabaan et al. [92] (Table 2.2). The custom mapping script described above was used to map the all-atom selections into coarse grained atoms. The clamp twisting angle was measured by computing the dihedral angle of Bead 1–Bead 2–Bead 3–Bead 4. For BSA calculations, we excluded noninteracting subunits of the mother actin filament (MA1, MA0, MA5, MA7, and MA8 using nomenclature in Ding et al. [90]) in our calculations to avoid high computational cost and memory. The Shrake and Rupley algorithm in the mdtraj library was used for computing the solvent accessible area (SASA) of each residue of the selected group [144, 147]. The SASAs were used to compute the approximate BSA using the equation:

$$BSA = \frac{1}{2}\left(SASA_1 + SASA_2 - SASA_{12}\right) \tag{2.1}$$

Where $SASA_1$ is the SASA of the first group (excluding the second), $SASA_2$ is the SASA of the second group (excluding the first), and $SASA_{12}$ is the SASA including both groups. Clashes in the models presented in Fig. 2.19 were calculated in ChimeraX using the default settings, and side chain minimization was carried out in Phenix [146, 148]. Root mean squared fluctuation calculations were carried out in VMD [149].

2.4.3 Data Availibility

Scripts and input files used for system construction, data collection and analysis, and figure preparation can be found in the Hocky group GitHub repository. Simulation data saved every 100 ns can also be found in the GitHub repository. All other input or data files are available upon request.

2.4.4 Supplemental Data

Captions for supplemental videos

Video S1: *Activation of Arp2/3 complex requires two major sets of conformational changes.* Morph of Arp2/3 complex using the inactive Arp2/3 complex crystal structure 4JD2 as a starting point and the branch junction structure of Arp2/3 complex (7TPT) as the end point. The two major types of activating conformational changes can be measured by dihedral angles (defined by centers of geometry of groups of backbone atoms in the complex – see Data Analysis). Clamp twisting is shown with a yellow dihedral, whereas subunit flattening is shown with the grey dihedral (Arp2) or the cyan dihedral (Arp3). Note that only residues present in both the structures are modeled in the morph. Arp3, orange; Arp2, red; ARPC1, green; ARPC2, cyan; ARPC3, magenta; ARPC4, blue; ARPC5, yellow.

Video S2: *In steered simulations the clamp bends into a conformation not observed in experimentally determined structures.* Video shows two views of the 60 ns pulling simulation overlaid onto the globular portion of ARPC2 in the branch junction structure. Arp3 from the branch junction structure is shown in orange. The Arp3 flattening/twisting angle (φ_{Arp3}) for the branch junction structure (7TPT), the inactive structure (4JD2), and the simulation are shown in red, yellow, and cyan, respectively. The ARPC4 subunit is shown for the branch junction structure (red), the inactive structure (yellow) and the simulation (cyan). Bending of the clamp repositions Arp3 so that it can retain contacts with the mother filament despite being in the twisted conformation.

Video S3: *Twisting of Arp2 closes the barbed end groove and weakens its interactions with the daughter filament.* Ribbon diagram of the branch junction trajectory showing the Arp2 subunit (red) and the actin D2 subunit (grey). The x_1 distance (cyan) measures the uncurling of the W-loop. The x_2 distance (green) measures the distance between M44 Cα in actin D2 and G172 Cα in the Arp2 W-loop. The x_2 distance increases after the sidechain of Met44 is ejected from the pocket in the side of the barbed end groove.

Video S4: *A steered simulation with 60 ns of pulling generates a splayed filament-bound Arp2/3 complex structure.* Video shows the 60 ns pulling simulation from start to finish. Frames were taken at approximately every 13 ns of the trajectory to make this video. The distance between the COG of subdomains 3 and 4 of Arp3 and subdomains 3 and 4 of Arp2 (grey spheres) is shown as a yellow dashed line. Arp3, orange; Arp2, red; ARPC1, green; ARPC2, cyan; ARPC3, magenta; ARPC4, blue; ARPC5, yellow; actin, gray.

Video S5: *Splayed Arp2/3 complex stays anchored to the mother filament and maintains contacts it makes in an activated conformation.* Ribbon representation of the 150 ns pulling simulation from start to finish. Frames were output every ~7.5 ns for this video. The distance between subdomains 3 and 4 of Arp3 and subdomains 3 and 4 of Arp2 is shown in white. The clamp twisting dihedral angle (α) is shown in red. Arp3, orange; Arp2, red; ARPC1, green; ARPC2, cyan; ARPC3, magenta; ARPC4, blue; ARPC5, yellow; actin, gray

Video S6: *Flexible segments in splayed Arp2/3 complex maintain contacts with the mother filament.* Video of ribbon representation of the 150 ns pulling simulation from start to finish. Frames were output every ~7.5 ns for this video. The distance between subdomains 3 and 4 of Arp3 and subdomains 3 and 4 of Arp2 is shown with a green dashed line. The distance between the Cα atom of Ile345 (yellow) on the surface of the mother filament (subunit M4) and the center of geometry of the globular domain of ARPC1 and shown with a magenta dashed line. The distance between the ARPC1 helix and the surface of the mother filament (ARPC1 Phe302 Cα to actin subunit M4 Ile345 Cα) is shown with a yellow dashed line. The ARPC1 insert and the ARPC2 C-terminal extension are shown in thick cartoon representation.

	Arp3	Arp2
Subdomain 1	6-32, 78-153, 375-408	7-33, 74-150, 352-387
Subdomain 2	33-37, 60-77	34-38, 55-73
Subdomain 3	154-196, 295-344, 362- 374	151-185, 277-326, 339- 351
Subdomain 4	197-282	186-265

Table 2.1: Definition of subdomains 1-4 in Bos taurus Arp3 and Arp2. Backbone atoms in the listed residues were used for center of geometry calculations.

	Subunit	Residues
Bead 1	ARPC4	14-50, 63-73, 117-132
Bead 2	ARPC4	3-13, 51-62, 74-77, 135-142
Bead 3	ARPC2	125-262
Bead 4	Arp3	6-32, 78-153, 375-408, 33-37, 60-77

Table 2.2: Definition of centers of geometry in Bos taurus Arp3, ARPC2 and ARPC4 used for calculating the clamp twist angle. Backbone atoms in the listed residues were used for center of geometry calculations.

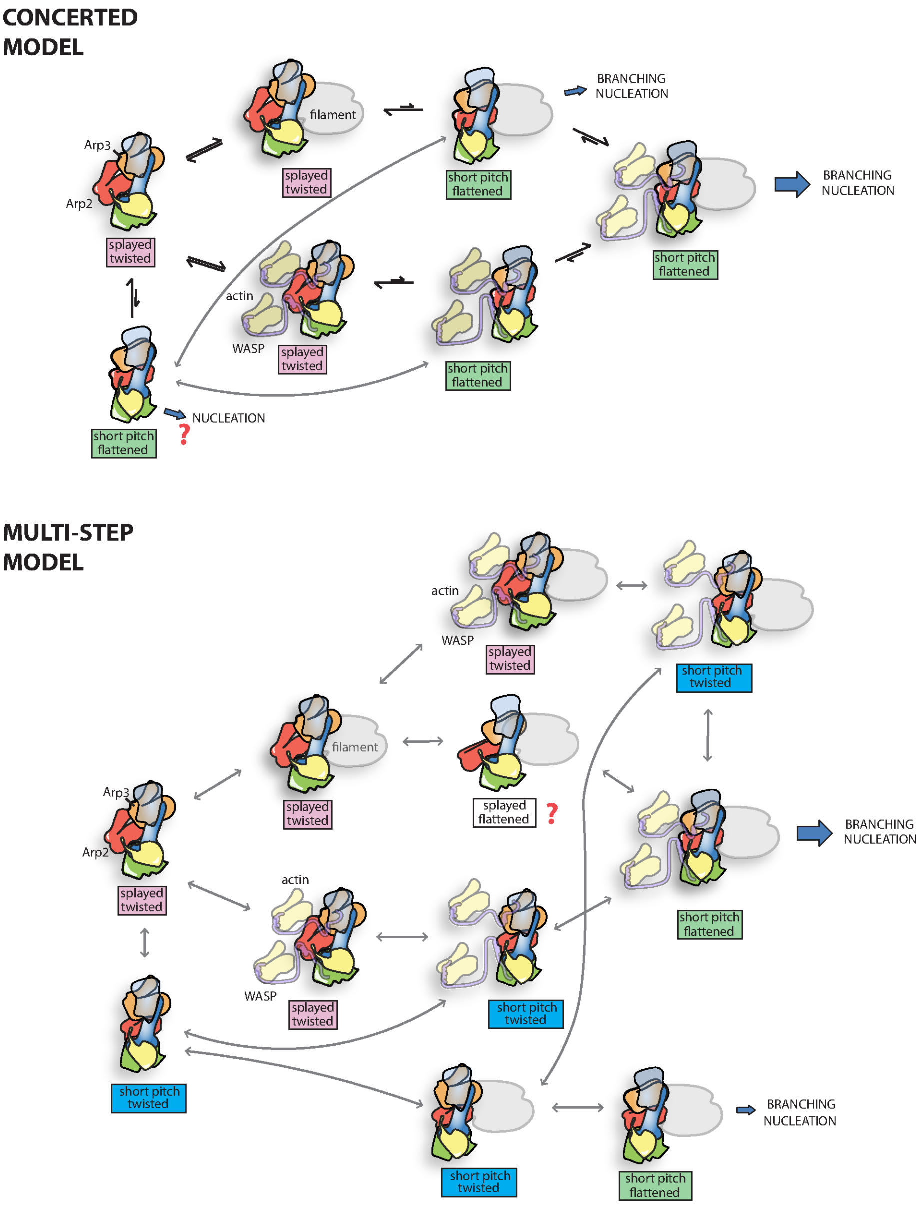

CONCERTED MODEL
Arp3
Arp2
splayed twisted
filament
splayed twisted
short pitch flattened
BRANCHING NUCLEATION
actin
WASP
splayed twisted
short pitch flattened
short pitch flattened
BRANCHING NUCLEATION
short pitch flattened
NUCLEATION
?
MULTI-STEP MODEL
actin
WASP
splayed twisted
short pitch twisted
filament
splayed twisted
splayed flattened
?
Arp3
Arp2
splayed twisted
short pitch flattened
BRANCHING NUCLEATION
actin
WASP
splayed twisted
short pitch twisted
short pitch twisted
short pitch twisted
short pitch flattened
BRANCHING NUCLEATION

Figure 2.16: These schematics show the relationship between conformation and binding state in more detail than in Fig. 2.1. Key additions relative to Fig. 2.1 are: the short pitch conformation can be weakly adopted even in the absence of WASP [77, 78], Arp2/3 complex can adopt the short pitch or splayed conformation when bound to WASP [77, 78], and Arp2/3 complex from some species shows NPF-independent activity [79]. It is unknown (red question mark, concerted model) whether the NPF-independent activity of Arp2/3 complex requires preformed filaments, though filaments activate short-pitch crosslinked S. cerevisiae Arp2/3 complex [80]. Neither schematic shows individual WASP binding steps. Biochemical data indicate WASP binds more tightly to the Arp2 site [81], and that there may be differences in the contribution of WASP binding at each site to stimulating the short pitch conformation [78]. The splayed/flattened state (red question mark, multi-state model) may not be adopted because of steric clash (see Discussion).

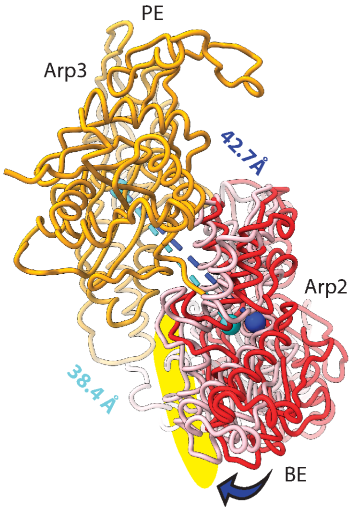


Figure 2.17: Ribbon representation of Arp3 (orange) and Arp2 (red) from the branch junction structure (7TPT). Arp2 from the MF-bound Arp2/3 complex simulation at 0.76 μs is shown in pink and is positioned by overlaying the Arp3 backbone from the simulation onto Arp3 from the branch junction structure. The centers of geometry of subdomains 3 and 4 of Arp3 and Arp2 from each of the structures are shown as spheres and connected with blue (branch junction) or cyan (simulation) dashed lines. The yellow region shows where actin D1 would clash with Arp2 in the tilted conformation observed in the simulation.

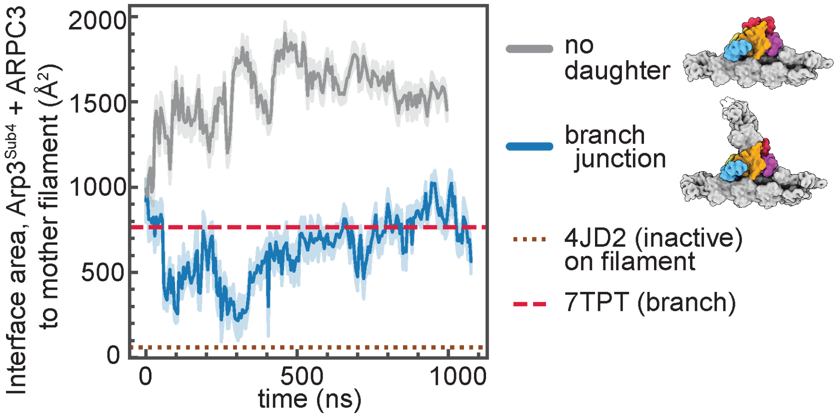


Figure 2.18: Plot of the area of subdomain 4 of Arp3 and ARPC3 buried upon interaction with the mother filament versus simulation time. Buried surface area of the same regions in the branch junction structure (7TPT) and a model of inactive Arp2/3 complex (4JD2) on an actin filament are shown as dashed or dotted lines, respectively.

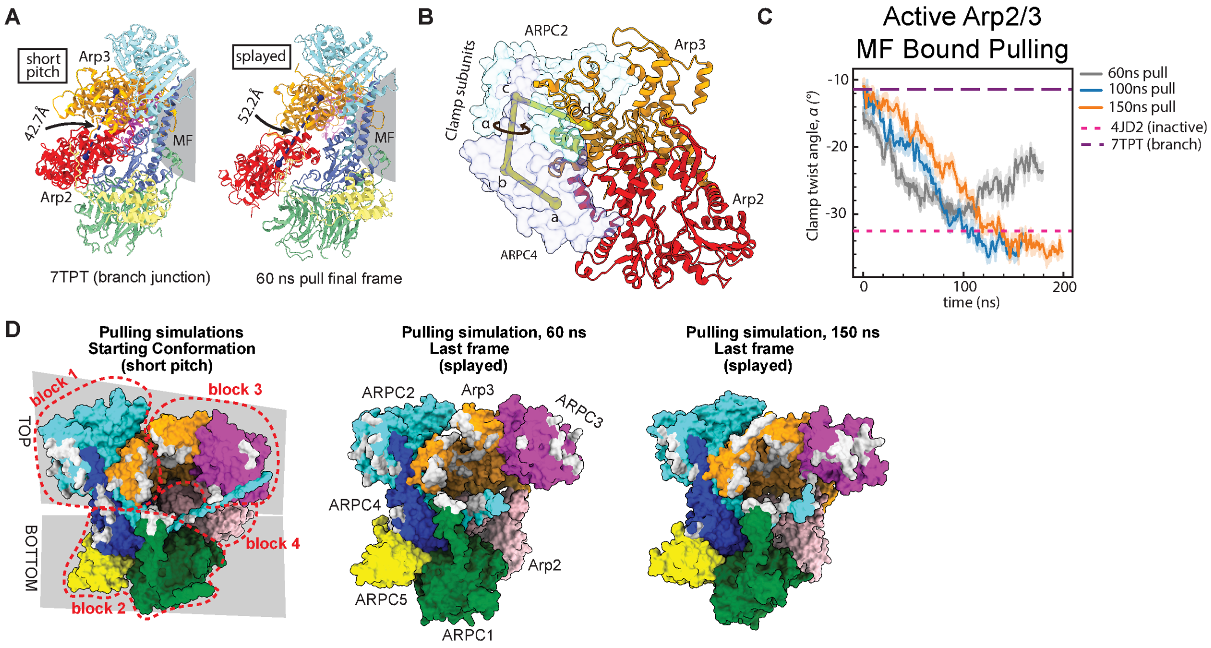
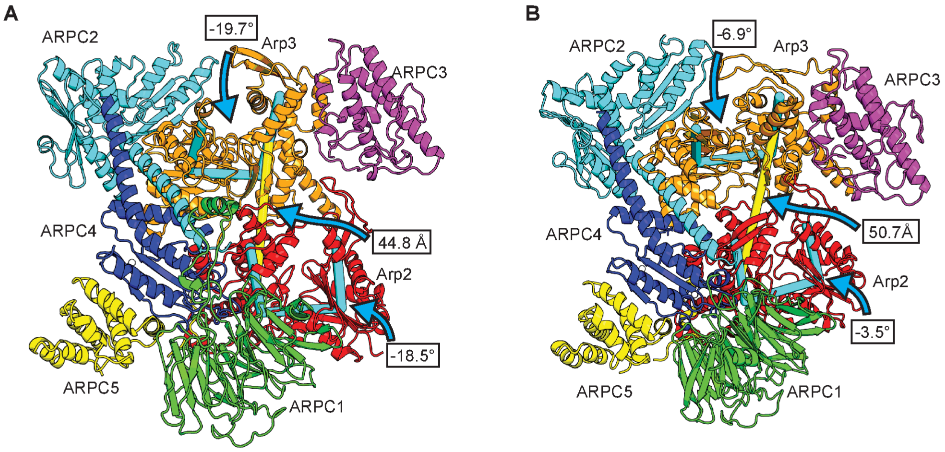

A
short pitch
Arp3
42.7Å
MF
Arp2
7TPT (branch junction)
splayed
52.2Å
MF
60 ns pull final frame
B
ARPC2
Arp3
Clamp subunits
c
α
d
b
a
Arp2
ARPC4
C
Active Arp2/3
MF Bound Pulling
Clamp twist angle, α (°)
-10
-20
-30
0
100
200
time (ns)
60ns pull
100ns pull
150ns pull
4JD2 (inactive)
7TPT (branch)
D
Pulling simulations
Starting Conformation
(short pitch)
block 1
block 3
TOP
BOTTOM
block 4
block 2
Pulling simulation, 60 ns
Last frame
(splayed)
ARPC2
Arp3
ARPC3
ARPC4
Arp2
ARPC5
ARPC1
Pulling simulation, 150 ns
Last frame
(splayed)

Figure 2.19: (A) Ribbon diagram of Arp2/3 complex from the branch junction structure and from the 60 ns pulling simulation showing the distances between subdomains 3 and 4 of Arp3 with subdomains 3 and 4 of Arp2. This distance measures movement of the complex into the short pitch position. The approximate position of the mother filament is shown in shaded grey. MF: Mother Filament. (B) Ribbon representation of Arp2 and Arp3 from the branch junction structure with the clamp subunits shown as semi-transparent surface. The dihedral angle (α) used to measure clamp twisting is shown in yellow. Residues used to define the centers of geometry marked a, b, c, and d are listed in the Supplementary Materials. (C) Plot of the clamp twisting angle (α) versus simulation time for all three pulling simulations. The clamp twisting angle for the branch junction structure (7TPT) and for an inactive Arp2/3 complex structure (4JD2) are shown as dashed lines. (D) (Left panel) Surface representation of Arp2/3 complex rendered using the starting coordinates of the MF-bound Arp2/3 complex 100 ns pulling simulation. The four rigid body blocks that move independently when the complex undergoes subunit flattening and clamp twisting (see Video S1) are indicated with red dashes. The "top" and "bottom" rigid body blocks that move independently when the clamp twists are indicated with grey boxes behind the complex. Residues of Arp2/3 complex that contact the mother filament at the start of the simulation (PyContact calculated contact score >1) are colored grey. (Right panel) Same as left panel except surface representation is rendered from the final frame of the simulations and residues that have an average contact score > 1 over the last 1 ns of the simulations are colored grey. Note that the complex is in the splayed conformation at the end of the simulations.



Short pitch, twisted
(theoretical model)

Splayed, flattened
(theoretical model)

C

structure	**Number of clashes**
7TPT (branch)	13
4JD2 (inactive)	5
short pitch, twisted (side chain min.)	8
splayed, flattened (side chain min.)	53

Figure 2.20: (A) Cartoon representation of Arp2/3 complex modeled into the short pitch, twisted conformation. Arp3 was constructed by superposing the backbone atoms of (well-ordered) residues in subdomains 1 and 2 of Arp3 from the inactive structure (4JD2, Arp3:6-32,78-153,375-408,33-37,60-77) with the same atoms in the branch junction structure (7TPT). Subdomains 3 and 4 from the superposed inactive structure along with ARPC3 were then used to replace the corresponding residues in the branch junction structure. Arp2 was constructed by superposing the backbone atoms of well-ordered residues in subdomains 3 and 4 of Arp2 in the inactive structure (4JD2, Arp2: 186-265, 151-185, 266-326,339-351) with the same atoms in the branch junction structure (7TPT). Subdomains 1 and 2 from the superposed inactive structure were then used to replace the corresponding residues in the branch junction structure. The dihedral angles that measure subunit twisting/flattening in the Arps are shown in cyan. The distance between the center of geometry of subdomains 3 and 4 of Arp3 and subdomains 3 and 4 of Arp2 is shown in yellow. (B) Cartoon representation of Arp2/3 complex modeled into the splayed and flattened conformation. Arp3 in this model was constructed by superposing the backbone atoms of (well-ordered) residues in subdomains 1 and 2 of Arp3 from the branch junction structure with the same atoms in the inactive structure (4JD2). Arp2 in this model was constructed by superposing the backbone atoms of well-ordered residues in subdomains 3 and 4 of Arp2 with the same atoms in the branch junction structure (7TPT). (C) Table showing the number of clashes in the experimentally determined versus modeled structures. Clashes were identified with the clash command in ChimeraX [82].

CHAPTER 3

INFINITE SWITCH SIMULATED TEMPERING IN FORCE (FISST)

This chapter is adapted from Ref. [60]

Abstract

Many proteins in cells are capable of sensing and responding to piconewton-scale forces, a regime in which conformational changes are small but significant for biological processes. In order to efficiently and effectively sample the response of these proteins to small forces, enhanced sampling techniques will be required. In this work, we derive, implement, and evaluate an efficient method to simultaneously sample the result of applying any constant pulling force within a specified range to a molecular system of interest. We start from Simulated Tempering in Force, whereby force is added as a linear bias on a collective variable to the system's Hamiltonian, and the coefficient is taken as a continuous auxiliary degree of freedom. We derive a formula for an average collective-variable-dependent force, which depends on a set of weights, learned on-the-fly throughout a simulation, that reflect the limit where force varies infinitely quickly. Simulation data can then be used to retroactively compute averages of any observable at any force within the specified range. This technique is based on recent work deriving similar equations for Infinite Switch Simulated Tempering in Temperature, that showed the infinite switch limit is the most efficient for sampling. Here, we demonstrate that our method accurately samples molecular systems at all forces within a user defined force range simultaneously, and show how it can serve as an enhanced sampling tool for cases where the pulling direction destabilizes states of low free-energy at zero-force. This method is implemented in, and will be freely-distributed with, the PLUMED open-source sampling library, and hence can be readily applied to problems using a wide range of molecular dynamics software packages.

3.1 INTRODUCTION

Mechanical forces acting on the molecular scale play crucial roles across biology, from driving essential processes such as cell migration to determining the emergent macroscale properties of biological materials [6, 20, 150–152]. While the response of macroscopic systems to force can be measured by rheological techniques and often matched to theories of elasticity or viscous flow [153, 154], understanding the response of microscopic systems to force is more challenging. Significant progress has been made through pioneering single-molecule force spectroscopy studies, which have given insight into the folding landscape of proteins, the kinetics of protein-protein interactions, and the behavior of molecular motors [155–157].

Molecular dynamics (MD) simulations are often capable of representing the equilibrium behavior of a system, and therefore are a key tool to elucidate the detailed, molecular-scale picture of what underlies important chemical and biological processes [138, 158, 159]. Schulten and others pioneered the use of Steered Molecular Dynamics (SMD) to predict the behavior of molecules in single molecule pulling experiments, where an external force is applied within a molecular dynamics simulation in a way that mimics common experimental setups [57, 160]. Although these SMD simulations and the experiments they mimic are performed out-of-equilibrium, it is in principle possible to use non-equilibrium fluctuation theorems to extract equilibrium information from an ensemble of trajectories [160–163]. However, in order to observe empirically relevant structural changes (e.g. protein unfolding) within an achievable simulation time scale, SMD pulling must be performed with unphysically large forces applied to the system [164]. These large forces then lead to poor agreement with the experiments they were designed to simulate and a dependence of the result on the pulling rate [165]. To overcome these limitations, one can reduce the number of degrees of freedom and artificially smooth the free energy landscape by using coarse-grained models, which effectively decreases the timescale of the targeted process [166]. Alternatively, one could use enhanced sampling simulations in conjunction with SMD at lower pulling forces/rates to more quickly sample a molecule's conformations [167, 168]. These combined approaches are challenging because they require techniques for sampling non-equilibrium trajectories, and typically are more difficult to converge for large systems than standard equilibrium sampling methods [169].

In this work, *we focus only on systems under a small constant mechanical load*, on the order of a few to tens of piconewtons (pN), a regime known to initiate and drive many important biological processes [15]. Although we typically imagine that applying a pulling force will drive a system out of equilibrium, thermodynamically speaking, applying a small constant force simply creates a new, tilted, energy landscape on which the system will equilibrate [170]. Moreover, small applied forces are expected to be near the linear-response regime, and would simply change the weight of the conformations observed at equilibrium rather than drive large conformational changes often studied by single molecule force probes and SMD simulations. Because of these factors, standard equilibrium sampling methods such as Parallel Tempering should adequately probe the effect of these forces on the resulting conformational ensemble. *Here, we investigate a method in which small applied forces can be used to simultaneously obtain equilibrium information about a molecular system while also accelerating sampling.*

Martinsson *et al.* have recently developed a useful enhanced sampling method called Infinite Switch Simulated Tempering (ISST) [171]. They show that the most efficient way of performing simulated tempering, where temperature is a dynamical variable in the simulation, occurs in the limit where the temperature can change infinitely quickly. In this regime, an effective configuration-dependent temperature is learned and used to propagate the dynamics. Information about the system at any temperature in the chosen temperature range between T_{min} and T_{max} can be obtained *post facto* using weights calculated on-the-fly during the simulation.

In this work, we derive the force-equivalent of the ISST method, which we term FISST. FISST allows us to run simulations over a user defined range of forces, and by learning the 'weights' for each force on-the-fly, quantitatively reconstruct the probability density function of a given observable at any force within the force range; we therefore effectively gain information about the N_f different forces one wants to study within a single simulation. *While this method is derived using the same logic as ISST, the two methods are fundamentally distinct.* Importantly, because FISST is a collective-variable-based method, it only depends on an intensive quantity of the system, hence its effectiveness does not deteriorate with system size or dimensionality. We illustrate the performance of FISST for a number of test systems ranging in complexity, including a simple 2D analytical potential, a chain of beads with i/i+4 interactions that favors a degenerate left and right handed helical configurations at zero force, and deca-alanine in water. In addition, we attempt to quantify the amount of information gained using FISST over traditional equilibrium sampling methods and comment on the prospect of FISST as an enhanced sampling method. FISST is implemented as a module in the open-source PLUMED package [143, 172], a plug-in for many of the most popular simulation packages, and can therefore be immediately applied to virtually any system of interest.

3.2 Theory and Methods

3.2.1 Simulations under constant force

Assume that the system under zero force has the Hamiltonian:

$$H(\boldsymbol{p},\boldsymbol{q}) = \tfrac{1}{2}\boldsymbol{p}^T M^{-1}\boldsymbol{p} + U(\boldsymbol{q}), \tag{3.1}$$

where $\boldsymbol{q}$ and $\boldsymbol{p}$ represent the position and momenta of the particles in the system, M is the mass matrix, and $U(\boldsymbol{q})$ the potential. Then the system's Hamiltonian with a force can be written as:

$$H_F(\boldsymbol{p},\boldsymbol{q}) = \tfrac{1}{2}\boldsymbol{p}^T M^{-1}\boldsymbol{p} + U(\boldsymbol{q}) - FQ(\boldsymbol{q}). \tag{3.2}$$

where $Q(\boldsymbol{q})$ is a collective variable (CV), defined here as a function of particle positions (although this could be generalized). In this equation, a positive F corresponds to pulling (i.e. a larger Q will be preferred for $F > 0$). It is evident that as long as F does not vary in time, then any standard constant-temperature MD, MC, or enhanced sampling method can be applied to sample configurations from the equilibrium Boltzmann distribution with density $\propto e^{-\beta H_F}$, where $\beta = (k_B T)^{-1}$, k_B is Boltzmann's constant, and T is the temperature of the system.

3.2.2 Simulated Tempering in Force

One way to implement tempering in force is to do Hamiltonian Replica Exchange [173], with discrete forces applied to collective variables. To study N_F different forces in the range $F_{\min} < F_i < F_{\max}$, we would simulate N_F copies of our system with Hamiltonians given by:

$$H_i(\boldsymbol{p}, \boldsymbol{q}) = \tfrac{1}{2}\boldsymbol{p}^T M^{-1} \boldsymbol{p} + U(\boldsymbol{q}) - F_i Q(\boldsymbol{q}). \tag{3.3}$$

Monte Carlo exchanges between replicas are done periodically, with a Metropolis acceptance rate of $P_{\text{exchange}} = \min\{1, \exp(-\beta(F_i - Fj)(Q_i - Qj))\}$.

Alternatively, one could perform the equivalent of a continuous version of simulated tempering [171, 174], in which case F becomes a continuous extra degree of freedom. It is then possible to perform Langevin Dynamics (LD) such that the following probability density for each configuration $(\boldsymbol{q}, F)$ is sampled:

$$\rho(\boldsymbol{q}, F) = C^{-1}(\beta)\omega(F)e^{-\beta U(\boldsymbol{q}) + \beta F Q(\boldsymbol{q})}. \tag{3.4}$$

Here $\omega(F)$ is a weight function to be specified (more on this below) which is positive for $F_{\min} <$

$F < F_{\max}$ and zero outside that range (such that these forces are not accessible), and

$$C(\beta) = \int_{F_{\min}}^{F_{\max}} dF\,\omega(F) \int d\boldsymbol{q}\, e^{-\beta U(\boldsymbol{q})+\beta F Q(\boldsymbol{q})} \tag{3.5}$$

$$\equiv \int_{F_{\min}}^{F_{\max}} dF\,\omega(F) Z_q(F). \tag{3.6}$$

where we have defined the partition function

$$Z_q(F) \equiv \int d\boldsymbol{q}\, e^{-\beta U(\boldsymbol{q})+\beta F Q(\boldsymbol{q})}. \tag{3.7}$$

3.2.3 The Infinite Switch Limit

As discussed above, the arguments of Ref. [171] suggest that the most efficient sampling scheme occurs in the infinite switch limit, i.e. when the mass of the fictitious “force-momentum” becomes 0. In brief, the validity of the infinite switch limit follows from the Large Deviation Principle, which asserts that rare events, such as those associated with large changes in the system, become exponentially less likely as their magnitude increases [175, 176]. To better examine the theoretical framework, consider the following Hamiltonian,

$$H(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2}\boldsymbol{p}^T M^{-1}\boldsymbol{p} + U(\boldsymbol{q}) - FQ(\boldsymbol{q}) + \frac{p_F}{m_F} + \phi(F) \tag{3.8}$$

Here, p_F is the fictitious "force-momentum", m_F is the corresponding "force-mass", and $\phi(F)$ is a confining potential to ensure F is confined within the "force-ladder", ie $F_{\min} < F < F_{\max}$. The following Langevin scheme can be used to sample configurations according to Eq. (3.8),

$$\frac{d\boldsymbol{q}}{dt} = M^{-1}\boldsymbol{p} \tag{3.9}$$

$$\frac{d\boldsymbol{p}}{dt} = -\nabla U + F\nabla Q - \gamma\boldsymbol{p} + \sqrt{2\gamma\beta^{-1}}M^{1/2}\boldsymbol{\eta} \tag{3.10}$$

To implement the Large Deviation theory [175, 176] we define ν_T as an empirical measure of the dynamics in Eq. 3.9 and 3.10, at time T as follows,

$$\nu_T(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{T} \int_0^T dt\ \delta(\boldsymbol{q} - \boldsymbol{q}(t))\delta(\boldsymbol{p} - \boldsymbol{p}(t)) \tag{3.11}$$

From Donsker–Varadhan theory [177, 178], the empirical (Eq. 3.11) must satisfy the Large Deviation principle as $T \longrightarrow \infty$ with the rate functional given by,

$$I(\mu) = \int d\mu\ \frac{\mathcal{L}f}{f} \tag{3.12}$$

where $I(\mu)$ is the rate functional, μ is the empirical measure, $\mathcal{L}$ is an infinitesimal operator to evolve the system of equations (3.9) and (3.10), and f is a smooth function in $\mathbb{R}$. By choosing $f = du/d\rho$ [179], where ρ is the density (Eq. 3.4), it can be shown that $I(\mu)$ decreases monotonically as the value of the fictitious mass m_F decreases. To satisfy the Large Deviation Principle, we hence choose $m_F \longrightarrow 0$, ensuring our sampling strategy aligns with theoretical principles and sets the stage for exploring phase space dynamics in the following section.

In this limit, we can write an alternative LD scheme to sample a phase space density for $\boldsymbol{q}$ where force has been integrated over by the fast dynamics of F,

$$\bar{\rho}(\boldsymbol{q}) = \frac{\int_{F_{\min}}^{F_{\max}} dF\, \omega(F) e^{-\beta U(\boldsymbol{q}) + \beta F Q(\boldsymbol{q})}}{\int_{F_{\min}}^{F_{\max}} dF\, Z_q(F)\omega(F)} \tag{3.13}$$

This scheme is given by the following equations,

$$\frac{d\boldsymbol{q}}{dt} = M^{-1}\boldsymbol{p} \tag{3.14}$$

$$\frac{d\boldsymbol{p}}{dt} = -\nabla U + \bar{F}(Q)\nabla Q - \gamma \boldsymbol{p} + \sqrt{2\gamma\beta^{-1}} M^{1/2} \boldsymbol{\eta} \tag{3.15}$$

where $\boldsymbol{\eta}$ is a white-noise with corelation $\langle \eta_i(t)\eta_j(s)\rangle = \delta_{i,j}\delta(t-s)$, γ is the friction coefficient, and the dynamical variable F from the extended LD scheme has been replaced by the average $\bar{F}(Q)$ given by,

$$\begin{aligned}\bar{F}(Q) &= \int_{F_{\min}}^{F_{\max}} dF\, F\rho(F|\boldsymbol{q}) \\ &\equiv \frac{\int_{F_{\min}}^{F_{\max}} dF\, e^{-\beta U(\boldsymbol{q})+\beta F Q(\boldsymbol{q})}\omega(F)F}{\int_{F_{\min}}^{F_{\max}} dF\, e^{-\beta U(\boldsymbol{q})+\beta F Q(\boldsymbol{q})}\omega(F)} \\ &= \frac{\int_{F_{\min}}^{F_{\max}} dF\, e^{\beta F Q(\boldsymbol{q})}\omega(F)F}{\int_{F_{\min}}^{F_{\max}} dF\, e^{\beta F Q(\boldsymbol{q})}\omega(F)}.\end{aligned} \tag{3.16}$$

At any point in the simulation we can compute $\bar{F}(Q)$ as an additional force to apply to our system and perform the LD scheme in Eq. 3.14 and 3.15. Note that we are free to choose the function $\omega(F)$, but its form will effect the efficacy of the sampling and statistical errors. Later, in Sec. 3.2.5, we describe a scheme to learn an efficient $\omega(F)$ on the fly. From these simulations, it is possible to recover the average of any observable A as if we had performed the simulation with a particular fixed applied force and taken the average over that fixed-force ensemble density:

$$\rho_F(\boldsymbol{q}) = Z_q^{-1}(F)e^{-\beta U(\boldsymbol{q})+\beta F Q(\boldsymbol{q})}. \tag{3.17}$$

We can see this by manipulating the equation for $\langle A\rangle_F$ in the following way, to introduce an average over $\bar{\rho}(\boldsymbol{q})$ rather than over $\rho_F(\boldsymbol{q})$:

$$\begin{aligned}\langle A\rangle_F &= \int d\boldsymbol{q}\, A(\boldsymbol{q})\rho_F(\boldsymbol{q}) \\ &\equiv \int d\boldsymbol{q}\, A(\boldsymbol{q})\bar{\rho}(\boldsymbol{q})W_F(\boldsymbol{q}) \\ &= \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\, A(\boldsymbol{q}(t))W_F(\boldsymbol{q}(t))\end{aligned} \tag{3.18}$$

where in Eq. 3.18 we have used the property of ergodicity to transform an ensemble average to a time average over the simulation, and $W_F(\boldsymbol{q}) = \rho_F(\boldsymbol{q})/\bar{\rho}(\boldsymbol{q})$ is the observable weight that we will use to recover the correct average of observable A from our simulation. $W_F(\boldsymbol{q})$ can be expressed in terms of $\omega(F)$ and the as yet unknown partition functions $Z_q(F)$:

$$\begin{aligned}
W_F(\boldsymbol{q}) &= \left[\frac{\rho_F(\boldsymbol{q})}{\bar{\rho}(\boldsymbol{q})}\right] && (3.19)\\
&= \frac{Z_q^{-1}(F)e^{-\beta U(\boldsymbol{q})+\beta FQ}\left[\int dF' Z_q(F')\omega(F')\right]}{\int dF'\omega(F')e^{-\beta U(\boldsymbol{q})+\beta F'Q(\boldsymbol{q})}} && (3.20)\\
&= \frac{Z_q^{-1}(F)\int dF' Z_q(F')\omega(F')}{\int dF'\omega(F')e^{\beta(F'-F)Q(\boldsymbol{q})}}. && (3.21)
\end{aligned}$$

Here, we have suppressed the integration range in force (F_{min} to F_{max}) for compactness.

Given a set of weights $\omega(F)$, we can find an expression for $Z_q(F)$ up to a constant factor:

$$\begin{aligned}
Z_q(F) &= \int d\boldsymbol{q}\, e^{-\beta U(\boldsymbol{q})+\beta FQ}\\
&= \int d\boldsymbol{q}\, e^{-\beta U(\boldsymbol{q})+\beta FQ}\frac{\bar{\rho}(\boldsymbol{q})}{\bar{\rho}(\boldsymbol{q})}\\
&= \int d\boldsymbol{q}\, \bar{\rho}(\boldsymbol{q})\frac{Ce^{\beta FQ(\boldsymbol{q})}}{\int dF'\omega(F')e^{\beta F'Q(\boldsymbol{q})}} && (3.22)\\
&= \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\frac{Ce^{\beta FQ(\boldsymbol{q}(t))}}{\int dF'\omega(F')e^{\beta F'Q(\boldsymbol{q}(t))}}, && (3.23)
\end{aligned}$$

where $C = \int_{F_{\text{min}}}^{F_{\text{max}}} dF\, Z_q(F)\omega(F)$, and Eq. 3.23 again follows from ergodicity—since only ratios like $Z_q(F)/Z_q(F')$ will matter for our purposes, it is not necessary to estimate C is irrelevant.

Being able to estimate $Z_q(F)$ from a simulation trajectory gives us a scheme for choosing the $\omega(F)$. If we want to be able to compute the average $\langle A\rangle_F$ for any F in our desired force range, then we can assert that an efficient sampling scheme will have all forces sampled with equal probability. This happens when $\omega(F) \propto Z_q^{-1}(F)$, in which case the PDF of F, found by integrating out position from Eq. 3.4, is given by

$$P(F) = \frac{Z_q(F)\omega(F)}{\int dF\, Z_q(F)\omega(F)} = \frac{1}{F_{\max} - F_{\min}}, \tag{3.24}$$

for $F \in [F_{\min}, F_{\max}]$ and $P(F) = 0$ otherwise. We can construct an adaptive scheme to simultaneously learn the weights and estimate the partition functions as in Ref. [171], the details of which are given below.

3.2.4 EFFECTIVE POTENTIAL

In the limit where this uniform sampling is achieved, then configurations of $\boldsymbol{q}$ will occur with probability $P(\boldsymbol{q}) \propto \bar{\rho}(\boldsymbol{q})$. From this probability density function, we can define an effective potential energy that the system samples up to an additive constant as $U_{\text{eff}}(\boldsymbol{q}) = -k_B T \log(P(\boldsymbol{q}))$, or

$$e^{-\beta U_{\text{eff}}(\boldsymbol{q})} \equiv \int_{F_{\min}}^{F_{\max}} dF'\, \omega(F') e^{-\beta U(\boldsymbol{q}) + \beta F' Q(\boldsymbol{q})}. \tag{3.25}$$

This expression is valid for any choice of $\omega(F)$, including the case $\omega(F) = Z_q^{-1}(F)$, for which the integrand becomes $\rho_F(\boldsymbol{q})$; it can be evaluated using numerical integration for any test potential, and hence we can use it as a reference to predict the expected behavior of our sampling method for those cases.

3.2.5 ALGORITHM FOR LEARNING WEIGHTS

Above, we discuss that in order to sample all forces with equal probability, the form of the weights are such that $\omega(F) \propto Z_q^{-1}(F)$. Here, we sketch the algorithm and implementation details used to learn the weights on-the-fly during sampling, following the same scheme as Ref. [171] as implemented in the MIST package [180].

A good numerical scheme for adapting the weights and performing integrals of the form

Eq. 3.16 and 3.23 is to learn the weights at a fixed set of M “node” points (f_i) placed at the roots of a Legendre polynomial between $F_{\min}$ and $F_{\max}$, and perform the integrals by Gauss-Legendre quadrature; for this we use the implementation of John Burkardt.[1] Here, each f_i has a corresponding weight B_i such that for a function g, $\sum_{i=1}^{M} g(f_i)B_i \approx \int_{F_{\min}}^{F_{\max}} g(F)dF$.

Having chosen an initial distribution of weights, we can begin computing a running average of $Z_q(f_i)$ (discretizing Eq. 3.23) up to sample number n,

$$z_{i,n} = \frac{1}{n}\frac{e^{\beta f_i Q(\boldsymbol{q}_n)}}{\sum_{j=1}^{M} B_j \omega_{j,n} e^{\beta f_i Q(\boldsymbol{q}_n)}} + z_{i,n-1}\frac{n-1}{n}. \tag{3.26}$$

We then update the weights at the discrete forces ω_i by a scheme such that ω_i converges towards $\omega_i \propto z_i^{-1}$,

$$\omega_{i,\diamond} = \omega_{i,n}(1-h) + \frac{h}{z_{i,n}}, \tag{3.27}$$

and then re-normalize the weights,

$$\omega_{i,n+1} = \frac{\omega_{i,\diamond}}{\sum_{j=1}^{M} B_j \omega_{j,\diamond}}. \tag{3.28}$$

Here, $h = dt/\tau$, where τ is a timescale parameter that controls how quickly the weights are adjusted.

We then compute the current average force (Eq. 3.16) using the discretized weights by:

$$\bar{F}(Q) = \frac{\sum_{j=1}^{M} B_j f_j \omega_j e^{\beta f_j Q(\boldsymbol{q})}}{\sum_{j=1}^{M} B_j \omega_j e^{\beta f_j Q(\boldsymbol{q})}}, \tag{3.29}$$

and the observable weights (Eq. 3.21) as

$$W_{f_i}(\boldsymbol{q}) = \frac{f_M - f_1}{z_i \sum_{j=1}^{M} B_j \omega_j e^{\beta (f_j - f_i) Q(\boldsymbol{q})}}. \tag{3.30}$$

---

[1] `https://people.sc.fsu.edu/~jburkardt/c_src/legendre_rule_fast/legendre_rule_fast.html`

Here we follow Ref. [171] and replace the numerator with its steady state value (a good approximation because Eq. 3.27 converges quickly).

3.3 RESULTS

In this study, we demonstrate the utility of FISST as a computationally efficient method to simulate force dependent dynamics for a range of systems. In addition, we show that FISST samples an averaged potential energy surface from the entire range of forces, and discuss how this feature allows FISST to be used as an enhanced sampling method. We present results for three different systems of varying complexity: a 2D, V-shaped analytical potential for proof-of-concept, a chain of 12 beads, where interbead interactions are chosen such that the global minimum configuration is degenerate between a left and right handed helix, and deca-alanine in explicit water, the last showing that FISST can immediately be applied to atomistic biological systems. Together, these systems will be used to illustrate what sampling data is accessible with FISST, quantify the accuracy of FISST results, and analyze the performance of FISST relative to alternative methods. Further simulation details for each model system are given in Sec. 3.5.

3.3.1 ANALYTICAL POTENTIAL

We first consider the situation of a particle undergoing LD on an analytical potential. Using an analytical potential allows us to easily calculate the exact, bias-dependent potential to compare with sampled data and determine the accuracy of FISST for each applied force. In addition, we can numerically calculate the effective potential (via Eq. 3.25) that is sampled for a given FISST force range. The potential we have crafted is a V-shaped analytical potential, parameterized such that there is a single minimum when a positive force is applied in the pulling coordinate $Q = y$ direction, and two minima separated by a barrier when a negative force is applied (Eq. 3.31).

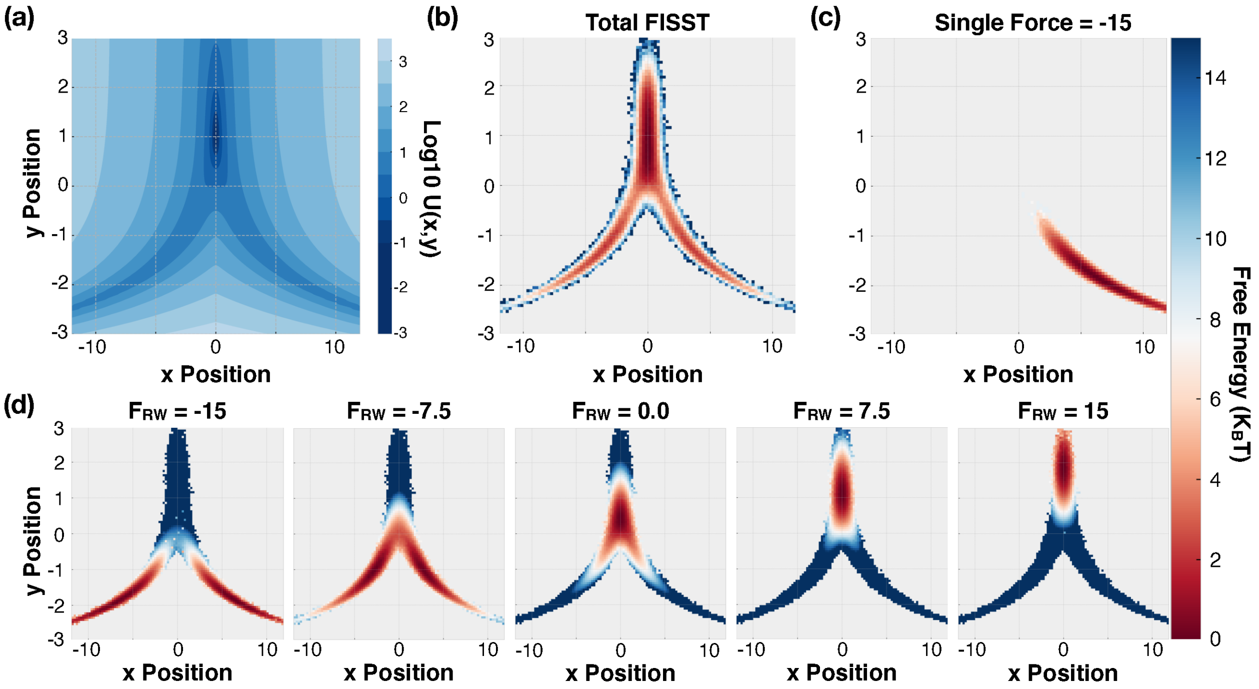


Figure 3.1: (a) Contour plot of the potential (Eq. 3.31) used for Langevin dynamics. (b) Estimated effective potential (U_{eff}, Sec. 3.2.4) sampled by FISST using a force range of [-15:15] applied to $Q = y$. (c) Failure of running a simulation at a single force of $F = -15$, where the position is stuck in the right minimum for the entirety of the simulation. (d) Reweighted FISST data predicts free energy surfaces at -15, -7.5, 0, 7.5, and 15. In all cases (a)-(d), the minimum value of the potential is shifted to be zero.

$$\beta U(x,y) = -8\ln\left[e^{-\frac{1}{2}(x-e^{-y})^2} + e^{-\frac{1}{2}(x+e^{-y})^2}\right] + \frac{1}{2}y^2 \tag{3.31}$$

This minimalist model is chosen to represent the case of a polymer/peptide with multiple possible conformations that collapse to a single extended state when pulled upon, as considered in Sec. 3.3.2 and 3.3.3 below.

In Fig. 3.1a, we show the effective potential sampled when FISST is applied in the y direction, that is $U_{\text{eff}}(\boldsymbol{q}) = -k_B T \log P(x,y)$, where $P(x,y)$ is a histogram of positions sampled in the FISST simulation. The effective potential sampled by FISST clearly differs from what would be expected of a standard simulation on this potential, where now the sampling of the minimum is elongated relative to the potential and two minima not present in the unbiased potential appear in the negative y direction (Fig. 3.1a and 3.1b). This expected deviation arises because FISST samples an effective potential defined by Eq. 3.25, which is an average over the whole force range. Qualitatively, the elongation emerges because applying a force essentially tilts the potential energy surface in the direction of the pulling coordinate [170]. In the case of the potential shown in Fig. 3.1, the effective potential has a contribution from each force that tilts the potential in both the y and $-y$ directions. Using the observable weights defined in Eq. 3.21, we can reweight the FISST probability distribution to forces, F = -15, -7.5, 0, 7.5, and 15 pN, and show the predicted free energy surface, given by $A(x, y) = -k_B T \log P_F(x, y)$, at each force in Fig. 3.1d (the Supplemental Movie shows how the density distribution changes with force). The effect of tilting in both directions is evident from each reweighted density, where large negative forces sample more in $-y$ and large positive forces sample more in $+y$.

This potential was chosen such that a strong force in the $-y$ direction would have two degenerate minima, and at the largest negative force considered, $F = -15$, FISST samples each of these minima approximately equally. However in trajectories with a single applied force, a force of $F = -15$ leads to only one of these wells being sampled (Fig. 3.1c). The large force in the $-y$ direction deepens each of these two minima and the particle is unable to escape the first minimum that it samples. This clearly shows a significant advantage of FISST, where sampling over an effective potential flattens force-specific energy barriers across the potential energy surface (including and perhaps especially those in directions orthogonal to the pulling coordinate) without limiting the ability to reweight the trajectory to a specific force, thereby improving sampling at each force.

While it is clear for this example that reweighting a FISST trajectory to different forces qualitatively reproduces the expected potential energy surface from individual simulations under a constant force, in Fig. 3.5 we also quantify the error of each reweighted density by calculating the Jensen-Shannon distances [181] between the reweighted probability distribution at each force, $P_F(x, y)$, against the exact Boltzmann distribution ($\rho_F(x, y) \propto \exp(-\beta U(x, y) + \beta F y)$). We then compare this error to results from separate simulations performed at $N_f = 20$ forces from -15 to 15. These data demonstrate that the error in the reweighted distribution from a single FISST run is comparable to that from a separate simulation of the same length run at each applied force (except that it performs much better in the high negative force regime where standard MD fails). Moreover, FISST in this case is much more efficient; a single FISST simulation has a much lower total error when compared to N_f single-force MD simulations that in total comprise the same amount of computational work.

3.3.2 BEADED HELIX

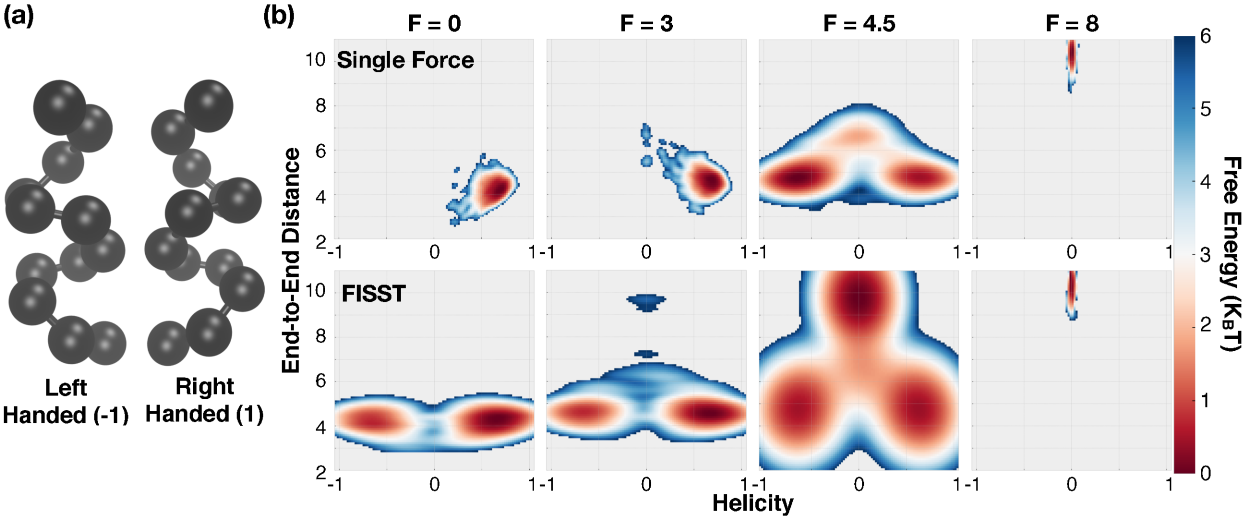


Figure 3.2: (a) Left and right handed helices used as reference structures to quantify helicity. (b) Estimated free-energy surface for end-to-end distance and helicity sampled from single force simulations (top) and reweighted from a single FISST simulation (bottom), for forces of F = 0, 3, 4.5, and 8 ($k_B T/\sigma$). In this case, probability histograms were computed by a two-dimensional Gaussian kernel-density-estimate as implemented in `scipy` [182]. For all cases, the lowest free energy is shifted to zero.

We now consider a toy model of an $\alpha-$helix, designed to be a higher-dimensional analog of the V-shaped potential considered above. This helix is composed of a 12 atom chain of beads, where each bead is connected by a harmonic spring of length $\sigma = 1$ (the unit of length) and spring constant $k = 100 k_B T/\sigma^2$. Additionally, we apply a Lennard-Jones interaction with an interaction distance $\sigma_{\mathrm{LJ}} = 1.5\sigma$ a strength of $\epsilon = 7.5\ k_B T$ between the ith and i+4th beads, and a purely-repulsive WCA potential [183] with $\epsilon = 3k_B T$ between all other bead pairs.

This setup leads to a model with two degenerate ground state conformations of a left or right handed helix (Fig. 3.2a), analogous to the analytical potential discussed above. In these simulations, we apply a pulling force to the terminal atoms of the helix and plot the 2D probability density of end-to-end distances ($d_{\text{end-end}}$) and helicity for the system (in this case, $Q = d_{\text{end-end}}$). The helicity is determined from the minimum RMSD of the frame against a left and right handed helix, where positive (negative) is a right (left) handed configuration, and 0 represents extended structures with RMSDs greater than 0.85σ for both references. Further details can be found in Sec. 3.5.

The sampling of the helix with no applied force shows the same behavior that was seen in the Fig. 3.1c for the analytical potential, where sampling of only one folded minimum is observed (Fig. 3.2b, top left). This behavior is expected because the interaction energy between beads is chosen to be high relative to the temperature, causing the system to get stuck in the initial right-handed configuration and at no point over the course of the trajectory does the helix unravel enough to switch handedness. As seen in the upper panel at $F = 4.5k_BT/\sigma$ of Fig. 3.2b, a pulling force that is applied to elongate the helix can aid the system in transitioning between the two differently handed helices.

We now apply FISST to the system, specifying a force range of $F_{\text{min}} = -2\ k_BT/\sigma$ to $F_{\text{max}} = 8\ k_BT/\sigma$. The effective potential from FISST provides an alternative free-energy surface that can enhance proper sampling of the unbiased PES of this helical system; high pulling forces are present to pull the helix out of its initial right handed configuration and negative restoring forces permit the helix to refold in either conformation with equal probability. Using FISST we observe a relatively balanced population of left and right handed helices when reconstructing the PES of the system from zero force up through intermediate forces (Fig. 3.2b bottom), matching what is expected from these degenerate configurations. Because FISST samples a wide range of configurations and is then reweighted to a particular force, we observe the appearance of an additional high free energy extended state at $F = 3k_BT/\sigma$, which is then stabilized at $F =$

$4.5k_BT/\sigma$, in contrast to the single force simulation, which is still stuck in compact structures. At high forces, both standard MD and reweighted FISST data agree on the expected result of only a single extended state. Importantly, we note that this helix example illustrates a hallmark of enhanced sampling methods, where sampling along a biased coordinate (end-end distance) enhances sampling along a collective variable that was not directly biased (helicity), one which contains a very high effective barrier between relevant states.

3.3.3 ALANINE-10

Up to this point we have considered toy models that illustrate how FISST can be used to both simultaneously sample dynamics at a range of forces through reweighting the trajectory as well as be used as an enhanced sampling method to 'unstick' the system from certain low energy conformations. The final system we consider is the deca-alanine peptide explicitly solvated in water, which provides a complex, biologically relevant test system for FISST. Here we wish to understand how the conformational landscape of a peptide in solution is affected by a pulling force applied between its two ends (as defined by the terminal α-carbons, with the pulling coordinate $Q = d_{\mathrm{end-end}}$).

To establish benchmark sampling data for alanine-10, we perform temperature replica exchange simulations at forces of -10, -5, 0, 5, and 10 pN, using 40 replicas between 300K and 400K (see Sec. 3.5). The probability density of end-to-end distances for each method at -5, 0, and 5 pN are shown in Fig. 3.3a (additional data is shown in Fig. 3.6), where all simulations were run for 160 ns. In Fig. 3.3a, a separate single force simulation was performed for each force and the FISST data acquired by reweighting with the observable weights determined on-the-fly during a simulation with force range [-10:10] pN. It is clear that there is a qualitative agreement between the replica exchange benchmarks and both the single force and FISST simulations.

In order to determine the error in the end-to-end distance distribution quantitatively for each method, we employ the Eigenvector method for Umbrella Sampling (EMUS) to interpolate the

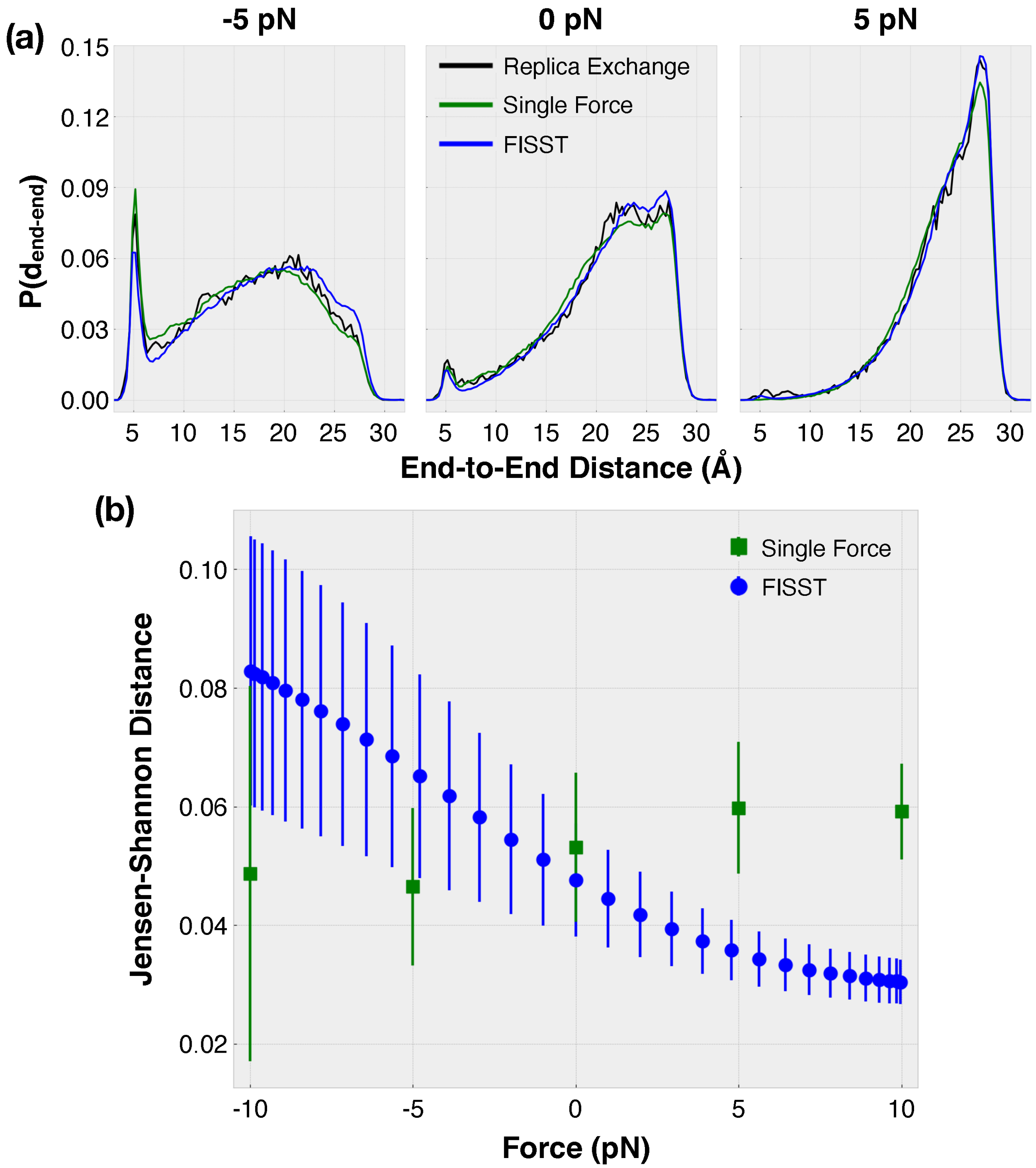


Figure 3.3: (a) Probability density function of end-to-end distances at $F = -5$, 0, and 5 pN calculated from replica exchange, single force, and FISST simulations. (b) Jensen-Shannon distances as a function of force for single force and FISST simulations. The reference density is determined from 5 replica exchange simulations at -10, -5, 0, 5, and 10 pN and interpolated to desired force using EMUS [184].

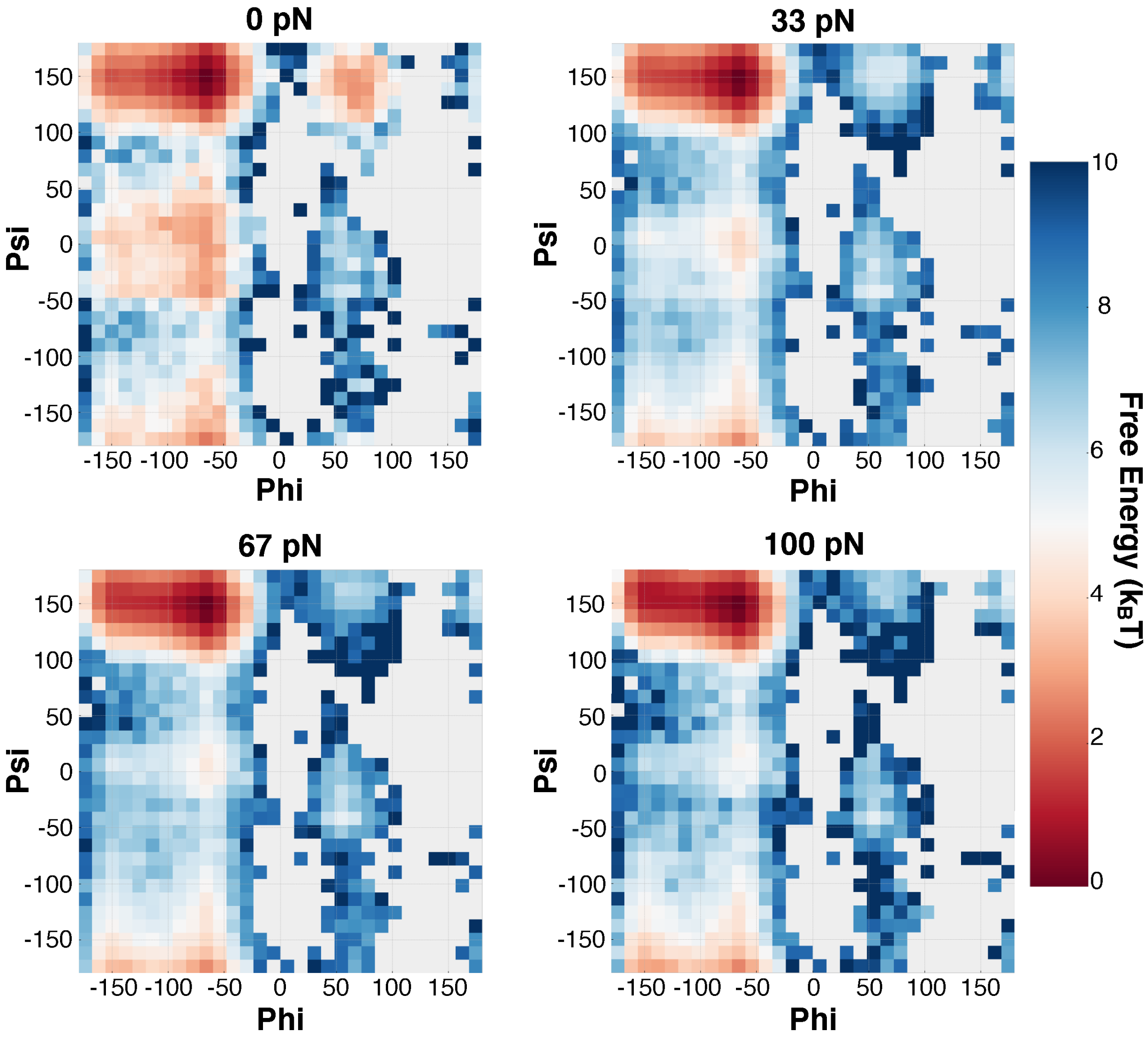


Figure 3.4: Ramachandran plots of alanine-10 peptide at 0, 33, 67, and 100 pN applied force. All plot were generated by reweighting data from a single 160 ns FISST simulation using a force range of [0:100] pN. The regions with the most significant change in free-energy upon upon pulling are the α/δ area at (-50,0), γ at (75,150), PPII at (-75,150) and β at (-140,150), using the nomenclature from Ref. [185].

replica exchange data to other forces for comparison with single force and FISST (Fig. 3.3b) [184]. This was done by using the replica exchange trajectories at $F = -10$, -5, 0, 5, and 10 pN as input distributions to EMUS and a target force between -10 and 10 pN was provided for EMUS to predict the equilibrium sampling at that force. This predicted density was then compared to end-to-end densities generated by both single force and FISST simulations at the target force (each point is average over 5 replicates, error bars are ± one standard deviation of the replicates).

In both FISST and single force simulations, the large standard deviations at negative forces reflect the fact that compressive forces ($F < 0$) are harder to sample accurately. This is due to a larger variety of transient conformations that can be formed when the peptide is being compressed, for example bending the chain in ways that cause uncommon geometries. However, similar to the results from the V-shaped analytical potential, the average error in single force calculations is relatively constant across the forces considered. The error in the reweighted FISST calculations decreases from about 0.08 to 0.03, and crosses the average error from running individual single force simulations (Fig. 3.3b). Each single force calculation had the same duration as each FISST simulation, hence the same computational time for single force would only directly give one data point in this range, whereas we get all points simultaneously with FISST. It is also possible to reweight single-force sampling to other forces; in Fig. 3.7 we show that FISST is still more accurate at equal cost, and that using single-force simulations, at least 3 different forces are probably required to equal or surpass the accuracy of FISST across the full force range.

Lastly, we wish to illustrate that the observable weights from FISST can allow one to reconstruct the probability densities of other observables, not just the one that was biased. In Fig. 3.4, we show the effect of pulling on the end-to-end distance on the dihedral angle densities of the alanine-10 peptide. Small alanine peptides are known to prefer the polyproline II (PPII) helix, a left handed helix that is present in many folded, unfolded, and amorphous biomolecules [186]. Fig. 3.4 shows the Ramachandran plots of alanine-10 at 0, 33, 67, and 100 pN calculated using FISST with a force range of [0:100] pN, where the lowest energy configuration is referenced to zero. Analogous Ramachandran plots calculated with single force and replica exchange simulations are shown in Fig. 3.8 and 3.9, and qualitatively agree with the FISST data. The global minimum conformation under each applied force is the polyproline II (PPII) region at (-75, 150) degrees, consistent with the literature of short alanine peptides [187]. Additionally, there is a local minimum in the α-helical region of the plot at (-50,0), consistent with previous work on alanine peptides, where the α-helix is known to be a stable conformation [188]. For all methods used, we observe that as force is increased, the helical basin decreases in population relative to the less compact PPII structure, as shown by previous theoretical and computational work [59, 189].

3.4 Discussion and conclusions

In this work, we present the FISST method for performing simulated tempering in force in the infinite switch limit. In FISST, observable weights are computed during the simulation which can be used to compute averages of any structural quantity at any force within the simulated force range. We benchmark this new method on a variety of model systems with varying complexity to evaluate the performance of FISST, including a simple analytical potential, a toy helix, and deca-alanine in water. For each system we showed that a FISST simulation was able to quantitatively reproduce the quality of sampling at each force across the force range at lower total computational cost, and in some cases accelerated sampling over barriers that could not be crossed in a standard MD simulation. This efficiency makes FISST a promising method for studying the response of larger and more complex biomolecular systems to small applied forces.

A key to successfully using FISST is to choose a force range that is relevant to the problem being studied. Through testing FISST, we have found that the quality of sampling can depend on the choice of force range, however simple intuition about the system is usually sufficient to overcome these difficulties. For example, we have observed that enhanced sampling at $F = 0$ can be aided using force ranges that extend into the negative (compression) region to include restoring forces. FISST causes the molecule to always feel some effect from every force in the force range. This can lead to some problems with very high force, especially before the weights have converged at the beginning of the simulation, where the large forces can contribute more that expected. However, after only a few nanoseconds, the weights converge and remain stable for the remainder of the simulation (Fig. 3.10). While running long enough simulations will eliminate this effect, another simple option for studying very wide force ranges is to run multiple FISST simulations using smaller force ranges that span the target force difference. The current implementation also includes the ability to select different initial weight distributions, such that large forces have small weights at the beginning, however that was not necessary for any of the examples in this current work.

As currently implemented and described above, FISST can only be applied to a single collective variable. The formulation herein can trivially be extended to higher dimensions, however as with many similar histogram methods, the need to learn the weights over a discrete set of points means that it is not likely to perform well for more than two dimensions. Rather than going to higher dimensions in FISST, we believe the most promising strategy is to apply FISST to study mechanical forces along a CV of interest, and combine that simulation with other methods that will accelerate the sampling of conformations along other degrees of freedom. With the weights fixed, FISST is a fully equilibrium sampling method, hence any other equilibrium method (such as various forms of tempering, umbrella sampling, metadynamics, variationally enhanced sampling, etc. [52, 67, 173, 174, 190–192]) can be used on top of the learned FISST potential. We are currently exploring which of these other enhanced sampling methods can be rigorously combined during the FISST simulation to accelerate the convergence of the sampling during the time when the weights are being learned. Our expectation is that FISST alone may suffice for studying the properties of disordered peptides under biologically relevant forces (e.g. Ref. [13]), as well as the small deformations of compact protein domains, cases where we don't expect other large changes to occur (such as an allosterically coupled folding/unfolding of a distal loop). For other cases, we believe that combination of FISST with an orthogonal method for enhancing sampling will be more effective than either alone.

Lastly, we note that, although we have targeted the problem of understanding the effect of mechanical forces along a collective variable, FISST could be used to accelerate the sampling across any range of linear coupling terms in a Hamiltonian. Because FISST is implemented in PLUMED, it can be immediately applied to any CV to enhance sampling over a range of couplings for that CV, whether or not it corresponds to a physical force. As one example, FISST could be applied along with the DIPOLE CV to enhance sampling over a range of electric fields. Additionally, recent studies have shown that experimental information can be directly incorporated into MD simulations with minimal bias using extra linear coupling terms in the Hamiltonian [193–197]. It will be interesting to explore whether this method, which flattens the probability of seeing a given coupling term, can be connected to understanding the optimal terms determined by those relative-entropy based methods.

3.5 Simulation details

3.5.1 Code and Data Availability

All code for the method is now available as a module in the PLUMED open source sampling library [172]. The version used for this work is available in the FISST-dev branch in our group's github repository. Input files and scripts for repeating the types of calculations in this work will be deposited in the PLUMED-NEST [143]. The data that support the findings of this study are also available from the corresponding author upon request.

Analysis was performed using `numpy` [198], `scipy` [182] and `mdtraj` [144]. Figures were made using Matplotlib [199], with some tools from `seaborn`.

3.5.2 Langevin dynamics on test potentials

The Langevin dynamics on the analytical V-shaped potential defined in Eq. 3.31 were performed using the `pesmd` module implementation in PLUMED. The simulation was initialized at position (0.0, 0.0), with a temperature of 1.0 $k_B T$, a timestep of 0.05, and friction of 1 was used. A total of 5,000,000 steps were collected, where every 20th frame of the simulation was used for the analysis. The FISST simulations were run with a force range of [-15:15], 21 quadrature points were used to discretize the force range, a uniform initial weight distribution was used, and the weights were updated every 200 steps.

3.5.3 Beaded helix simulations

Beaded helix simulations were run in LAMMPS [2] Simulations were performed with a timestep of 0.005 at constant temperature of $T = 1.0$ (with $k_B = 1.0$ such that $k_BT = 1.0$) using `fix_nvt` with a damping time of 0.5. Lennard-Jones interactions between i and $i+4$ beads had $\epsilon_{\mathrm{LJ}} = 7.5k_BT$ and $\sigma_{\mathrm{LJ}} = 1.5\sigma$, and were cut off at a distance of $2.0\sigma_{LJ}$. WCA interactions between all other pairs of beads were formed using a shifted Lennard-Jones interaction with $\epsilon_{\mathrm{WCA}} = 3.0k_BT$ a cutoff at length $2^{1/6}\sigma_{\mathrm{LJ}}$. Bond lengths were maintained using a harmonic spring constant of $k = 100k_BT/\sigma$. Initial structures for left and right handed helices were formed by starting multiple initial seeds from a straight line structure, with helical structures forming in very short MD simulations. A right-handed starting structure was used for the subsequent simulations in this work. Single force and FISST simulations were then run for $t = 10^7$ steps. FISST was applied with a force range of [-2:8] k_BT/σ, using 31 quadrature points for force interpolation and a weight update period of 200 steps.

3.5.4 Atomistic molecular dynamics on peptides

All atomistic molecular dynamics simulations were done using GROMACS version 2018.3 [200, 201]. Alanine-10 was set up in explicit TIP3P water and parameterized using the CHARMM36 all atom forcefield [202]. A cubic box with edge length 57.0 Åwas used and periodic boundary conditions were used in all three directions. The initial structure was minimized using the steepest descent algorithm for a maximum of 50000 steps. The cutoff for short range interactions was chosen to be 1.0 Å. Longer range coloumbic interactions were coupled using the Particle Mesh Ewald method and constraints for hydrogen bonds were computed using the LINCS algorithm [203]. Minimization was first followed by a NVT equilibration using the Berendsen thermostat [204], then a NPT equilibration where pressure was maintained at 1.0 atm using a Parinello-Rahman Barostat [205] and Bussi-Parinello thermostat [206]. The timestep used in all cases was 2 fs.

---

[2] https://lammps.sandia.gov/

Constant pressure FISST simulations of deca-alanine with force ranges of [-10:10] pN and [0:100] pN were run to collect the data in Fig. 3.3 and 3.4 respectively, with force being applied between the terminal α-carbons. In each case, 31 quadrature points were used, the weights were initiated with a uniform distribution, and a period of 200 steps was used to update the weights. The effect of changing this learning rate is assessed in Fig. 3.11.

For comparison, MD simulations applying a single constant force were performed. The constant force on the peptide was implemented using PLUMED's RESTRAINT function, where force was applied on the terminal α-carbons.

The total simulation time for production data at a single-force or using FISST was 160 ns. For both single force and FISST simulations, we performed 5 replicates, in order to compute average and standard deviations of error from a reference simulation.

In order to compute reference end-end distance histograms, we performed temperature replica exchange simulations at a set of reference forces. Each force dependent replica exchange simulation was performed in NPT, with 40 different temperatures spanning 300 and 400 K. Exchanges were attempted between alternating adjacent pairs of temperatures every 5 ps. Replica exchange simulations at each force were performed for at least 60 ns.

3.6 SUPPORTING INFORMATION

3.6.1 V-SHAPE POTENTIAL

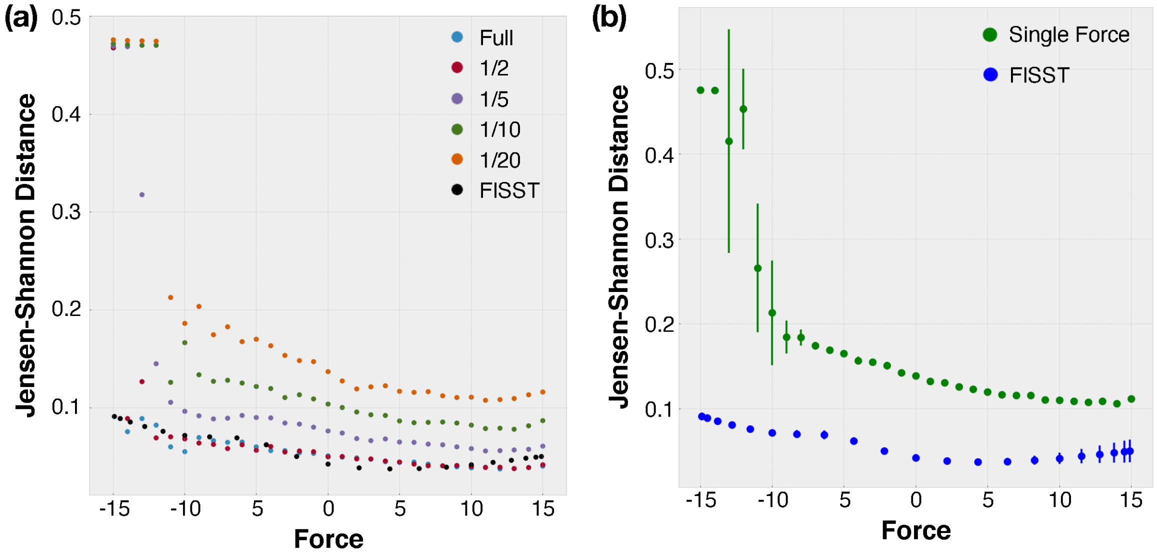


Figure 3.5: (a) Error of FISST [-15:15] and individual single force simulations run relative to the exact density (Eq. 3.4), using fractions of total work. In the 1/20 data set the error is very high compared to FISST despite the number of total steps across all forces being equal to that of the full FISST trajectory. Simulations on the order of the same length of FISST are required at each single force to replicate the same level of accuracy. (b) Average error relative to exact density over 5 replicates for FISST [-15:15] and single force simulations run at 20 different forces, each with 1/20th the number of steps.

For the V-shape analytical potential considered in the main text, we are able to compute the error in the sampled probability distribution relative to the true Boltzmann distribution at a given force. To do so, the Jensen-Shannon distance [181] was computed between the sampled and exact probability distribution of position. This error metric shown in Fig. 3.5 was computed on a discrete grid, with 40 bins spanning $y \in (-2, 2)$ and 200 bins spanning $x \in (-10, 10)$.

In Fig. 3.5, we compute the error in sampling relative to the exact distribution for FISST, and the standard alternative of performing a single simulation at each applied force of interest. In this case, this brute force approach leads to error that, other than large negative forces, is more or less constant across the force range studied, decreasing slightly at larger positive forces. In order to evaluate how efficient FISST is over traditional approaches, we ran single force simulations at a fraction of the length of FISST and plotted the error as the trajectory is shortened (Fig. 3.5a). Because we are considering 20 different forces, the $\frac{1}{20}$th dataset (orange points) contains the same total work as the FISST trajectory (black points). In addition to getting the large negative force regime qualitatively correct, this data shows that FISST is much more efficient than standard methods while not compromising the accuracy of the simulation.

Fig. 3.5b plots the average error for a constant amount of work for both FISST and single force simulations averaged over 5 simulations. Here, the quality of the simulations are constant over multiple independent simulations. In the single force case, larger deviations between trials show up at negative forces (< -10), which is the point at which these simulations begin to fail due to getting stuck in one of the arms of the potential. *In none of the 5 independent trials at $F = -14, -15$ did a single force simulation sample both minima.*

3.6.2 Alanine-10

In total, we performed replica exchange simulations at forces of {-10,-7.5, -5, -2.5, 0, 2.5, 5, 7.5,10} pN to get a set of benchmark simulations to compare to FISST and single force simulations run over the same force range. Fig. 3.6 shows a comparison between FISST and a single force simulation for five of these forces (data from only one of the replicates). At each force to which the FISST data is reweighted, we clearly match the end-to-end distance probability density from the replica exchange benchmark simulations. Each of the single force simulations are run for the same amount of time as the single FISST run, however a separate simulation was used to calculate the probability density at each force.

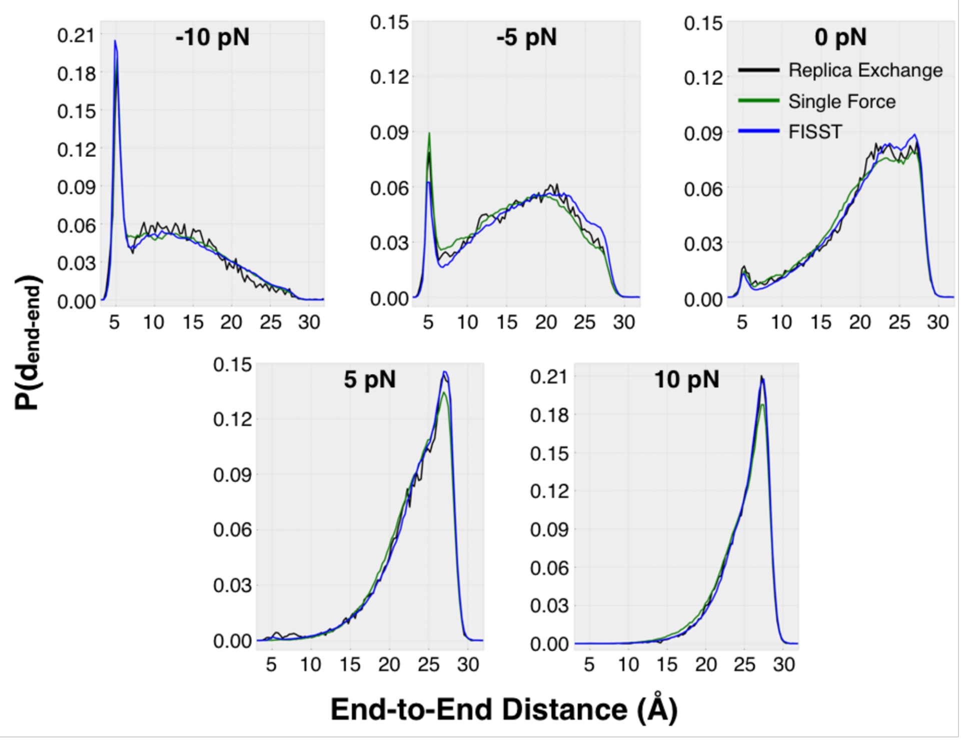


Figure 3.6: Probability density of end-to-end distances at F = -10, -5, 0, 5, and 10 pN calculated from replica exchange, single force, and FISST simulations.

Error for single force and FISST is computed by taking the sum of the Jensen-Shannon error (as in Fig. 3.3) comparing reweighted histograms to reference replica exchange simulations performed at {-10,-7.5, -5, -2.5, 0, 2.5, 5, 7.5,10} pN. Single force data came from four replicates performed at {-10,-5,0,5,10} pN. For a given set of data, error was computed for each of four replicates, and averaged. Error bars are computed by $\sqrt{\sum_{F_i} \sigma_{F_i}^2}$, where σ_{F_i} is the standard deviation across the four replicates at each of the 9 reference forces. For FISST, histograms are computed by reweighting data directly to each reference force. The error for FISST serves as a reference value for the single force data, and is divided out on the y-axis.

The error for single-force was computed by reweighting different sets of single-force data. First, the error at each reference force was computed using a simple weighted histogram, by taking the data from 0 force and giving each distance value $d_{\text{end-end}}(t)$ from the simulation a weight of $w_t^F = e^{\beta F d_{\text{end-end}}(t)}$. This results in fairly accurate predictions for the distance histograms, so apparently 10 pN is still close to the linear-response regime. The error from this approach is 32% higher than from FISST using the same amount of MD data. Error for larger amounts of work is computed by combining data from multiple single-force simulations using FISST. In particular, error with twice as much computational effort shown combines data from single force runs at {-10,10} pN, at 3x from {-10,0,10} pN, at 4x from {-10,-5,5,10} pN, and at 5x from {-10,-5,0,5,10} pN. For alanine-10 in this force range, at least 3 different forces using equivalent amounts of MD work per simulation are required to reach the same level of accuracy as one FISST simulation. Adding additional forces does not substantially improve the results, in comparison to the three-force case.

In Fig. 3.10 we plot the distribution of the weights as they are being learned during a 160 ns deca-alanine simulation. The initial weight distribution is chosen to be uniform and it quickly changes to an approximately exponential distribution at short times. During the first 10 ns of the simulation, both the amplitude and decay rate of the distribution change relatively quickly. After this initial learning stage, the weight distribution does not significantly change for the remainder of the simulation.

In Fig. 3.11 we plot the force dependent error in FISST simulations calculated with 5 different learning rates. The chosen learning rate does not significantly impact sampling over the range of learning rates considered.

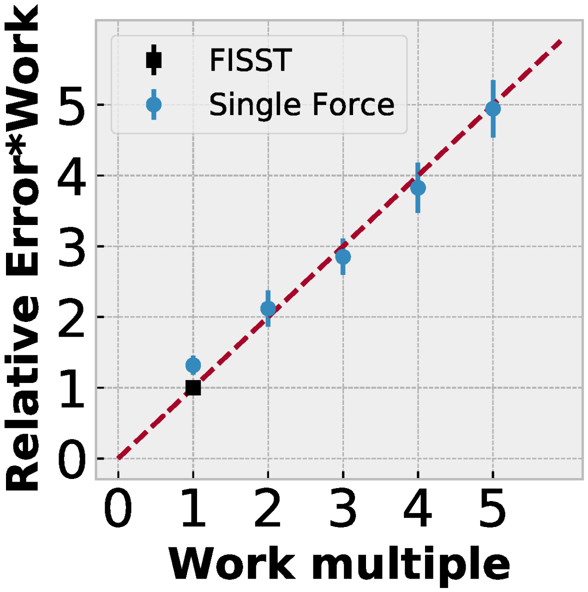


Figure 3.7: Relative error of single force simulations compared to FISST vs amount of computational work, as explained in the text above. The dashed line shows accuracy equal to FISST.

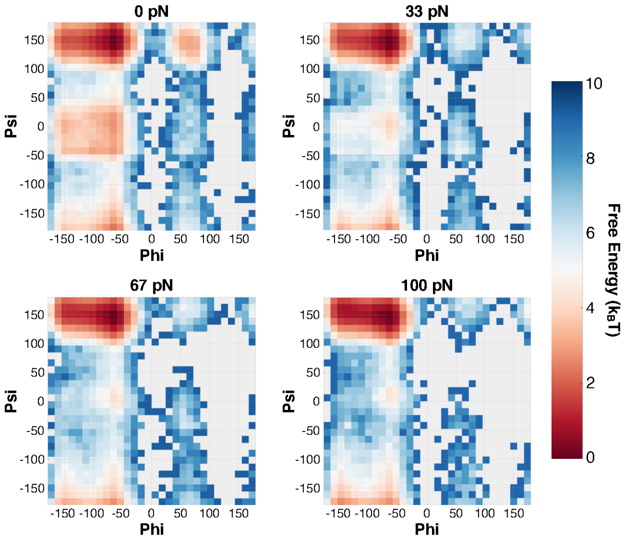


Figure 3.8: Ramachandran plots of alanine-10 peptide at 0, 33, 67, and 100 pN applied force, each simulated at a single applied force.

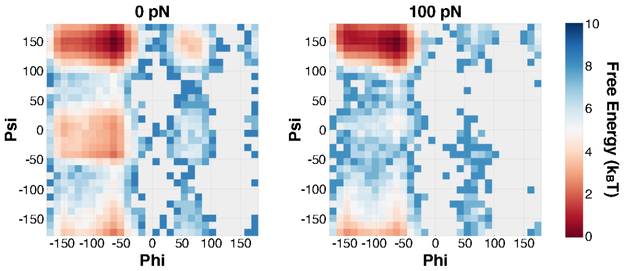


Figure 3.9: Ramachandran plots of alanine-10 peptide at 0 and 100 pN applied force, each calculated from Replica Exchange simulations.

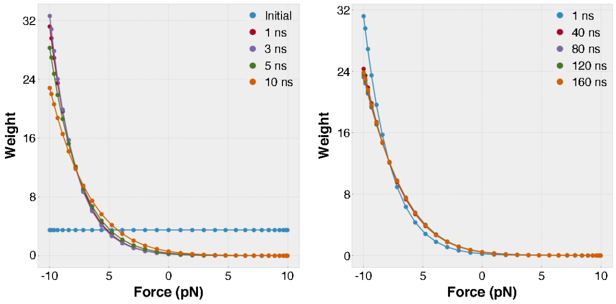


Figure 3.10: Distribution of weights at various stages of a 160 ns deca-alanine simulation. (a) At short times, the weight distribution changes significantly. (b) At longer times, the weight distribution is essentially constant for the remainder of the simulation.

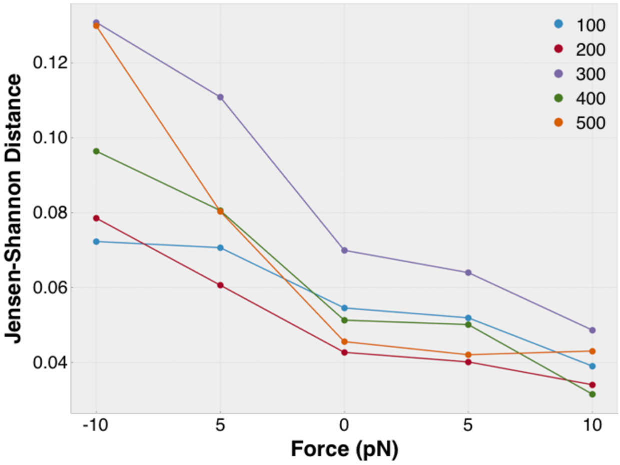


Figure 3.11: Error in FISST for a range of learning rates. A 160 ns FISST simulation was run with the weights updated every 100, 200, 300, 400, and 500 steps. We calculate the end-to-end distance probability densities at each force and calculate the error against results from replica exchange simulations. In the range considered, there is no obvious dependence on learning rate. An update period of 200 was used for all data discussed in the main text.

CHAPTER 4

IMPROVED PREDICTION OF MOLECULAR RESPONSE TO PULLING BY COMBINING FORCE TEMPERING WITH REPLICA EXCHANGE METHODS

This chapter is adapted from Ref. [207]

Abstract

Small mechanical forces play important functional roles in many crucial cellular processes, including in the dynamical behavior of the cytoskeleton and in the regulation of osmotic pressure through membrane-bound proteins. Molecular simulations offer the promise of being able to design the behavior of proteins that sense and respond to these forces. However, it is difficult to predict and identify the effect of the relevant piconewton (pN) scale forces due to their small magnitude. Previously, we introduced the Infinite Switch Simulated Tempering in Force (FISST) method which allows one to estimate the effect of a range of applied forces from a single molecular dynamics simulation, and also demonstrated that FISST additionally accelerates sampling of a molecule's conformational landscape. For some problems, we find that this acceleration is not sufficient to capture all relevant conformational fluctuations, and hence here we demonstrate that FISST can be combined with either temperature replica exchange or solute tempering approaches to produce a hybrid method that enables more robust prediction of the effect of small forces on molecular systems.

4.1 INTRODUCTION

Biological systems must have mechanisms for being able to sense and respond to mechanical forces from their environment and those that are generated internally through the action of molecular machines [4, 6, 208, 209]. Cells can employ proteins to sense and respond to these forces using a wide range of molecular mechanisms which we previously reviewed [4]. Perhaps the simplest such mechanism is the use of a single disordered peptide domain at the locus of a mechanical process, whose change from a collapsed to an extended conformation with single piconewtons of force could be sufficient to change the behavior of a larger protein machine. This kind of behavior has been identified in polymerization factors called formins through a combination of in vitro and in vivo biochemistry with simple modeling [13, 210–213], but a precise molecular mechanism for such behavior which explains the differences between homologous proteins in different species has not yet been shown [212, 213].

While these formin disordered domains are very large and the effect of force on their activity is complex, the effect of a pulling force on simple peptides has been exploited for the development of molecular sensors termed tension sensor modules (TSMs) [14, 15, 41]. These TSMs consist of a short protein or peptide with donor and acceptor dye molecules that can undergo fluorescence resonance energy transfer (FRET) on the termini [14, 15]. Because FRET energy transfer is highly sensitive to distance, the FRET signal can be used to infer the distance between the ends of the molecule; this distance can be converted into a force through calibration experiments performed with molecular tweezers, if a specially selected molecule is chosen which does not exhibit hysteresis [15]. Genetically encoded TSMs can then be used to measure the forces felt by certain proteins in living cells, such as those within focal adhesion complexes, which serve as the connection between the internal cytoskeleton and the exterior environment of a cell [42, 214]. These measurements were used to confirm the relevance of 1-20 pN forces in focal adhesion behavior [46]. Through experimentation, different peptides or small proteins have been found that exhibit peak force sensitivity over different ranges [15, 41, 215]. Our ultimate goal is to advance molecular simulation approaches such that we can predict *in silico* the sensitivity of a disordered peptide sequence or small folded protein to pN scale forces.

Molecular Dynamics simulations (MD) can reveal highly detailed molecular-level information about a wide range of biomolecular systems [49, 112]. To explore a biomolecule's conformational landscape using a reasonable amount of computational expense, it is often necessary to employ enhanced sampling techniques that bias the system's behavior in such a way that it can more readily cross barriers in its free energy landscape [2, 216]. As such, a wide range of techniques have been developed, most of which can be categorized by either heating part or all of the system, or adding a bias potential along some or many coordinates termed collective variables (CVs) [217]. MD simulations combined with enhanced sampling techniques can be used to explore the behavior of a system experiencing a constant or time varying mechanical force [4, 218]. In much of our work, we have focused on the *constant force* paradigm, in which case a force applied along a CV such as the end-end distance (d_{end}) of a protein produces a simple modification to the system's Hamiltonian,

$$H(\mathbf{q}, F) = H(\mathbf{q}) - FQ(\mathbf{q}), \tag{4.1}$$

where $Q(\mathbf{q})$ is a CV that depends on $\mathbf{q}$, the configurational degrees of freedom of the system. The negative sign convention is taken such that a positive F corresponds to a pulling force, i.e. which promotes an increase in Q.

Motivated by the problem of computing the force-extension behavior of peptides such as disordered formin domains or peptide tension sensors, we previously developed the method Infinite Switch Simulated Tempering in Force (FISST) [60]. There, we demonstrated that it is possible to sample the effect of a range of forces on a system using a single simulation which includes a combination of (a) a special CV-dependent force, and (b) an observable weight function that allows one to reweight samples to any intermediate force, as described in the next section. We also demonstrated that FISST can promote transitions between otherwise kinetically inaccessible states of a system due to the action of the additional bias potential. This method was implemented and released as a module in the PLUMED open source sampling library [143, 172], and we also described its use in a PLUMED masterclass.[1]

However, in some cases, we find that when applying FISST to peptides or proteins for which small forces should result in a population of extended states, the system remains trapped near its initial configuration. We therefore wish to combine the efficiency of FISST for sampling many simultaneous forces with a method that is more effective at exploring conformational states of the molecule.

Here, we demonstrate that the performance of FISST can be improved by coupling it with Replica Exchange (RE) approaches [2, 216, 219] using three benchmark systems of increasing difficulty (Fig. 4.1). After giving a theoretical overview of FISST and how it is naturally coupled with RE, we demonstrate that FISST combined with temperature replica exchange accelerates sampling for our previous test case of an alanine decamer [60]. We then give the example of the achiral Aib_9 helical peptide, where FISST alone is not enough to destabilize the folded state, but FISST combined with temperature or solute tempering allows robust sampling of the $F = 0$ free energy landscape, and prediction of the force extension curve for this molecule. Finally, we show data computing the force-extension behavior for a more complicated molecule, a villin headpiece mutant; this system is both well characterized in MD simulations and is a variant of a protein whose force-extension behavior has been measured experimentally as a TSM [14, 15, 31, 41].

---

[1] Number 22-15, `https://www.plumed.org/masterclass`

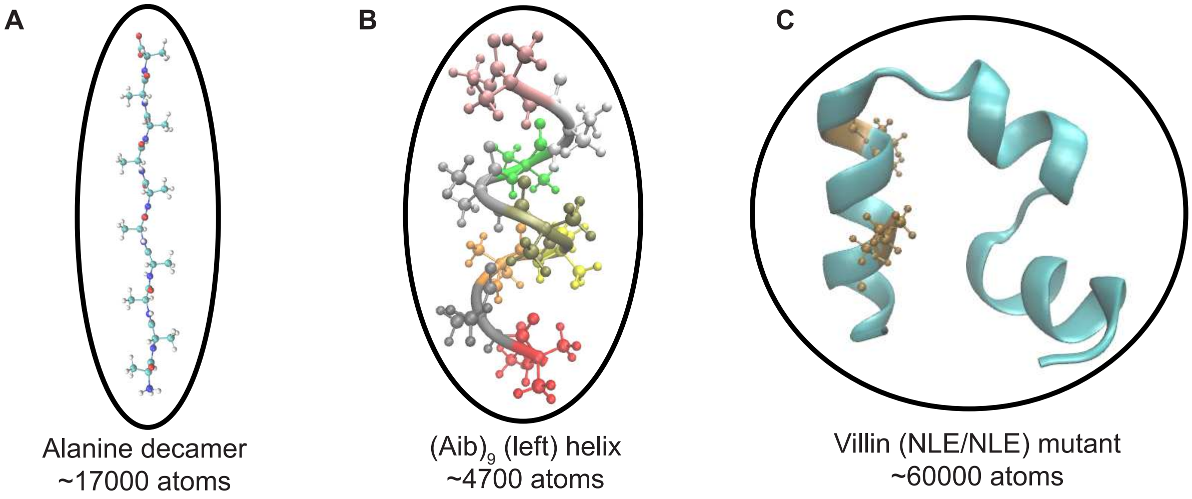


Figure 4.1: Systems probed in this study, shown without solvent for clarity. (A) Solvated alanine decamer starting in the extended state. (B) Solvated Aib_9 molecule starting from the left-handed helical state. Each residue is colored according to the residue ID number. (C) Solvated villin (NLE/NLE) mutant starting in the folded state. Locations of residue mutations are colored in ochre. In all cases, pulling forces are applied to the terminal C_α atoms.

4.2 THEORY

4.2.1 FISST OVERVIEW

The aim of FISST is to compute averages of observables $O(\mathbf{q})$ when a constant force F is applied along a collective variable $Q(\mathbf{q})$. At constant temperature, this corresponds to

$$\langle O \rangle_F = \frac{\int d\mathbf{q} O(\mathbf{q}) e^{-\beta U(\mathbf{q}) + \beta F Q(\mathbf{q})}}{Z_q(F)}, \tag{4.2}$$

where $\beta = 1/(k_\mathrm{B} T)$, U is the potential energy function for the system, and $Z_q(F) \equiv \int d\mathbf{q} e^{-\beta U(\mathbf{q}) + \beta F Q(\mathbf{q})}$ is the configurational partition function for a given F.

In Ref. [60], we showed that averages of this type can be obtained from a single simulation with a modified applied force $\bar{F}(Q)$ that is derived from the infinitely fast switching limit which would arise if sampling a ladder of applied forces from F_{min} to F_{max}.

In this limit, the probability density that would be sampled is a weighted average over all forces, with weights $\omega(F)$ that say how important each force is,

$$\bar{\rho}(\mathbf{q}) = \frac{\int_{F_{\min}}^{F_{\max}} dF'\omega(F')e^{-\beta U(\mathbf{q})+\beta F' Q(\mathbf{q})}}{\int_{F_{\min}}^{F_{\max}} dF' Z_q(F')\omega(F')} \tag{4.3}$$

From this distribution, we get the potential of mean force up to an additive constant through

$$e^{-\beta A(\mathbf{q})} \equiv \int_{F_{\min}}^{F_{\max}} dF'\omega(F')e^{-\beta U(\mathbf{q})+\beta F' Q(\mathbf{q})} \tag{4.4}$$

The FISST algorithm attempts to learn $\omega(F)$ "on-the-fly" such that forces are sampled evenly, which occurs when $\omega(F) \propto 1/Z_q(F)$, and this is accomplished in an iterative manner. After doing so, the integral of their product becomes a constant $C \equiv \int_{F_{\min}}^{F_{\max}} dF'\omega(F')Z_q(F')$

From the potential $A(\mathbf{q})$ in Eq. 4.4, we can get the forces to apply in an MD simulation that will sample from this probability density by taking the negative gradient with respect to atomic positions,

$$-\nabla A(\mathbf{q}) = -\nabla U + \bar{F}(Q)\nabla Q, \tag{4.5}$$

where

$$\bar{F}(Q) = \frac{\int_{F_{\min}}^{F_{\max}} dF'\omega(F')F' e^{\beta F' Q(\mathbf{q})}}{\int_{F_{\min}}^{F_{\max}} dF'\omega(F')e^{\beta F' Q(\mathbf{q})}}. \tag{4.6}$$

The FISST module in PLUMED works by computing $\bar{F}(Q)$ for any choice of CV Q, and then modifying the forces used in any compatible MD engine by adding $\bar{F}(Q)\nabla Q$. We note that this need not be a simple force/distance pair, but could be a more general quantity, e.g. a tension and an area or an electric field and a dipole moment.

After simulating with this modified potential, averages of observables at different forces can be computed from a weighted average over N_t snapshots by including 'observable weights' $W_F(\mathbf{q})$

computed on the fly [60],

$$\langle O\rangle_F = \frac{1}{N_t}\sum_{i=1}^{N_t} W_F(\mathbf{q}(t))O(\mathbf{q}(t)), \tag{4.7}$$

where

$$W_F(\mathbf{q}) = \frac{Ce^{\beta U(\mathbf{q})+\beta FQ(\mathbf{q})}}{Z_q(F)\int dF'\omega(F')e^{\beta(F'-F)Q(\mathbf{q})}} \tag{4.8}$$

4.2.2 OVERVIEW OF REPLICA EXCHANGE METHODS

In replica exchange simulations, a Markov Chain Monte Carlo procedure is carried out, with detailed balance in exchanges ensuring that each replica maintains a particular equilibrium distribution [2, 216].

In Hamiltonian replica exchange, each replica is simulated via its own Hamiltonian H_i, which could be simulated at inverse temperature β_i. Within each copy of the simulation, configurations appear with probability $P_i(\mathbf{q}) \propto \exp(-\beta_i H_i(\mathbf{q}))$ [173]. Ensuring detailed balance of exchange between configurations $\mathbf{q}$ and $\mathbf{q'}$ generated from Hamiltonians H_i and H_j respectively using a Metropolis criterion requires that [173, 220].

$$P_{\text{accept}} = \min\left(1, \frac{P_i(\mathbf{q'})P_j(\mathbf{q})}{P_j(\mathbf{q'})P_i(\mathbf{q})}\right) \tag{4.9}$$

In this case,

$$\frac{P_i(\mathbf{q'})P_j(\mathbf{q})}{P_j(\mathbf{q'})P_i(\mathbf{q})} = \frac{e^{-\beta_i H_i(\mathbf{q'})}e^{-\beta_j H_j(\mathbf{q})}}{e^{-\beta_i H_i(\mathbf{q})}e^{-\beta_j H_j(\mathbf{q'})}} \tag{4.10}$$

$$= e^{-\beta_i(H_i(\mathbf{q'})-H_i(\mathbf{q}))+\beta_j(H_j(\mathbf{q'})-H_j(\mathbf{q}))} \tag{4.11}$$

Below, when we combine FISST with temperature replica exchange (TRE), β_i will be different for each replica, but we will still be performing a form of Hamiltonian exchange due to the different $\bar{F}_i(Q)$ computed in each replica. When performing solute tempering, β_i will be identical for all replicas, however in addition to the different bias applied, the potential energy function will also be different between replicas in such a way as to represent effectively higher solute temperatures.

The partial tempering variants that we employ here are based on the Replica Exchange with Solute Tempering (REST) idea [67, 221, 222]. In this work, we use REST3 [68] which scaled interactions between solute and solvent in a way that was shown to not suppress extended configurations of peptides at higher effective temperatures as could occur with earlier REST variants. REST2 and REST3 are formulated such that for N replicas, the target temperature of replica i is given by $T_i = T_0(\frac{T_{\max}}{T_0})^{i/(N-1)}$, for i from 0 to $N-1$ [67, 68]. The potential energy function of each replica $U_i(\mathbf{q})$ scales the protein-protein and protein-water interactions by factors λ_i^{pp} and λ_i^{pw} respectively [67, 68]. REST3 introduces an additional scaling factor for the non-electrostatic contributions to the protein-water interactions κ_i. The total potential energy in replica i is then given by,

$$\begin{aligned} U_i(\mathbf{q}) = \lambda_i^{pp} U_{pp}(\mathbf{q}) + \lambda_i^{pw} U_{pw}^{elec}(\mathbf{q}) \\ + \kappa_i \lambda_i^{pw} U_{pw}^{non-elec}(\mathbf{q}) + U_{ww}(\mathbf{q}), \end{aligned} \tag{4.12}$$

with $\lambda_i^{pp} = T_0/T_i$, $\lambda_i^{pw} = \sqrt{T_0/T_i}$, and $\kappa_i = 1 + 0.005(m-3)(m > 3)$, with REST2 being recovered if κ_i is set to unity for all i [67, 68].

4.3 COMBINING FISST WITH REPLICA EXCHANGE

Hamiltonian RE is implemented in GROMACS through the PLUMED plugin library [143, 172, 220]. In general, PLUMED functions by computing at every step the values of one or several CVs, and then a 'bias' energy and forces which is a function of the current CV or CV values. When performing Hamiltonian exchange with GROMACS, PLUMED can use different sets of parameters corresponding to each replica to compute the bias function [220]. GROMACS computes the forcefield energy for the original configurations and the swapped configurations. The bias and forcefield energies are combined, and the total potential energy before and after a proposed swap are compared, with the swap accepted or rejected using Eq. 4.11. In this way, very generic replica exchange schemes can be implemented, such as the combination of solute tempering and FISST implemented here.

To enable the combination of FISST with RE, we had to modify our PLUMED implementation such that statistics gathered for computing the quantities $\omega_i(F)$ and $Z_q^i(F)$ which are needed for computing the on-the-fly force $\bar{F}_i(Q)$ are properly computed during the exchange procedure (prior code would update statistics every time the bias is computed, which occurs three times during the exchange procedure). This revised code is available from the github page for this paper (see data availability statement) and will be contributed to our FISST module in the public PLUMED library soon.

Computing these quantities using data from the parallel simulations should improve convergence of the weights, however as discussed in Ref. [171] and [60] the observable weights are correct even before these quantities are converged, and in practice the weights assigned to each force, $\omega(F)$, can converge quickly so this often does not have a major effect. We also perform simulations where the weights are fixed after an initial equilibration phase.

4.4 METHODS

In this section we provide an overview of the systems studied and simulation protocol. Specific details regarding system setup and simulation parameters are provided in the Supporting Information.

4.4.1 SYSTEM DETAILS

In this section, we describe the three systems that we will study in this paper. Further simulation data are provided in the Supporting Information.

1. **Ala$_{10}$**—This system is the same as that used in our previous FISST study [60]. In summary, the system consists of a cubic box of size 56.0 Å, solvated using TIP3P water [223] and parameterized using the CHARMM36 forcefield. The total system size is 17293, including 5730 water molecules. The system is simulated at 300K.

2. **Aib$_9$**—GROMACS [200] inputs were provided by the authors of Ref. [224]. The system consists of a cubic box of length 35.0 Å, solvated using TIP3P water molecules [223], and parameterized using the CHARMM36m forcefield [225]. The total size of the system was 4749 atoms including 1540 water molecules. The net charge of the system was neutral with no additional ions added. The system is simulated at 400K.

3. **Villin Mutant**—Inputs for this system were those generated according to the protocol in Ref. [226]. The 35-residue Villin headpiece "HP35" mutant (PDB ID: 2F4K [227]) was constructed in a cubic box of length 86.80 Å, solvated using TIP3P water molecules [223], and parameterized using Amberff99SB*-ILDN forcefield [228]. This fast folding mutant has two lysines replaced with the non-natural amino acid norleucine [227]. The total system size is 60392 atoms including 19928 water molecules. The system was neutralized and ions were added to bring the system to a 40 mM salt concentration (15 Na$^+$ ions, 16 Cl$^-$ ions). The system is simulated at 298K and 360K. We note that these are the same simulation parameters as described in Ref. [229] and have also provided those details in our Supporting Information.

4.4.2 PRODUCTION RUNS

4.4.2.1 OVERVIEW

Production data were collected using the GROMACS MD engine [200]. All single process MD were run in GROMACS 2020.4, while Hamiltonian exchange simulations were run in GROMACS 2019.6 patched with PLUMED version 2.7.0 [143]. Simulations performed at constant force employed the `RESTRAINT` feature in PLUMED [172].

4.4.2.2 FISST DETAILS

The FISST [60] algorithm and single force calculations (applied with the RESTRAINT keyword) were performed using PLUMED [143]. In all cases, the bias is applied along a collective variable that is the distance between the first and last C_α atoms of the peptides. The FISST force range chosen for Ala_{10} was [-10pN:10pN] and for all other simulations [-10pN:20pN], discretized over 121 gridpoints to perform the integrals [60]. An initially uniform distribution for the force weights was used. For Aib_9, weights were updated every 500 steps (1 ps) and both observable and restart data were also saved every 500 steps. For Ala_{10} and HP35, the weights were updated every 1000 steps (2 ps), and the observable data and restart data were also saved for the same number of steps.

4.4.2.3 REST3 SIMULATIONS

We implemented the REST3 [68] algorithm for all of our multiple process MD runs. For Ala_{10} we choose tempering parameters λ and κ parameters using the script provided by Ref. [68] to simulate a solute temperature range of 300K to 600K over 10 replicas, and for Aib_9 we chose 400K to 800K. For Villin mutant we ran two sets of λ and κ values; with one set of 8 replicas from 298K to 450K and another set of 8 replicas from 360K to 500K. Exchanges were attempted every 5 ps. Our REST3 inputs, scripts, and instructions to set up GROMACS topologies for REST3 simulations can be found on the manuscript GitHub (see below).

4.4.3 Data Analysis

All trajectory files were analyzed using the PLUMED driver and `mdtraj` [144] in python 3.8.0 Trajectory and structure files were visualized in VMD 1.9.3 [149].

4.5 Results and Discussion

4.5.1 Simulations of polyalanine validate implementation of hybrid sampling approach

In our previous work, we demonstrated using the alanine decamer (Ala_{10}) that FISST could accurately compute the end-end distance distribution at a range of forces from a single simulation, as compared to a reference TRE simulation [60]. Taking this as a stand-in for the more complicated peptides that we wish to probe in the future, we chose this as a benchmark to check that combining FISST with RE does not degrade performance.

Here, we compute the end-end distance probability distribution functions for forces ranging from −10 pN to 10 pN using a combination of FISST and RE approaches. We first combined TRE and FISST by running 40 parallel FISST simulations at the same temperatures as our reference TRE simulation, using 100 ns for each replica and a force range of [-10pN:10pN]. As benchmarks, we also show previously obtained results for a 500 ns FISST calculation at $T = 300K$, and TRE calculations at individual forces computed using 40 replicas of 160 ns per window (6.4 μs total simulation time) with temperatures ranging from 300K to 400K [60]. For all analyses presented here, we compute results using the bottom replica.

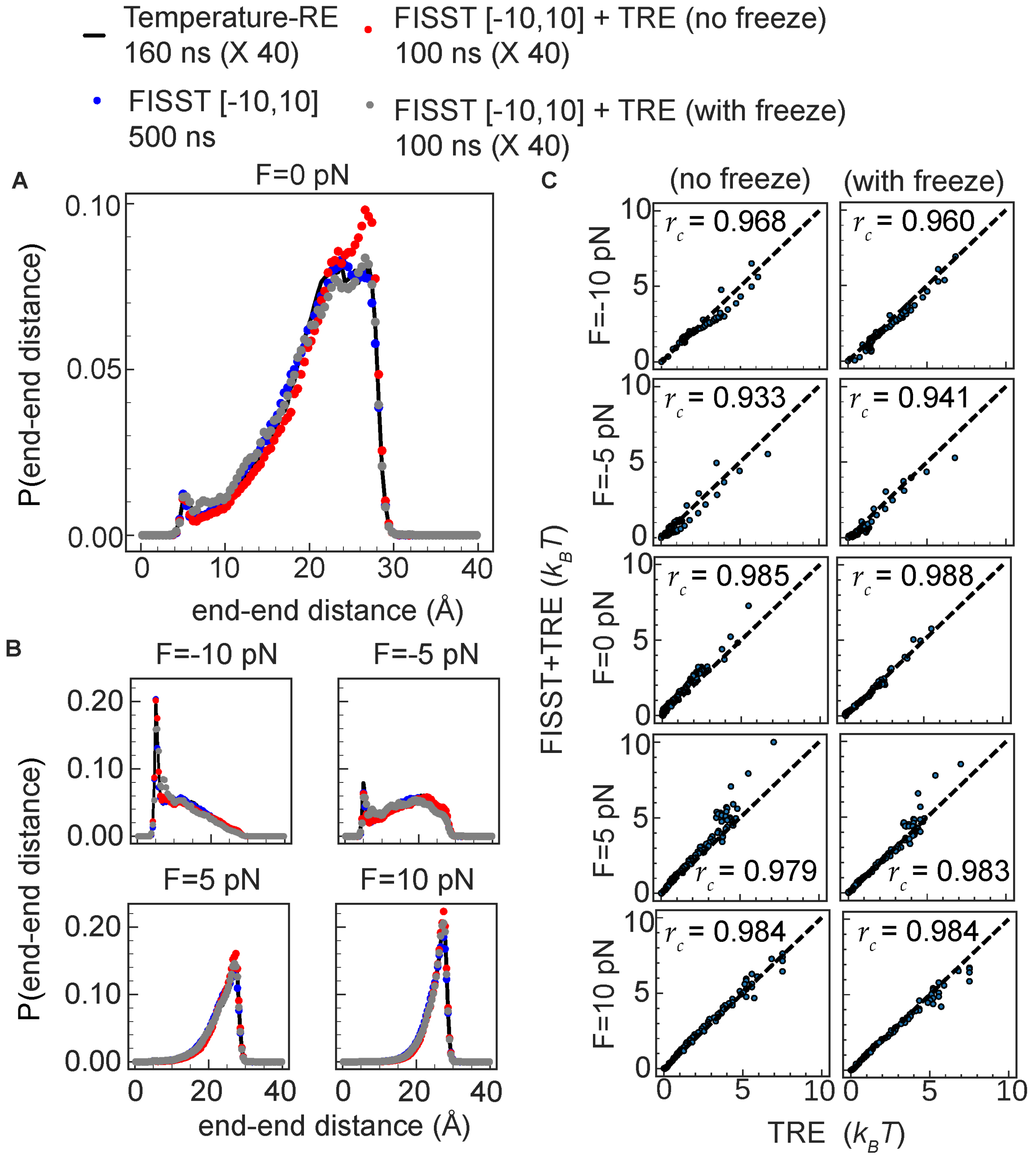


Figure 4.2: (A) End-end distance distributions for Ala_{10} at $F = 0$ for TRE (black solid line), FISST (blue spheres), FISST+TRE without freezing weights (red spheres), and FISST+TRE with freezing (gray spheres). (B) End-end distributions for F =-10, -5, 5, and 10 pN. (C) Comparison of free energies (see Fig. 4.7) computed from end-end distribution functions, comparing FISST+TRE with and without freezing to corresponding reference TRE data at -10, -5, 0, 5, 10 pN forces.

In Fig. 4.2, we show a comparison of these methods for fixed forces of -10, -5, 0, 5, and 10 pN. We find a reasonable visual agreement from all our simulation methods relative to our previous FISST and Temperature replica exchange data at all forces, including a peak representing a collapsed state at ~5 Å for 0, -5, and -10 pN (Fig. 4.2A,B).

However, our initial FISST+TRE run for which we did not freeze the FISST weights shows a slightly higher peak at ~27 Å compared to the FISST and Temperature RE runs at the zero force (Fig. 4.2A, red spheres). We repeated these runs with frozen weights obtained by simulating the parallel replicas without any exchange attempts for 20 ns each, and then continuing with FISST+TRE with those weights fixed using the `FREEZE` option in the FISST code. In this case, the data with freezing (Fig. 4.2A, gray spheres) has a more accurate peak at ~27 Å and better qualitative agreement with the TRE reference.

To check our results quantitatively, we computed the free energy profiles $A(Q)$ from the probability distribution functions at the different forces shown in Fig. 4.2A,B by taking $A(Q) \equiv -k_{\mathrm{B}}T\ln(P(d_{\mathrm{End}}))$ and subtracting an offset such that the minimum in all cases was zero (see Fig. 4.7). We then constructed scatter plots of the two sets of FISST+TRE free energies (with and without freezing the weights) versus the TRE data at each of the corresponding forces, with results shown in Fig. 4.2C. For each of the free energy scatter plots we computed the Spearman's rank correlation coefficient [230] (r_c) using `stats.spearmanr` function implemented in `scipy` [54, 182]. Although both sets of data gave relatively high r_c values when averaged over the 5 forces (0.9698 and 0.9712 without and with freezing the weight distributions, respectively), we observe a slight improvement when freezing the weights for the cases of -5, 0, and 5 pN forces.

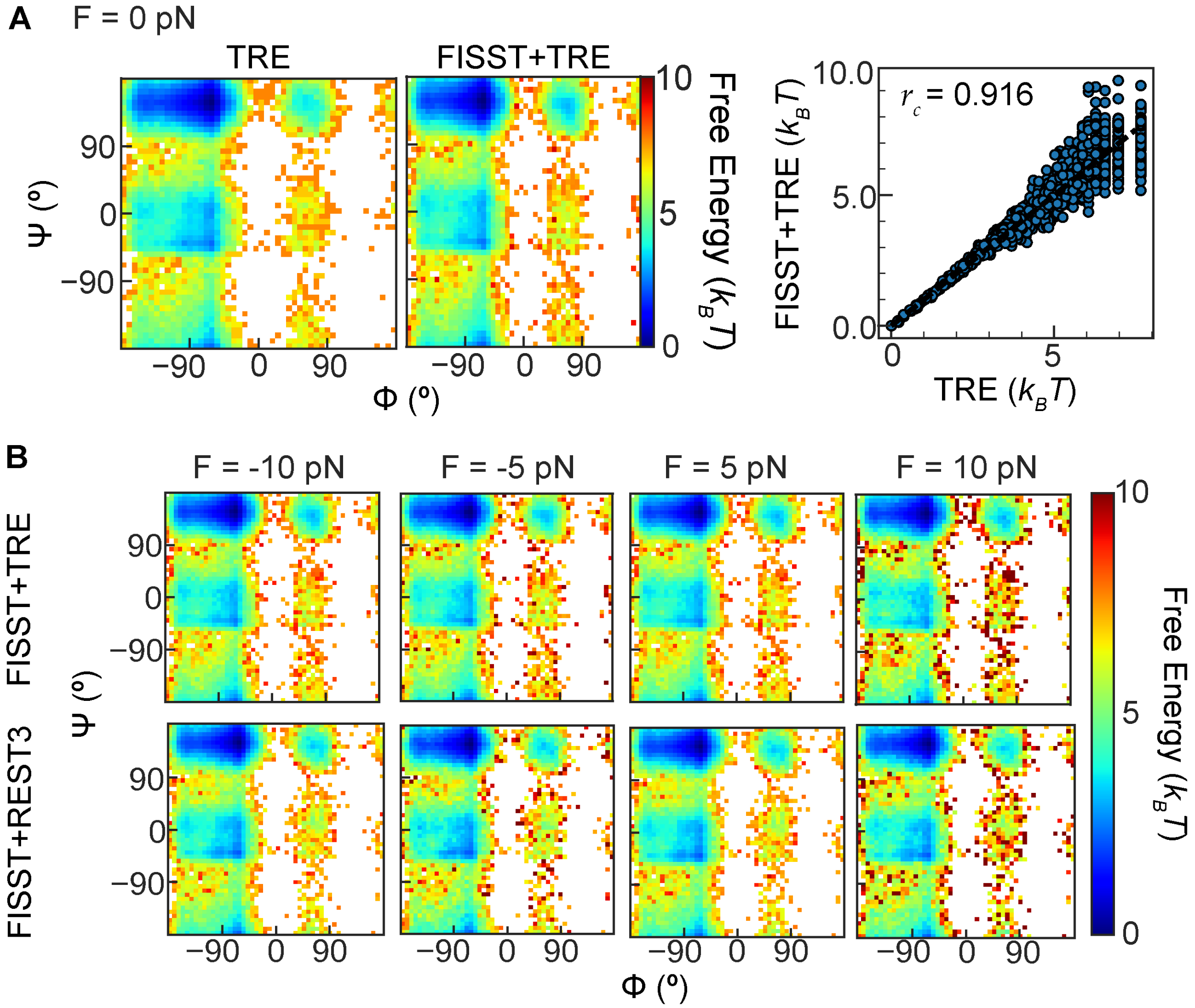


Figure 4.3: (A) (left) TRE and FISST+TRE Ramachandran plot at zero force and (right) scatter plot comparing FISST+TRE and TRE free energies at zero force. (B) Ramachandran plots at -10, -5, 5, and 10 pN forces for FISST+TRE (top row) and FISST+REST3 (bottom row).

We emphasized in our previous work [60] that using the observable weights (Eq. 4.8) we are able to reconstruct averages of other observables at any force besides the one that was biased. In addition to end-end distance, we also reconstructed Ramachandran plots for Ala_{10} FISST+TRE simulations, reweighting at the zero force shown in Fig. 4.3A. We also computed r_c between the free energies in both sets of data and found a relatively high value of 0.916. In Fig. 4.3B we show the result of reweighting the backbone dihedral angles at -10, -5, 5, and 10 pN forces and noted the expected strengthening of the PPII basin (top left) at high force and destabilization of the alpha-helical basin right center as force is increased [59].

To demonstrate the accuracy and generality of our implementation, we also performed FISST simulations coupled to REST3 and repeated our end-end distance and Ramachandran angle analysis for the -10, 0, and 10 pN forces, using REST3 simulation data collected at those forces. We carried an initial FISST+REST3 run without freezing the weights and another FISST+REST3 by freezing the weights after 20 ns in an analogous fashion, and compared the results of each run to our reference data (Fig. 4.8, Fig. 4.9). In Fig. 4.8A we again find reasonable qualitative agreement in the probability distribution functions at all the forces shown. Qualitative analysis in Fig. 4.8A finds relatively high r_c values for both FISST+REST3 runs, an average of 0.982 and 0.977 when freezing the weights and without, indicating a slight improvement with respect to the reference REST3 calculations when freezing. We also reweighted the Ramachandran angles calculated from FISST+REST3 (Fig. 4.3B (bottom row)) at the forces shown to visually demonstrate that the combination of FISST+REST3 gives equivalent results to FISST+TRE, with quantitative analysis shown in Fig. 4.9.

4.5.2 Simulations of Aib$_9$ show improved performance from hybrid FISST+RE sampling

While our results on Ala$_{10}$ show that we are able to combine FISST with replica exchange techniques, they do not demonstrate an obvious improvement that requires such a hybrid method. In this section, we show that FISST alone may not be able to sample the free energy landscape of a structured peptide, necessitating the additional sampling from tempering.

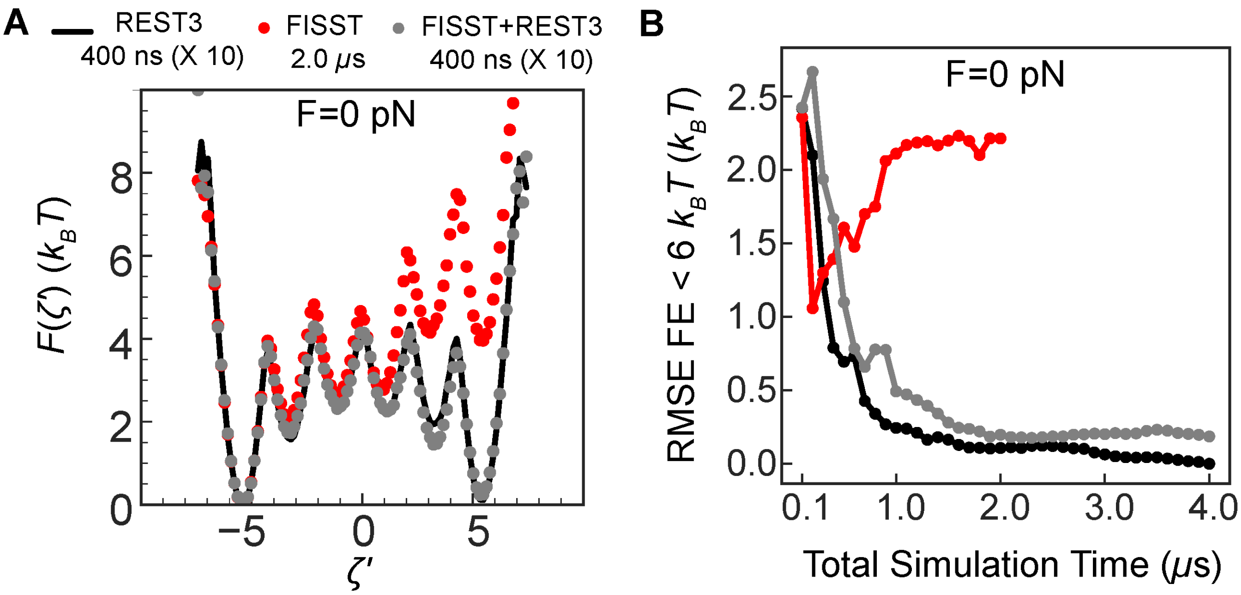


Figure 4.4: (A) $F(\zeta')$ at zero force calculated from REST3 (black solid line), FISST (red spheres), and FISST+REST3 (gray spheres). (B) RMSE of $F(\zeta')$ for values below 6 k_BT using data points from different simulation time windows.

Here we analyze results for the achiral Aib$_9$ system starting from a left-handed configuration at $T = 400K$, for which we initially performed a 4.0 μs unbiased simulation and a 2.0 μs FISST simulation for the force range [-10pN:20pN]. While the unbiased simulation shows a transition rate of approximately 1 inversion per 2 μs (Fig. 4.10), the FISST simulation actually does not, showing a case where FISST can impede conformational exploration. Hence we felt this is an ideal test system for demonstrating the effectiveness of FISST+RE.

We measured the chirality transition of the Aib$_9$ helix using the ζ' coordinate defined as the negative sum of the five inner ϕ dihedral angles shown in Fig. 4.10, as done in previous studies.[224, 226, 231] With this definition using angles in radians, the left handed helical configuration takes on a value of $\zeta' = -5$ and the right handed $\zeta' = 5$. We constructed $F(\zeta')$ at zero force for each of our simulations using 100 equally spaced windows starting from $\zeta' = -7.5$ to $\zeta' = +7.5$ (Fig. 4.4A).

When performing 4μs (400 ns × 10 replicas) of REST3 simulations spanning 400K–800K, a symmetric free energy profile is obtained (black solid line) from the lowest replica in the ladder, showing that solute tempering is an effective sampling approach for this model problem. This is in contrast to the FISST data (red spheres) which fails to sample the right-helix basin even after 2 μs of simulation time.

We then proceeded to combine FISST with REST3 by using the same 10 replicas but adding FISST sampling on the range [-10pN:20pN]. We find that the FISST+REST3 data (gray spheres) not only overcomes the poor sampling from the FISST method alone but also converges with the benchmark REST3 data. Snapshots depicting some molecular configurations observed in this process are shown in Fig. 4.11. To quantify the accuracy of the combined sampling, in Fig. 4.4B we computed the root-mean-squared error (RMSE) of $F(\zeta')$ for free energies below $6k_\mathrm{B}T$ (chosen to encompass all of the metastable states based on Fig. 4.4A). We computed the RMSE at $F = 0$ for progressively longer time windows starting with 100 ns. At short times, the simulation does not adequately sample the entire $\zeta' = [-7.5, 7.5]$ range as it remains near the left-handed state, resulting in a high error. While as previously noted, the FISST alone simulation does not converge, the FISST+REST3 converges towards the REST3 reference to less 0.5 kcal/mol (~0.63 $k_\mathrm{B}T$ for $T = 400\mathrm{K}$) in approximately 300 ns of total sampling. The same trends hold when using all bins for the RMSE calculation (Fig. 4.12). While the FISST+REST3 curve does not approach zero, this appears to be due to simply finite sampling resulting in slightly more data in the almost-right metastable state for FISST+REST3 and slightly more data in the almost-left state in the reference calculation.

It should also be emphasized that from a practical point of view, if many processors are available, the FISST+REST3 may be faster in wall clock time than running a single long trajectory, where for example we needed a microsecond or more of Metadynamics simulation to converge a good free energy profile for this system, even with a good reaction coordinate [226] (see Table 4.9 for simulation times). We also note that the FISST+REST3 data also contains additional information about all forces from -10 to 20 pN, which we will discuss next, making it much more efficient when this data is needed.

In Fig. 4.5A we show the force-extension curve obtained from our simulations, by reweighting the end-end distance data to compute a mean distance $\langle d_{\text{end}} \rangle$ as a function of force. This is an academic exercise since experimental data for this system is not available. Here we compare the force-extension curve for reference REST3 simulations performed at different forces of 0, 10, and 20 pN with our FISST alone or FISST+REST3 simulations. We assess accuracy in Fig. 4.5B as we did in Ref. [60] by computing the Jenson-Shannon Distance[181] between the reweighted end-end probability distributions. While this analysis shows that our FISST+REST3 result is accurate, it also appears that the FISST alone result is accurate. This is because, evidently, the Aib$_9$ helix is quite resistant to extensional force, and the response is predicted correctly even when trapped in only one helical state.

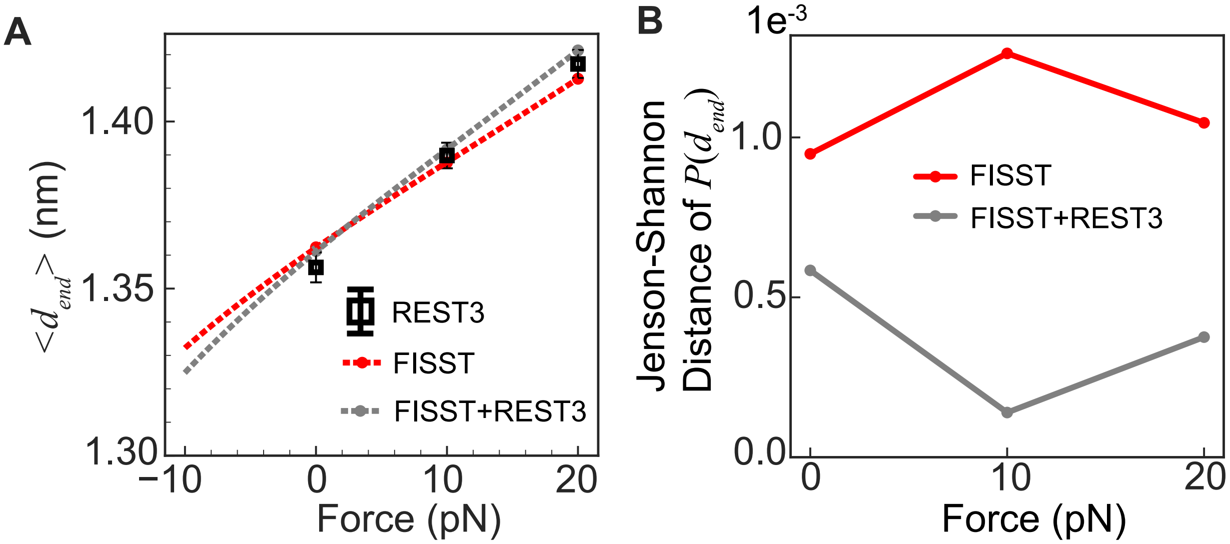


Figure 4.5: (A) Theoretical Force vs average End-to-end distance $\langle d_{End} \rangle$ curve for Aib$_9$ calculated from FISST (red dashed line) and FISST+REST3 (gray dashed line). End-to-end distance values calculated from REST3 simulations at 0, 10, and 20 pN forces (black), and unbiased MD (blue) are embedded for comparison. (B) Jenson-Shannon distance of $P(d_{End})$ calculated for FISST (red) and FISST+REST3 (gray) for 0, 10, 20 pN forces using REST3 simulation data as the reference.

4.5.3 Villin (NLE/NLE) mutant simulations allow us to assess performance on a TSM-like molecule

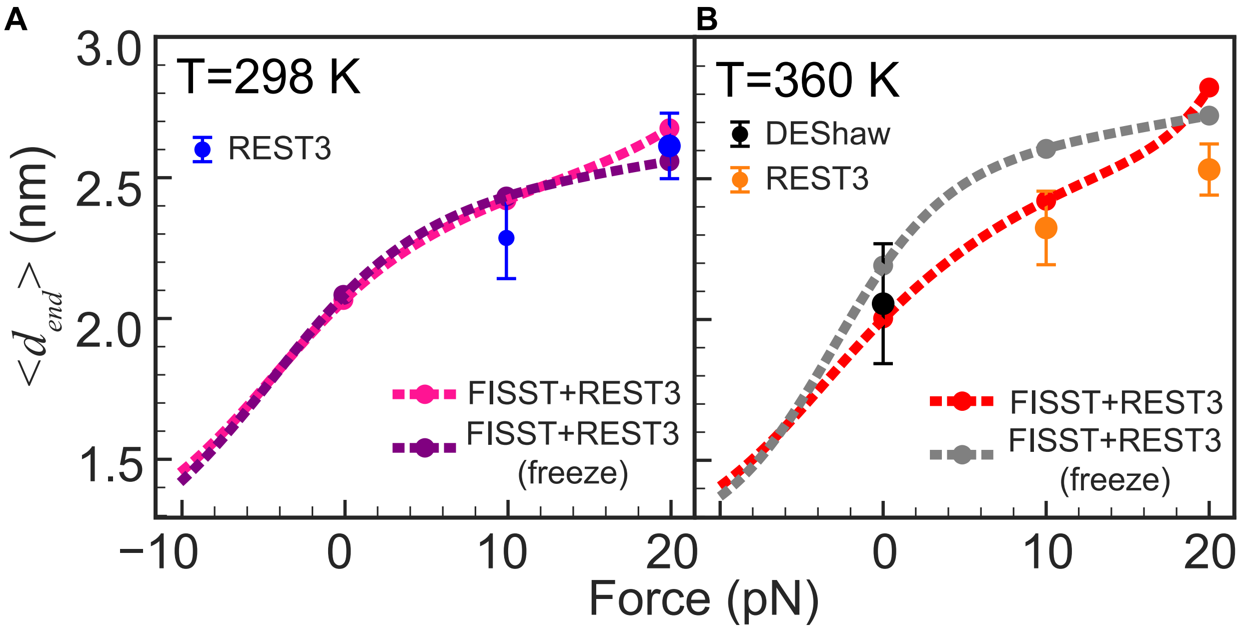


Figure 4.6: (A) Villin Mutant Force vs average End-to-end distance $\langle d_{End} \rangle$ calculated for FISST + REST3 without freezing the weights and freezing the weights at 298K. $\langle d_{End} \rangle$ values calculated from REST3 simulations at $F = 10$ and 20 pN. Error bars represent 1/3 of the standard deviation in length at that force. (B). Similar data as (A) for simulations where the lowest replica is at $T = 360K$, which is close to the melting temperature. Also shown is data from a 310 μs trajectory at $T = 360K$ from Ref. [229].

The resistance of Aib$_9$ to pulling prevents us from showing the full extent of FISST+REST3's performance on force-extension curves. We now wanted to test our approach for a protein used in a TSM. Many such peptides do not have known structures (because they are not ordered), making it difficult to know if we are using a good starting structure or forcefield. We therefore decided to study the villin headpiece domain (HP35) since this protein is both well characterized in experiments and probed as a tension sensor module [31].

Initial test simulations we performed using wild type HP35 showed little stretching within available simulation time, even with parallel tempering approaches, which we attributed to potential forcefield over-stabilization of collapsed states [232, 233]; forcefield choice has also been shown to have a very strong effect on the predicted stability of villin in solution [234]. For this work, we therefore elected to study the Villin (NLE/NLE) mutant whose behavior has been extensively characterized and studied across many simulation studies, and in particular was exhaustively sampled by the DE Shaw Research group [229].

We analyze data collected from two sets of REST3 simulations consisting of 8 replicas, one with a solute temperature range from 298K to 450K and another ranging from 360K to 500K. For each solute temperature range, we ran FISST+REST3 using a force range [-10pN:20pN] for ∼200 ns each (1.6 μs total simulation time). We then repeated these simulations restarting from the point of 20 ns of simulation with weights frozen. We also collected data for $F = 10$ and 20 pN for both solute temperature ranges. An additional reference that we include in our analysis is the ∼310 μs unbiased simulation of Villin (NLE/NLE) at 360K, identical to the simulation parameters used by the authors of Ref. [229]. Consistent with analysis of our other systems, we analyze only the bottom replica.

Fig. 4.6 shows our computed force extension curves for the two temperatures selected. The lower is room temperature, where single molecule pulling experiments are performed on TSMs, and 360K is close to but below the melting temperature for the mutant using this forcefield, so that many more folding/unfolding events are observed in long unbiased simulations [229].

Here we observe an elastic regime where the force extension curve is linear from around −5 pN to +5 pN of pulling force at both temperatures. In both cases, there is a turnover to an inextensible regime, although at each temperature, one of the two data sets shows an indication of entering another stretching regime. At $T = 360K$, the FISST+REST3 curves lie above the reference calculations, which may be an indication of additional sampling of unlikely extended states due to additional sampling from using the hybrid method. We argue that this is due to enhanced sampling rather than hindered sampling, because if anything we would naively expect the FISST method to promote spending times at smaller extensions due to the need to sample the full force range from -10 to 20 pN, but the opposite is observed here.

The data shown here contain a discrepancy between simulations of the same length with and without frozen weights. This could be a consequence of the exchange protocol improving the weight calculation such that the FISST calculation in each replica becomes more efficient, resulting in the non-frozen weight data to lie closer to the reference single-force results.

In Fig. 4.13 we show how the histograms are transformed as the force on the ends of villin is increased. For both 298K and 360K, there is a prominent peak at shorter distances (~ 1.2 nm) for low force which is shifted to a prominent peak at larger lengths (~ 2.5 nm). The high force distributions are unimodal, although there is some evidence for a shoulder developing at 3.0 nm at the highest forces. The linear increase in average length due to a shift between two populations is something we discussed as the most likely scenario for the low force regime when there are two possible states [4], however here we are still remaining within compact states, meaning that predominantly unfolded states are not being accessed here. In contrast, experimental data on wild type HP35 shows a full unfolding with a change in length of 7 nm over this force range. For this situation, we previously speculated based on geometric arguments that the force extension curve would have the behavior like we observe here up through ~ 5 pN, followed by a separation of the folded state into three independent helices up to around 10 pN, at which point the helices begin to populate fully extended states [4], commensurate with the discussion on folded TSMs in Ref. [41]. We hypothesize that our lack of observation of this behavior in the experimentally probed force regime still corresponds to over-stabilization of the folded state or collapsed partially unfolded states by the forcefield/water model.

4.6 CONCLUSIONS

In this work, we demonstrated that our force tempering method can be enhanced through combination with replica exchange approaches. Combination with solute tempering showed definitive improvement for a test case where FISST alone failed. The combined approach is much more efficient than running many individual simulations at different fixed forces when attempting to compute a full force-extension curve as in our final example of the HP35 protein. Also, when combining FISST with TRE, the full force extension profile at all temperatures is obtained simultaneously.

We chose to combine force and temperature sampling by employing a replica exchange approach, which we did because the implementation via a Monte Carlo scheme was practically realizable due to the efforts of the developers of PLUMED and GROMACS [200, 220]. However, we would also like to note that it should be possible to combine infinite switch simulated tempering in force with the infinite switch simulated tempering in temperature (ISST), upon which FISST was originally based [171]. This may be more effective than our approach here, since at least on paper the infinite switch limit is the most efficient choice for parallel tempering [171, 235, 236]. While ISST is implemented in the MIST library [180], combining the two approaches would require efficient implementation of estimating partition functions and weights using two dimensional integrals over both inverse temperature and force which could pose a numerical challenge, hence we chose not to pursue that effort at this time.

Finally, even with our improved sampling method, we have not yet computed a force extension curve that matches one measured experimentally. While it is possible that the difference is due to the difference in solvent conditions (experiments mostly performed in phosphate-buffered saline or similar), or that the experimental curve is not quite right, given the complex setup using tethering molecules and the need to significantly postprocess data from many pulling runs [31, 41], for now we presume that the larger error comes from the simulation side. Given that we have implemented and then improved an effective force sampling approach, this points us towards considering alternative water models and protein forcefields (or modifying terms in the current ones) to find one that best matches the known behavior for a molecule like HP35. We hope that combining our effective sampling approach with the appropriate forcefield will allow us to design *in silico* new tension-sensing peptide molecules.

4.7 DATA AVAILABILITY

All input files, scripts, and output files are available from a GitHub repository for this manuscript. Any additional files will be made available upon request.

4.8 SUPPORTING INFORMATION

4.8.1 SIMULATION DETAILS

4.8.1.1 SYSTEM CONSTRUCTION

	Ala_{10}	Aib_9	HP35	HP35 (Ref. [229])
Forcefield	Charmm36 [237]	Charmm36m [225]	Amberff99SB-ILDN [228]	Amberff99SB*-ILDN [228]
Water model	TIP3P [223]	TIP3P [223]	TIP3P [223]	TIP3P [223]
Box size (Å)	56.0	35.0	86.80	54.0
Salt Concentration (mM)	0	0	40	40
Simulated Temperature (K)	300	400	298, 360	360

Table 4.1: Overivew of system construction details for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35). We have also included system setup details for Villin (NLE/NLE) mutant described in Ref. [229], which we use as a reference for our results in Sec. 4.5.3.

4.8.1.2 MINIMIZATION

	Ala_{10}	HP35
Integrator	Steepest Descent	Steepest Descent
Maximum Force (kj/mol/nm)	1000	1000
Maximum Steps	50000	50000
Nearest Neighbour list type	grid	grid
Cutoff-scheme	Verlet	Verlet
Method for computing long-range Electrostatcs	PME	PME
Short range interactions cutoff (nm)	1	1

Table 4.2: System minimization details provided for Alanine decamer (Ala_{10}) and Villin (NLE/NLE) mutant (HP35). No entries for Left-handed AIB9 helix (Aib_9) have been provided as the equilibrated inputs were directly obtained from the authors of Ref. [224].

4.8.1.3 EQUILIBRATION AND PRODUCTION RUN INPUTS

	Ala_{10}	Aib_9	HP35
Timestep (fs)	2	2	2
Integrator	Leap-frog	Leap-frog	Leap-frog
Thermostat	Berendsen	Nose-Hoover	Berendsen
Barostat	Parrinello-Rahman	Parrinello-Rahman	Parrinello-Rahman
Nearest Neighbour list update frequency	10	20	10
Cutoff-scheme	Verlet	Verlet	Verlet
Method for computing long-range Electrostatcs	PME	PME	PME
Short range interactions cutoff (nm)	1	1.2	1
Algorithm for computing constraints	LINCS	LINCS	LINCS

Table 4.3: Equilibration and Production run parameters used for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35).

4.8.1.4 PRODUCTION RUNS

FISST

	Ala_{10}	Aib_9
Force range (pN)	[-10,10]	[-10,20]
Total simulation time (μs)	0.5	2
Total number of runs	1	1

Table 4.4: Summary of all single-process simulations with FISST. Force range and total simulation time collected provided for Alanine decamer (Ala_{10}) and Left-handed AIB9 helix (Aib_9). For Aib_9, weights were updated every 500 steps (1 ps) and both observable and restart data were also saved every 500 steps. For Ala_{10}, the weights were updated every 1000 steps (2 ps), and the observable data and restart data were also saved for the same number of steps.

REST3

	Ala_{10}	Aib_9	HP35
Solute temperature range (K)	[300,600]	[400,500]	[298,450], [360,500]
Forces (pN)	0, 10, 20	0, 10, 20	0, 10, 20
Number of replicas	10	10	8
Exchange frequency (Number of steps)	2500	2500	2500
Total Simulation Time (μs)	4	4	1.6
Total number of runs	3	3	6

Table 4.5: Summary of all REST3 simulations without FISST implemented. Forces used, Replica Exchange setup, and total simulation time collected provided for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35).

FISST+REST3

	Ala_{10}	Aib_9	HP35
Solute temperature range (K)	[300,600]	[400,500]	[298,450], [360,500]
Number of replicas	10	10	8
Exchange frequency (Number of steps)	2500	2500	2500
Force range (pN)	[-10,10]	[-10,20]	[-10,20]
Total Simulation Time (μs)	4	4	1.6
Ran with frozen weights	Yes	Yes	Yes
Total number of runs	2	1	4

Table 4.6: Summary of all FISST+REST3 simulations without FISST implemented. Force ranges, Replica Exchange setup, and total simulation time collected provided for Alanine decamer (Ala_{10}), Left-handed AIB9 helix (Aib_9), and Villin (NLE/NLE) mutant (HP35). We also specified if we ran additional simulations with the updated PLUMED source code, which freezes the FISST weights. For Aib_9, weights were updated every 500 steps (1 ps) and both observable and restart data were also saved every 500 steps. For Ala_{10} and HP35, the weights were updated every 1000 steps (2 ps), and the observable data and restart data were also saved for the same number of steps.

TRE

	Ala_{10}
System temperature range (K)	[300,400]
Number of replicas	40
Exchange frequency (number of steps)	2500
Forces used (pN)	0
Total simulation time (μs)	4
Total number of runs	1

Table 4.7: Summary of Temperature Replica Exchange (TRE) performed for Alanine decamer (Ala_{10}). Forces used, Replica Exchange setup, and total simulation time collected are provided.

FISST+TRE

	Ala_{10}
System temperature range (K)	[300,400]
Number of replicas	40
Exchange frequency (number of steps)	2500
Forces range (pN)	[0,10]
Total simulation time (μs)	4
Ran with frozen weights	Yes
Total number of runs	2

Table 4.8: Summary of FISST+TRE runs performed for Alanine decamer (Ala_{10}). Forces used, Replica Exchange setup, and total simulation time collected are provided. For Ala_{10}, the weights were updated every 1000 steps (2 ps), and the observable data and restart data were also saved for the same number of steps.

4.8.2 Alanine Decamer

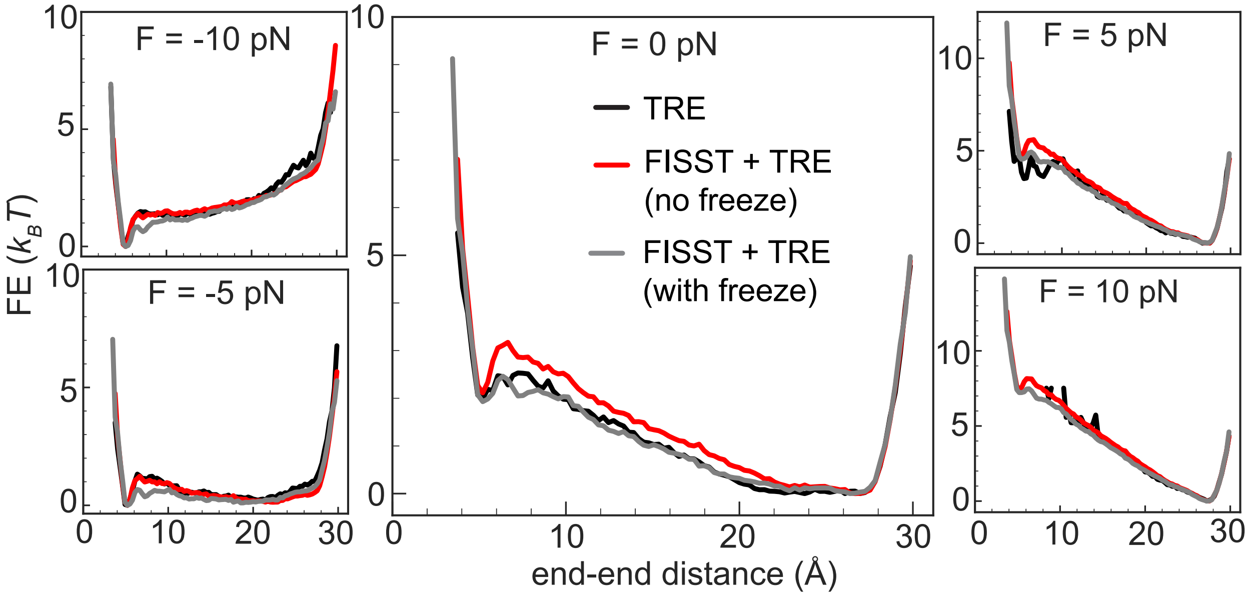


Figure 4.7: Alanine Decamer Free energy profiles for TRE (black solid line), FISST+TRE without freezing weights (red solid line), and FISST+TRE with freezing weights (gray solid line) calculated from the End-to-end distance probability distributions at -10, -5, 0, 5, and 10 pN forces shown in Fig. 4.2A and Fig. 4.2B.

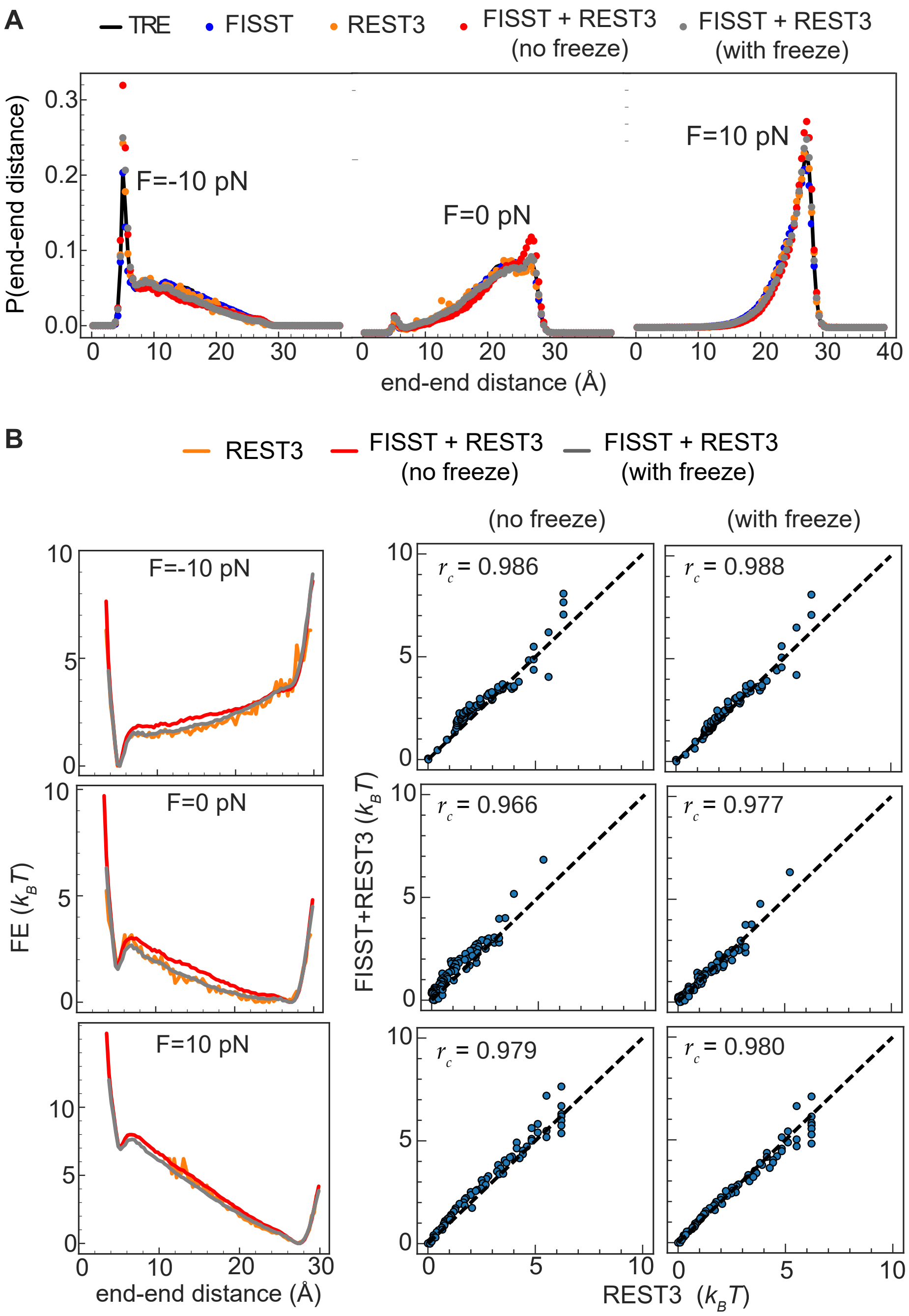


Figure 4.8: (A) Alanine decamer End-to-end distance probability distribution functions calculated for TRE (black solid line), FISST (blue spheres), REST3 (orange spheres), FISST+REST3 without freezing the weights (red spheres), and FISST+REST3 with freezing the weights (gray spheres) at -10, 0, and 10 pN forces. (B) (left) Corresponding free energy profiles. (right) Free energy scatter plots comparing FISST+TRE and TRE data without and with freezing of the weights.

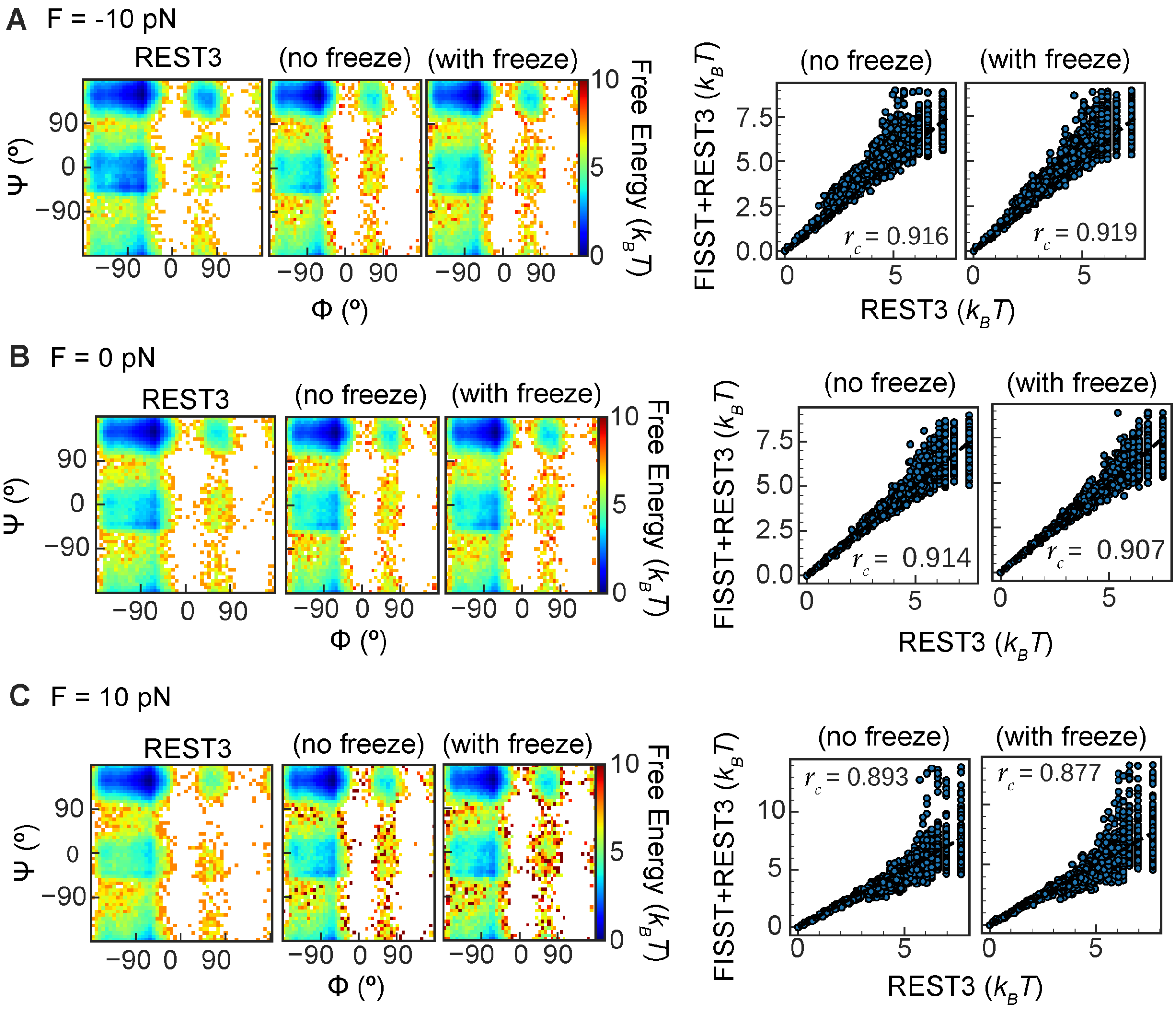


Figure 4.9: (left to right) Ramachandran plots calculated from REST3, FISST+REST3 without freezing the weights, FISST+REST3 with freezing the weights and corresponding free energy scatter plots comparing FISST+REST3 without and with freezing of the weights for (A) -10, (B) 0, and (C) 10 pN forces.

4.8.3 AIB9

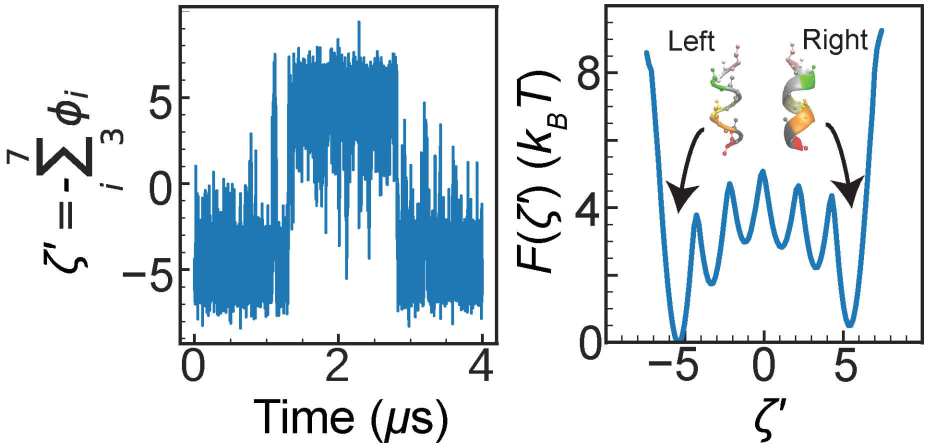


Figure 4.10: (left) Time series plot for ζ' coordinate generated from unbiased MD simulation of Aib$_9$. (right) Corresponding free energy profile of ζ'. (inset) Left and right-handed Aib$_9$ helices are marked in their respective basins.

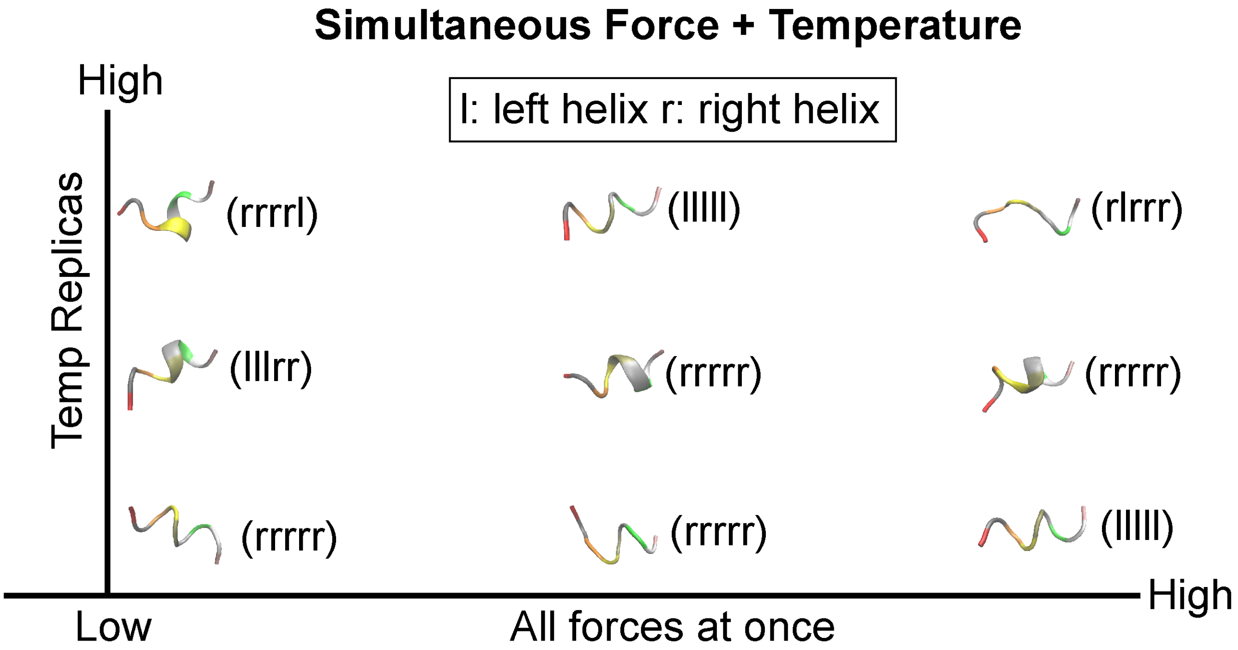

Simultaneous Force + Temperature
High
l: left helix r: right helix
Temp Replicas
(rrrrl)
(lllll)
(rlrrr)
(lllrr)
(rrrrr)
(rrrrr)
(rrrrr)
(rrrrr)
(lllll)
Low
All forces at once
High

Figure 4.11: Snapshots from Aib$_9$ FISST+REST3 trajectory depicting helical compositions at different forces and solute temperature replicas.

	Simulation Time (μs)	Wall time (days:hours)
Unbiased MD	4	14:19
FISST (single MD)	2	8:9
RE (×10)	4	1:16
FISST+RE (×10)	4	1:13

Table 4.9: Summary of total simulation and wall times of all Aib$_9$ runs.

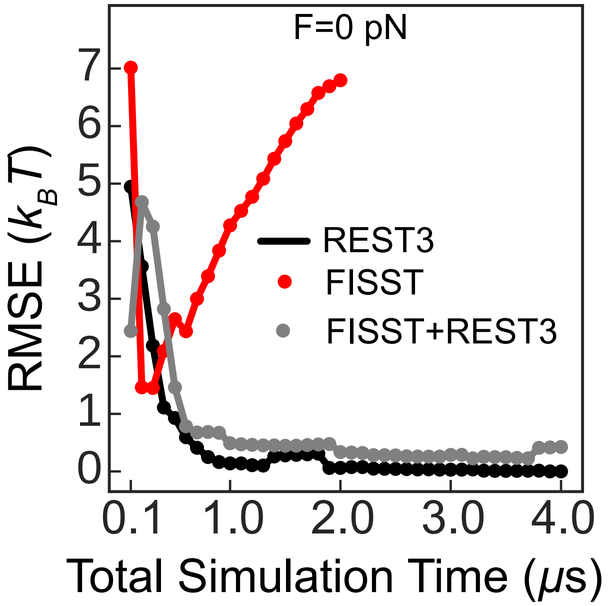


Figure 4.12: RMSE of $F(\zeta')$ at zero force calculated from full free energy profiles from REST3 (black), FISST (red), and FISST+REST3 (gray) using data points from different simulation time windows.

4.8.4 Villin (NLE/NLE) Mutant

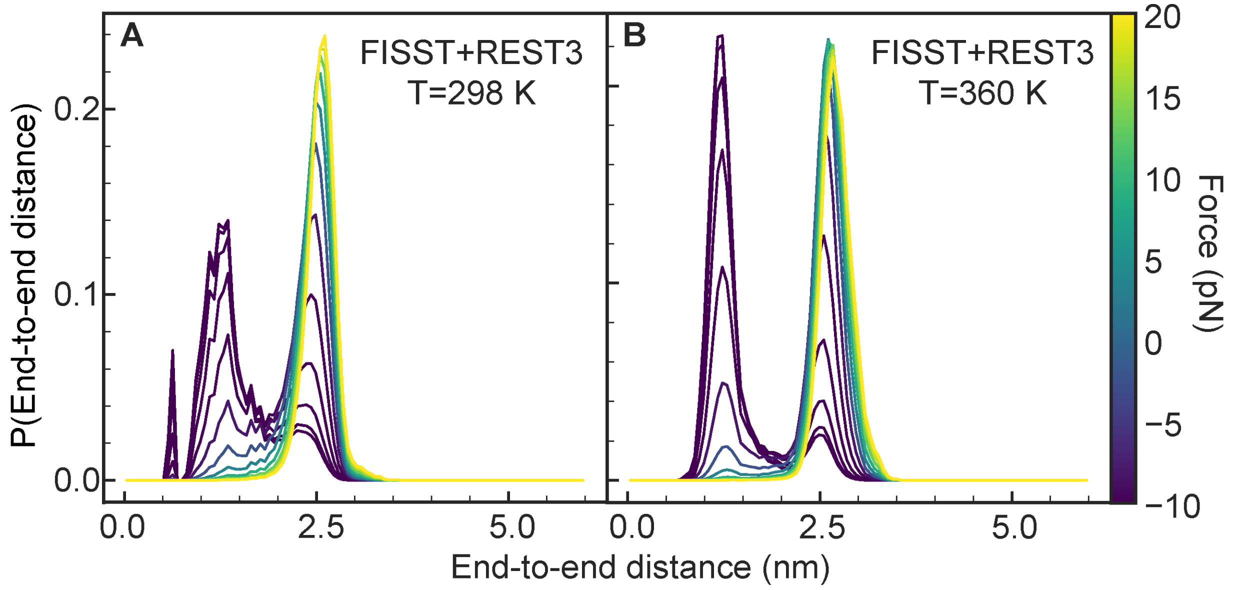


Figure 4.13: Villin Mutant End-to-end distance probability distribution reweighted to forces in the range [-10 pN:20 pN] from FISST+REST3 simulations at (A) Solute temperature of 298K and (B) Solute temperature of 360K. The color bar distinguishes the forces to which the distributions are reweighted.

CHAPTER 5

INFLUENCE OF LINKER SEQUENCE ON TENSION SENSOR MODULE BEHAVIOUR: FORCE INDUCED STRUCTURAL INSIGHTS

This chapter summarizes unpublished work in collaboration with the Hoffman group at Duke University.

5.1 Introduction

Linker peptides, occurring naturally, primarily serve as spacers between protein domains to prevent undesired interactions [238–240]. These linkers often constitute polar and non-polar residues such as Proline, Serine, Glycine, Glutamine, and Arginine. Depending on their amino acid composition and length, diverse linkers can be engineered for constructing *Fusion proteins*, amalgamating the functionalities of two or more protein domains. For instance, **Flexible Linkers**, rich in Glycine, enhance flexibility, facilitating movement for protein domains to interact when necessary [240]. Incorporating polar or charged residues like Serine, Lysine, and Glutamic acid increases hydrogen bonding with water, reducing non-native contact with protein domains. A notable example is the design of a fusion protein with antigen-binding and marker activity, enhancing its robustness for single-step immunodetection [241]. Conversely, **Rigid Linkers**, constructed with non-polar residues like Proline and Alanine, restrict linker mobility and serve as effective spacers to prevent unwanted interactions between protein domains *in vivo*, preserving biological activity [242–244]. Another category is ***In Vivo* Cleavable Linkers**, designed to be protease-sensitive or contain a readily reducible disulfide bond. This allows the fusion protein to be cleaved into separate protein domains *in vivo* when specific sites necessitate individual functionality [245, 246]. Tailoring linkers according to amino acid composition and length offers a diverse library of linkers applicable in various scenarios, highlighting linker design as a relevant yet under-explored domain in Chemical Engineering.

Given the diverse landscape of linker design explored earlier, numerous linker databases have emerged, housing vast repositories of linker peptides tailored for engineering Fusion Proteins. Notable examples include the web-based program managed by the Centre for Integrative Bioinformatics VU (IBIVU) at Vrije University[1] and LINKER [247], collectively containing thousands of linker peptides. These databases serve as valuable resources, offering a foundation for a more rational linker design strategy. Users can define various parameters in the search engine, including linker length (measured in either amino acid residues or Angstroms, Å), amino acid sequence, or secondary structure. However, despite the available databases, a more refined and nuanced characterization of linkers, incorporating considerations of functionality and their respective fusion proteins, holds the potential to significantly advance linker design methodologies. Moving beyond length and sequence would offer a more comprehensive understanding and utilization of linkers in diverse applications.

---

[1] `https://www.ibi.vu.nl/programs/linkerdbwww/`

As we attempt to understand the intricacies of linker design, efforts to illuminate the connection between linker structure and function have been propelled by various Molecular Dynamics (MD) and Monte Carlo (MC) simulation studies. These endeavors seek to unravel the stability of different secondary structures of short peptides in both polar and non-polar solvents [248–253]. Notably, PolyProline II (PPII) helices emerge as promising secondary structures for linkers, standing behind $\alpha-$helices and $\beta-$sheets in versatility. Importantly, despite the nomenclature, the presence of Proline is not a prerequisite for the formation of PPII helices (see Sec. 5.2). These helices not only provide a desirable degree of flexibility but also possess the capability to form hydrogen bonds with water, a crucial feature for mitigating undesirable protein-linker interactions [186]. This becomes particularly pertinent in the realm of Tension-Sensor Module (TSM) engineering, where precision in linker design is paramount.

In this work, we present Monte Carlo and Molecular Dynamics simulations of 15 linkers (Fig. 5.1) with varying amino acid composition and stiffness to predict the secondary structure of potential Tension Sensor Module candidates with and without mechanical force. First, we show the results of Monte Carlo simulations of 6 of the 15 linkers performed in the CAMPARI[2] engine with ABSINTH [254] implicit solvent and forcefield in the absence of mechanical force. After our initial Monte Carlo runs, we proceed to show results from more detailed GROMACS [142] MD simulations in TIP4P-D [232] explicit solvent model and DES-AMBER (a99sb-disp) [234] forcefield on the complete set of 15 linkers, with and without forces. Our CAMPARI MC and GROMACS MD results provide valuable insights into the stability of different secondary structures exhibited by linker peptides. Notably, our simulations hint at the formation of PolyProline II helices, a significant observation in the context of linker design for Tension Sensor Modules. Further details regarding the Monte Carlo and Molecular Dynamics simulation setups can be found in Sec. 5.4.

---

[2] `https://campari.sourceforge.net/`

5.2 PRELIMINARY RESULTS

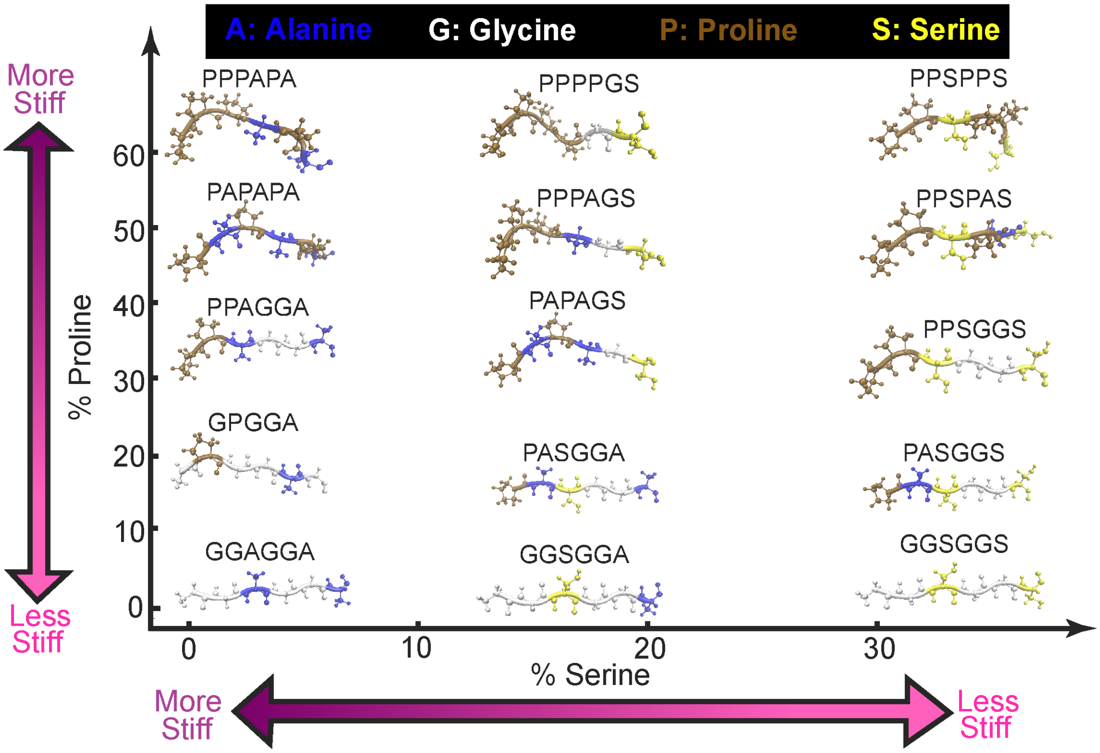

A: Alanine
G: Glycine
P: Proline
S: Serine
More Stiff
Less Stiff
% Proline
0
10
20
30
40
50
60
PPPAPA
PPPPGS
PPSPPS
PAPAPA
PPPAGS
PPSPAS
PPAGGA
PAPAGS
PPSGGS
GPGGA
PASGGA
PASGGS
GGAGGA
GGSGGA
GGSGGS
0
10
20
30
% Serine
More Stiff
Less Stiff

Figure 5.1: Linker peptide repeats probed for MC and MD studies, plotted according to their percentage of Serine and Proline to illustrate their effect on linker stiffness. The color scheme for amino acid residues are Alanine: Blue, Glycine: White, Proline: Brown, Serine: Yellow.

Description of linkers. For our computational studies, we constructed a set of 15 linkers $(\mathrm{X})_N$, where X is the linker repeat (Fig. 5.1) and N is the number of repeats ($N = 3, 5$). All linkers contain only a combination of Alanine, Glycine, Proline, and Serine. In particular, the fraction of Proline and Serine serve as our primary tuning parameters as shown in Fig. 5.1. The presence of non-polar Proline increases stiffness and reduces intramolecular hydrogen bonding due to the absence of amide hydrogen. In contrast, the presence of polar Serine improves flexibility and the side-chain hydroxyl group promotes hydrogen bonding and hence the formation of multiple secondary structure motifs. We also acknowledge that additional variables, including the positioning of Proline and the proportion of Alanine within each repeat, also play integral roles in determining the overall stiffness and the variety of secondary structure motifs that the linkers will explore during simulation.

5.2.1 RESULTS FROM CAMPARI SIMULATIONS

For our initial CAMPARI MC simulations we only constructed linkers and collected data for the following repeats GGSGGS, GPGGA, PAPAGS, PASGGS, PPPAGS, and PPPPGS for $N = 3$ at 298K and 310K. The analysis presented here uses roughly 26000 frames after discarding an initial 1000 frames during equilibration.

We began our analysis by histogramming the End-to-end distance probability distribution functions for each linker simulation. Here, we define the End-to-end distance as the distance between the terminal C_α atoms of the peptide. These distributions are shown in Fig. 5.2 along with the corresponding Worm-like Chain fitted distribution based on the model in Ref. [255]. Collectively, the linker composition, particularly the amount of Serine and Proline affects End-to-end distance sampling and overall linker stiffness. The data for the $(\mathrm{GGSGGS})_3$ shows that the configurations from the simulation tend to a more collapsed state which can be visualized by the snapshot of ensemble configurations. The sampling of more collapsed states alludes to a considerable degree of flexibility in $(\mathrm{GGSGGS})_3$ and intramolecular hydrogen bonding which is given by the relatively low persistence length calculated from the model fit. Conversely, the simulation of the $(PPPPGS)_3$ linker which contains the highest Proline content out of the six shown here, samples more extended configurations. The $(PPPPGS)_3$ linker also yielded the highest persistence length value implying a higher stiffness supported by the snapshot of ensemble configurations, which appear almost entirely extended. From Fig. 5.2, we see an increase in Proline content leads to an increase in persistence length and hence overall stiffness which also correlates with the more extended structures in each snapshot of ensemble configurations. We note that the $(GPGGA)_3$ deviates slightly from the trend even though it contains a Proline residue in its repeat. Given that each repeat is made up of 5 residues as opposed to 6 in the others, the calculated End-to-end distance values would be slightly lower. An important detail concerning the $(GPGGA)_3$ linker is the position of the Proline within the repeat rather than at one of the ends, which does not hinder the linker's flexibility as significantly. The position of Proline is another important factor to be considered in linker design (see Sec. 5.3).

While the End-to-end distance analysis of our CAMPARI simulations verifies how residues, particularly Serine and Proline impact linker rigidity, we also examined if the configurations we predicted in our implicit solvent MC simulations gave us relevant structural motifs that are desirable for linker design. To gain insights into linker composition and secondary structure formation, we constructed Ramachandran plots from each simulation using the backbone dihedrals at both temperatures (Fig. 5.3). The Ramachandran regions we reference are according to the definitions presented in Ref. [256]. In general, an increase in Proline content in the linker is associated with the expected strengthening of the PolyProline II (PPII) helix basin (Φ at (-110°, -20°) and Ψ at (-180°, -120°) and (90°, 180°)). Simultaneously we see a weakening in the right-handed $\alpha-$helix, left-handed $\alpha-$helix, and $\beta-$sheet basins. Analogous to the End-to-end distance analysis of the $(GPGGA)_3$ linker, we see more comparable sampling to the $(GGSGGS)_3$ linker rather than the other 4 linkers whose repeats contain Proline at the N-terminus, providing data to demonstrate the effect of Proline position on linker secondary structure formation.

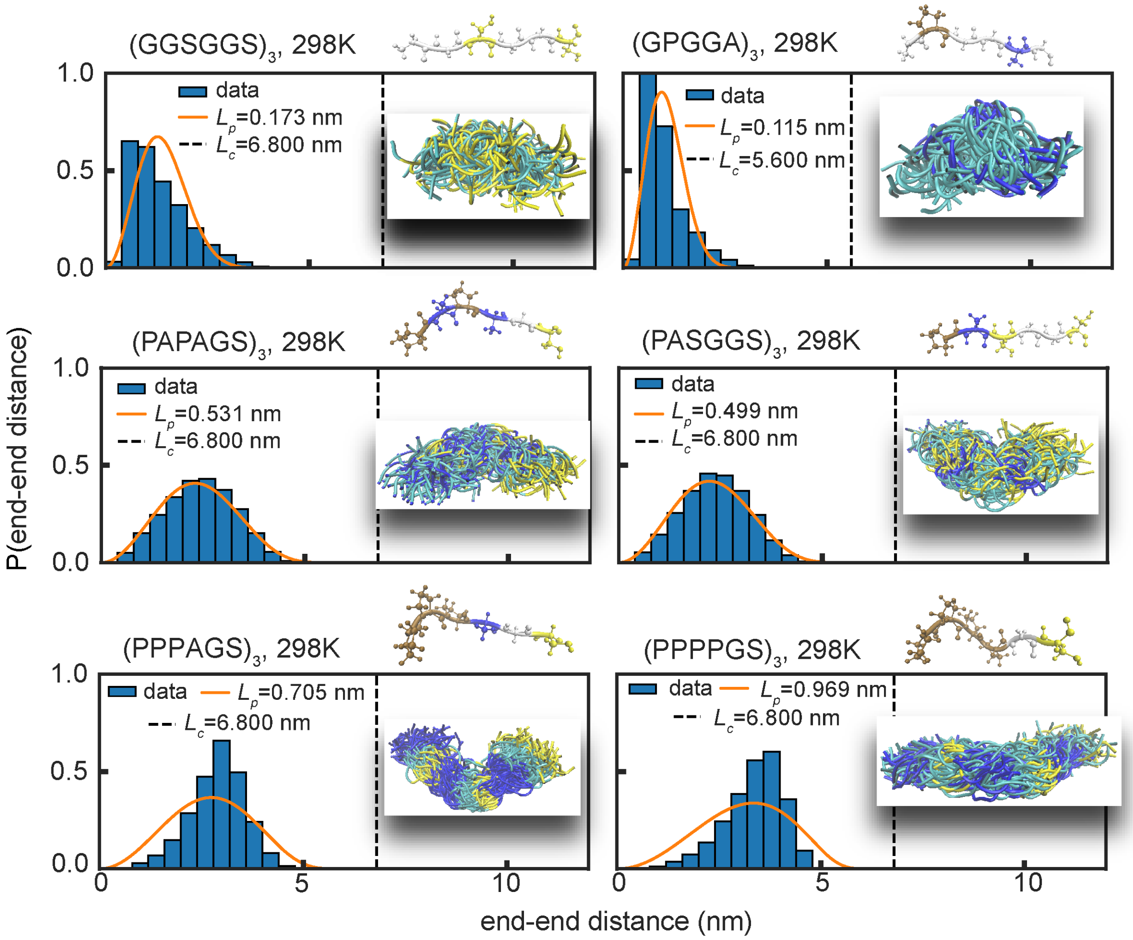


Figure 5.2: End-to-end Distance probability distribution functions for each Linker CAMPARI simulation at zero force along with an embedded snapshot of 100 configurations across all frames. For each subplot, the blue histograms are calculated from roughly 25000 frames for each simulation, the orange solid line denotes the Worm Like Chain model fit described by the model in Ref. [255], and the black dashed line represents the linker contour length calculated by multiplying the approximate length of each amino acid with the total number of residues in each peptide. For each embedded snapshot, Proline residues are colored in blue, serine residues in yellow, and the rest of the peptide in cyan.

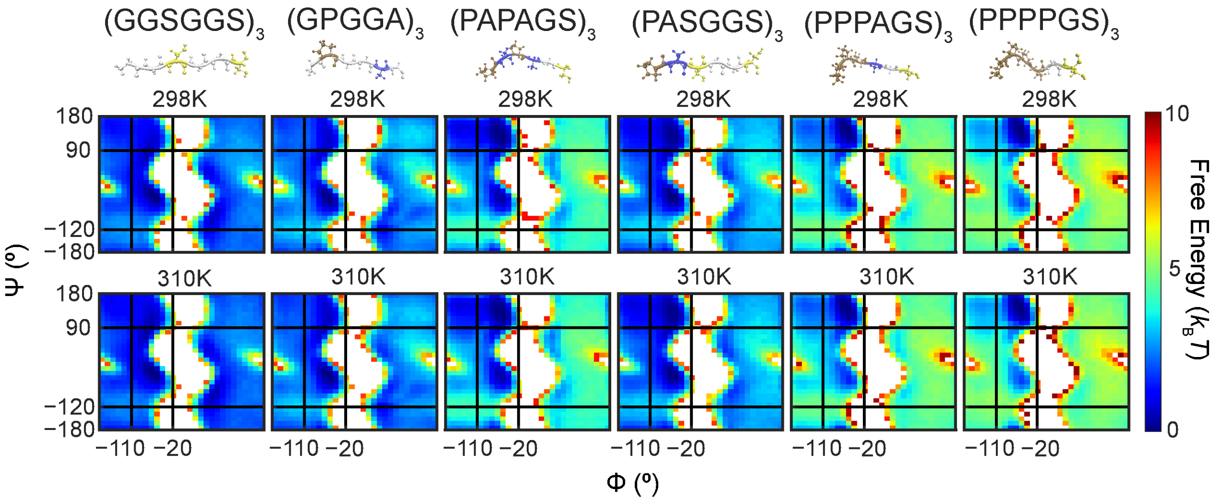


Figure 5.3: Linker Ramachandran plots generated from CAMPARI simulation for zero force at 298K (top row) and 310K (bottom row).

5.2.2 RESULTS FROM GROMACS SIMULATIONS

Although the CAMPARI simulations verified analysis from what is generally known in the field of linker design, such as an increase in Proline content increasing linker stiffness and polyproline II helix propensity, we shifted to a more complicated setup to mimic *in vitro* linker design experiments. In addition, we also implement pulling forces at 0, 5, 10, 15, and 20 pN along the z−component of the End-to-end distance while restraining the backbone atoms of the the N-terminus residue. We present simulation data for all 15 linker repeats shown in Fig. 5.1 for $N = 5$ at 298K and 310K. All linker analysis presented here uses 100ns of simulation time after discarding an initial 10ns for equilibration.

The first part of our analysis was to verify the Worm-like Chain behavior of linkers. We computed the mean extension along the z−axis at 5, 10, 15, and 20 pN forces by calculating the mean End-to-end distance at the zero force and subtracting from the mean values obtained at each simulation with force. We then normalized by the contour length L_c, which was calculated by multiplying the length of each amino acid residue with the total number of residues in the peptide

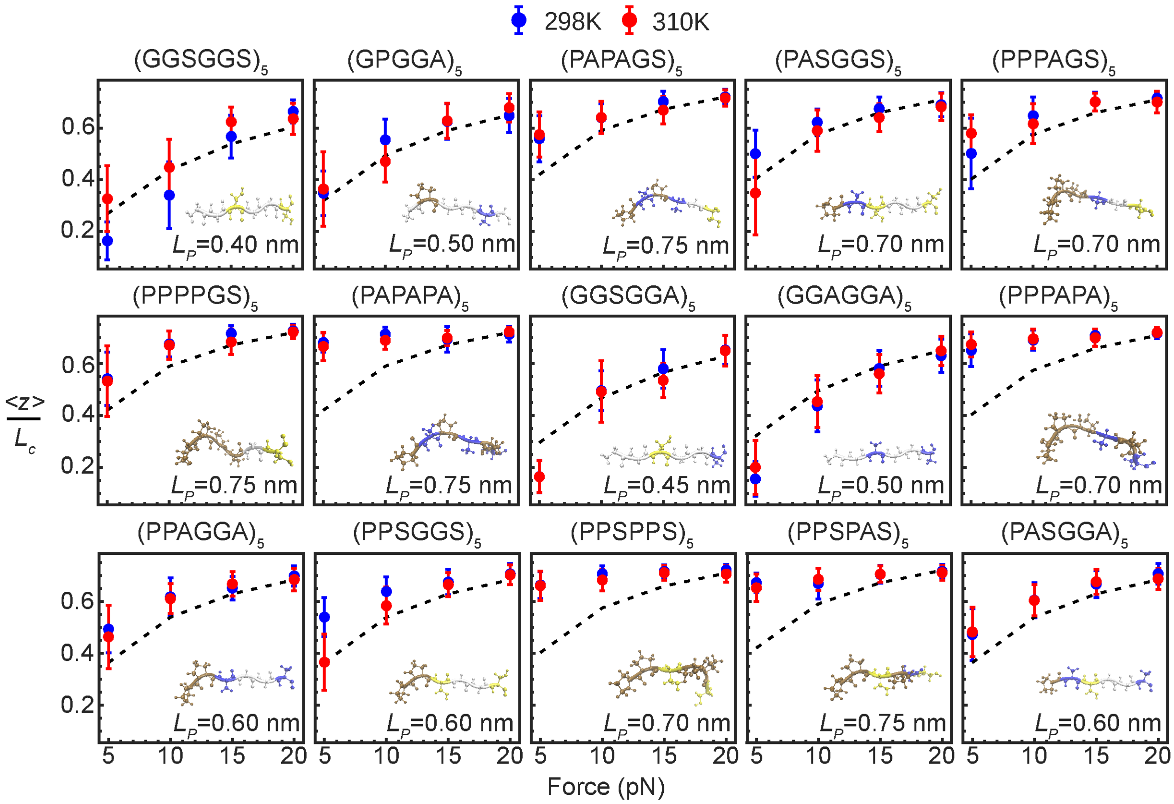


Figure 5.4: Mean z/L_c calculated for each linker trajectory, with error bars representing the standard deviation. The linker contour length L_c was calculated by multiplying the approximate length of each amino acid with the total number of residues in each peptide. Simulation data is from GROMACS MD runs at 0, 5, 10, 15, and 20 pN forces for 298K (blue) and 310K (red). The black dashed line represents Worm Like Chain (WLC) Model fitting by iteratively solving Marko-Sigio's equation [257] by defining an initial Persistence Length L_p for each linker.

for each linker. We presented the trends for $\langle z \rangle / L_c$ at each force for both temperatures in Fig. 5.4. Consistent with our previous analyses, linkers such as (PPPAPA)$_5$, (PPSPPS)$_5$, and (PPPPGS)$_5$ with a higher proportion of proline in their repeats demonstrate greater rigidity ($\langle z \rangle / L_c$>0.5). The data shown in Fig. 5.4 illustrates that for the more Proline rice linkers, the value of $\langle z \rangle$ with increasing pulling force magnitude does not fluctuate considerably, further emphasizing the greater stiffness of these linkers. In contrast, the linkers with no or less Proline content, such as (GGSGGS)$_5$, (GPGGA)$_5$, or (PASGGA)$_5$ show more discernible trends in $\langle z \rangle / L_c$ with increasing force, which points to their heightened flexibility and ability to adopt more collapsed states.

Similar to our analysis of the CAMPARI MC simulations, it was imperative to examine the

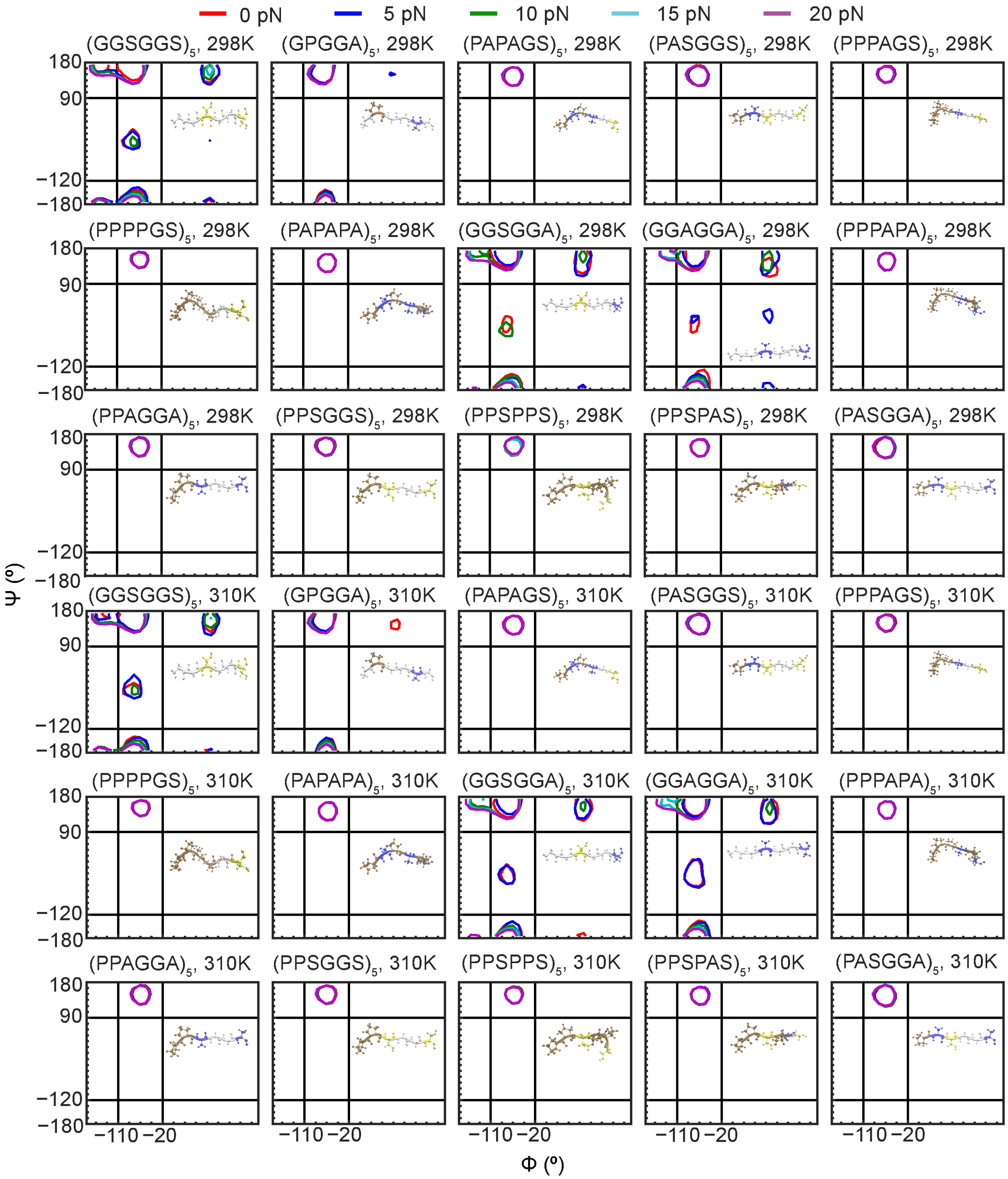


Figure 5.5: Linker contour plots for ϕ and ψ dihedrals calculated from GROMACS MD runs for 0 (red solid line), 5 (blue solid line), 10 (green solid line), 15 (cyan solid line), and 20 pN (magenta solid line) forces at 298K and 310K. Enclosed regions represent conformations within one free energy unit of the lowest energy configuration at each force.

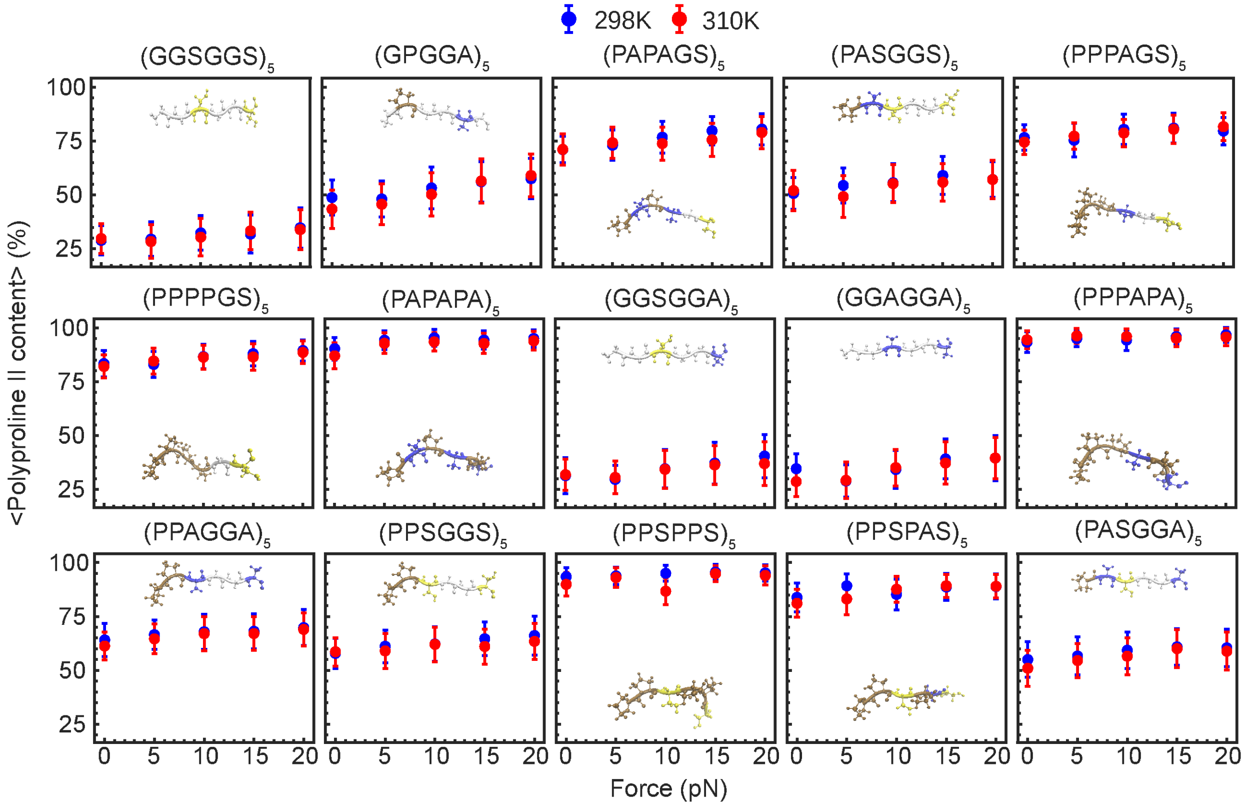


Figure 5.6: Mean PolyProline II content calculated for each linker trajectory as a percentage, with error bars representing the standard deviation. Simulation data is from GROMACS MD runs at 0, 5, 10, 15, and 20 pN forces for 298K (blue) and 310K (red).

secondary structure formation that correlates with experimental findings. We began by constructing free energy contour plots of the backbone dihedrals from our linker simulations at each force for both temperatures (Fig. 5.5). To visualize the effect of force on the sampling of the most probable conformations, we only considered those configurations within one free energy unit of the lowest energy configuration in each simulation. As expected, linkers with minimal to no Proline such as (GGSGGS)$_5$, (GGSGGA)$_5$, and (GGAGGA)$_5$ yielded more probable conformations that populated more regions of the Ramachandran plot including the $\beta-$sheet, right-handed $\alpha-$ helix, and Polyproline basins. In contrast, the linkers whose sequences begin with a Proline residue or contain a greater fraction of Proline mainly populated the Polyproline II basins. In Fig. 5.5, we also see that increasing the force applied increases the likelihood of sampling the PPII basin, even in the linkers without any Proline.

Given the significance of the Polyproline II helix in linker design, we moved to quantify the sampling of the PPII basin more analytically. To this end, we computed the average PPII content of our linker simulations at each force for both temperatures. Since the PPII helix is not an entry in the secondary structure database, we referred to the definition from Ref. [256] and calculated the PPII content directly from the backbone dihedrals. Although the data in Fig. 5.6 cannot be directly correlated with the data in Fig. 5.5 as we are only considering the lowest energy conformations in the latter, we are also able to observe increasing force leading to a monotonic increase in PPII content and higher proline content linkers are more likely to contain PPII in their sequence ($> 50\%$).

5.3 DISCUSSION AND ONGOING WORK

For this work, we conducted Monte Carlo (MC) and Molecular Dynamics (MD) simulations to characterize the mechanical and biochemical properties of a set of linker peptides as potential candidates in Tension Sensor Module engineering. Our simulations were able to provide molecular-level insights into the structural changes and conformational ensembles that can be adopted by these linker peptides of varying compositions of Alanine, Glycine, Proline, and Serine. For the first half of this study, we performed CAMPARI MC simulations with an implicit solvent model and ABSINTH forcefield as an initial approach to probe peptides similar to Intrinsically Disordered Proteins (IDPs). However, it was important to gain molecular-level insights into linker behavior in a setup more similar to experimental protocols. To this end, we increased the complexity of our setup by including explicit solvent and ions and implemented pulling forces for our GROMACS MD runs.

The data from both our CAMPARI and GROMACS simulations provided relevant insights into the secondary structure formation and mechanical behavior of a diverse set of linker peptides with and without force. More specifically, both sets of simulation data support the effect of Proline content on linker stiffness and propensity to populate the PPII basin of the Ramachandran plot, we emphasize the persistence lengths calculated from our simulations are not within a margin of error of persistence length values of peptides containing Proline as previous studies [258, 259]. We also note that our analysis of the GROMACS simulations in Fig. 5.5 for verifying Worm-Like Chain behavior, our simulation is not in reasonable agreement for the 5 and 10 pN forces, especially for the proline-rich linkers particularly $(PPPAPA)_5$, $(PPSPPS)_5$, and $(PAPAPA)_5$. For now, we attribute any inconsistencies with our Simulation protocol and ongoing efforts will involve implementing more sophisticated Enhanced Sampling Methodologies to address sampling bottlenecks. In particular, our work in Ref. [207] details a protocol for combining force tempering with replica exchange methods, which essentially extracts the benefits of conformational sampling from Replica Exchange [219] and force response prediction from the Infinite Switch Simulated Tempering in Force (FISST) method [60]. Given the intrinsically disordered behavior of these linker peptides, we rationalize a more rigorous enhanced sampling protocol would be required to achieve more robust sampling and observable prediction.

From a more Biochemical perspective, an alternate area to explore is utilizing all 20 natural amino acids for designing linker sequences, such as Arginine, and Glutamine which are also prominent linker residues that were not present in any linker peptide we studied here. Importantly, a direction to consider is determining the contribution a particular residue in a linker has on secondary structure formation. Specifically, an ongoing study as part of our collaboration in progress is determining how the secondary structure character of an amino acid is influenced by its neighboring residue. For example, preliminary data from a model developed using our GROMACS simulations has shown Glycine and Serine decrease the Polyproline II propensity of their neighbors while Proline and Alanine have an increasing effect. Identifying how these trends change under force has also been considered. Broadly speaking, predicting force response on secondary structure formation of linker peptides from simulation data that correlates well with experimental findings will help develop a more theoretical framework for linker design in Fusion Protein and Tension Sensor Module engineering.

5.4 SUPPORTING INFORMATION

5.4.1 SIMULATION SETUP AND DATA COLLECTION

CAMPARI Monte Carlo (MC) All CAMPARI simulations were run using CAMPARI Version 4. We carried implicit solvent simulations using ABSINTH [254] forcefield. We implemented the same MC moveset as Ref. [233] provided in Table 5.1. An equilibration run was carried out for 10000000 steps which was discarded during the trajectory analysis stage, followed by an additional data collection for 260000000 timesteps.

GROMACS Molecular Dynamics (MD). We employed GROMACS 2020.4 [200] for all GROMACS system construction, equilibration, and production runs, patched with PLUMED 2.7.0 [143]. Raw linker sequences were fed in AMBER MD [260] engine to generate starting PDB structures, which were then converted into GROMACS compatible files. All linker systems were constructed with TIP4P-D [232] explicit solvent, neutralized and ionized with Na^+ and Cl^- ions to a 15mM salt concentration. To account for stretchability under force, we constructed each linker in a cubic box with a length one nanometer greater than the contour length, which we calculated by multiplying the approximate length of each amino acid (~0.38 nm) with the number of residues in the peptide. Restraints were applied to the backbone atoms of the N-terminus residue by setting the force constant of each atom along all three cartesian coordinates to 10000 kJ/mol/nm. We followed the minimization equilibration protocol according to the standard GROMACS simulation protocol[3]. Production runs were carried out at 298K and 310K at 0, 5, 10, 15, and 20 pN forces. The pulling protocol was implemented through the `RESTRAINT` module in PLUMED along

---

[3] `http://www.mdtutorials.com/gmx/lysozyme/index.html`

FMCSC_RIGIDFREQ	0.05
FMCSC_TRANSSTEPSZ	2.0
FMCSC_ROTSTEPSZ	10.0
FMCSC_CLURBFREQ	0.2
FMCSC_CLURBMAX	3
FMCSC_COUPLERIGID	1
FMCSC_RIGIDRDFREQ	0.5
FMCSC_CHIFREQ	0.19
FMCSC_CHIRDFREQ	0.4
FMCSC_NRCHI	4
FMCSC_CHISTEPSZ	30.0
FMCSC_OMEGAFREQ	0.0684
FMCSC_OMEGARDFREQ	0.1
FMCSC_OMEGASTEPSZ	5.0
FMCSC_PIVOTRDFREQ	0.3
FMCSC_PIVOTSTEPZ	10.0
FMCSC_PKRFREQ	0.1231
FMCSC_PKRRDFREQ	0.2
FMCSC_PUCKERSTEP_DI	4.0
FMCSC_PUCKERSTEP_AN	2.0
FMCSC_CRFREQ	0.076
FMCSC_ALIGN	4

Table 5.1: CAMPARI Monte Carlo Moveset parameters for implicit solvent simulations with ABSINTH forcefield.

the $z-$component of the End-to-end distance. A total of 110 ns of simulation time was collected for each linker simulation and the first 10 ns were discarded during trajectory analysis.

5.4.2 TRAJECTORY ANALYSIS AND VISUALIZATION

All trajectory snapshots were visualized in VMD version 1.9.3 [149]. Snapshots in Fig. 5.2 were generated by selecting every 250th frame starting from the very first to the last frame in the trajectory (100 configurations). Trajectory analysis was carried out using the `mdtraj` library [144] in Python version 3.8.0.

Chapter 6

Concluding Remarks and Future Studies

The primary goal of this thesis was to probe the connection between the structural changes of Force-sensing machines studied in Mechanobiology and their respective functions during cellular processes, using various approaches in Computational Chemistry and Statistical Mechanics to predict structural changes at the molecular level. We began by presenting different Molecular Machines called Force Sensors and the structural changes they undergo in response to piconewton magnitude forces, which facilitate essential cellular functions including motility, division, and material transport. After exploring the substantial progress and insights from experimental techniques implemented in Mechanobiology such as Single Molecule Force Spectroscopy techniques and Tension Sensor Modules, we underscored some limitations and open-ended questions that require Computational and Theoretical frameworks. We structured this thesis to study two major Mechanoresponsive systems: Actin Cytoskeleton and Molecular Springs. For each of these systems, we identified a focused research question that could be addressed using computational techniques and gain novel molecular-level insights that contribute to the knowledge base of Mechanobiology and open new areas for Theoretical Chemistry research.

6.1 FUTURE DIRECTIONS IN ARP2/3 ACTIVATION

We began our Computational Chemistry approach to study Mechanobiology phenomena starting with the Actin Cytoskeleton in Chapter 2. Specifically, we focused on Arp2/3 complex, an actin nucleator unique for its ability to form branched actin filaments which is a critical component in meiosis, maintaining cell-cell junctions, and endocytosis. Leveraging a recent high-resolution crystal structure of *Bos Taurus* Arp2/3 in the branch junction, we performed unbiased and biased Molecular Dynamics (MD) studies on Arp2/3 complex in different bound states which support a Multi-step or unconcerted activation pathway as opposed to a Single-step or Concerted activation pathway. The data from the simulations tells us that the two major activation changes undergone by actin-related proteins Arp2 and Arp3; movement into the shortpitch or filament-like arrangement and subunit flattening are not tightly linked but rather can occur independently. Unconcerted activation changes hint at the existence of stable intermediates the complex can adopt, which could offer insights into novel mechanisms for actin filament binding, monomer recruitment, and the release of Nucleation Promoting Factors (NPFs). Future computational studies for understanding Arp2/3 mechanisms would involve determining an energetic coupling between the two activation changes (movement into short pitch and subunit flattening) and calculating rate constants associated with each activation change, which would require more sophisticated Enhanced Sampling and Non-equilibrium free energy calculation approaches. In addition to these long-term research goals, we have also acquired preliminary simulation results that follow from our work in Ref. [76]; namely, estimating a free energy barrier between the inactive and active states and understanding the effect of Arp subunit nucleotide state (ADP or ATP) during activation.

6.1.1 Inactive and Active Arp2/3 free energy barrier

The results in section 2.2.2 of Chapter 2 alluded to a considerable free energy barrier that had to be crossed for Arp2/3 to transition from the inactive to active state. Rudimentary stages to estimate the free energy between the two states involved leveraging the steered MD trajectories of Active Arp2/3 bound to the mother filament and Free Active Arp2/3 complex and running umbrella sampling simulations followed by Weighted Histogram Analysis Method (WHAM, [2]) to construct 1D and 2D free energy profiles in Fig. 6.1. We also compared our WHAM calculations with the Eigenvector Method for Umbrella Sampling [184]. Our protocol for umbrella simulations began with choosing 16 configurations along the steered MD trajectory (Fig. 2.12 of Chapter 2), specifically the trajectory for which the bias was applied for 150 ns. We postulated that the configurations along this trajectory would give a reasonably approximate pathway from the active to inactive state as the Arps adopt a splayed arrangement. We then used each of these configurations as our umbrella windows, biasing the Arp2-Arp3 COG distance (defined in section 2.2.2 of Chapter 2) and collected 5 ns of simulation data for each window. From the literature, we know the energy between the inactive and active states is several $k_{\mathrm{B}}T$ [120] and from experimental findings, Free Inactive Arp2/3 is more stable than Free Active Arp2/3, unless Arp2/3 is in the branch junction or bound to a mother filament [134]. Evaluating our preliminary analysis in Fig. 6.1 based on these precursors we acknowledge several inconsistencies and emphasize that additional efforts need to be made to reproduce more accurate results. Firstly, we note that we only bias one activating structural change as our CV in our umbrella simulations, which in its own is insufficient to capture other major structural changes that contribute to the thermodynamics. For the 1D free energy plots calculated from the Active Arp2/3 bound to the mother filament and Free Active Arp2/3, although we can see a difference in the two local minima for the inactive and active states, the magnitude of energies is far less compared to experimental findings. Along with these 1D free energy plots, we constructed a 2D free energy surface for sets of Umbrella runs using the Arp2-Arp3 COG distance and the twisting of clamp subunits ARPC2 and ARPC4, another proponent structural change that occurs during activation [90, 91, 111, 113]. We also note these 2D FES plots do not accurately capture a significant free energy difference between the two states. Efforts to address these consistencies would involve calculating a Path Collective Variable (CV) [261] which would provide a better representation of Arp2/3 transition pathway and hence a better biasing coordinate as opposed to biasing a single CV.

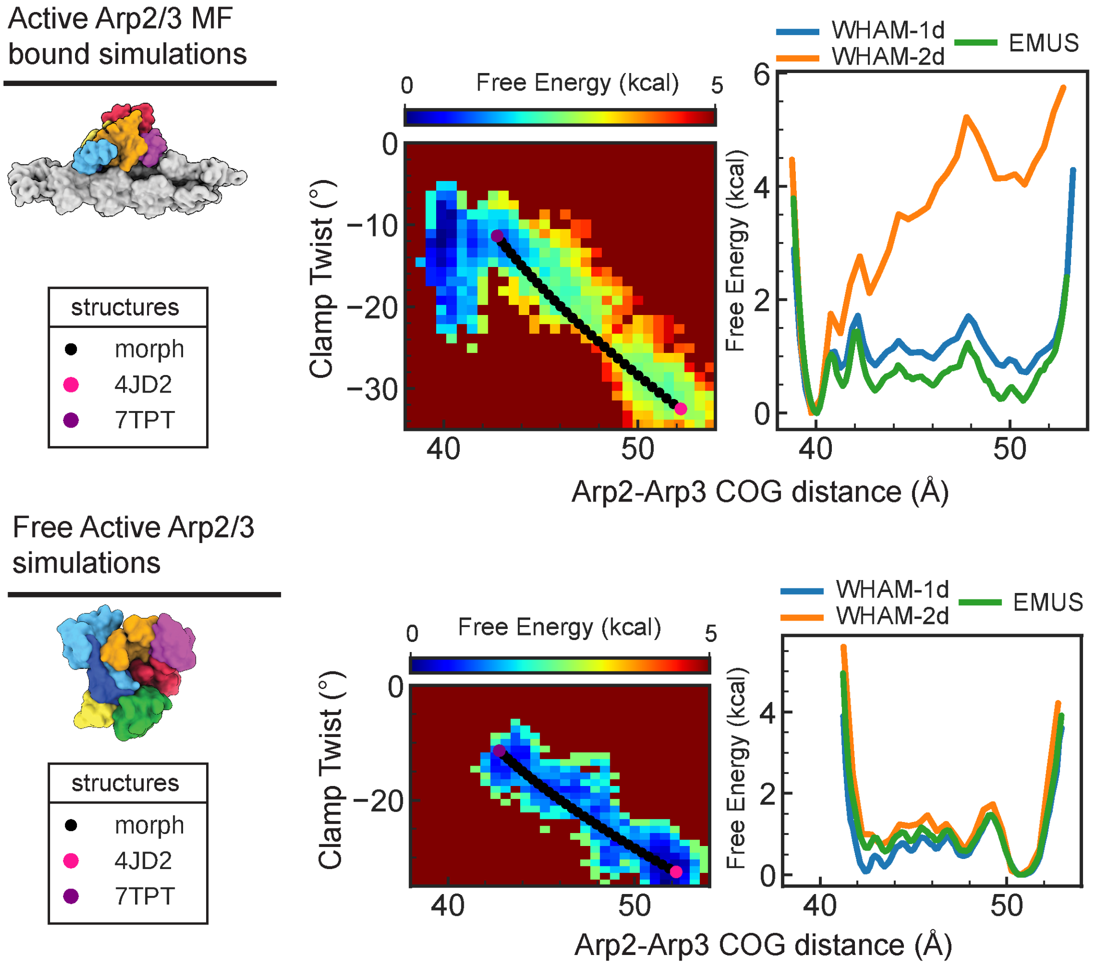


Figure 6.1: 2D and 1D free energy plots calculated from umbrella simulations for Active Arp2/3 Mother filament bound 150 ns pulling trajectory (top) and Free Active Arp2/3 150 ns pulling trajectory (bottom). Morph structures (black spheres) were generated from PyMOL [262] using PDB: 7TPT and PDB: 4JD2 as the initial and final structures respectively.

6.1.2 IMPORTANCE OF ARP SUBUNIT NUCLEOTIDE STATE DURING ACTIVATION

In section 2.2.1 of Chapter 2, we outlined four different Arp2/3 systems representing the different bound states (Active Arp2/3 complex in the branch junction, Active Arp2/3 complex bound to the mother filament only, Free Active Arp2/3 complex, and Free Inactive Arp2/3 complex) that we constructed and simulated. We mentioned the nucleotide clefts of the Arp subunits are bound to ATP consistent with a pre-activation state for every system studied (except for the system with Active Arp2/3 complex in the branch junction). Additionally, systems of Active Arp2/3 bound to the mother filament and Free active Arp2/3 complex containing ADP in the nucleotide clefts of the Arp subunits were constructed and simulated, and the resulting data from these systems were analyzed. Similar to our analysis described in section 2.2.3 of Chapter 2, we constructed free energy contour plots for the Arp2-Arp3 COG distance (defined in section 2.2.2) and each Arp subunit dihedral (φ_{Arp2} and φ_{Arp3}) shown in Fig. 6.2, which not only support the existence of intermediates for a multi-step model of Arp2/3 activation but also underscore considerable differences between the ADP and ATP nucleotide states. For the simulations of Active Arp2/3 bound to the mother filament, we observe a higher degree of twisting of the arps for the ADP-bound state as opposed to the ATP-bound state. Another observation consistent with our other data is the “tilting” of Arp2 subunit described in section 2.2.2 and Fig. 2.17 of Chapter 2 that occurs for the ADP-bound state. The simulation of Free Active Arp2/3 complex in the ADP-bound state also shows more twisting of both Arp subunits as compared to the simulation with the ATP-bound state. In addition, the contour plot for Arp2-Arp3 COG distance and Arp3 dihedral for the ADP-bound Free Active Arp2/3 complex simulation has a broader spread compared to the corresponding plot of the ATP-bound state, with a roughly symmetric distribution of flatter and more twisted configurations. The differences observed for the ADP and ATP-bound simulations point to a more intricate role ATP has in providing stability to the Arps in the activated state that requires further simulation analysis for a more comprehensive study.

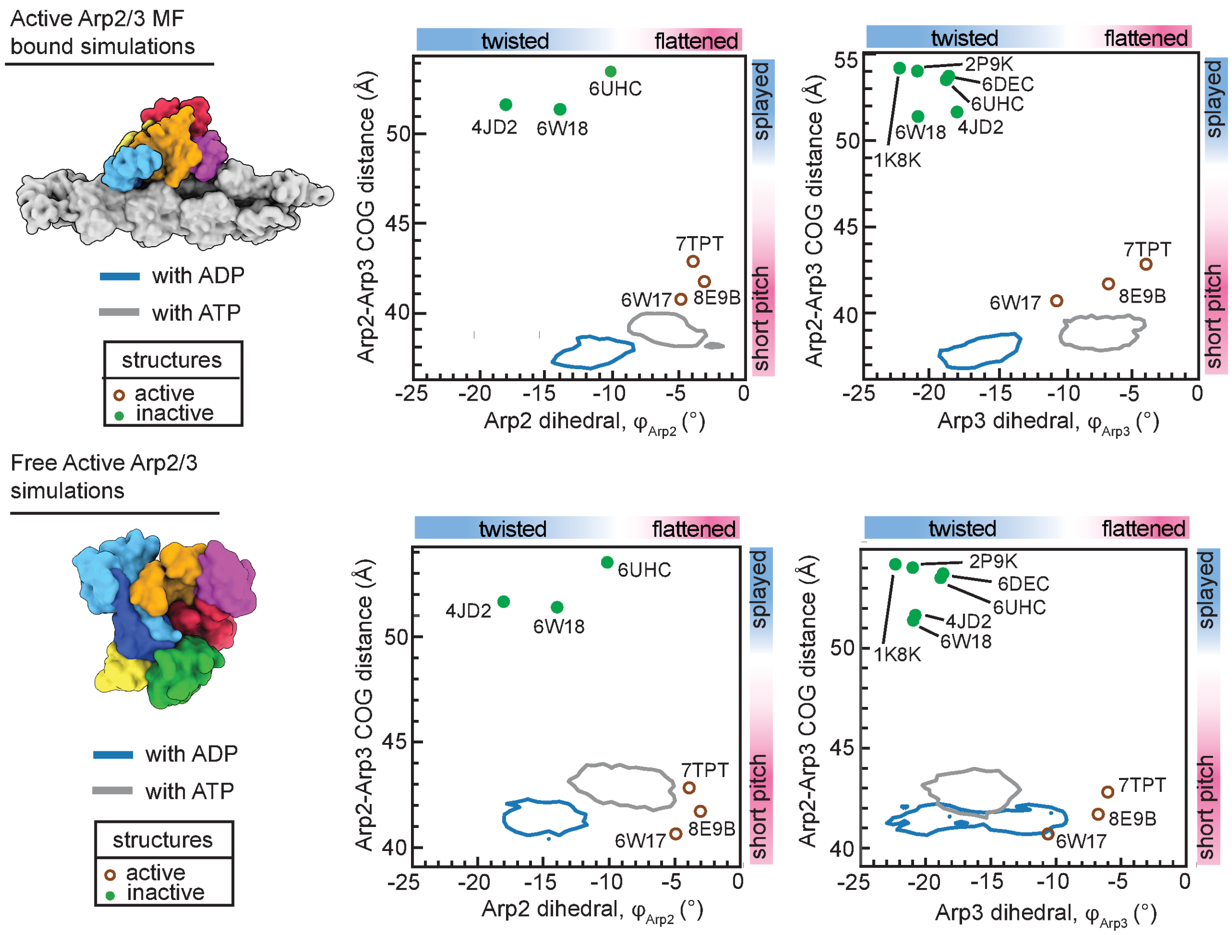


Figure 6.2: Contour plots for Arp2-Arp3 COG distance vs φ_{Arp2} and Arp2-Arp3 COG distance vs φ_{Arp3} from Active Arp2/3 Mother filament bound (top) and Free Active Arp2/3 Complex (bottom) simulations. Enclosed regions represent configurations within one free energy unit of the lowest energy conformation for each nucleotide state: ADP (blue solid liner) and ATP (gray solid line).

6.2 FUTURE DIRECTIONS FOR TENSION SENSOR PREDICTION

After a detailed computational study of Arp2/3 complex within the Actin Cytoskeleton system, we transitioned to Molecular Springs from Chapter 3 onwards, where we presented the FISST method [60], a theoretical framework for predicting the effect of forces in our MD simulation. The method follows from the formulation defined in Ref. [171], where we show an average force is calculated from a predefined force range which facilitates the crossing of energy barriers, aiding conformational sampling. Specifically, the method linearly biases a Collective Variable (CV) of interest, which is useful for investigating the effect of force along a crucial degree of freedom. Each force is assigned a weight function that describes the contribution it has during the calculation of the average force and is updated "on the fly" and used in the reconstruction of any observable at any given force, from a single simulation. We then implemented the FISST formulation and assessed its performance on three different systems of increasing complexity: A V-shape analytic potential, a "Toy" Beaded Helix system, and an atomistic Alanine decamer in explicit water system. The method exhibited improved sampling efficacy for the first two systems and demonstrated satisfactory results for the final system, thereby validating our approach. The FISST method is now accessible to the public through its inclusion in the PLUMED repository [143].

While we obtained promising results from our initial FISST implementation, in Chapter 4 we asked if our method could be combined with another enhanced sampling methodology, namely Replica Exchange, thereby incorporating the benefits of force tempering and conformational sampling simultaneously. Following an approach consistent with our work in Ref. [60] we tested the FISST and Replica Exchange hybrid methodology on three systems of increasing complexity: Alanine Decamer from our original work, Left-handed AIB9 helix, and Villin (NLE/NLE) mutant, all in explicit solvent. We find in this subsequent study that FISST alone fails to completely sample the AIB9 helix energy landscape. It was only after concurrent sampling with FISST and Replica Exchange we were able to recover a more complete free energy profile along the chirality coordinate ξ'. The data obtained from the more complicated folded Villin (NLE/NLE) mutant system indicates our system setup and protocol require more fine-tuning to recover a force-extension curve that correlates better with experimental findings and for predicting force response when designing Tension Sensor Modules.

Chapter 5 follows from Chapters 3 and 4, and also falls under Molecular Springs, wherein our overarching goal is the *in silico* prediction of amino acid sequence force response to provide a more rational basis for Tension Sensor Module and Fusion Protein engineering. Unlike the previous Chapters where we have studied more defined atomistic systems, Chapter 5 features mainly intrinsically disordered linker peptides of varying amino acid composition, which in turn alters the mechanical and biochemical properties. Although we have only presented preliminary data, we assert that our simulation data offers useful insights into the secondary structure formation of linker peptides that correlate with experimental findings, more specifically proline-rich linkers forming stable Polyproline II helices and proline increasing linker stiffness. We also acknowledge some of our results, namely the persistence length calculations are not within a reasonable margin of error with previous studies. Therefore, we propose that the FISST and Replica Exchange hybrid methodology introduced in Chapter 4 would address sampling issues we may have encountered in our initial simulation protocol.

BIBLIOGRAPHY

[1] David W Oxtoby, H Pat Gillis, and Laurie J Butler. *Principles of Modern Chemistry*. Cengage AU, 2016.

[2] Mark E Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford university press, 2023.

[3] Rob Phillips et al. *Physical Biology of the Cell*. Garland Science, 2012.

[4] David Gomez et al. "Molecular paradigms for biological mechanosensing". In: *The Journal of Physical Chemistry B* 125.44 (2021), pp. 12115–12124.

[5] Peter William Atkins, Julio De Paula, and James Keeler. *Atkins' Physical Chemistry*. Oxford university press, 2023.

[6] Pere Roca-Cusachs, Vito Conte, and Xavier Trepat. "Quantifying forces in cell biology". In: *Nature Cell Biology* 19.7 (2017), pp. 742–751.

[7] SO MECHANOBIOLOGY. "Forces in cell biology". In: *Nature Cell Biology* 19.6 (2017), p. 579.

[8] Aashish Manglik et al. "Crystal structure of the μ-opioid receptor bound to a morphinan antagonist". In: *Nature* 485.7398 (2012), pp. 321–326.

[9] Weijiao Huang et al. "Structural insights into μ-opioid receptor activation". In: *Nature* 524.7565 (2015), pp. 315–321.

[10] Olga Yakovenko et al. "FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation". In: *Journal of Biological Chemistry* 283.17 (2008), pp. 11596–11605.

[11] Debasis Banik et al. "Single molecule force spectroscopy reveals distinctions in key biophysical parameters of $\alpha\beta$ T-cell receptors compared with chimeric antigen receptors directed at the same ligand". In: *The Journal of Physical Chemistry Letters* 12.31 (2021), pp. 7566–7573.

[12] Hyun-Kyu Choi et al. "Catch bond models may explain how force amplifies TCR signaling and antigen discrimination". In: *Nature Communications* 14.1 (2023), p. 2616.

[13] Dennis Zimmermann et al. "Mechanoregulated inhibition of formin facilitates contractile actomyosin ring assembly". In: *Nature Communications* 8.1 (2017), pp. 1–13.

[14] Lisa S Fischer et al. “Molecular force measurement with tension sensors”. In: *Annual Review of Biophysics* 50 (2021), pp. 595–616.

[15] Andrea Freikamp, Anna-Lena Cost, and Carsten Grashoff. “The piconewton force awakens: quantifying mechanics in cells”. In: *Trends in Cell Biology* 26.11 (2016), pp. 838–847.

[16] Jing Liu. “Intracellular Force Measurements in Live Cells With Förster Resonance Energy Transfer–Based Molecular Tension Sensors”. In: *Mechanobiology*. Elsevier, 2020, pp. 161–171.

[17] GP Talwar. *Textbook of Biochemistry, Biotechnology, Allied and Molecular Medicine*. PHI Learning Pvt. Ltd., 2015.

[18] Ilka Böhme and Annette G Beck-Sickinger. “Illuminating the life of GPCRs”. In: *Cell Communication and Signaling* 7 (2009), pp. 1–22.

[19] Shu Z Wiley et al. “GPR68: an emerging drug target in cancer”. In: *International Journal of Molecular Sciences* 20.3 (2019), p. 559.

[20] Jie Xu et al. “GPR68 Senses Flow and is Essential for Vascular Physiology”. In: *Cell* 173.3 (2018), pp. 762–775.

[21] George I Bell. “Models for the specific adhesion of cells to cells: a theoretical framework for adhesion mediated by reversible bonds between cell surface molecules.” In: *Science* 200.4342 (1978), pp. 618–627.

[22] Maximilian M Sauer et al. “Catch-bond mechanism of the bacterial adhesin FimH”. In: *Nature Communications* 7.1 (2016), p. 10738.

[23] Willmor J Peña Ccoa and Glen M Hocky. “Assessing models of force-dependent unbinding rates via infrequent metadynamics”. In: *The Journal of Chemical Physics* 156.12 (2022).

[24] Geoffrey M Cooper. “Structure and organization of actin filaments”. In: *The cell: a molecular approach* 2 (2000).

[25] Tatyana Svitkina. “The actin cytoskeleton and actin-based motility”. In: *Cold Spring Harbor Perspectives in Biology* 10.1 (2018), a018267.

[26] Diana C Muñoz-Lasso et al. “Much more than a scaffold: cytoskeletal proteins in neurological disorders”. In: *Cells* 9.2 (2020), p. 358.

[27] Fatemah Mukadum, Willmor J. Pena Ccoa, and Glen M. Hocky. *Molecular simulation approaches to probing the effects of mechanical forces in the actin cytoskeleton*. 2023.

[28] Ning Wang. “Review of cellular mechanotransduction”. In: *Journal of Physics D: Applied Physics* 50.23 (2017), p. 233002.

[29] Andrew R Harris, Pamela Jreij, and Daniel A Fletcher. “Mechanotransduction by the actin cytoskeleton: converting mechanical stimuli into biochemical signals”. In: *Annual Review of Biophysics* 47 (2018), pp. 617–631.

[30] Thang K Chiu et al. “High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.21 (2005), pp. 7517–7522.

[31] Katharina Austen et al. “Extracellular rigidity sensing by talin isoform-specific mechanical linkages”. In: *Nature Cell Biology* 17.12 (2015), pp. 1597–1606.

[32] Prakash Shrestha, Shankar Mandal, and Hanbin Mao. “Mechanochemical Sensing”. In: May 2015, pp. 241–258. ISBN: 978-3-319-17304-7. DOI: `10.1007/978-3-319-17305-4_12`.

[33] Keir C Neuman and Attila Nagy. “Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy”. In: *Nature Methods* 5.6 (2008), p. 491.

[34] Aleksandr Noy and Raymond W Friddle. “Practical single molecule force spectroscopy: How to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope”. In: *Methods* 60.2 (2013), pp. 142–150.

[35] Arthur Ashkin et al. “Observation of a single-beam gradient force optical trap for dielectric particles”. In: *Optics Letters* 11.5 (1986), pp. 288–290.

[36] Michelle D Wang et al. “Force and velocity measured for single molecules of RNA polymerase”. In: *Science* 282.5390 (1998), pp. 902–907.

[37] Jeff Gore et al. “Mechanochemical analysis of DNA gyrase using rotor bead tracking”. In: *Nature* 439.7072 (2006), pp. 100–104.

[38] Hiroyasu Itoh et al. “Mechanically driven ATP synthesis by F1-ATPase”. In: *Nature* 427.6973 (2004), pp. 465–468.

[39] Ingo Schwaiger et al. “A mechanical unfolding intermediate in an actin-crosslinking protein”. In: *Nature Structural & Molecular Biology* 11.1 (2004), pp. 81–85.

[40] Surena Vahabi, Bahareh Nazemi Salman, and Anahita Javanmard. “Atomic force micro–scopy application in biological research: a review study”. In: *Iranian Journal of Medical Sciences* 38.2 (2013), p. 76.

[41] Trevor R Ham, Kasie L Collins, and Brenton D Hoffman. “Molecular tension sensors: moving beyond force”. In: *Current Opinion in Biomedical Engineering* 12 (2019), pp. 83–94.

[42] Anna-Lena Cost et al. “How to measure molecular forces in cells: a guide to evaluating genetically-encoded FRET-based tension sensors”. In: *Cellular and Molecular Bioengineering* 8 (2015), pp. 96–105.

[43] Bryce T Bajar et al. “A guide to fluorescent protein FRET pairs”. In: *Sensors* 16.9 (2016), p. 1488.

[44] Joseph R Lakowicz. *Principles of fluorescence spectroscopy.* Springer, 2006.

[45] Theodor Förster. “Energy migration and fluorescence”. In: *Journal of Biomedical Optics* 17.1 (2012), pp. 011002–011002.

[46] Carsten Grashoff et al. “Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics”. In: *Nature* 466.7303 (2010), pp. 263–266.

[47] Fanjie Meng and Frederick Sachs. “Visualizing dynamic cytoplasmic forces with a compliance - matched FRET sensor”. In: *Journal of Cell Science* 124.2 (2011), pp. 261–269.

[48] Adam Hospital et al. “Molecular dynamics simulations: advances and applications”. In: *Advances and Applications in Bioinformatics and Chemistry* (2015), pp. 37–47.

[49] Tamar Schlick and Stephanie Portillo-Ledesma. “Biomolecular modeling thrives in the age of technology”. In: *Nature Computational Science* 1.5 (2021), pp. 321–331.

[50] Martin Karplus and J Andrew McCammon. “Molecular dynamics simulations of biomolecules”. In: *Nature Structural & Molecular Biology* 9.9 (2002), pp. 646–652.

[51] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (2002), pp. 12562–12566.

[52] Alessandro Laio and Francesco L Gervasio. “Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science”. In: *Reports on Progress in Physics* 71.12 (2008), p. 126601.

[53] Giovanni Bussi and Alessandro Laio. “Using metadynamics to explore complex free-energy landscapes”. In: *Nature Reviews Physics* 2.4 (2020), pp. 200–212.

[54] Gregory W Corder and Dale I Foreman. *Nonparametric Statistics: A Step-by-Step Approach.* John Wiley & Sons, 2014.

[55] Pratyush Tiwary and Michele Parrinello. “From metadynamics to dynamics”. In: *Physical Review Letters* 111.23 (2013), p. 230602.

[56] Pratyush Tiwary. “Molecular determinants and bottlenecks in the dissociation dynamics of biotin–streptavidin”. In: *The Journal of Physical Chemistry B* 121.48 (2017), pp. 10841–10849.

[57] Barry Isralewitz, Mu Gao, and Klaus Schulten. “Steered molecular dynamics and mechanical functions of proteins”. In: *Current Opinion in Structural Biology* 11.2 (2001), pp. 224–230.

[58] Sergei Izrailev et al. “Molecular dynamics study of unbinding of the avidin-biotin complex”. In: *Biophysical Journal* 72.4 (1997), pp. 1568–1581.

[59] Guillaume Stirnemann et al. “Elasticity, structure, and relaxation of extended proteins under force”. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.10 (2013), pp. 3847–3852.

[60] Michael J Hartmann et al. “Infinite switch simulated tempering in force (FISST)”. In: *The Journal of Chemical Physics* 152.24 (2020).

[61] Don Van Ravenzwaaij, Pete Cassey, and Scott D Brown. “A simple introduction to Markov Chain Monte–Carlo sampling”. In: *Psychonomic Bulletin & Review* 25.1 (2018), pp. 143–154.

[62] Martin Feinberg and Martin Feinberg. “Detailed Balancing”. In: *Foundations of Chemical Reaction Network Theory* (2019), pp. 293–307.

[63] Yuji Sugita and Yuko Okamoto. “Replica-exchange molecular dynamics method for protein folding”. In: *Chemical Physics Letters* 314.1-2 (1999), pp. 141–151.

[64] Ruxi Qi et al. “Replica exchange molecular dynamics: a practical application protocol with solutions to common problems and a peptide aggregation and self-assembly example”. In: *Peptide Self-Assembly: Methods and Protocols* (2018), pp. 101–119.

[65] Zachary A Levine et al. “Regulation and aggregation of intrinsically disordered peptides”. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.9 (2015), pp. 2758–2763.

[66] Jianhan Chen, Wonpil Im, and Charles L Brooks. “Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field”. In: *Journal of the American Chemical Society* 128.11 (2006), pp. 3728–3736.

[67] Lingle Wang, Richard A Friesner, and BJ Berne. “Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2)”. In: *The Journal of Physical Chemistry B* 115.30 (2011), pp. 9431–9438.

[68] Yumeng Zhang, Xiaorong Liu, and Jianhan Chen. “Re-Balancing Replica Exchange with Solute Tempering for Sampling Dynamic Protein Conformations”. In: *Journal of Chemical Theory and Computation* 19.5 (2023), pp. 1602–1614.

[69] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* " O’Reilly Media, Inc.", 2022.

[70] Soumil Y Joshi and Sanket A Deshmukh. “A review of advancements in coarse-grained molecular dynamics simulations”. In: *Molecular Simulation* 47.10-11 (2021), pp. 786–803.

[71] Sebastian Kmiecik et al. “Coarse-grained protein models and their applications”. In: *Chemical reviews* 116.14 (2016), pp. 7898–7936.

[72] Bonnie A Merchant and Jeffry D Madura. “A review of coarse-grained molecular dynamics techniques to access extended spatial and temporal scales in biomolecular simulations”. In: *Annual Reports in Computational Chemistry* 7 (2011), pp. 67–87.

[73] Gaurav Mitra et al. “A coarse-grained simulation model for colloidal self-assembly via explicit mobile binders”. In: *Soft Matter* (2023).

[74] PC Agu et al. “Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management”. In: *Scientific Reports* 13.1 (2023), p. 13398.

[75] Xuan-Yu Meng et al. “Molecular docking: a powerful approach for structure-based drug discovery”. In: *Current computer-aided drug design* 7.2 (2011), pp. 146–157.

[76] Yuvraj Singh, Glen M Hocky, and Brad J Nolen. “Molecular dynamics simulations support a multistep pathway for activation of branched actin filament nucleation by Arp2/3 complex”. In: *Journal of Biological Chemistry* 299.9 (2023).

[77] Orit Siton-Mendelson and Anne Bernheim-Groswasser. “Functional actin networks under construction: the cooperative action of actin nucleation and elongation factors”. In: *Trends in biochemical sciences* 42.6 (2017), pp. 414–430.

[78] Ikuko Fujiwara, Dimitrios Vavylonis, and Thomas D Pollard. “Polymerization kinetics of ADP-and ADP-Pi-actin determined by fluorescence microscopy”. In: *Proceedings of the National Academy of Sciences of the United States of America* 104.21 (2007), pp. 8827–8832.

[79] Roberto Dominguez. "The WH2 domain and actin nucleation: necessary but insufficient". In: *Trends in biochemical sciences* 41.6 (2016), pp. 478–490.

[80] Kurt J Amann and Thomas D Pollard. "The Arp2/3 complex nucleates actin filament branches from the sides of pre-existing filaments". In: *Nature Cell Biology* 3.3 (2001), pp. 306–310.

[81] Nicolas Molinie and Alexis Gautreau. "The Arp2/3 regulatory system and its deregulation in cancer". In: *Physiological reviews* 98.1 (2018), pp. 215–238.

[82] Kenneth G Campellone, Nadine M Lebek, and Virginia L King. "Branching out in different directions: Emerging cellular functions for the Arp2/3 complex and WASP-family actin nucleation factors". In: *European Journal of Cell Biology* 102.2 (2023), p. 151301.

[83] Peter Bieling and Klemens Rottner. "From WRC to Arp2/3: Collective molecular mechanisms of branched actin network assembly". In: *Current Opinion in Cell Biology* 80 (2023), p. 102156.

[84] Vérane Achard et al. "A "primer"-based mechanism underlies branched actin filament network formation and motility". In: *Current Biology* 20.5 (2010), pp. 423–428.

[85] Jean-Baptiste Marchand et al. "Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex". In: *Nature Cell Biology* 3.1 (2001), pp. 76–82.

[86] R Dyche Mullins, John A Heuser, and Thomas D Pollard. "The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments". In: *Proceedings of the National Academy of Sciences of the United States of America* 95.11 (1998), pp. 6181–6186.

[87] Robert C Robinson et al. "Crystal structure of Arp2/3 complex". In: *Science* 294.5547 (2001), pp. 1679–1684.

[88] Yifan Cheng. "Single-particle cryo-EM at crystallographic resolution". In: *Cell* 161.3 (2015), pp. 450–457.

[89] Florian Fäßler et al. "Cryo-electron tomography structure of Arp2/3 complex in cells reveals new insights into the branch junction". In: *Nature Communications* 11.1 (2020), p. 6437.

[90] Bojian Ding et al. "Structure of Arp2/3 complex at a branched actin filament junction resolved by single–particle cryo–electron microscopy". In: *Proceedings of the National Academy of Sciences of the United States of America* 119.22 (2022), e2202723119.

[91] Steven Z Chou, Moon Chatterjee, and Thomas D Pollard. "Mechanism of actin filament branch formation by Arp2/3 complex revealed by a high-resolution cryo-EM structure of the branch junction". In: *Proceedings of the National Academy of Sciences of the United States of America* 119.49 (2022), e2206722119.

[92] Mohammed Shaaban, Saikat Chowdhury, and Brad J Nolen. "Cryo-EM reveals the transition of Arp2/3 complex from inactive to nucleation-competent state". In: *Nature Structural & Molecular Biology* 27.11 (2020), pp. 1009–1016.

[93] Max Rodnick-Smith et al. “Role and structural mechanism of WASP-triggered conformational changes in branched actin filament nucleation by Arp2/3 complex”. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.27 (2016), E3834–E3843.

[94] Austin Zimmet et al. “Cryo-EM structure of NPF-bound human Arp2/3 complex and activation mechanism”. In: *Science Advances* 6.23 (2020), eaaz7651.

[95] Steven Z Chou and Thomas D Pollard. “Mechanism of actin polymerization revealed by cryo-EM structures of actin filaments with three different bound nucleotides”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.10 (2019), pp. 4265–4274.

[96] Toshiro Oda et al. “The nature of the globular-to fibrous-actin transition”. In: *Nature* 457.7228 (2009), pp. 441–445.

[97] Felipe Merino et al. “Structural transitions of F-actin upon ATP hydrolysis at near-atomic resolution revealed by cryo-EM”. In: *Nature Structural & Molecular Biology* 25.6 (2018), pp. 528–537.

[98] Shih-Chieh Ti et al. “Structural and biochemical characterization of two binding sites for nucleation-promoting factor WASp-VCA on Arp2/3 complex”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.33 (2011), E463–E471.

[99] Benjamin A Smith et al. “Pathway of actin filament branch formation by Arp2/3 complex revealed by single-molecule imaging”. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.4 (2013), pp. 1285–1290.

[100] Heidy Y Narvaez-Ortiz and Brad J Nolen. “Unconcerted conformational changes in Arp2/3 complex integrate multiple activating signals to assemble functional actin networks”. In: *Current Biology* 32.5 (2022), pp. 975–987.

[101] Thomas D Pollard and Gary G Borisy. “Cellular motility driven by assembly and disassembly of actin filaments”. In: *Cell* 112.4 (2003), pp. 453–465.

[102] Alexander Mogilner and George Oster. “Cell motility driven by actin polymerization”. In: *Biophysical Journal* 71.6 (1996), pp. 3030–3045.

[103] Johanna Funk et al. “A barbed end interference mechanism reveals how capping protein promotes nucleation in branched actin networks”. In: *Nature Communications* 12.1 (2021), p. 5329.

[104] R Dyche Mullins, Peter Bieling, and Daniel A Fletcher. “From solution to surface to filament: actin flux into branched networks”. In: *Biophysical Reviews* 10 (2018), pp. 1537–1551.

[105] Lauren E Burianek and Scott H Soderling. “Under lock and key: spatiotemporal regulation of WASP family proteins coordinates separate dynamic cellular processes”. In: *Seminars in Cell & Developmental Biology in cell & developmental biology.* Vol. 24. 4. Elsevier. 2013, pp. 258–266.

[106] Jim Pfaendtner and Gregory A Voth. “Molecular dynamics simulation and coarse-grained analysis of the Arp2/3 complex”. In: *Biophysical Journal* 95.11 (2008), pp. 5324–5333.

[107] Shuting Zhang and Dimitrios Vavylonis. “Steps of actin filament branch formation by Arp2/3 complex investigated with coarse-grained molecular dynamics”. In: *Biophysical Journal* 122.3 (2023), 26a–27a.

[108] Sara Laporte and Alessandra Magistrato. “Deciphering the molecular terms of Arp2/3 allosteric regulation from all-atom simulations and dynamical network theory”. In: *The Journal of Physical Chemistry Letters* 12.22 (2021), pp. 5384–5389.

[109] Paul Dalhaimer, Thomas D Pollard, and Brad J Nolen. “Nucleotide-mediated conformational changes of monomeric actin and Arp3 studied by molecular dynamics simulations”. In: *Journal of Molecular Biology* 376.1 (2008), pp. 166–183.

[110] Jim Pfaendtner et al. “Key structural features of the actin filament Arp2/3 complex branch junction revealed by molecular simulation”. In: *Journal of Molecular Biology* 416.1 (2012), pp. 148–161.

[111] Paul Dalhaimer and Thomas D Pollard. “Molecular dynamics simulations of Arp2/3 complex activation”. In: *Biophysical Journal* 99.8 (2010), pp. 2568–2576.

[112] Ron O Dror et al. “Biomolecular simulation: a computational microscope for molecular biology”. In: *Annual Review of Biophysics* 41 (2012), pp. 429–452.

[113] Qing Luan and Brad J Nolen. “Structural basis for regulation of Arp2/3 complex by GMF”. In: *Nature Structural & Molecular Biology* 20.9 (2013), pp. 1062–1068.

[114] Mark J Dayel and R Dyche Mullins. “Activation of Arp2/3 complex: addition of the first subunit of the new filament by a WASP protein triggers rapid ATP hydrolysis on Arp2”. In: *PLOS Biology* 2.4 (2004), e91.

[115] Elena Ingerman, Jennifer Ying Hsiao, and R Dyche Mullins. “Arp2/3 complex ATP hydrolysis promotes lamellipodial actin network disassembly but is dispensable for assembly”. In: *Journal of Cell Biology* 200.5 (2013), pp. 619–633.

[116] Adam C Martin, Matthew D Welch, and David G Drubin. “Arp2/3 ATP hydrolysis catalysed branch dissociation is critical for endocytic force generation”. In: *Nature Cell Biology* 8.8 (2006), pp. 826–833.

[117] Laurent Blanchoin and Thomas D Pollard. “Hydrolysis of ATP by polymerized actin depends on the bound divalent cation but not profilin”. In: *Biochemistry* 41.2 (2002), pp. 597–602.

[118] Matthias Buck et al. “Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme”. In: *Biophysical Journal* 90.4 (2006), pp. L36–L38.

[119] Qing Luan et al. “Structure of the nucleation-promoting factor SPIN 90 bound to the actin filament nucleator Arp2/3 complex”. In: *The EMBO Journal* 37.22 (2018), e100005.

[120] Katherine Henzler-Wildman and Dorothee Kern. “Dynamic personalities of proteins”. In: *Nature* 450.7172 (2007), pp. 964–972.

[121] Kenneth E Prehoda and Wendell A Lim. “How signaling proteins integrate multiple inputs: a comparison of N-WASP and Cdk2”. In: *Current Opinion in Cell Biology* 14.2 (2002), pp. 149–154.

[122] Sudeepa Rajan, Jonathan R Terman, and Emil Reisler. “MICAL-mediated oxidation of actin and its effects on cytoskeletal and cellular dynamics”. In: *Frontiers in Cell and Developmental Biology* 11 (2023), p. 1124202.

[123] Jieming Chen, Nicholas Sawyer, and Lynne Regan. “Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area”. In: *Protein Science* 22.4 (2013), pp. 510–515.

[124] Byron Hetrick et al. “Small molecules CK-666 and CK-869 inhibit actin-related protein 2/3 complex by blocking an activating conformational change”. In: *Chemistry & Biology* 20.5 (2013), pp. 701–712.

[125] Arjun Narayanan et al. “Phosphorylation of the Arp2 subunit relieves auto-inhibitory interactions for Arp2/3 complex activation”. In: *PLOS Computational Biology* 7.11 (2011), e1002226.

[126] Mark J Dayel, Elizabeth A Holleran, and R Dyche Mullins. “Arp2/3 complex requires hydrolyzable ATP for nucleation of new actin filaments”. In: *Proceedings of the National Academy of Sciences of the United States of America* 98.26 (2001), pp. 14871–14876.

[127] Christophe Le Clainche, Dominique Pantaloni, and Marie-France Carlier. “ATP hydrolysis on actin-related protein 2/3 complex causes debranching of dendritic actin arrays”. In: *Proceedings of the National Academy of Sciences of the United States of America* 100.11 (2003), pp. 6337–6342.

[128] Adam C Martin et al. “Effects of Arp2 and Arp3 nucleotide-binding pocket mutations on Arp2/3 complex function”. In: *Journal of Cell Biology* 168.2 (2005), pp. 315–328.

[129] Benjamin A Smith et al. “Three-color single molecule imaging shows WASP detachment from Arp2/3 complex triggers actin filament branch formation”. In: *Elife* 2 (2013), e01008.

[130] S Vorobiev et al. “The structure of nonvertebrate actin: implications for the ATP hydrolytic mechanism”. In: *Proceedings of the National Academy of Sciences of the United States of America* 100.10 (2003), pp. 5760–5765.

[131] Meghal Gandhi et al. “GMF is a cofilin homolog that binds Arp2/3 complex to stimulate filament debranching and inhibit actin nucleation”. In: *Current Biology* 20.9 (2010), pp. 861–867.

[132] Chao Xie et al. “Actin filament debranching regulates cell polarity during cell migration and asymmetric cell division”. In: *Proceedings of the National Academy of Sciences of the United States of America* 118.37 (2021), e2100805118.

[133] Sofia Espinoza-Sanchez et al. “Conformational changes in Arp2/3 complex induced by ATP, WASp-VCA, and actin filaments”. In: *Proceedings of the National Academy of Sciences of the United States of America* 115.37 (2018), E8642–E8651.

[134] Max Rodnick-Smith et al. "Identification of an ATP-controlled allosteric switch that controls actin filament nucleation by Arp2/3 complex". In: *Nature Communications* 7.1 (2016), p. 12226.

[135] Paul Emsley et al. "Features and development of Coot". In: *Acta Crystallographica Section D: Biological Crystallography* 66.4 (2010), pp. 486–501.

[136] Benjamin Webb and Andrej Sali. "Comparative protein structure modeling using MODELLER". In: *Current Protocols in Bioinformatics* 54.1 (2016), pp. 5–6.

[137] Courtney M Schroeder et al. "A burst of genetic innovation in Drosophila actin-related proteins for testis-specific function". In: *Molecular Biology and Evolution* 37.3 (2020), 757 –772.

[138] Glen M. Hocky et al. "Cations stiffen actin filaments by adhering a key structural element to adjacent subunits". In: *The Journal of Physical Chemistry B* 120 (2016), pp. 4558–4567.

[139] Marissa G Saunders and Gregory A Voth. "Water molecules in the nucleotide binding cleft of actin: effects on subunit conformation and implications for ATP hydrolysis". In: *Journal of Molecular Biology* 413.1 (2011), pp. 279–291.

[140] James C Phillips et al. "Scalable molecular dynamics on CPU and GPU architectures with NAMD". In: *The Journal of Chemical Physics* 153.4 (2020).

[141] Sunhwan Jo et al. "CHARMM-GUI: a web-based graphical user interface for CHARMM". In: *Journal of Computational Chemistry* 29.11 (2008), pp. 1859–1865.

[142] David Van Der Spoel et al. "GROMACS: fast, flexible, and free". In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1701–1718.

[143] Massimiliano Bonomi et al. "Promoting transparency and reproducibility in enhanced molecular simulations". In: *Nature Methods* 16.8 (2019), pp. 670–673.

[144] Robert T McGibbon et al. "MDTraj: a modern open library for the analysis of molecular dynamics trajectories". In: *Biophysical Journal* 109.8 (2015), pp. 1528–1532.

[145] Maximilian Scheurer et al. "PyContact: rapid, customizable, and visual analysis of noncovalent interactions in MD simulations". In: *Biophysical Journal* 114.3 (2018), pp. 577–583.

[146] Eric F Pettersen et al. "UCSF ChimeraX: Structure visualization for researchers, educators, and developers". In: *Protein Science* 30.1 (2021), pp. 70–82.

[147] Andrew Shrake and John A Rupley. "Environment and exposure to solvent of protein atoms. Lysozyme and insulin". In: *Journal of Molecular Biology* 79.2 (1973), pp. 351–371.

[148] Dorothee Liebschner et al. "Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix". In: *Acta Crystallographica Section D: Structural Biology* 75.10 (2019), pp. 861–877.

[149] William Humphrey, A Dalke, and K Schulten. "VMD: Visual molecular dynamics". In: *Journal of Molecular Graphics and Modelling* 14 (), pp. 27–38.

[150] Gang Bao and Subra Suresh. "Cell and molecular mechanics of biological materials". In: *Nat. Mater.* 2.11 (2003), p. 715.

[151] Douglas A Davis et al. "Force-induced activation of covalent bonds in mechanoresponsive polymeric materials". In: *Nature* 459.7243 (2009), p. 68.

[152] Edouard Hannezo and Carl-Philipp Heisenberg. "Mechanochemical Feedback Loops in Development and Disease". In: *Cell* 178 (2019), pp. 12–25.

[153] Caroline E. Wagner et al. "A Rheological Study of the Association and Dynamics of MUC–5AC Gels". In: *Biomacromolecules* 18 (2017), pp. 3654–3664.

[154] David Saintillan. "Rheology of Active Fluids". In: *Annual Review of Fluid Mechanics* 50 (2018), pp. 563–592.

[155] Julio M Fernandez and Hongbin Li. "Force-clamp spectroscopy monitors the folding trajectory of a single protein". In: *Science* 303.5664 (2004), pp. 1674–1678.

[156] Andrés Vera and Mariano Carrión-Vázquez. "Direct Identification of Protein-Protein Interactions by Single-Molecule Force Spectroscopy". In: *Angewandte Chemie International Edition* 55 (2016), pp. 13970–13973.

[157] Floriane Sluysmans Damienand Devaux et al. "Dynamic Force Spectroscopy of Synthetic Oligorotaxane Foldamers". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.38 (2018), pp. 9362–9366.

[158] Emelie Flood et al. "Atomistic Simulations of Membrane Ion Channel Conduction, Gating, and Modulation". In: *Chemical Reviews* 119 (2019), pp. 7737–7832.

[159] Valentina Corradi et al. "Emerging Diversity in Lipid-Protein Interactions". In: *Chemical Reviews* 119 (2019), pp. 5775–5848.

[160] Sanghyun Park et al. "Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality". In: *The Journal of Chemical Physics* 119.6 (2003), pp. 3559–3566.

[161] Gerhard Hummer and Attila Szabo. "Free energy reconstruction from nonequilibrium single-molecule pulling experiments". In: *Proceedings of the National Academy of Sciences of the United States of America* 98.7 (2001), pp. 3658–3661.

[162] Jan Liphardt et al. "Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality". In: *Science* 296.5574 (2002), pp. 1832–1835.

[163] Qi Zhang, Jasna Brujić, and Eric Vanden-Eijnden. "Reconstructing free energy profiles from nonequilibrium relaxation trajectories". In: *J. Stat. Phys.* 144.2 (2011), pp. 344–366.

[164] Anil K. Sahoo, Biman Bagchi, and Prabal K. Maiti. "Unfolding Dynamics of Ubiquitin from Constant Force MD Simulation: Entropy-Enthalpy Interplay Shapes the Free-Energy Landscape". In: *The Journal of Physical Chemistry B* 123 (2019), pp. 1228–1236.

[165] Steven Sheridan, frauke Grater, and Csaba Daday. "How Fast is Too Fast in Force-Probe Molecular Dynamics Simulations". In: *The Journal of Physical Chemistry B* 123.17 (2019), pp. 3658–3664.

[166] Helgi I. Ingolfsson et al. “The Power of Coarse Graining in Biomolecular Simulations”. In: *WIREs Comput. Mol. Sci.* 4 (2014), pp. 225–248.

[167] DK Klimov and D Thirumalai. “Stretching single-domain proteins: phase diagram and kinetics of force-induced unfolding”. In: *Proceedings of the National Academy of Sciences of the United States of America* 96.11 (1999), pp. 6166–6170.

[168] Christian Mücksch and Herbert M Urbassek. “Accelerating steered molecular dynamics: toward smaller velocities in forced unfolding simulations”. In: *Journal of Chemical Theory and Computation* 12.3 (2016), pp. 1380–1384.

[169] Alex Dickson et al. “Flow-dependent unfolding and refolding of an RNA by nonequilibrium umbrella sampling”. In: *Journal of Chemical Theory and Computation* 7.9 (2011), pp. 2710–2720.

[170] Carlos Bustamante et al. “Mechanical Processes in Biochemistry”. In: *Annual Review of Biochemistry* 73 (2004), pp. 705–748.

[171] Anton Martinsson et al. “The simulated tempering method in the infinite switch limit with adaptive weight learning”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.1 (2019), p. 013207.

[172] Gareth A Tribello et al. “PLUMED 2: New feathers for an old bird”. In: *Computer Physics Communications* 185.2 (2014), pp. 604–613.

[173] Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. “On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction”. In: *The Journal of Chemical Physics* 116.20 (2002), pp. 9058–9067.

[174] Enzo Marinari and Giorgio Parisi. “Simulated tempering: a new Monte Carlo scheme”. In: *Europhysics Letters* 19.6 (1992), p. 451.

[175] Paul Dupuis et al. “On the infinite swapping limit for parallel tempering”. In: *Multiscale Modeling & Simulation* 10.3 (2012), pp. 986–1022.

[176] Nuria Plattner et al. “An infinite swapping approach to the rare-event sampling problem”. In: *The Journal of chemical physics* 135.13 (2011).

[177] Monroe D Donsker and SR Srinivasa Varadhan. “Asymptotic evaluation of certain Markov process expectations for large time, I”. In: *Communications on pure and applied mathematics* 28.1 (1975), pp. 1–47.

[178] Jean-Dominique Deuschel and Daniel W Stroock. *Large deviations, volume 137 of Pure and Applied Mathematics.* 1989.

[179] Otton Nikodym. “Sur une généralisation des intégrales de MJ Radon”. In: *Fundamenta Mathematicae* 15.1 (1930), pp. 131–179.

[180] Iain Bethune et al. “MIST: A simple and efficient molecular dynamics abstraction library for integrator development”. In: *Computer Physics Communications* 236 (2019), pp. 224–236.

[181] Dominik M. Endres and Johannes E. Schindelin. “A New Metric for Probability Distributions”. In: *IEEE Transactions on Information Theory* 49.7 (2003), pp. 1858–1860.

[182] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature Methods* 17.3 (2020), pp. 261–272.

[183] John D Weeks, David Chandler, and Hans C Andersen. “Role of repulsive forces in determining the equilibrium structure of simple liquids”. In: *The Journal of Chemical Physics* 54.12 (1971), pp. 5237–5247.

[184] Erik Thiede et al. “Eigenvector method for umbrella sampling enables error analysis”. In: *The Journal of Chemical Physics* 145.8 (2016), p. 084115.

[185] Scott A Hollingsworth and P Andrew Karplus. “A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins”. In: *Biomolecular Concepts* 1.3-4 (2010), pp. 271–283.

[186] Alexei A Adzhubei, Michael JE Sternberg, and Alexander A Makarov. “Polyproline-II helix in proteins: structure and function”. In: *Journal of Molecular Biology* 425.12 (2013), pp. 2100–2132.

[187] Zhengshuang Shi et al. “Polyproline II structure in a sequence of seven alanine residues”. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.14 (2002), pp. 9190–9195.

[188] Jurgen Graf et al. “Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study”. In: *Journal of the American Chemical Society* 129 (2007), pp. 1179–1189.

[189] Marcelo E Guerin, Guillaume Stirnemann, and David Giganti. “Conformational entropy of a single peptide controlled under force governs protease recognition and catalysis”. In: *Proceedings of the National Academy of Sciences of the United States of America* 115.45 (2018), pp. 11525–11530.

[190] Glenn M Torrie and John P Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of Computational Physics* 23.2 (1977), pp. 187–199.

[191] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-tempered metadynamics: a smoothly converging and tunable free-energy method”. In: *Physical Review Letters* 100.2 (2008), p. 020603.

[192] Omar Valsson and Michele Parrinello. “Variational approach to enhanced sampling and free energy calculations”. In: *Physical Review Letters* 113.9 (2014), p. 090601.

[193] Jed W Pitera and John D Chodera. “On the use of experimental observations to bias simulated ensembles”. In: *Journal of Chemical Theory and Computation* 8.10 (2012), pp. 3445–3451.

[194] Andrew D White and Gregory A Voth. “Efficient and minimal method to bias molecular simulations with experimental data”. In: *Journal of Chemical Theory and Computation* 10.8 (2014), pp. 3023–3030.

[195] Glen M Hocky, Thomas Dannenhoffer-Lafage, and Gregory A Voth. "Coarse-grained directed simulation". In: *Journal of Chemical Theory and Computation* 13.9 (2017), pp. 4593–4603.

[196] Andrea Cesari, Sabine Reißer, and Giovanni Bussi. "Using the maximum entropy principle to combine simulations and solution experiments". In: *Computation* 6.1 (2018), p. 15.

[197] Dilnoza B Amirkulova and Andrew D White. "Recent advances in maximum entropy biasing techniques for molecular dynamics". In: *Molecular Simulation* (2019), pp. 1–10.

[198] Travis E Oliphant. *A guide to NumPy*. Vol. 1. Trelgol Publishing USA, 2006.

[199] John D Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95.

[200] Mark James Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1 (2015), pp. 19–25.

[201] David van der Spoel et al. "GROMACS: Fast, Flexible, and Free". In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1701–1718.

[202] B. Best Robert et al. "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles". In: *Journal of Chemical Theory and Computation* 8 (2012), pp. 3257–3273.

[203] Berk Hess et al. "LINCS: A Linear Constraint Solver for Molecular Simulations". In: *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472.

[204] Herman J. C. Berendsen et al. "Molecular Dynamics with Coupling to and External Bath". In: *The Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690.

[205] A. Rahman and M. Parrinello. "Crystal Structure and Pair Potentials: A Molecular Dynamics Study". In: *Physical Review Letters* 45.14 (1980), pp. 1196–1199.

[206] Giovanni Bussi, Davide Donadio, and Michele Parinello. "Canonical Sampling Through Velocity Rescaling". In: *The Journal of Chemical Physics* 126 (2007), p. 014101.

[207] Yuvraj Singh and Glen M Hocky. "Improved prediction of molecular response to pulling by combining force tempering with replica exchange methods". In: *The Journal of Physical Chemistry B* (2024).

[208] Brenton D Hoffman, Carsten Grashoff, and Martin A Schwartz. "Dynamic molecular processes mediate cellular mechanotransduction". In: *Nature* 475.7356 (2011), pp. 316–323.

[209] Thomas Iskratsch, Haguy Wolfenson, and Michael P Sheetz. "Appreciating force and shape-the rise of mechanotransduction in cell biology". In: *Nature Reviews Molecular Cell Biology* 15.12 (2014), pp. 825–833.

[210] Naomi Courtemanche et al. "Tension modulates actin filament polymerization mediated by formin and profilin". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.24 (2013), pp. 9752–9757.

[211] Antoine Jégou, Marie-France Carlier, and Guillaume Romet-Lemonne. “Formin mDia1 senses and generates mechanical forces on actin filaments”. In: *Nature Communications* 4.1 (2013), p. 1883.

[212] Dennis Zimmermann and David R Kovar. “Feeling the force: formin’s role in mechanotransduction”. In: *Current Opinion in Cell Biology* 56 (2019), pp. 130–140.

[213] Pekka Lappalainen et al. “Biochemical and mechanical regulation of actin dynamics”. In: *Nature Reviews Molecular Cell Biology* 23.12 (2022), pp. 836–852.

[214] Andrew S LaCroix et al. “Tunable molecular tension sensors reveal extension-based control of vinculin loading”. In: *eLife* 7 (2018), e33927.

[215] Michael D Brenner et al. “Spider silk peptide is a compact, linear nanospring ideal for intracellular tension sensing”. In: *Nano Letters* 16.3 (2016), pp. 2096–2102.

[216] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

[217] Jérôme Hénin et al. “Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1. 0]”. In: *Living Journal of Computational Molecular Science* 4.1 (2022), pp. 1583–1583.

[218] Guillaume Stirnemann. “Recent advances and emerging challenges in the molecular modeling of mechanobiological processes”. In: *The Journal of Physical Chemistry B* 126.7 (2022), pp. 1365–1374.

[219] Robert H Swendsen and Jian-Sheng Wang. “Replica Monte Carlo simulation of spin–glasses”. In: *Physical Review Letters* 57.21 (1986), p. 2607.

[220] Giovanni Bussi. “Hamiltonian replica exchange in GROMACS: a flexible implementation”. In: *Molecular Physics* 112.3-4 (2014), pp. 379–384.

[221] Pu Liu et al. “Replica exchange with solute tempering: A method for sampling biological systems in explicit water”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.39 (2005), pp. 13749–13754.

[222] Xuhui Huang et al. “Replica exchange with solute tempering: efficiency in large scale systems”. In: *The Journal of Physical Chemistry B* 111.19 (2007), pp. 5405–5410.

[223] William L Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935.

[224] Shams Mehdi et al. “Accelerating all-atom simulations and gaining mechanistic understanding of biophysical systems through state predictive information bottleneck”. In: *Journal of Chemical Theory and Computation* 18.5 (2022), pp. 3231–3238.

[225] Jing Huang et al. “CHARMM36m: an improved force field for folded and intrinsically disordered proteins”. In: *Nature Methods* 14.1 (2017), pp. 71–73.

[226] Subarna Sasmal, Martin McCullagh, and Glen M Hocky. “Reaction Coordinates for Conformational Transitions Using Linear Discriminant Analysis on Positions”. In: *Journal of Chemical Theory and Computation* (2023).

[227] Jan Kubelka et al. "Sub-microsecond protein folding". In: *Journal of Molecular Biology* 359.3 (2006), pp. 546–553.

[228] Kresten Lindorff-Larsen et al. "Improved side-chain torsion potentials for the Amber ff99SB protein force field". In: *Proteins: Structure, Function, and Bioinfor- matics* 78.8 (2010), 1950–1958.

[229] Stefano Piana, Kresten Lindorff-Larsen, and David E Shaw. "Protein folding kinetics and thermodynamics from atomistic simulation". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.44 (2012), pp. 17845–17850.

[230] Claude Sammut and Geoffrey I Webb. *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.

[231] Sebastian Buchenberg, Norbert Schaudinnus, and Gerhard Stock. "Hierarchical biomolecular dynamics: Picosecond hydrogen bonding regulates microsecond conformational transitions". In: *Journal of Chemical Theory and Computation* 11.3 (2015), pp. 1330–1336.

[232] Stefano Piana et al. "Water dispersion interactions strongly influence simulated structural properties of disordered protein states". In: *The Journal of Physical Chemistry B* 119.16 (2015), pp. 5113–5123.

[233] Sarah Rauscher et al. "Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment". In: *Journal of Chemical Theory and Computation* 11.11 (2015), pp. 5513–5524.

[234] Paul Robustelli, Stefano Piana, and David E Shaw. "Developing a molecular dynamics force field for both folded and disordered protein states". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.21 (2018), E4758–E4766.

[235] Daniel Sindhikara, Yilin Meng, and Adrian E Roitberg. "Exchange frequency in replica exchange molecular dynamics". In: *The Journal of Chemical Physics* 128.2 (2008).

[236] Daniel J Sindhikara, Daniel J Emerson, and Adrian E Roitberg. "Exchange often and properly in replica exchange molecular dynamics". In: *Journal of Chemical Theory and Computation* 6.9 (2010), pp. 2804–2808.

[237] Robert B Best and Gerhard Hummer. "Optimized molecular dynamics force fields applied to the helix- coil transition of polypeptides". In: *The Journal of Physical Chemistry B* 113.26 (2009), pp. 9004–9015.

[238] Vishnu Priyanka Reddy Chichili, Veerendra Kumar, and Jayaraman Sivaraman. "Linkers in the structural biology of protein–protein interactions". In: *Protein Science* 22.2 (2013), pp. 153–167.

[239] Joshua S Klein et al. "Design and characterization of structured protein linkers with differing flexibilities". In: *Protein Engineering, Design & Selection, Design & Selection* 27.10 (2014), pp. 325–330.

[240] Xiaoying Chen, Jennica L Zaro, and Wei-Chiang Shen. "Fusion protein linkers: property, design and functionality". In: *Advanced Drug Delivery Reviews* 65.10 (2013), pp. 1357–1369.

[241] Robert E Bird et al. "Single-chain antigen-binding proteins". In: *Science* 242.4877 (1988), pp. 423–426.

[242] Nurmamet Amet, Hsin-Fang Lee, and Wei-Chiang Shen. "Insertion of the designed helical linker led to increased expression of tf-based fusion proteins". In: *Pharmaceutical Research* 26 (2009), pp. 523–528.

[243] Ryoichi Arai et al. "Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering". In: *Proteins: Structure, Function, and Bioinfor- matics* 57.4 (2004), pp. 829–838.

[244] Ryoichi Arai et al. "Design of the linkers which effectively separate domains of a bifunctional fusion protein". In: *Protein Engineering, Design & Selection* 14.8 (2001), pp. 529–532.

[245] Philip E Thorpe et al. "New coupling agents for the synthesis of immunotoxins containing a hindered disulfide bond with improved stability in vivo". In: *Cancer Research* 47.22 (1987), pp. 5924–5931.

[246] Hong Liang Zhao et al. "Balancing the pharmacokinetics and pharmacodynamics of interferon−α2b and human serum albumin fusion protein by proteolytic or reductive cleavage increases its in vivo therapeutic efficacy". In: *Molecular Pharmaceutics* 9.3 (2012), pp. 664–670.

[247] Chiquito J Crasto and Jin-an Feng. "LINKER: a program to generate linker sequences for fusion proteins". In: *Protein Engineering, Design & Selection* 13.5 (2000), pp. 309–312.

[248] Chi-Duen Poon et al. "Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution?" In: *Journal of the American Chemical Society* 122.23 (2000), pp. 5642–5643.

[249] Paul E Smith. "The alanine dipeptide free energy surface in solution". In: *The Journal of Chemical Physics* 111.12 (1999), pp. 5568–5579.

[250] Parag Mukhopadhyay, Gérard Zuber, and David N Beratan. "Characterizing aqueous solution conformations of a peptide backbone using Raman optical activity computations". In: *Biophysical Journal* 95.12 (2008), pp. 5574–5586.

[251] Narasimha Sreerama and Robert W Woody. "Molecular dynamics simulations of polypeptide conformations in water: A comparison of α, β, and poly (pro) II conformations". In: *Proteins: Structure, Function, and Bioinfor- matics* 36.4 (1999), pp. 400–406.

[252] Peter B Law and Valerie Daggett. "The relationship between water bridges and the polyproline II conformation: a large-scale analysis of molecular dynamics simulations and crystal structures". In: *Protein Engineering, Design & Selection, Design & Selection* 23.1 (2010), pp. 27–33.

[253] Mihaly Mezei et al. "Polyproline II helix is the preferred conformation for unfolded polyalanine in water". In: *Proteins: Structure, Function, and Bioinfor- matics* 55.3 (2004), pp. 502–507.

[254] Andreas Vitalis and Rohit V Pappu. "Methods for Monte Carlo simulations of biomacromolecules". In: *Annual Reports in Computational Chemistry* 5 (2009), pp. 49–76.

[255] NB Becker, A Rosa, and R Everaers. "The radial distribution function of worm-like chains". In: *The European Physical Journal E* 32 (2010), pp. 53–69.

[256] Mahmoud Moradi et al. "Are long-range structural correlations behind the aggregration phenomena of polyglutamine diseases?" In: *PLOS Computational Biology* 8.4 (2012), e1002501.

[257] JF Marko and ED Siggia. "Statistical mechanics of supercoiled DNA". In: *Physical Review E* 52.3 (1995), p. 2912.

[258] Martijn Van Rosmalen, Mike Krom, and Maarten Merkx. "Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins". In: *Biochemistry* 56.50 (2017), pp. 6565–6574.

[259] Shanmei Cheng, Murat Cetinkaya, and Frauke Gräter. "How sequence determines elasticity of disordered proteins". In: *Biophysical Journal* 99.12 (2010), pp. 3863–3869.

[260] David A Case et al. "AmberTools". In: *Journal of Chemical Information and Modeling* 63.20 (2023), pp. 6183–6191.

[261] Davide Branduardi, Francesco Luigi Gervasio, and Michele Parrinello. "From A to B in free energy space". In: *The Journal of Chemical Physics* 126.5 (2007).

[262] LLC Schrödinger and Warren DeLano. *PyMOL*. Version 2.4.0. May 20, 2020.