Nucleotide and Polymerization Effects on Actin Structure and Dynamics

by

Lauren Jepsen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in The University of Michigan
2018

Doctoral Committee:

Professor David Sept, Chair
Professor Charles Brooks III
Professor Daniel Burns Jr
Professor Peter Freddolino
Professor Ann Miller
Professor Janet Smith

Lauren Jepsen

ljepsen@umich.edu

ORCID iD: 0000-0003-3621-0483

To my wonderful husband.

Your unconditional love and support means the world to me.

Acknowledgements

It is no secret that graduate school is hard, for many reasons. Obtaining a Ph.D. is a process that can be made immeasurably more difficult if you do not have the right people to surrounding you. In this regard, I have been lucky. I have had the pleasure of working for Professor Dave Sept the last six years. Your understanding, guidance and mentorship have made the last six years enjoyable. I could not have asked for a better advisor. I look forward to continuing to work for you. Thank you also for putting up with my pervasive use of the color purple. I have also been lucky to have the full support of my family. I would especially like to thank my husband Nathan. With you by my side, I feel like I can accomplish anything.

Table of Contents

List of Figures

List of Tables

Abstract

Actin is one of the most highly conserved and abundant proteins found in eukaryotic cells, essential for determining cell shape, polarity and motility, as well as protein transport and even mitosis and muscle contraction. These functions depend on actin's ability to exist – under the strict control of nucleotide hydrolysis and interaction with actin binding partners – as either a monomer (G-actin) or a polymer (F-actin). Over the last seven decades of research, much has been determined in the way of actin function, but key questions still remain. First, the affinity of actin for the >150 actin binding proteins, and for polymerization, is reliant on upon actin's nucleotide state but the structural changes that occur upon nucleotide hydrolysis are not so clear. Next, the two ends of the actin filament have different properties, with incoming ATP-actin subunits preferentially adding to the barbed end of the filament and ADP-actin subunits preferentially dissociating from the pointed end. Much like with the structural changes that occur upon hydrolysis, the physical and biochemical bases for these differences are unknown. Finally, as the cell relies on the strict regulation of the actin filament, alterations to actins sequence are poorly tolerated. Although many mutations prove to be lethal to the organism, over 140 disease causing actin mutants have been reported, with a large subset clustering on actin's pathogenic helix (residues 113-125). Little is known about the structural consequences of these mutations and how they relate to disease. In my thesis, I use molecular dynamics simulations of both G- and F-actin to probe these questions.

Actin itself has two clefts, the nucleotide-binding cleft at the center of the protein, and the target-binding cleft at the bottom of the protein between subdomains 1 and 3, where the majority of actin binding proteins dock. I show that changes within the nucleotide-binding cleft propagate down to the target-binding cleft through the intermediary C-terminal hinge (A331-Y337). Within the target-binding cleft itself, I identify a new loop at the profilin binding site (FQQ-loop: S348-W356) that moves by nearly 5 Å in the ATP state to partially obstruct the target-binding cleft. All of these changes help explain nucleotide state specificity for actin binding proteins. My work also reveals that ATP G-actin takes on a flatter conformation that is structurally similar to F-actin's barbed end protomer, explaining the observation that ATP G-actin polymerizes faster than its ADP counterpart. I find that the pointed end of the filament takes on a conformation that is divergent from remainder of the filament and monomer simulations, effectively raising the conformational energy barrier for the addition of actin protomers.

I also looked at the structural consequences of the deafness causing mutations K118M/N. The mutations to K118 result in changes in the structure and dynamics of the D-loop, alterations in the structure of the H73-loop as well as the sidechain orientations of W79 and W86, changes in nucleotide exchange rates, and significant shifts in the twist of the actin monomer. With K118N the twist of the monomer is nearly identical to the F-actin protomer, and *in vitro* polymerization assays show that this mutation results in faster polymerization. Taken together, it is evident that mutations at this site give rise to a series of small changes that can be tolerated *in vivo*, but result in misregulation of actin assembly and dynamics.

Chapter 1

Introduction

1.1 Actin in the Cytoskeleton

Actin is one of the most highly conserved and abundant proteins. Nearly all eukaryotic cells contain actin, and all eukaryotic organisms produce at least one actin isoform - all of which are conserved in both sequence and fold. Within cells, actin participates in a wide array of protein-protein interactions. It essential for determining cell shape, polarity and motility, as well as protein transport and even mitosis and muscle contraction (Pollard & Cooper, 2009; Carlier, 2010). Central to its function is its ability to polymerize into filaments. Actin monomers are referred to as globular-actin or G-actin, and actin filaments are called F-actin. The actin filament, along with microtubules and intermediate filaments, is one of the three pillars of the cytoskeleton (Pollard & Cooper, 2009).

Actin filaments form in unidirectional fashion and as such are polar, with end dependent behavior. The two ends of the filament are called the barbed end and the pointed end. Under steady state conditions, in a process called filament treadmilling, there is a net addition of actin subunits onto the barbed end of the filament and a net dissociation of subunits from the pointed end. The

cell takes advantage of this behavior. Within the cell, the filament's barbed end is typically in contact with the plasma membrane and it serves as the track for myosin motors (Pollard & Cooper, 2009). As myosin moves away from the barbed end of the filament, putting tension on the filaments. The force produced by this action is used for cell locomotion, to alter cell shape, and to divide the cell in two during mitosis. This force can also be transmitted to adjacent cells or to substrates.

Cellular movement occurs through a cyclical process of protrusion, attachment, and retraction of the tailing edge. These are actin driven processes. As the barbed end of the filament is in contact with the plasma membrane, filament polymerization pushes the membrane outward leading to cellular protrusion (Svitkina, 2018). At the tailing end of the cell, the forces of myosin pulling on actin helps to retract the tailing end of the cell.

At the leading edge of the cell, large branched networks of actin filaments, called lamellipodia, form and generate the large forces that push the cellular membrane forward (Svitkina, 2018). The dendritic nucleation hypothesis for the assembly of F-actin networks describes this process (Figure 1.1). First, within the cell, a large pool of ATP G-actin monomers is maintained by proteins such as profilin and thymosin-β4. Both of these proteins allow the cell to accumulate ATP G-actin at concentrations much greater than that needed for polymerization (Dominguez & Holmes, 2011). Next, when the cell receives an extracellular signal, a family Rho GTPases activate WASp/Scar proteins (Pollard, Blanchoin, & Mullins, 2001). These proteins in turn activate the Arp2/3 complex to nucleate new filament formation (Pollitt & Insall, 2009). The Arp2/3 complex is made up of actin-related proteins 2 and 3 as well as 5 other proteins. The Arp2/3 complex forms the branches of actin filaments by nucleating new filaments off other actin filaments. New branches form at actin's barbed end at an angle of 70° off of the mother filament

(Mullins, Heuser, & Pollard, 1998). The growth of these actin filaments towards the plasma membrane pushes the membrane forward. Within the filament, ATP is hydrolyzed to ADP-P_i and the phosphate is eventually released, leaving ADP subunits. ADF/cofilin then comes in and binds to the ADP region of the filaments, leading to dissociation of ADP actin. Freely released ADP G-actin binds to profilin which catalyzes the exchange of ADP for ATP, and then helps to maintain the pool of ATP G-actin, bringing us full circle.

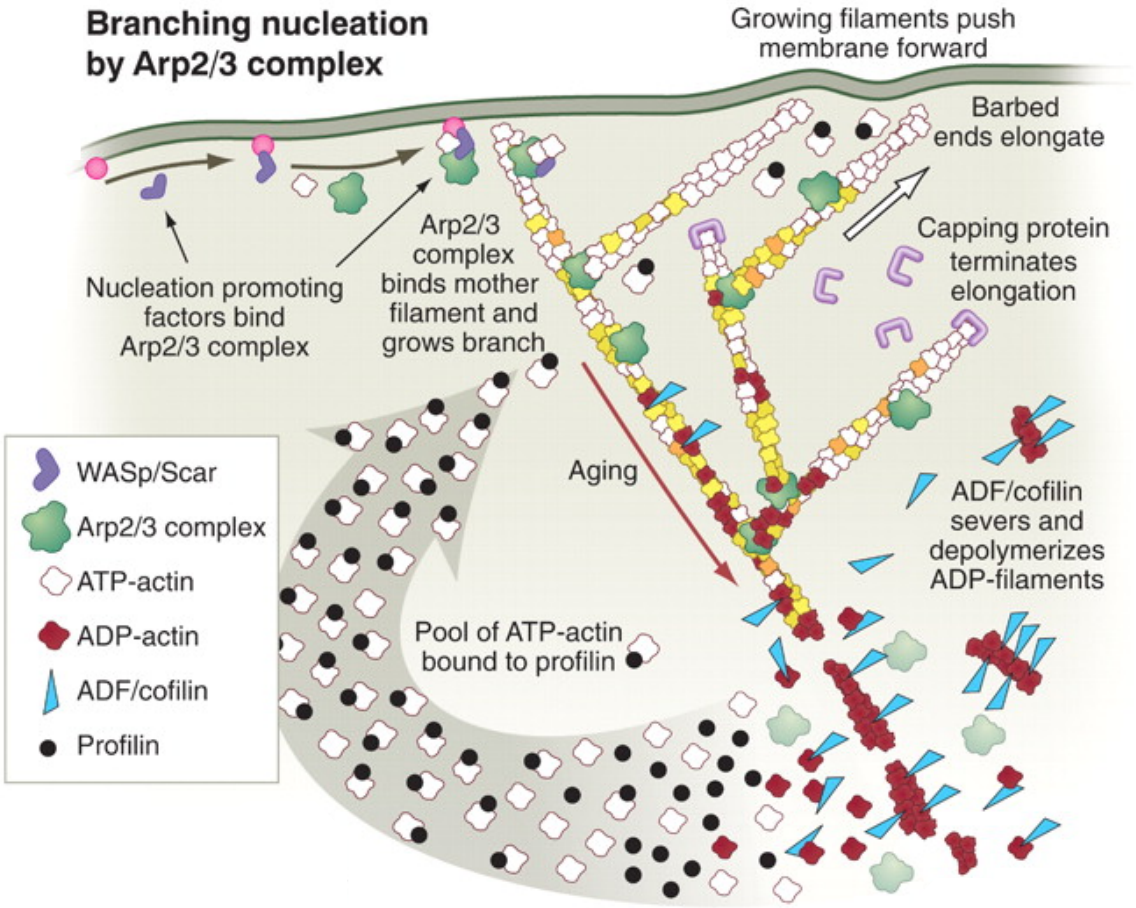


Figure 1.1 Dendritic Nucleation Model The dendritic nucleation model explains the formation of Arp2/3 mediated branched filament networks needed for cellular movement. Reprinted from Pollard & Cooper, 2009

1.2 G-Actin Structure

All actin and actin related proteins (ARP) adopt the same basic fold, forming a relatively flat protein (55 Å x 55 Å X 35 Å), comprised of two major domains, or four subdomains (Figure 1.2). Connecting subdomains 1 and 3 are two short polypeptides (Q137-S145, and A331-Y337), these divide the protein into two clefts. The nucleotide-binding cleft sits in the center of the protein, and the target-binding cleft sits at the bottom of the protein between subdomains 1 and 3. It is in this bottom cleft where actin preferentially interacts with its binding partners. Within the central, nucleotide-binding cleft, actin can bind a cation and ATP, ADP-P_i, or ADP. Actin's function and structure is dependent upon both the bound cation and the bound nucleotide (Strzelecka-Gołaszewska, Moraczewska, Khaitlina, & Mossakowska, 1993).

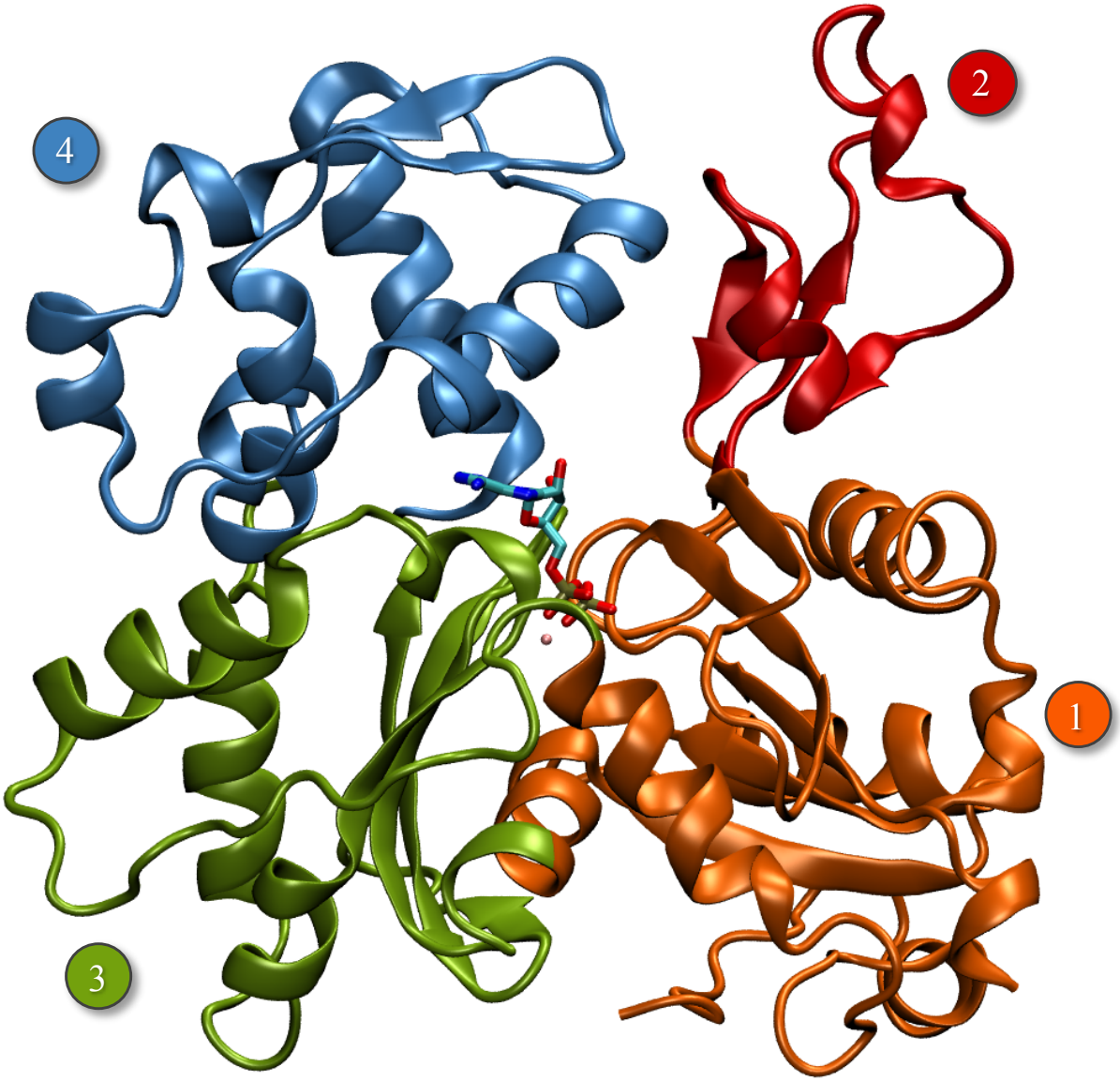


Figure 1.2 Actin Subdomain Arrangement Actin is composed of two major domains, or four subdomains. Subdomains 1 (residues 1-31, 70-144, and 338-375) and 2 (residues 32-69) make up the first major domain and subdomains 3 (residues 145-180, and 270-337) and 4 (residues 181-269) the second. In the center of the protein, between all subdomains, sits the nucleotide binding cleft which binds either ATP or ADP along with a high affinity cation. At the bottom of the protein, between subdomains 1 and 3, is the target binding cleft.

1.2.1 Nucleotide Effects on G-Actin

Actin subunits function differently depending on the nucleotide bound. ATP-actin and ADP-bound actin have different affinities for various actin binding proteins (dos Remedios et al., 2003), as well as for the filament with ATP-actin polymerizing faster than its ADP-counterpart (Pollard, 1986; Fujiwara, Vavylonis, & Pollard, 2007). Many structural studies have been done to

help elucidate the structural differences between the two nucleotide states that lead to their unique properties. Most useful is the comparison of actin crystal structures in different nucleotide states. In order to crystalize G-actin polymerization must be inhibited. Inhibition of polymerization can be accomplished by binding small molecules or other proteins, mutating key residues, or by covalent modification. Actin's first crystal structure was obtained in 1990 by binding rabbit muscle actin with DNase-I (Kabsch, Mannherz, Suck, Pai, & Holmes, 1990). In the following sections, we discuss actin crystalized by the binding of a tetramethylrhodamine (TMR) to the C-terminus of the protein (Otterbein, Graceffa, & Dominguez, 2001; Graceffa & Dominguez, 2003) and structures crystalized through the mutation of two key residues (A204E/P243K) to produce polymerization incompetent structures (Rould, Wan, Joel, Lowey, & Trybus, 2006). These crystal structures as well as molecular-dynamics simulations of G-actin in different nucleotide states, have revealed a few regions of differing structure.

1.2.1.1 Changes within the Nucleotide Binding Cleft

Within the nucleotide binding cleft, we see the first structural shifts due to nucleotide state. Three key state state-sensing loops are the S-loop (residues 11-16), G-loop (residues 155-160) and H73-loop (residues 70-78). The S- and G-loops receive their names from the S14 and G158 residues respectively. These are conserved residues in NTPases (Vetter & Wittinghofer, 2001; Zheng, Diraviyam, & Sept, 2007). Both the S- and G-loops form β-hairpin structures and shift positions to accommodate different nucleotide structures (Otterbein et al., 2001; Graceffa & Dominguez, 2003; Rould et al., 2006; Zheng et al., 2007). In ADP G-actin, the S-loop makes hydrogen-bonding interactions with ADP's β-phosphate. When the nucleotide is exchanged for ATP, the S- and G-loop rock with respect to each other and the S-loop moves to make contacts with the γ-phosphate (Figure 1.3). The H73-loop, named after the H73 which is methylated in the

majority of eukaryotes (Kalhor et al., 1999; Yao, Grade, Wriggers, & Rubenstein, 1999), changes conformation as a result of the shifts of the S- and G-loops. Unlike the former loops, the H73-loop changes its secondary structure in response to the nucleotide (Figure 1.3). In ADP actin, the H73-loop remains unstructured, but upon introduction of the additional phosphate group, either in ADP-P_i actin or ATP actin, it forms a β-sheet (Otterbein et al., 2001; Graceffa & Dominguez, 2003; Rould et al., 2006; Zheng et al., 2007).

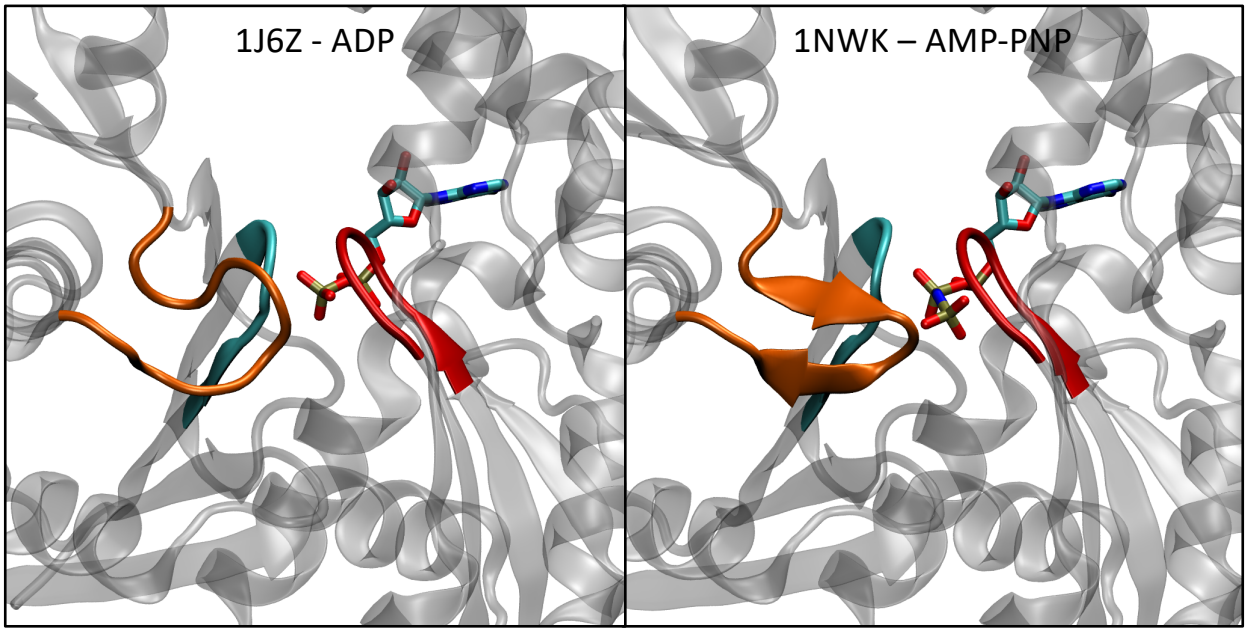


Figure 1.3 S-, G-, and H73-Loops in G-Actin The nucleotide binding clefts of crystal structures 1J6Z (Otterbein et al., 2001) and 1NWK (Graceffa & Dominguez, 2003) representing the ADP and ATP G-actin states respectively. 1NWK uses AMP-PNP as a nonhydrolyzable ATP analog. We see the S-loop (cyan) and G-loop (red) rocking with respect to each other and the H73-loop (orange) moving towards the nucleotide and becoming structured when the nucleotide is switched from ADP.

1.2.1.2 D-Loop

The DNase-I loop, or D-loop (R39-S52), is a mobile loop located at the top of subdomain 2 (Figure 1.4). In ADP-TMR-actin, the D-loop forms a short helix but is unresolved in the ATP structure (Otterbein et al., 2001; Graceffa & Dominguez, 2003). The nucleotide dependent structure of this loop is debated within the literature. The current hypothesis is that the D-loop undergoes a loop-to-helix transition upon nucleotide hydrolysis. Tryptophan cleavage data

supports a more compact D-loop in the ADP state. In ADP G-actin, the D-loop is less susceptible to cleavage between residues 47 and 48 indicating it is in a more compact state, consistent with a helical structure (Strzelecka-Gołaszewska et al., 1993; Kudryashov & Reisler, 2013). Molecular dynamics have captured the loop-to-helix transition (Zheng et al., 2007), but have also indicated that the loop is unstable regardless of the nucleotide state (Dalhaimer, Pollard, & Nolen, 2008; Splettstoesser, Noé, Oda, & Smith, 2009; Durer et al., 2012). Within the F-actin filament, the D-loop makes several inter-protomer contacts (Holmes, Popp, Gebhard, & Kabsch, 1990; Kim et al., 2000; Oda, Iwasa, Aihara, Maéda, & Narita, 2009; Fujii, Iwane, Yanagida, & Namba, 2010; Oztug Durer, Diraviyam, Sept, Kudryashov, & Reisler, 2010; Durer et al., 2012; Kudryashov & Reisler, 2013), indicating that it may play a role in the nucleotide dependent affinity for the polymerization.

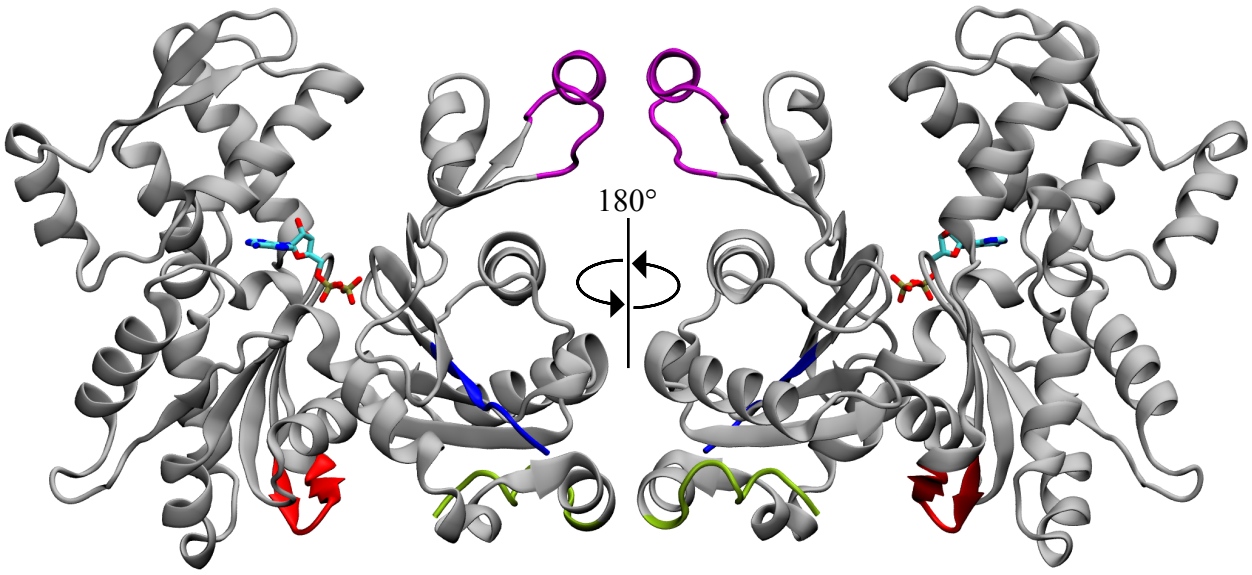


Figure 1.4 D-loop, C-terminus, N-terminus, W-loop. The D-loop (purple) sits at the top of subdomain 2. Both the N- and C-terminus (blue and green respectively) are located within subdomain 1, and the W-loop (red) is positioned with the target-binding cleft at the bottom of subdomain 3.

1.2.1.3 W-Loop

The WH2-loop, or the W-loop, sits at the bottom of subdomain 3 (residues 165-172) within the target-binding cleft of actin (Figure 1.4). This loop has been shown to be the interaction site

for proteins with a WH2 domain; such proteins include Wiskott-Aldrich syndrome protein (WASp), spire and thymosin-β4 (Zheng et al., 2007; Dominguez, 2016). The WH2 domain is one of the most abundant actin-binding folds (Dominguez & Holmes, 2011). A comparison of crystal structures shows no discernible difference in the structure of the W-loop between nucleotide states (Otterbein et al., 2001; Graceffa & Dominguez, 2003). Molecular dynamics simulations, reveal that in the ADP or ADP-P_i actin states that the W-loop can form a stable β-sheet and that it takes on an altered conformation over ATP-actin (Zheng et al., 2007). The shift in conformation was experimentally observed by Kudryashov *et al.* (Kudryashov, Grintsevich, Rubenstein, & Reisler, 2010). They attached fluorescent probes to residues in the W-loop mutated to cystine and observed differing patterns of fluorescence between nucleotide states. The altered structure of the loop within ATP-actin matches well with the observation that WH2 domain containing proteins preferentially bind ATP-actin (Carlier, Jean, Rieger, Lenfant, & Pantaloni, 1993; Marchand, Kaiser, Pollard, & Higgs, 2001; Mattila, Salminen, Yamashiro, & Lappalainen, 2003; Chereau et al., 2005; Zheng et al., 2007; Hild, Bugyi, & Nyitrai, 2010). While the differing structure of the W-loop helps explain the nucleotide preference of WH2-domain containing proteins, it is insufficient to fully explain the nucleotide state sensitivity of all actin binding proteins. The question still remains as to how the changes within the nucleotide binding cleft get allosterically propagated down to the W-loop.

1.2.1.4 C-Terminus

Actin's N- and C-terminal regions, located in subdomain 1 (Figure 1.4), both have a high degree of structural flexibility. In TMR-actin, ADP G-actin has a more resolved N- and C-terminus than its ATP equivalent, suggesting that perhaps these regions are more mobile in the ATP state (Otterbein et al., 2001; Graceffa & Dominguez, 2003). In AP-actin, which contains two mutations

(A204E/P243K) to make it non-polymerizable, the two regions are equivalent in either nucleotide state (Rould et al., 2006). While the crystal structures do not provide much insight into the differences between nucleotide states of these regions, experimental results show that the C-terminus takes on nucleotide specific structure and conformation. A tryptic cleavage site exists between residues 372 and 373. Cleavage of this site is affected by nucleotide state (Strzelecka-Gołaszewska et al., 1993) suggesting some differences within the final helix formed within the crystals from residue 369 to 373. Further, molecular dynamics simulations show that the C-terminus from 349 to 375 shifts by 5Å between the ADP, and ATP states (Zheng et al., 2007). As the C-terminus sits at the bottom of subdomain 1, it is quite possible that changes to the C-terminus in connection to the nucleotide play a role in actin binding proteins selectivity. Further, the 5 Å shift in conformation of C-terminus observed by Zheng *et al.* (2007), matches what is observed when profilin is bound to actin (Chik, Lindberg, & Schutt, 1996).

1.2.1.5 Subdomain Arrangement and Dynamics

The discussion up until now has focused on how actin's secondary structure is altered between nucleotide states, but now we will focus on how nucleotide state influences both tertiary structure and dynamics. In this regard, we will look at data from crosslinking experiments, and molecular dynamics simulations. One hypothesis that has been proposed, is that the nucleotide binding cleft is more open in the ADP G-actin sate, in other words that subdomains 3 and 4 are more separated (Splettstoesser et al., 2009). This came about when the crystal structure of profilin bound to actin in the open state was released. After performing cysteine-crosslinking experiments with yeast actin's nucleotide cleft mouth, Kudryashov and Reisler (Kudryashov & Reisler, 2013) concluded that the nucleotide binding cleft was more open in the ADP G-actin state. However, contradictory to their conclusion, examination of their data shows that the nucleotide cleft mouth

in ATP G-actin was more efficiently crosslinking for every length of crosslinker used (Kudryashov & Reisler, 2013). This would imply that the cleft mouth is wider, or at the very least more dynamic, when ATP is bound. Molecular dynamics simulations have shown that there is no difference in the openness of the nucleotide binding cleft (Zheng et al., 2007). Simulations have also shown that ATP G-actin has a smaller conformational space than its ADP counterpart, suggesting that smaller domain fluctuations exist in the ATP monomer (Saunders et al., 2014). The smaller fluctuations in ATP G-actin indicate that an entropic argument could be made for its faster polymerization than ADP G-actin. Since there are smaller domain movements, ATP G-actin is more likely to be in a conformation suitable for addition onto the F-actin monomer.

1.3 F-Actin

Filament polymerization is dependent upon the cellular concentration of G-actin, and occurs in a head-to-tail fashion leading to polar filaments with unique end geometry. The two ends of actin are referred to as the barbed or pointed end. The barbed end of the filament is where subdomains 1 and 3 are exposed, and the pointed end where subdomains 2 and 4 are exposed. Under physiological conditions, nucleation of new filaments is slow and begins with the formation of unstable dimers and trimers. However, once filaments are formed, polymerization advances swiftly (Pollard & Cooper, 2009), with growth primarily occurring at the barbed end through addition of ATP G-actin subunits (Figure 1.5). The addition of ATP G-actin onto the barbed end is diffusion limited, whereas the addition of ADP subunits is slower than diffusion limited (Pollard, 2017). At the pointed end of the filament, ADP protomers dissociate 5x faster than ATP protomers (Figure 1.5).

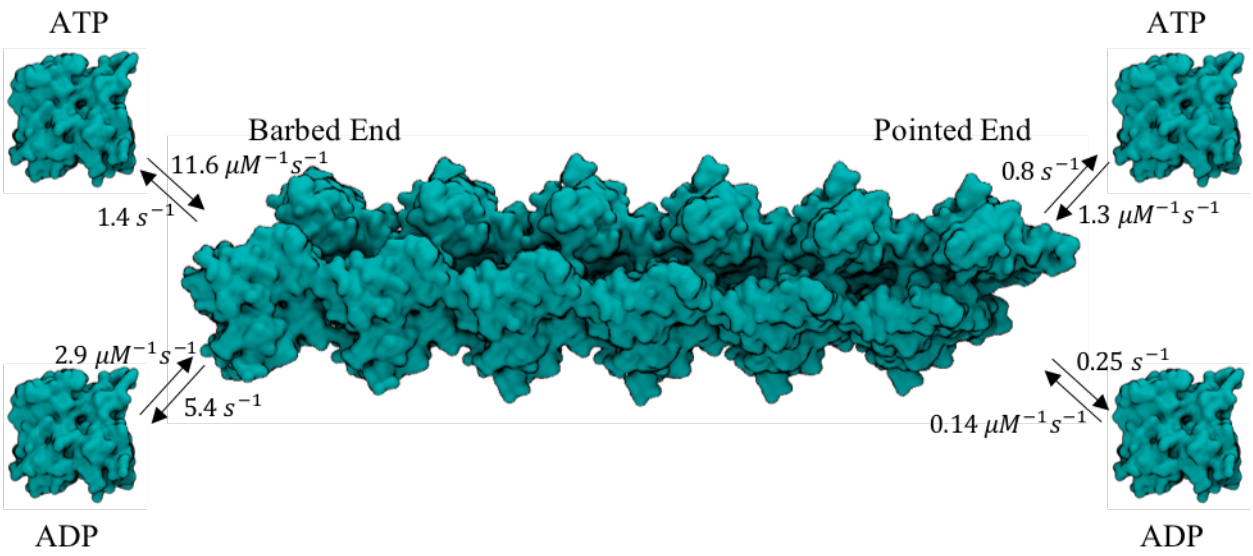


Figure 1.5 Filament Kinetics Polymerization and depolymerization of filament ends is dependent upon nucleotide state of incoming protomers (Pollard, 1986; Fujiwara et al., 2007; Pollard, 2017).

Shortly after ATP subunits get incorporated into the filament, ATP's γ-phosphate is cleaved and the nucleotide becomes ADP-P_i. Towards the pointed end of the filament, the phosphate gets released. This, paired with the fact that during filament treadmilling there is net addition of ATP subunits onto the barbed end of the filament and net dissociation of ADP protomers from the pointed end of the filament, leads to filaments with distinct patches of ATP, ADP-P_i, and ADP protomers. These patches serve to 'age' the filament. The unique conformations of each patch of protomers is key for cytoskeleton remodeling (Merino et al., 2018). The remodeling of the actin cytoskeleton is dependent upon actin binding proteins which cap, sever, and nucleate the formation of new filaments. For example, the Arp2/3 complex binds to the ATP filament, whereas ADF/cofilin prefer ADP F-actin interactions.

1.3.1 Filament Structure

The question still remains as to how the filament changes its structure upon nucleotide hydrolysis. To date, no high-resolution crystal structure exists for the F-actin filament. Researchers have had to rely on structures derived from x-ray fiber diffraction (Oda et al., 2009) and cryo-

electron microscopy (Fujii et al., 2010; Galkin, Orlova, Schröder, & Egelman, 2010; Galkin, Orlova, Vos, Schröder, & Egelman, 2015; Von Der Ecken et al., 2015; Merino et al., 2018). These techniques have been successful in giving us filament structures with resolutions in the 3.3-7 Å range but until recently they have all been solved in the ADP state. The newest set of filament models derived by Merino *et al.* (2018), give hints about the structural changes induced by nucleotide hydrolysis.

They developed six F-actin models derived from rabbit skeletal muscle, all in different nucleotide states (Merino et al., 2018). Again, as with the 'ATP' G-actin crystal structures, they used a non-hydrolyzable ATP analog AMP-PNP. Their models show, that helical symmetry is essentially unchanged upon nucleotide hydrolysis, with twist values of -166.63° in ATP F-actin and -166.58° in the ADP state. Protomer offset was also unchanged. As is expected, they noticed some small changes within the nucleotide binding cleft, largely consistent with what was observed in G-actin crystals. Larger changes were observed between nucleotide states at the interprotomer surface. Within the F-actin filament, the D-loop has been suggested to make lateral interprotomer contacts with the target-binding cleft of adjacent protomers. Merino *et al*. (2018) found that these contacts were nucleotide dependent, with the D-loop in the ATP protomer being more flexible.

1.3.2 End Dependent Filament Dynamics

As outlined above, the two ends of the filament are functionally different. During filament treadmilling, there is net addition of protomers onto the barbed end and net dissociation of protomers from the pointed end of the filament. The reasons for the drastic differences in kinetics between filaments ends is yet unknown. The higher-resolution structures that we discussed above are strictly for the middle of the filament. One structure 23Å structure has been solved for the pointed end, of the filament (Narita, Oda, & Maéda, 2011). This structure showed that the pointed

end protomers were tilted by 12° and blocked the addition of incoming actin protomers. Many of the F-actin simulations have been done using periodic boundary conditions, creating infinite filaments (Chu & Voth, 2005; Chu & Voth, 2006; Pfaendtner, Lyman, Pollard, & Voth, 2010; Saunders et al., 2014). In this thesis, I use molecular dynamics simulations of filaments without periodic boundary conditions to help elucidate the structural differences in the filament ends.

1.3.3 Monomer-to-Protomer Transition

Finally, of interest is the structural changes that occur to the actin subunit upon polymerization. Comparison of filament models with available G-actin crystals, shows us that the monomer-to-protomer transition is accompanied by a flattening of the actin subunit (Fujii et al., 2010; Oda & Maéda, 2010; Dominguez & Holmes, 2011). This leads me to the hypothesis that the ATP monomer is flatter than its ADP counterpart - providing a conformational argument for the differences in polymerization rates.

1.4 Actin Mutations and the Pathogenic Helix

To date, over 144 different disease-causing actin mutations have been identified, many of which are autosomal-dominant. The disease associated with an actin mutation is dependent upon which actin isoform they are present in. Humans have six different isoforms of actin which are tissue specific. In our striated muscle, we possess both α-muscle and α-cardiac actin. In the cytoplasm, we have both β- and γ-cytoplasmic actin. Finally, we also have both α- and γ-smooth muscle actin. Our cellular actin is highly conserved, deviating by at most 7% (α-muscle vs β-cytoplasmic), with most of the deviations occurring at the N-terminus of the protein (Figure 1.6).

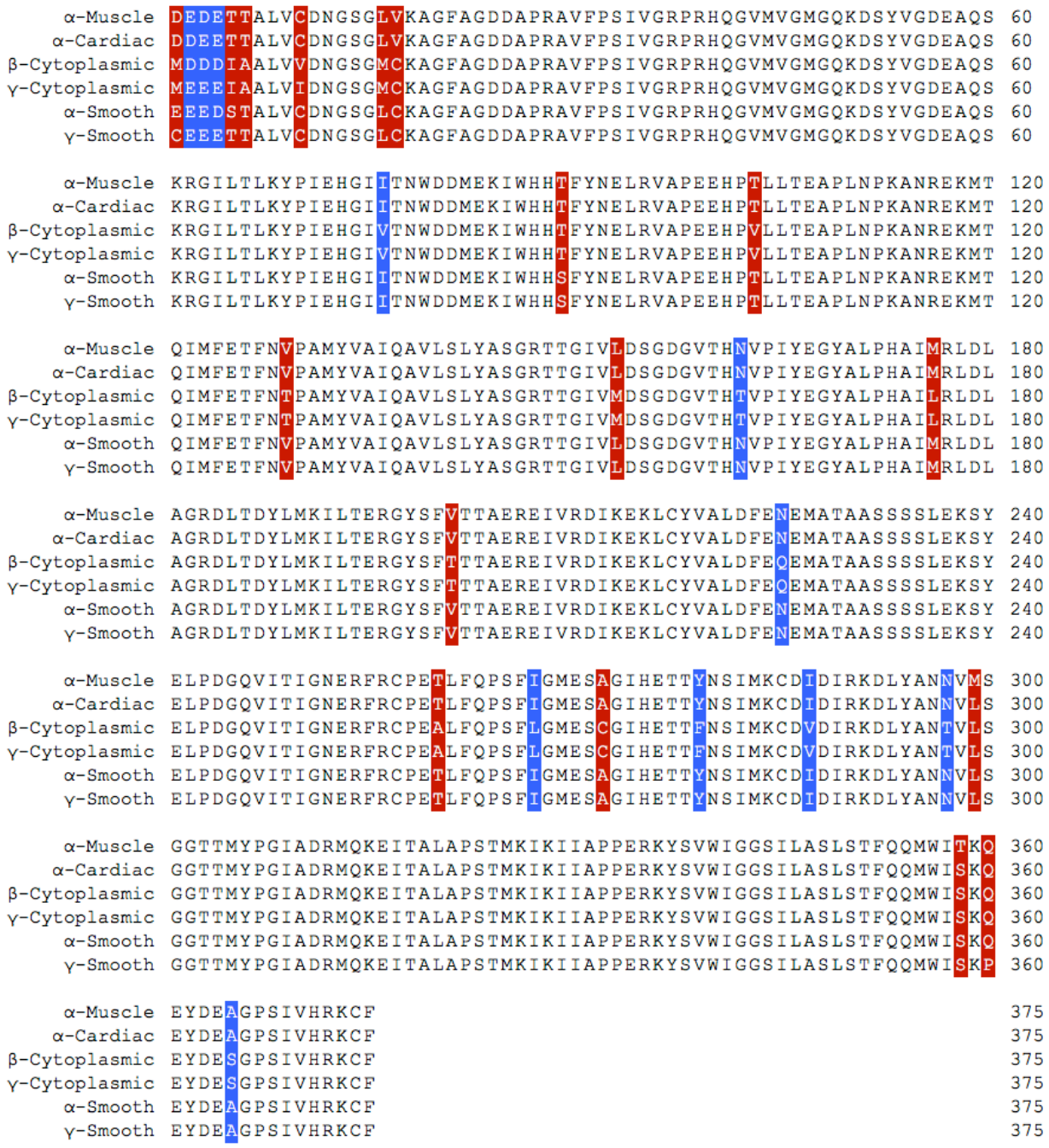


Figure 1.6 Sequence Alignment Sequence alignment of human actin isoforms reveals a high degree of sequence identity, with deviations occurring mostly near the N-terminus. Highlighted residues show divergence between isoforms with blue representing similar amino acids.

In each actin isoform, there exists the possibility for disease causing mutations. The disease caused is dependent upon which actin isoform contains the mutations. Mutations to α-skeletal muscle cause nemaline myopathies (Ilkovski et al., 2004). Mutations to α-cardiac muscle cause

both hypertrophic and dilated cardiomyopathies, as well as atrial septal defects (Olson, Michels, Thibodeau, Tai, & Keating, 1998; Vang et al., 2005; Matsson et al., 2008). Mutation to either cytoplasmic actin can lead to developmental malformations such as Baraitser-Winter Syndrome (Rivière et al., 2012; Johnston et al., 2013). γ-cytoplasmic actin mutations may also cause deafness (Zhu et al., 2003; Rendtorff et al., 2006; Morín & Bryan, F Goodyear, R Mencía, A Modamio-Høybjør, S del Castillo, I Cabalka, JM Richardson, G Moreno, F Rubenstein, PA Moreno-Pelayo, MA, 2009). Mutations to α- and γ-smooth muscle actins cause TAAD and visceral myopathies respectively (Guo et al., 2007; Guo et al., 2009; Wangler et al., 2014). Two differing actin isoforms can contain the same mutation but they may result in differing phenotypes. For example, the N115T mutation causes TAAD aneurysms if found in α-smooth muscle actin or nemaline myopathy if made in α-muscle actin (Guo et al., 2007; Rubenstein & Wen, 2014). This implies there may be a common structural etiology responsible for both phenotypes. The residue N115 lies within a helix called the pathogenic helix (residues 113-126). This helix is a hot-spot for actin mutations (Figure 1.7). If we examine all actin mutations reported in UniProtKB, nearly 10% of them map to this helix.

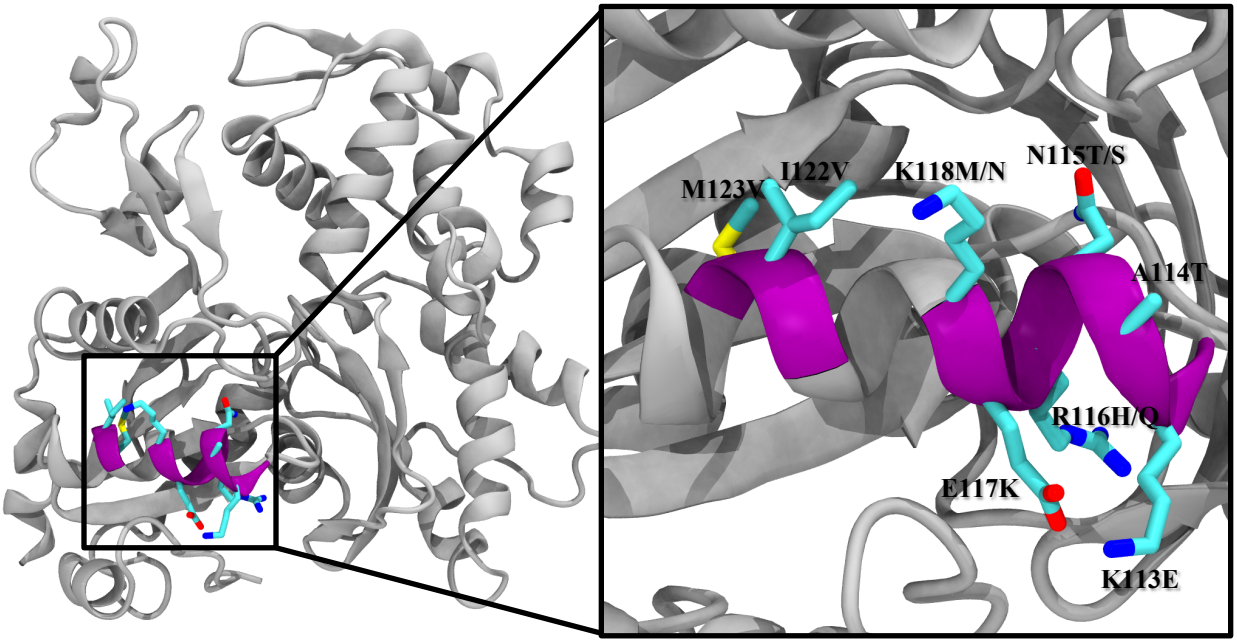


Figure 1.7 Actin's Pathogenic Helix Actin's pathogenic helix (residues 113-126) resides in the middle of subdomain 1, at the surface of the protein. Nearly every residue within the helix has an associated disease-causing mutation (see *Table 1.1*).

Table 1.1 Mutations to Actin's Pathogenic Helix

Residue	Disease	Actin Isoform	References
K113E	Nemaline Myopathy	α-skeletal muscle	Laing et al., 2009; Malloy et al., 2012
A114T	Nemaline Myopathy	α-skeletal muscle	
N115T/S	Nemaline Myopathy	α-skeletal muscle	Nowak et al., 1999; Ilkovski et al., 2001; Ilkovski et al., 2004
N115T	Aortic Aneurysm	α-smooth muscle	Guo et al., 2007; Guo et al., 2009
R116H	Nemaline Myopathy	α-skeletal muscle	
R116Q	Aortic Aneurysm	α-smooth muscle	Guo et al., 2007; Guo et al., 2009
E117K	Baraitser-Winter Syndrome	β-cytoplasmic	Johnston et al., 2013
K118M/N	DFNA20/26	γ-cytoplasmic	Zhu et al., 2003; Rendtorff et al., 2006; Morín & Bryan, 2009
T120I	Baraitser-Winter Syndrome	γ-cytoplasmic	Rivière et al., 2012
I122V	DFNA20/26	γ-cytoplasmic	Liu et al., 2008
M123V	Atrial Septal Defects	α-cardiac muscle	Matsson et al., 2008

The pathogenic helix has been shown to play an important role in longitudinal protomer dynamics, as well as sits at the surface for actin binding proteins. It seems that actinopathies

associated with this helix result from disruption of either polymerization, filament stability, or from altered interaction with regulatory actin-binding proteins. The helix's N-terminal residue, K113, is proposed to make and interstrand contacts with E195. The K113E mutation, found in α-muscle actin to cause nemaline myopathy, leads to polymerization defects (Malloy et al., 2012; Rubenstein & Wen, 2014). Function can be recovered with by a compensatory mutation, E195K – showing the importance of this helix in filament stability. The pathogenic helix is at the binding surface for Arp2/3 complex (Goley et al., 2010). The K118M mutation, for example, has been shown to disrupt Arp2/3 complex branch formation (Kruth & Rubenstein, 2012). We believe that the cluster of seemingly unrelated diseases caused by mutations to this helix, may actually be variations of the same disease resulting from similar alterations to actin structure but expressed in the context of different cells with different functions.

The K118M/N mutations provides an interesting opportunity to study this hypothesis. Methionine and asparagine are two dissimilar amino acids, the first is similar in size to lysine but hydrophobic in nature and the second is smaller but still hydrophilic. The two mutations lead to the same disease phenotype, but seem to disrupt the cells in different ways (Kruth & Rubenstein, 2012). Residue K118 resides at the binding surface for Arp2/3 complex (Goley et al., 2010; Kruth & Rubenstein, 2012), and unsurprisingly both K118M and K118N actin displayed altered behavior when exposed to Arp2/3 (Kruth & Rubenstein, 2012). K118N actin also polymerized at twice the rate of WT actin. Thus, mutations to this residue provide a good opportunity to examine how mutations to the pathogenic helix alter actin structure in such a way as to alter both polymerization and the interaction with actin binding proteins.

1.5 Thesis Outline and Goals

The goal of my thesis is to gain a thorough understanding of how nucleotide state, polymerization, and changes to actin's sequence alter its structure and dynamics. I accomplish this through the use of multiple molecular dynamics simulations. The following details the specific goals and questions addressed in this body of work.

1.5.1 Effects of Nucleotide on G-Actin Structure and Dynamics

My first goal was to gain an understanding of the effects of nucleotide on G-actin. As outlined above, actin's nucleotide state is key in determining its affinity for different actin binding proteins as well as for polymerization; with the nucleotide state determining the architecture of the nucleotide-binding cleft, W-loop and potentially the D-loop. The questions then become: *What is the structural basis for the differences in affinity for actin binding proteins and for polymerization? How do changes within the nucleotide binding cleft get propagated throughout the protein resulting in these changes?* To answer these questions, I use extensive, multi-microsecond molecular dynamics simulations of G-actin in both the ATP and ADP states. The results of this work are presented in Chapter 3.

1.5.2 Effects of Polymerization on Actin Dynamics

Within the actin community there are many unanswered questions related to filament structure and dynamics (Pollard, 2017) such as: *Why does ATP G-actin polymerize so much faster than its ADP counterpart? What are the structures of the filament ends that lead to their unique properties - namely why does the pointed end of the filament polymerize so much slower than the barbed end?* To get at these questions we again turn to molecular dynamics simulations. I performed extended simulations of the Oda *et al.* and Fujii *et al.* filament models without periodic boundary conditions to allow for freely fluctuating ends. This analysis is presented in Chapter 4.

1.5.3 Comparison of the Structure and Dynamics of Yeast vs Muscle Actin

Yeast has only one actin isoform with 86.7% sequence identity with α-muscle actin, making it an ideal model system. The high degree of sequence similarity between the two actins should equate to similar structure and dynamics. Knowledge of the similarities and differences between yeast and muscle actin will increase its usefulness as a model system for the *in vitro* study of disease causing actin mutations. Results of this analysis are presented in Chapter 5.

1.5.4 Effects of Mutations to Actin's Pathogenic Helix

Nearly 10% of the mutations identified in actin cluster on the pathogenic helix suggesting that diseases caused by these mutations have common structural etiologies. Simulation of actin mutants will be applicable not only to the disease they are associated with, but will also provide further knowledge of how actin's sequence effects its dynamics. To this end, we partnered with collaborators to evaluate the effects of two dissimilar mutations to residue K118 which cause deafness (K118M/N). This work has been published (Jepsen, Kruth, Rubenstein, & Sept, 2016) and is presented in Chapter 5.

1.6 References

Belmont, L. D., Orlova, A., Drubin, D. G., & Egelman, E. H. (1999). A change in actin conformation associated with filament instability after P(i) release. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(1), 29-34.

Carlier, M., Jean, C., Rieger, K. J., Lenfant, M., & Pantaloni, D. (1993). Modulation of the interaction between G-actin and thymosin β4 by the ATP/ADP ratio: Possible implication in the regulation of actin dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(11), 5034-5038.

Carlier, M. (2010). *Actin-based motility: Cellular, molecular and physical aspects*.

Chereau, D., Kerff, F., Graceffa, P., Grabarek, Z., Langsetmo, K., & Dominguez, R. (2005). Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and

the implications for filament assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16644-16649.

Chik, J. K., Lindberg, U., & Schutt, C. E. (1996). The structure of an open state of beta-actin at 2.65 Å resolution. *J. Mol. Biol.*, *263*(4), 607-623.

Chu, J., & Voth, G. A. (2005). Allostery of actin filaments: Molecular dynamics simulations and coarse-grained analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(37), 13111-13116.

Chu, J., & Voth, G. A. (2006). Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophysical journal*, *90*(5), 1572-1582.

Dalhaimer, P., Pollard, T. D., & Nolen, B. J. (2008). Nucleotide-Mediated Conformational Changes of Monomeric Actin and Arp3 Studied by Molecular Dynamics Simulations. *J. Mol. Biol.*, *376*(1), 166-183.

Dominguez, R., & Holmes, K. C. (2011). Actin structure and function. *Annu. Rev. Biophys.*, *40*, 169-168.

Dominguez, R. (2016). The WH2 Domain and Actin Nucleation: Necessary but Insufficient. *Trends in biochemical sciences*, *41*(6), 478-490.

dos Remedios, C. G., Chhabra, D., Kekic, M., Dedova, I. V., Tsubakihara, M., Berry, D. A. et al. (2003). Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol. Rev.*, *82*(2), 433-473.

Durer, Z. A., Kudryashov, D. S., Sawaya, M. R., Altenbach, C., Hubbell, W., & Reisler, E. (2012). Structural states and dynamics of the D-loop in actin. *Biophys J.*, *103*(5), 930-939.

Fujii, T., Iwane, A. H., Yanagida, T., & Namba, K. (2010). Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, *467*(7316), 724-728.

Fujiwara, I., Vavylonis, D., & Pollard, T. D. (2007). Polymerization kinetics of ADP- and ADP-Pi-actin determined by fluorescence microscopy. *Proc. Natl. Acad. Sci. USA*, *104*(21), 8827-8832.

Galkin, V. E., Orlova, A., Schröder, G. F., & Egelman, E. H. (2010). Structural polymorphism in F-actin. *Nature Structural and Molecular Biology*, *17*(11), 1318-1323.

Galkin, V. E., Orlova, A., Vos, M. R., Schröder, G. F., & Egelman, E. H. (2015). Near-Atomic Resolution for One State of F-Actin. *Structure*, *23*(1), 173-182.

Goley, E. D., Rammohan, A., Znameroski, E. A., Firat-Karalar, E. N., Sept, D., & Welch, M. D. (2010). An actin-filament-binding interface on the Arp2/3 complex is critical for nucleation and branch stability. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(18), 8159-8164.

Graceffa, P., & Dominguez, R. (2003). Crystal structure of monomeric actin in the ATP state. Structural basis of nucleotide-dependent actin dynamics. *J. Biol. Chem.*, *278*, 34172-34180.

Guo, D. C., Pannu, H., Tran-Fadulu, V., Papke, C. L., Yu, R. K., Avidan, N. et al. (2007). Mutations in smooth muscle alpha-actin (ACTA2) lead to thoracic aortic aneurysms and dissections. *Nature Genetics*, *39*(12), 1488-1493.

Guo, D., Papke, C. L., Tran-Fadulu, V., Regalado, E. S., Avidan, N., Johnson, R. J. et al. (2009). Mutations in Smooth Muscle Alpha-Actin (ACTA2) Cause Coronary Artery Disease, Stroke, and Moyamoya Disease, Along with Thoracic Aortic Disease. *American Journal of Human Genetics*, *84*(5), 617-627.

Hild, G., Bugyi, B., & Nyitrai, M. (2010). Conformational dynamics of actin: Effectors and implications for biological function. *Cytoskeleton*, *67*(10), 609-629.

Holmes, K. C., Popp, D., Gebhard, W., & Kabsch, W. (1990). Atomic model of the actin filament. *Nature*, *347*(6288), 44-49.

Ilkovski, B., Cooper, S. T., Nowak, K., Ryan, M. M., Yang, N., Schnell, C. et al. (2001). Nemaline myopathy caused by mutations in the muscle α-skeletal-actin gene. *American Journal of Human Genetics*, *68*(6), 1333-1343.

Ilkovski, B., Nowak, K. J., Domazetovska, A., Maxwell, A. L., Clement, S., Davies, K. E. et al. (2004). Evidence for a dominant-negative effect in ACTA1 nemaline myopathy caused by abnormal folding, aggregation and altered polymerization of mutant actin isoforms. *Human molecular genetics*, *13*(16), 1727-1743.

Jepsen, L., Kruth, K. A., Rubenstein, P. A., & Sept, D. (2016). Two Deafness-Causing Actin Mutations (DFNA20/26) Have Allosteric Effects on the Actin Structure. *Biophysical journal*, *111*(2), 323-332.

Johnston, J. J., Wen, K. K., Keppler-Noreuil, K., McKane, M., Maiers, J. L., Greiner, A. et al. (2013). Functional analysis of a de novo ACTB mutation in a patient with atypical Baraitser-Winter syndrome. *Human Mutation*, *34*(9), 1242-1249.

Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., & Holmes, K. C. (1990). Atomic structure of the actin: DNase I complex. *Nature*, *347*(6288), 37-44.

Kalhor, H. R., Niewmierzycka, A., Faull, K. F., Yao, X., Grade, S., Clarke, S. et al. (1999). A highly conserved 3-methylhistidine modification is absent in yeast actin. *Archives of Biochemistry and Biophysics*, *370*(1), 105-111.

Kim, E., Wriggers, W., Phillips, M., Kokabi, K., Rubenstein, P. A., & Reisler, E. (2000). Cross-linking constraints on F-actin structure. *Journal of Molecular Biology*, *299*(2), 421-429.

Kruth, K. A., & Rubenstein, P. A. (2012). Two deafness-causing (DFNA20/26) actin mutations affect Arp2/3-dependent actin regulation. *Journal of Biological Chemistry*, *287*(32), 27217-27226.

Kudryashov, D. S., Grintsevich, E. E., Rubenstein, P. A., & Reisler, E. (2010). A nucleotide state-sensing region on actin. *J. Biol. Chem.*, *285*(33), 25591-25601.

Kudryashov, D. S., & Reisler, E. (2013). ATP and ADP actin states. *Biopolymers*, *99*(4), 245-256.

Laing, N. G., Dye, D. E., Wallgren-Pettersson, C., Richard, G., Monnier, N., Lillis, S. et al. (2009). Mutations and polymorphisms of the skeletal muscle α-actin gene (ACTA1). *Human mutation*, *30*(9), 1267-1277.

Liu, P., Li, H., Ren, X., Mao, H., Zhu, Q., Zhu, Z. et al. (2008). Novel ACTG1 mutation causing autosomal dominant non-syndromic hearing impairment in a Chinese family. *Journal of Genetics and Genomics*, *35*(9), 553-558.

Malloy, L. E., Wen, K., Pierick, A. R., Wedemeyer, E. W., Bergeron, S. E., Vanderpool, N. D. et al. (2012). Thoracic Aortic Aneurysm (TAAD)-causing Mutation in Actin Affects Formin Regulation of Polymerization. *Journal of Biological Chemistry*, *287*(34), 28398-28408.

Marchand, J., Kaiser, D. A., Pollard, T. D., & Higgs, H. N. (2001). Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nature cell biology*, *3*(1), 76-82.

Matsson, H., Eason, J., Bookwalter, C. S., Klar, J., Gustavsson, P., Sunnegårdh, J. et al. (2008). Alpha-cardiac actin mutations produce atrial septal defects. *Human molecular genetics*, *17*(2), 256-265.

Mattila, P. K., Salminen, M., Yamashiro, T., & Lappalainen, P. (2003). Mouse MIM, a tissue-specific regulator of cytoskeletal dynamics, interacts with ATP-actin monomers through its C-terminal WH2 domain. *Journal of Biological Chemistry*, *278*(10), 8452-8459.

Merino, F., Pospich, S., Funk, J., Wagner, T., Küllmer, F., Arndt, H. et al. (2018). Structural transitions of F-actin upon ATP hydrolysis at near-atomic resolution revealed by cryo-EM. *Nature Structural and Molecular Biology*, *25*(6), 528-537.

Morín, M., & Bryan, K. E. M.-M., F Goodyear, R Mencía, A Modamio-Høybjør, S del Castillo, I Cabalka, JM Richardson, G Moreno, F Rubenstein, PA Moreno-Pelayo, MA. (2009). In vivo and in vitro effects of two novel gamma-actin (ACTG1) mutations that cause DFNA20/26 hearing impairment. *Human Molecular Genetics*, *18*(16), 3075-3089.

Mullins, R. D., Heuser, J. A., & Pollard, T. D. (1998). The interaction of Arp2/3 complex with actin: Nucleation, high affinity pointed end capping, and formation of branching networks of filaments. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(11), 6181-6186.

Narita, A., Oda, T., & Maéda, Y. (2011). Structural basis for the slow dynamics of the actin filament pointed end. *EMBO Journal*, *30*(7), 1230-1237.

Nowak, K. J., Wattanasirichaigoon, D., Goebel, H. H., Wilce, M., Pelin, K., Donner, K. et al. (1999). Mutations in the skeletal muscle α-actin gene in patients with actin myopathy and nemaline myopathy. *Nature genetics*, *23*(2), 208-212.

Oda, T., & Maéda, Y. (2010). Multiple Conformations of F-actin. *Structure*, *18*(7), 761-767.

Oda, T., Iwasa, M., Aihara, T., Maéda, Y., & Narita, A. (2009). The nature of the globular- to fibrous-actin transition. *Nature*, *457*(7228), 441-445.

Olson, T. M., Michels, V. V., Thibodeau, S. N., Tai, Y., & Keating, M. T. (1998). Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. *Science*, *280*(5364), 750-752.

Otterbein, L. R., Graceffa, P., & Dominguez, R. (2001). The crystal structure of uncomplexed actin in the ADP state. *Science*, *293*, 708-711.

Oztug Durer, Z. A., Diraviyam, K., Sept, D., Kudryashov, D. S., & Reisler, E. (2010). F-Actin Structure Destabilization and DNase I Binding Loop Fluctuations. Mutational Cross-Linking and Electron Microscopy Analysis of Loop States and Effects on F-Actin. *Journal of Molecular Biology*, *395*(3), 544-557.

Pfaendtner, J., Lyman, E., Pollard, T. D., & Voth, G. A. (2010). Structure and dynamics of the actin filament. *Journal of Molecular Biology*, *396*(2), 252-263.

Pollard, T. D. (1986). Rate constants for the reactions of ATP- and ADP-actin with the ends of actin filaments. *J. Cell Biol.*, *103*, 2747-2754.

Pollard, T. D. (2017). *What we know and do not know about actin* (235).

Pollard, T. D., Blanchoin, L., & Mullins, R. D. (2001). Actin dynamics. *Journal of cell science*, *114*(1), 3.

Pollard, T. D., & Cooper, J. A. (2009). Actin, a central player in cell shape and movement. *Science*, *326*(5957), 1208-1212.

Pollitt, A. Y., & Insall, R. H. (2009). WASP and SCAR/WAVE proteins: The drivers of actin assembly. *Journal of cell science*, *122*(15), 2575-2578.

Rendtorff, N. D., Zhu, M., Fagerheim, T., Antal, T. L., Jones, M. P., Teslovich, T. M. et al. (2006). A novel missense mutation in ACTG1 causes dominant deafness in a Norwegian DFNA20/26 family, but ACTG1 mutations are not frequent among families with hereditary hearing impairment. *European Journal of Human Genetics*, *14*(10), 1097-1105.

Rivière, J., Van Bon, B. W. M., Hoischen, A., Kholmanskikh, S. S., O'Roak, B. J., Gilissen, C. et al. (2012). De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. *Nature genetics*, *44*(4), 440-444.

Rould, M. A., Wan, Q., Joel, P. B., Lowey, S., & Trybus, K. M. (2006). Crystal structures of expressed non-polymerizable monomeric actin in the ADP and ATP states. *Journal of Biological Chemistry*, *281*(42), 31909-31919.

Rubenstein, P. A., & Wen, K. K. (2014). Insights into the effects of disease-causing mutations in human actins. *Cytoskeleton (Hoboken)*, *71*(4), 211-229.

Saunders, M. G., Tempkin, J., Weare, J., Dinner, A. R., Roux, B., & Voth, G. A. (2014). Nucleotide regulation of the structure and dynamics of G-actin. *Biophysical journal*, *106*(8), 1710-1720.

Splettstoesser, T., Noé, F., Oda, T., & Smith, J. C. (2009). Nucleotide-dependence of G-actin conformation from multiple molecular dynamics simulations and observation of a putatively polymerization-competent superclosed state. *Proteins*, *76*(2), 353-364.

Strzelecka-Gołaszewska, H., Moraczewska, J., Khaitlina, S. Y., & Mossakowska, M. (1993). Localization of the tightly bound divalent-cation-dependent and nucleotide-dependent conformation changes in G-actin using limited proteolytic digestion. *Eur. J. Biochem.*, *211*(3), 731-742.

Svitkina, T. (2018). The actin cytoskeleton and actin-based motility. *Cold Spring Harbor Perspectives in Biology*, *10*(1).

Vang, S., Corydon, T. J., Børglum, A. D., Scott, M. D., Frydman, J., Mogensen, J. et al. (2005). Actin mutations in hypertrophic and dilated cardiomyopathy cause inefficient protein folding and perturbed filament formation. *FEBS Journal*, *272*(8), 2037-2049.

Vetter, I. R., & Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. *Science*, *294*(5545), 1299-1304.

Von Der Ecken, J., Müller, M., Lehman, W., Manstein, D. J., Penczek, P. A., & Raunser, S. (2015). Structure of the F-actin-tropomyosin complex. *Nature*, *519*(7541), 114-117.

Wangler, M. F., Gonzaga-Jauregui, C., Gambin, T., Penney, S., Moss, T., Chopra, A. et al. (2014). Heterozygous De Novo and Inherited Mutations in the Smooth Muscle Actin (ACTG2) Gene Underlie Megacystis-Microcolon-Intestinal Hypoperistalsis Syndrome. *PLoS Genetics*, *10*(3).

Yao, X., Grade, S., Wriggers, W., & Rubenstein, P. A. (1999). His73, often methylated, is an important structural determinant for actin. A mutagenic analysis of His73 of yeast actin. *Journal of Biological Chemistry*, *274*(52), 37443-37449.

Zheng, X., Diraviyam, K., & Sept, D. (2007). Nucleotide effects on the structure and dynamics of actin. *Biophysical Journal*, *93*(4), 1277-1283.

Zhu, M., Yang, T., Wei, S., DeWan, A. T., Morell, R. J., Elfenbein, J. L. et al. (2003). Mutations in the γ-Actin Gene (ACTG1) Are Associated with Dominant Progressive Deafness (DFNA20/26). *American Journal of Human Genetics*, *73*(5), 1082-1091.

Chapter 2

Methods

In this chapter, I present the computational methods that are used throughout my work. The biochemical methods presented in Chapter 5, performed by our collaborators Karina A. Kurth and Peter A. Rubenstein, can be found in our published paper (Jepsen, Kruth, Rubenstein, & Sept, 2016). All analysis was done in R, and VMD. Sample functions and analysis scripts may be found in the Appendix.

2.1 Molecular Dynamics Simulations

2.1.1 ADP and ATP G-Actin

The PDB structure for ADP G-actin was taken from 1J6Z (Otterbein, Graceffa, & Dominguez, 2001) and the structure for ATP G-actin was taken from 1NWK (Graceffa & Dominguez, 2003). The backbones of the missing residues were filled in using VMD and missing atoms were filled in using AmberTools (Humphrey, Dalke, & Schulten, 1996; Case et al., 2016).

TMR residues were removed and acetylation of the N-terminus, as well as methylation of H73 were added using VMD. The high-affinity cation was set to Mg^{2+}.

2.1.2 Building Filament Models

Two filaments models were built, one each from the PDB structures 3MFP (Fujii, Iwane, Yanagida, & Namba, 2010) and 2ZWH (Oda, Iwasa, Aihara, Maéda, & Narita, 2009). Each was prepped by adding an N-terminal acetylation and methylated H73. The nucleotide in each was set to be ADP with Mg^{2+}. Within VMD, filaments were built by generating 8 identical protomers using the translation matrix provided with the PDB structure. Finally, ATP filaments were built by taking the last frame from our conventional molecular dynamics simulation from our Oda model (2ZWH) and converting ADP to ATP by adding the γ-phosphate. To obtain proper geometry, the α- and β -phosphates in ADP were fit to ATP taken from our G-actin simulation. When the best fit (measured by RMSD) was obtained, the γ-phosphate from the ATP was appended onto the ADP molecule.

2.1.3 Yeast G-Actin and the K118M/N Mutants

The PDB structure for yeast G-actin was taken from 1YAG (Vorobiev et al., 2003). Missing atoms and/or residues were completed using VMD Humphrey et al., 1996) and the nucleotide was set to be ADP with Mg^{2+}. Models for the K118M/N mutants were created from this structure again using VMD and AmberTools (Humphrey et al., 1996; Case et al., 2016).

2.1.4 Simulation

Preparation of all actin systems were done using AmberTools (Case et al., 2016). Each system was solvated using TIP3 water boxes with 10 Å padding and Na^+ and Cl^- added to both neutralize the system charge and set the ionic strength to 50 mM. Molecular dynamics (MD) simulations were carried out using NAMD (Phillips et al., 2005). Following heating and

equilibration steps, we performed simulations using the AMBER99SB forcefield (Hornak et al., 2006) at 300 K in an NpT ensemble with 1 atm pressure. Bonded hydrogens were fixed to allow us to use 2 fs time steps. We employed Particle Mesh Ewald (Essmann et al., 1995) for long-range electrostatics and used a 10 Å cut-off and 8.5 Å switch distance for van der Waals interactions. It should be noted that the force filed parameters for both nucleotides were obtained from the AMBER parameter database (http://research.bmh.manchester.ac.uk/bryce/amber/) and were parameterized by Heather Carlson with the College of Pharmacy at the University of Michigan (Meagher, Redman, & Carlson, 2003).

Molecular dynamics simulations were carried out in several stages. First, an initial 10 ns conventional MD (cMD) simulation was used as a further equilibration step. This simulation was not analyzed, nor was the simulation time included in our final count. Next, to enhance the sampling of conformational space and remove potential system artifacts we used accelerated MD (aMD) (Wang, Harrison, Schulten, & McCammon, 2011). aMD is a technique that applies a boost potential to the dihedrals of the system. To calculate an appropriate boost potential, we used the following equations (Pierce, Salomon-Ferrer, Augusto F. De Oliveira, McCammon, & Walker, 2012):

$$E_{dihedrals} = V_{0,dihedrals} + 3\frac{kcal}{mol * residues} * X\ residues$$

$$\alpha_{dihedrals} = \frac{E_{dihedrals}}{5}$$

In this equation, $V_{0,dihedrals}$ is the average potential energy for dihedrals taken from the initial 10 ns cMD simulation, and X is the number of residues in the system. The simulations were run for varying lengths of times (see Table 2.1) from 100 ns to 250 ns. All aMD G-actin simulations were run for approximately 100 ns before new starting velocities were generated and simulations were

repeated. We found that while our aMD simulations provided good sampling of conformational spaces and accurate tertiary structures, it was not as good for things such as secondary structure or hydrogen bonds. It is for this reason that we turned to cMD, for the final simulation phase, to thoroughly sample equilibrium conformations. All analysis, except initial principal component analysis, was performed on our cMD simulations unless specifically stated. With the exception of our F-actin simulations which were started from the end of the aMD simulations, starting points for the cMD simulations were taken from the equilibrium conformations in our aMD simulations as determined by PCA. The only deviation from this protocol was for our WT yeast G-actin simulation. This was the first simulation run and instead of running cMD simulations after our aMD, it was run first.

Table 2.1 Simulation Details In all cases but WT yeast G-actin (*), aMD simulations were run and then followed by cMD.

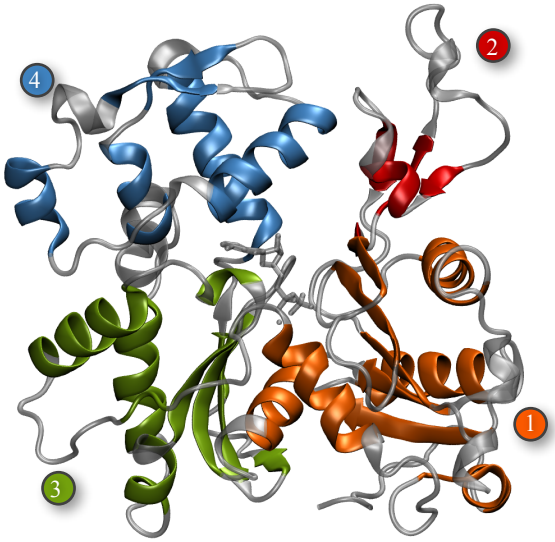
System	Starting PDB	aMD Time	cMD Time	Total Time
ADP G-actin	1J6Z	1.0 μs	1.55 μs	2.55 μs
ATP G-Actin	1NWK	0.50 μs	1.663 μs	2.163 μs
ADP F-actin	2ZWH	0.25 μs	0.40 μs	0.65 μs
ADP F-actin	3MFP	0.25 μs	0.30 μs	0.55 μs
ATP F-actin	2ZWH	0.04 μs	0 μs	0.04 μs
WT Yeast G-actin (ADP)	1YAG	0.98 μs *	0.30 μs	1.28 μs
K118M	1YAG	0.1 μs	0.05 μs	0.15 μs
K118N	1YAG	0.1 μs	0.05 μs	0.15 μs
				7.533 μs

2.2 Secondary Structure

Secondary structure was calculated using DSSP (Joosten et al., 2011; Kabsch & Sander, 1983). DSSP calculates the secondary structure of the protein based on its 3D structure, assigning each residue to one of eight categories representing various helical structures, β-bridges and ladders, turns, bends, and loops/irregular elements. For our purpose, we combined these into three

different categories, helical structures, β-structures, and other. The other category includes, turns, loops, and irregular elements.

Secondary structure was used to define the core of the protein. For consistency, this was done once using our aMD simulation of ADP G-actin. The core of the protein was defined as residues that were part of either helical or β-structures for at least 80% of the frames of our simulation. Total, 185 residues fit this description (Figure 2.1).



Subdomain 1	8-11, 16-21, 29-31, 79-91, 103-106, 113-124, 131-135, 137-144, 338-347, 357-364
Subdomain 2	32, 35-37, 53-54, 56-59, 66-68
Subdomain 3	142-144, 150-154, 161-166, 169-170, 176-177, 274-283, 290-294, 297-300, 303-304, 309-320, 329-330
Subdomain 4	182-195, 203-215, 223-230, 238-241, 247-250, 253-256, 258-259

Figure 2.1 Subdomain Cores The cores of the subdomains are defined as residues that are part of either helical or β-structures for at least 80% of the frames of the aMD simulation of ADP G-actin.

2.3 Principal Component Analysis

Principal component analysis (PCA) was used in two different ways using the bio3D package in R (Grant, Rodrigues, ElSawy, McCammon, & Caves, 2006; Skjaerven, Yao, Scarabelli, & Grant, 2014; http://www.r-project.org). First, to determine the large-scale motions of our simulations, we built a PCA basis set using the C_α of our aMD ADP G-actin simulation that were part of secondary structure elements (helices or β-sheets). These residues were defined as the core of our protein (Figure 2.1). Other simulations were then fit to the PCA mean and projected onto this basis set. By limiting our analysis to these residues, we ensured that our covariance matrix

was not dominated by the motion of the highly mobile loop regions, allowing us to get at the core dynamics of the protein.

Next, to look at small scale motions within our simulation, we did PCA on loops and other structural elements. Fitting was performed as described above but then a combined trajectory was created using the cMD simulations of two or more structural states. A new combined PCA basis set was then created by performing PCA on this combined fitted trajectory, but only residues within a loop region were used for analysis. The combined basis set maximizes the differences between the states. Further, by fitting to the core of the protein but then limiting analysis to small loop regions, we can see how these loops move with respect to the protein as a whole.

2.4 Projection of PDB Files onto PCA Space

Projection of actin and actin related structures onto our PCA space is non-trivial. First, PDBs of actin and related structures needed to be identified. This was done using the bio3d webserver (Skjærven, Jariwala, Yao, & Grant, 2016; Jariwala, Skjærven, Yao, & Grant, 2017) searching against the PDB 1J6Z (Otterbein et al., 2001). All hits with a BitScore of 665 or less were taken on for further analysis. Next, since our PCA space was built using 185 residues of the protein core, projection of PDBs onto this space needed to have those residues. In R using bio3d (Grant et al., 2006; Skjaerven et al., 2014; http://www.r-project.org), each PDB file from our initial search was downloaded and aligned to our reference structure using muscle (Edgar, 2004). Only structures that contained all 185 necessary residues were fit to, and projected on our PCA basis set.

2.5 Propeller-Twist, Phosphate Clamp, and Nucleotide Cleft Mouth

Actin's two dominant rotations are a propeller-twist and scissor-like motion (Dalhaimer, Pollard, & Nolen, 2008; Dominguez & Holmes, 2011). We measured the propeller-twist as the dihedral angle between the cores of subdomains 2, 1, 3, and 4 (Figure 2.2). The scissor like motion was measured in using the phosphate clamp size, or the C_α separation of G15 and D157, and the nucleotide cleft mouth (C_α positions of Q59 and E207).

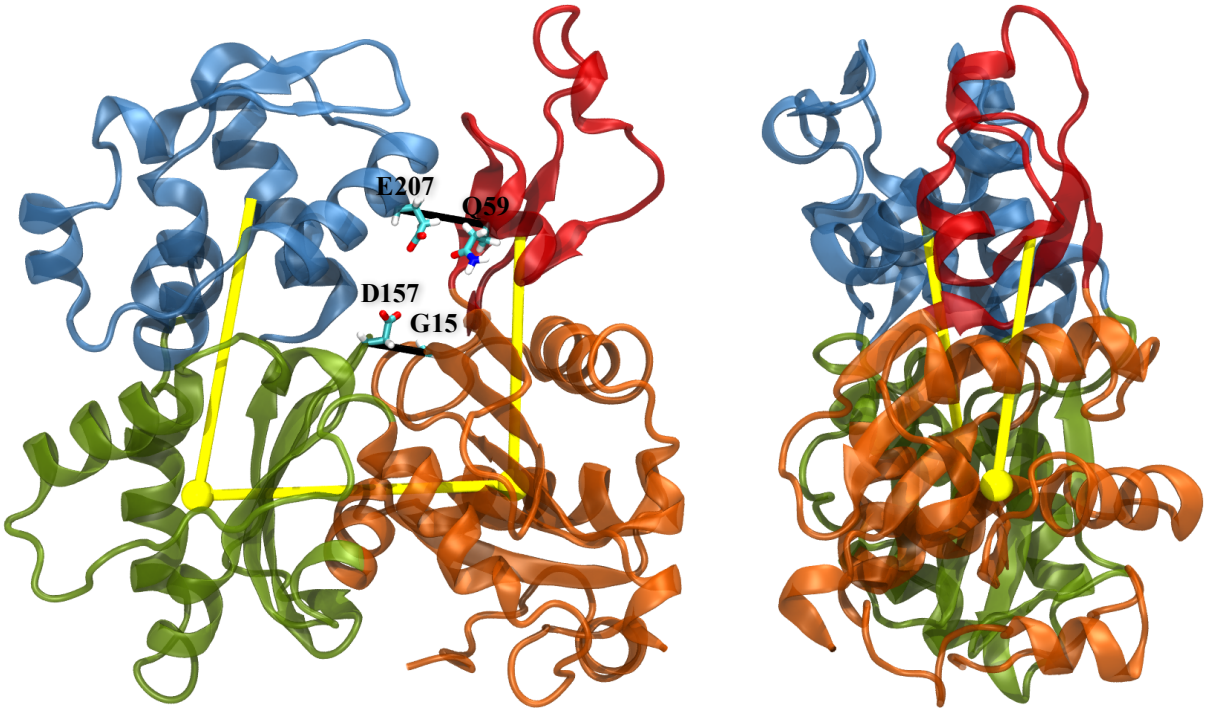


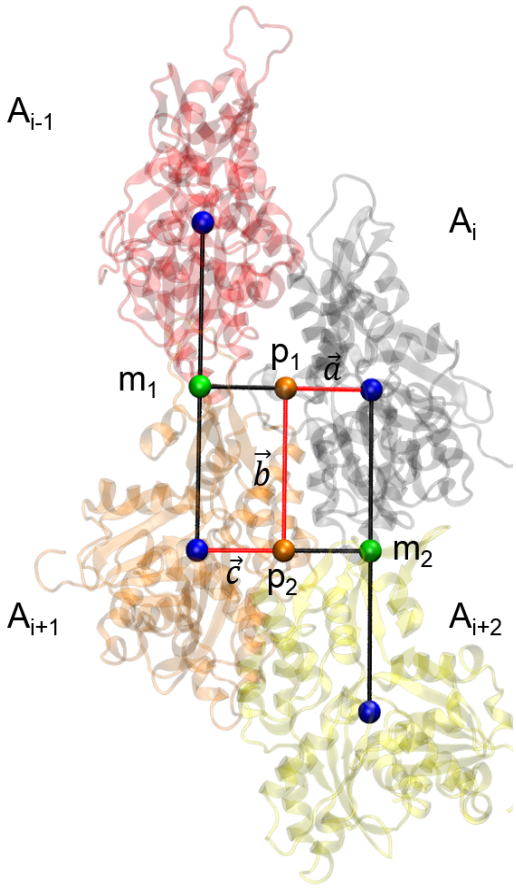
Figure 2.2 Propeller-twist, phosphate clamp, and nucleotide cleft mouth size

2.6 Hydrogen Bonds and Salt Bridges

Hydrogen bonds and salt-bridges within actin systems, were calculated in VMD using their hydrogen bonds and salt bridges plugins (Humphrey et al., 1996). Hydrogen bonds were defined using only polar atoms (N, O, S, F) with a distance cutoff of 3.2 Å and an angle cutoff of 20°. For salt bridge measurement, we set the oxygen-nitrogen cutoff distance to 3.5 Å.

2.7 Filament Twist and Offset

Previously, to measure filament twist and offset between two protomers (A_i and A_{i+1}), the centers of mass, and distances between groupings of two consecutive trimers (A_{i-1}, A_i, A_{i+1}; A_i, A_{i+1}, A_{i+2}) were necessary (Figure 2.3). This method leads to poor fit between protomers and the measurement of both twist and offset are biased by the positioning of protomers A_{i-1} and A_{i+2}.



$$\vec{a} = \overrightarrow{p_1 A_\iota}$$
$$\vec{b} = \overrightarrow{p_2 p_1}$$
$$\vec{c} = \overrightarrow{A_{\iota+1} p_2}$$

$$twist\ \theta = \tan^{-1}\left(\frac{|\vec{b}|\vec{a}\cdot(\vec{b}\times\vec{c})}{-|\vec{b}|^2\vec{a}\cdot\vec{c} + (\vec{a}\cdot\vec{b})(\vec{b}\cdot\vec{c})}\right)$$

$$offset = |\overrightarrow{p_1 p_2}|$$

Figure 2.3 Filament Twist and Offset – Old Method To measure the twist and offset between two consecutive protomers A_i and A_{i+1}, two consecutive trimers were necessary. Here the blue dots represent the protomer centers of mass, and the green dots represent midpoints between longitudinal protomers. To measure twist and offset, the vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ (in red) are used with the equations above.

To overcome these short comings, I developed an unbiased measure of the filament axis and protomer twist using a modified Kabsch algorithm (Kabsch, 1978). To calculate the twist and filament axis between two adjacent protomers (A_i and A_{i+1}) the following steps were performed:

(1) Set the dimer center of mass to the origin, then move the centers of mass of both protomers to the origin.

(2) Calculate the cross-covariance matrix between protomers:

$$X = {A_{i+1}}^T A_i$$

(3) The optimal rotation matrix *U* can then be found using a singular value decomposition of the covariance matrix *X:*

$$U = (X^T X)^{1/2} X^{-1}$$

(4) The eigenvector associated with the real eigenvalue of *U* is the filament axis:

$$filament\ axis = \langle u, v, w \rangle$$

Now, the rotation matrix *U* can be re-written as:

$$U = \begin{bmatrix} u^2 + (1-u^2)\cos\theta & uv(1-\cos\theta) - w\sin\theta & uw(1-\cos\theta) + v\sin\theta \\ uv(1-\cos\theta) + w\sin\theta & v^2 + (1-v^2)\cos\theta & vw(1-\cos\theta) - u\sin\theta \\ uw(1-\cos\theta) - v\sin\theta & vw(1-\cos\theta) + u\sin\theta & w^2 + (1-w^2)\cos\theta \end{bmatrix}$$

(5) The rotation angle θ can then be found from the trace of the rotation matrix:

$$z = 0.5 * (Tr(U) - 1)$$

$$\theta = \text{atan}^2\left(\sqrt{1-z^2}, z\right)$$

Finally, to calculate the offset between protomers, project the centers of mass of protomers A_i and A_{i+1} onto the filament axis. The length of the corresponding vector is the protomer offset. This method represents a significant improvement over the old method, providing better fits (as measured by RMSD) and thus a more accurate measurement of filament twist and offset.

2.8 Helix Bend Angles

Pathogenic helix bend angles were calculated in VMD using the Bendix plugin (Humphrey et al., 1996; Dahl, Chavent, & Sansom, 2012).

2.9 Tryptophan Solvent Accessible Surface Area

Measurement of the solvent accessible surface area for the tryptophan residues (W79, W86, W340, and W356) was done in VMD (Humphrey et al., 1996). Analysis was limited to tryptophan sidechains and a 1.4 Å sphere was used.

2.10 Significance Testing with Filament Parameters

Student t-tests were used to measure the significance of the filament parameters presented in Chapter 4. These were done by calculating the mean and standard deviation over the whole trajectory, but limiting n to the number of ns of conventional simulation performed. For example, when comparing the propeller-twist of the barbed end of the filament to that of the middle, n values of 800 and 1600 were used respectively.

2.11 References

Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E. et al. (2016). AMBER 2016. *University of California, San Francisco.*

Dahl, A. C. E., Chavent, M., & Sansom, M. S. P. (2012). Bendix: Intuitive helix geometry analysis and abstraction. *Bioinformatics*, *28*(16), 2193-2194.

Dalhaimer, P., Pollard, T. D., & Nolen, B. J. (2008). Nucleotide-Mediated Conformational Changes of Monomeric Actin and Arp3 Studied by Molecular Dynamics Simulations. *J. Mol. Biol.*, *376*(1), 166-183.

Dominguez, R., & Holmes, K. C. (2011). Actin structure and function. *Annu. Rev. Biophys.*, *40*, 169-168.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577-8593.

Fujii, T., Iwane, A. H., Yanagida, T., & Namba, K. (2010). Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, *467*(7316), 724-728.

Graceffa, P., & Dominguez, R. (2003). Crystal structure of monomeric actin in the ATP state. Structural basis of nucleotide-dependent actin dynamics. *J. Biol. Chem.*, *278*, 34172-34180.

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., & Caves, L. S. D. (2006). Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*, *22*(21), 2695-2696.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Genetics*, *65*(3), 712-725.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33-38.

Jariwala, S., Skjærven, L., Yao, X., & Grant, B. J. (2017). Investigating protein sequence-structure-dynamics relationships with Bio3D-web. *Journal of Visualized Experiments*, *2017*(125).

Jepsen, L., Kruth, K. A., Rubenstein, P. A., & Sept, D. (2016). Two Deafness-Causing Actin Mutations (DFNA20/26) Have Allosteric Effects on the Actin Structure. *Biophysical journal*, *111*(2), 323-332.

Joosten, R. P., Te Beek, T. A. H., Krieger, E., Hekkelman, M. L., Hooft, R. W. W., Schneider, R. et al. (2011). A series of PDB related databases for everyday needs. *Nucleic acids research*, *39*(SUPPL. 1), D411-D419.

Kabsch, W. (1978). A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, *34*(5), 827-828.

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577-2637.

Meagher, K. L., Redman, L. T., & Carlson, H. A. (2003). Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry*, *24*(9), 1016-1025.

Oda, T., Iwasa, M., Aihara, T., Maéda, Y., & Narita, A. (2009). The nature of the globular- to fibrous-actin transition. *Nature*, *457*(7228), 441-445.

Otterbein, L. R., Graceffa, P., & Dominguez, R. (2001). The crystal structure of uncomplexed actin in the ADP state. *Science*, *293*, 708-711.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E. et al. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781-1802.

Pierce, L. C. T., Salomon-Ferrer, R., Augusto F. De Oliveira, C., McCammon, J. A., & Walker, R. C. (2012). Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of Chemical Theory and Computation*, *8*(9), 2997-3002.

Skjaerven, L., Yao, X.-Q., Scarabelli, G., & Grant, B. J. (2014). Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics*, *15*(1).

Skjærven, L., Jariwala, S., Yao, X., & Grant, B. J. (2016). Online interactive analysis of protein structure ensembles with Bio3D-web. *Bioinformatics*, *32*(22), 3510-3512.

Vorobiev, S., Strokopytov, B., Drubin, D. G., Frieden, C., Ono, S., Condeelis, J. et al. (2003). The structure of nonvertebrate actin: Implications for the ATP hydrolytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 5760-5765.

Wang, Y., Harrison, C. B., Schulten, K., & McCammon, J. A. (2011). Implementation of accelerated molecular dynamics in NAMD. *Computational Science and Discovery*, *4*(1).

Chapter 3

Monomeric G-Actin and the Effects of Nucleotide on Actin Structure and Dynamics

Actin's function is highly dependent upon its nucleotide state. The affinity of actin for its related binding proteins, as well as for polymerization, is reliant on upon nucleotide state. In this chapter, we will examine the effects of nucleotide on the structure and dynamics of the actin monomer. Through the use of molecular dynamics simulations of ADP and ATP G-actin, we show that changes to the nucleotide lead to altered dynamics of several loops within the nucleotide-binding cleft – including the C-terminal hinge which sits at the interface of the nucleotide and target-binding clefts. We find evidence that the C-terminal hinge communicates changes within the nucleotide binding cleft down to the target binding cleft. Our simulations also reveal several loops with the target-binding cleft that show nucleotide dependent dynamics, accounting for actin binding proteins nucleotide sensitivity. Further, we show that ATP G-actin is flatter that its crystal structure with a more open nucleotide cleft mouth, putting it in a conformation more similar to the F-actin protomer.

3.1 Introduction

The nucleotide state of actin is essential to its function - determining its affinity for actin binding proteins (dos Remedios et al., 2003; Dominguez, 2004; Dominguez & Holmes, 2011), as well as for polymerization (Pollard, 1986; Fujiwara, Vavylonis, & Pollard, 2007). However, the structural differences between the ADP and ATP G-actins states that account for these changes are not as apparent. A comparison of the ADP muscle G-actin crystal structure (1J6Z: Otterbein, Graceffa, & Dominguez, 2001) to its ATP counterpart (1NWK: Graceffa & Dominguez, 2003), reveals little in the way of nucleotide effect. The two crystals are nearly identical, with a backbone RMSD of just 0.6 Å and only one region has a slightly altered secondary structure - the H73-loop right next to the nucleotide (Figure 3.1). The crystal structures show that an additional phosphate group in the ATP state leads to differential contacts between the nucleotide and the S-loop (residues 11-16 that contains Ser14), the G-loop (residues 154-172 that contains Gly158), and by extension the H73-loop (residues 70-78 that contains the methylated histidine at position 73) (Graceffa & Dominguez, 2003; Otterbein et al., 2001; Dominguez & Holmes, 2011). The question then becomes, *How do changes within the nucleotide cleft get propagated out to the surface of the protein and result in changes in affinity for actin binding proteins and for polymerization?*

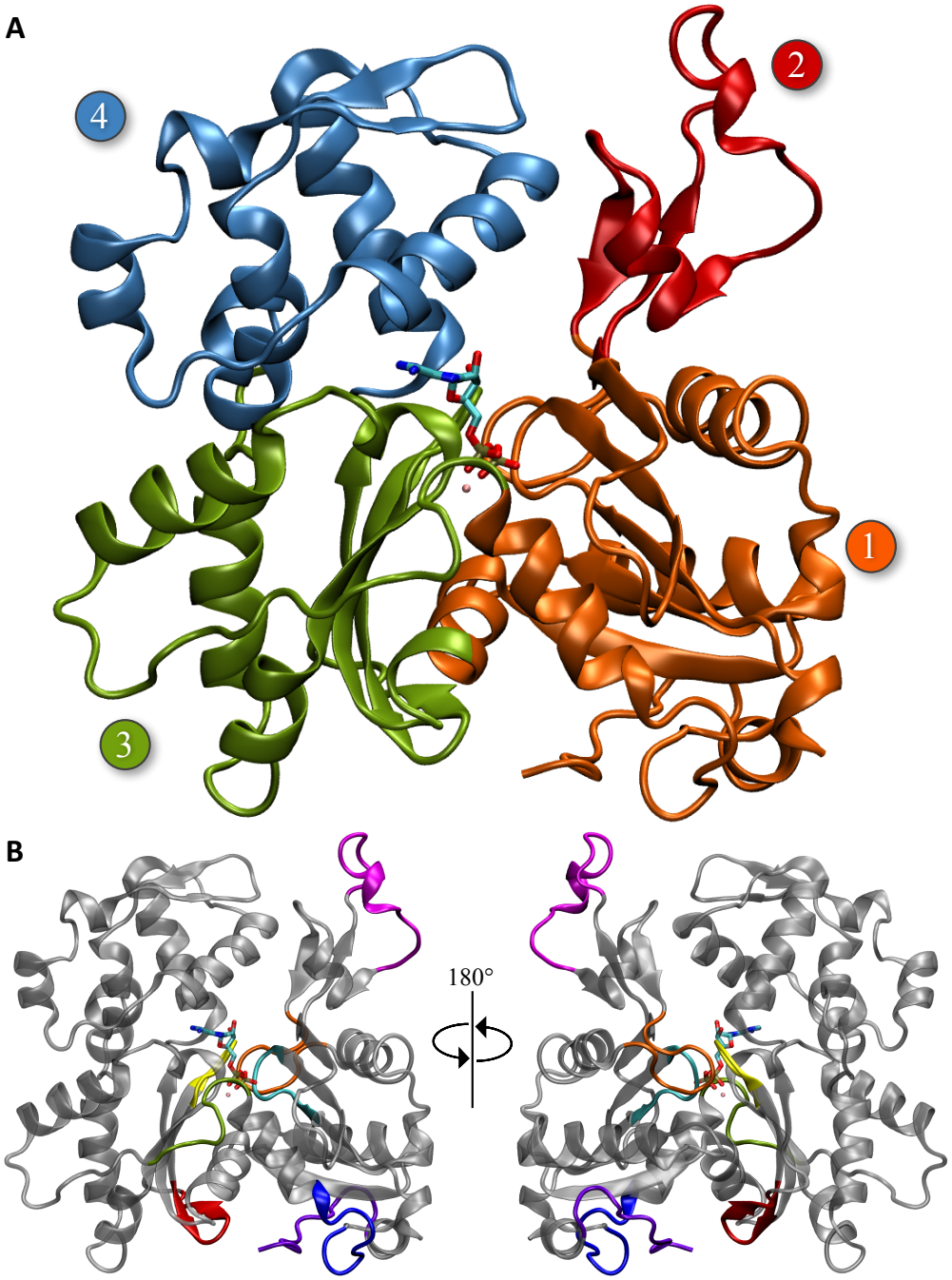


Figure 3.1 Actin Structure (A) Actin is composed of two major domains, or four subdomains. Subdomains 1 and 2 make up the first major domain and subdomains 3 and 4 the second. In the center of the protein, between all subdomains, sits the nucleotide binding cleft. At the bottom of the protein, between subdomains 1 and 3, is the target binding cleft. (B) Front and back views of the actin monomer with several structurally important regions highlighted. Within the nucleotide-binding cleft we have the S-, G-, and H73-loops (cyan, yellow, and orange). The nucleotide-binding and target-binding clefts are separated by two short polypeptides, one of which we call the C-terminal hinge, shown in green. Within the target-binding cleft the W-loop, FQQ-loop and C-terminus are all shown (red, blue, and violet). Finally, at the top of subdomain 2 is the D-loop (magenta).

The vast majority of actin binding proteins interact within actin's target-binding cleft (aka the barbed end groove) which sits at the bottom of the protein between subdomains 1 and 3 (Figure 3.1 and Figure 3.2) (Dominguez, 2004; Dominguez & Holmes, 2011; Kudryashov & Reisler, 2013). The target-binding cleft is separated from the nucleotide binding cleft by two hinge regions connecting subdomains 1 and 3 (Q137-S145, A331-Y337). This suggests that some inter-cleft communication must be present to relay actin's nucleotide state to the target-binding cleft and thus control the affinity of these binding proteins. Within the target-binding cleft, the W-loop (residues 165-172) has been identified as having nucleotide dependent secondary structure and conformation (Zheng, Diraviyam, & Sept, 2007; Kudryashov, Grintsevich, Rubenstein, & Reisler, 2010). This behavior explains some of differences in affinity for actin binding proteins, however, the minimal change of the W-loop between states seems unlikely to fully account for the drastic differences in binding affinities and polymerization rates.

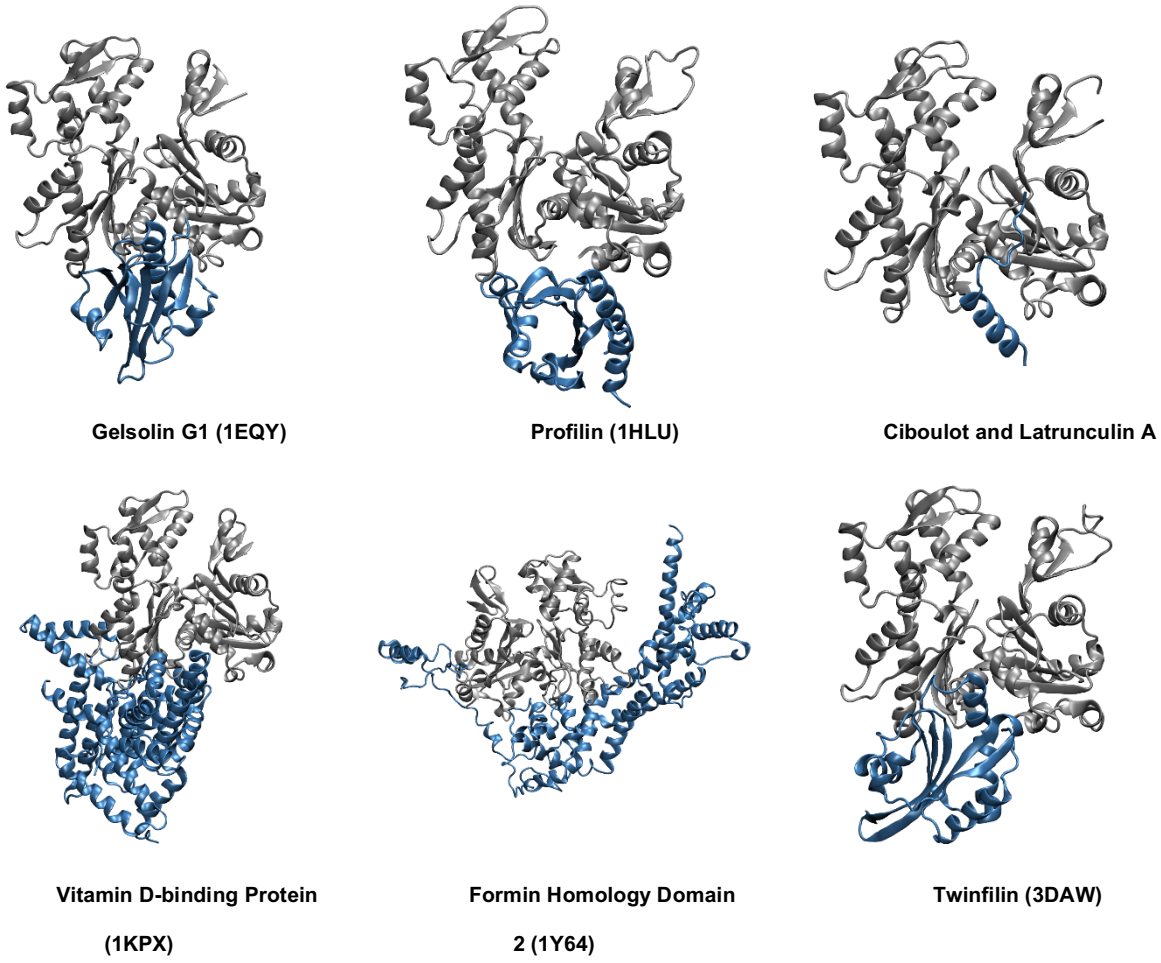


Figure 3.2 Actin Binding Proteins A majority of actin binding proteins dock in what is known as the target binding cleft which is located at the bottom of the protein between subdomains 1 and 3. Here we see a variety of different actin binding proteins that all bind in this region.

To explore the differences in nucleotides states and the allosteric connection between the nucleotide and target-binding clefts, we performed more than 4.7 μs of simulation on the ADP and ATP α-muscle G-actin. The goal of these simulations is to elucidate differences in both dynamics and structure that account for the changes in binding affinities and polymerization between nucleotide states, looking specifically at inter-cleft communication and changes within the nucleotide binding cleft.

3.2 Results

3.2.1 ADP G-Actin Tertiary Structure and Comparison to Crystal Structures

Our first objective was to characterize the overall tertiary structure of the ADP actin monomer. In order to meaningfully analyze the tertiary structure of G-actin, we turned to principal component analysis (PCA) to reduce the dimensionality of the system and extract the large-scale motions of the protein. To further improve the quality of this analysis, we only analyzed residues that were in stable α-helices or β-strands (> 80% of the time) so that the covariance calculation was not dominated by the motion of unstructured loops. Our initial PCA analysis, with the basis set built using the aMD G-Actin simulation, converged to two preferred states which were explored more thoroughly using cMD simulations, and further converged to one equilibrium structure (Figure 3.3). This state represents a significant shift of the ADP G-actin structure from its crystal starting point (RMSD of 2.69 ± 0.25 Å with the equilibrium structure). Importantly, this structural shift is almost entirely due to rearrangement of the four subdomains of G-actin and the changes within each of the subdomains, with the exception of subdomain 2, are very small (1.60 ± 0.17 Å, 3.06 ± 0.90 Å, 1.32 ± 0.14 Å, 1.78 ± 0.24 Å).

The two primary PCA modes represent a mixture of propeller-twist of the monomer, as measured by the dihedral angle between the cores of subdomains 2-1-3-4, and changes in the opening of the nucleotide binding cleft (see Figure 3.3 D). The 1J6Z structure has a propeller-twist angle of -22.2°, but our equilibrium solution structure is much flatter with a twist angle of -15.6° (note that this is strictly for the equilibrium structure, not the full simulation). The nucleotide binding cleft is typically characterized by the distance between the C_α positions of Q59 and E207 (Dalhaimer, Pollard, & Nolen, 2008; Oda & Maéda, 2010), and we find our solution structure for ADP G-actin more closed than the crystal (1J6Z:11.07 Å vs. Equil.: 6.48 Å).

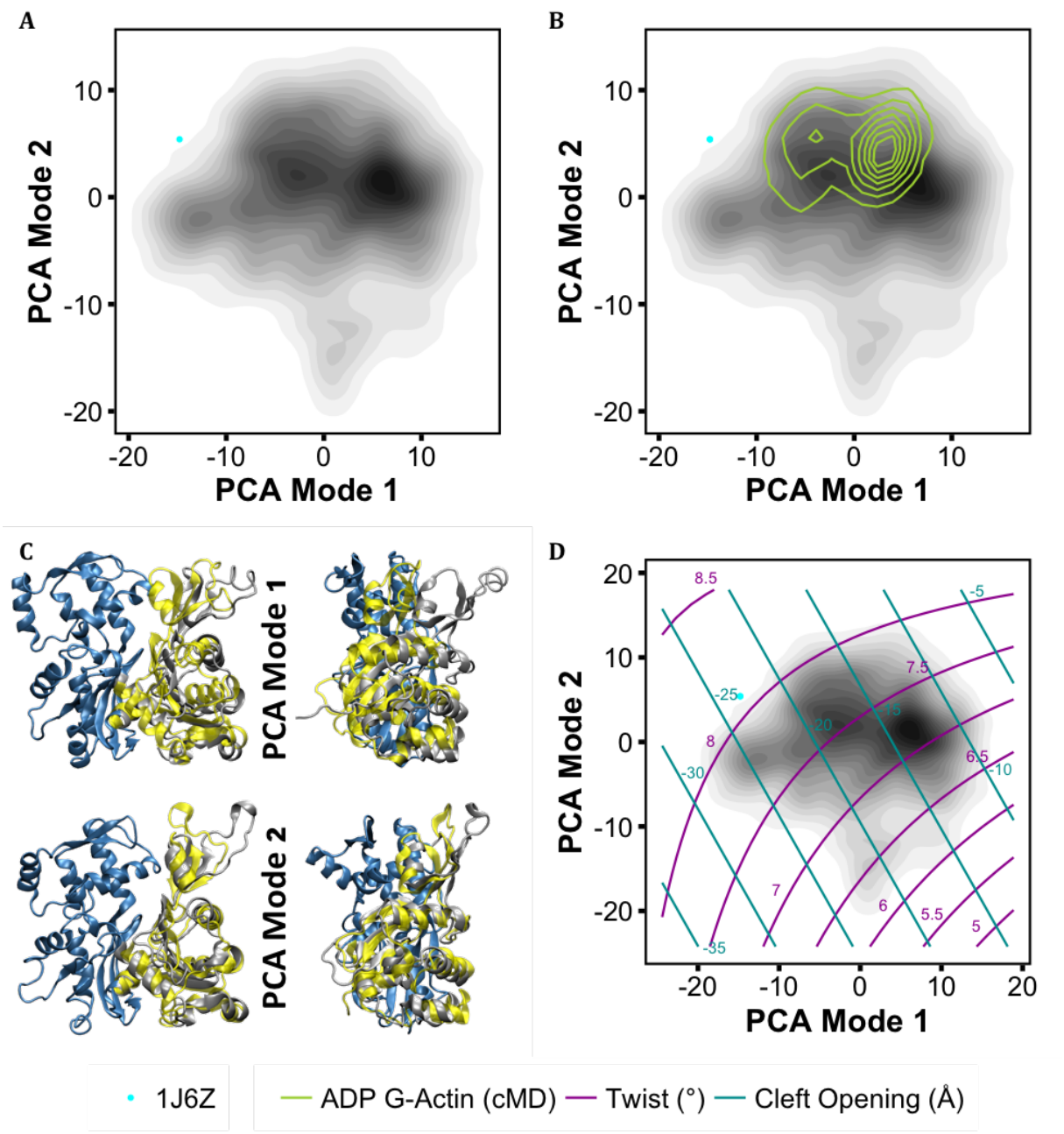


Figure 3.3 ADP G-Actin's PCA Space Our PCA basis set (A) was built using the C_α residues that were part of secondary structure elements for at least 80% of the time in our aMD ADP G-Actin simulation. This led to the formation of two peak structures, one at about PC1 = 6.2 and PC2 = 2.2, and a second more minor peak at PC1 = -2.8 and PC2 = 2.1. It was from these peak structures that our cMD simulations were started. In our cMD simulation, we see a convergence of those two peaks, to a third peak at about PC1 = 2.4 and PC2 = 3.3 (B). In panel C, we see the first motion of the first two PC modes. Here, subdomains 3 and 4 were fixed (blue) and the extreme positions of subdomains 1 and 2 are shown in yellow and gray. These motions seem to represent a combination of a propeller-twist like motion and the opening and closing of the nucleotide binding cleft. (D) We measured these motions as the dihedral angle between subdomains 2-1-3-4 and the C_α distance of Q59 and E207 and fit them using linear regression to PC modes 1 and 2. PC1, and to a lesser degree PC2, is strongly correlated with the propeller twist ($R^2 = 0.94$). The first two modes are also correlated with cleft mouth size ($R^2 = 0.24$).

To explore this conformational difference further, we projected more 170 additional actin and actin related structures onto our PCA basis set to see how these structures compare with our simulation results (Figure 3.4). Consistent with what was earlier observed by Oda and Maeda (Oda & Maéda, 2010), we see a broad range of propeller-twist values, ranging from -22.26° to +3.18°, and some variation in the nucleotide cleft. Considering the forces within protein crystals, the difference between our equilibrium structure and the range of crystal structures is not that surprising, and many of the crystal structures still fall within the margins of the space sampled by our simulations. Nevertheless, we believe that our structure represents a more physiologically relevant form of the actin monomer.

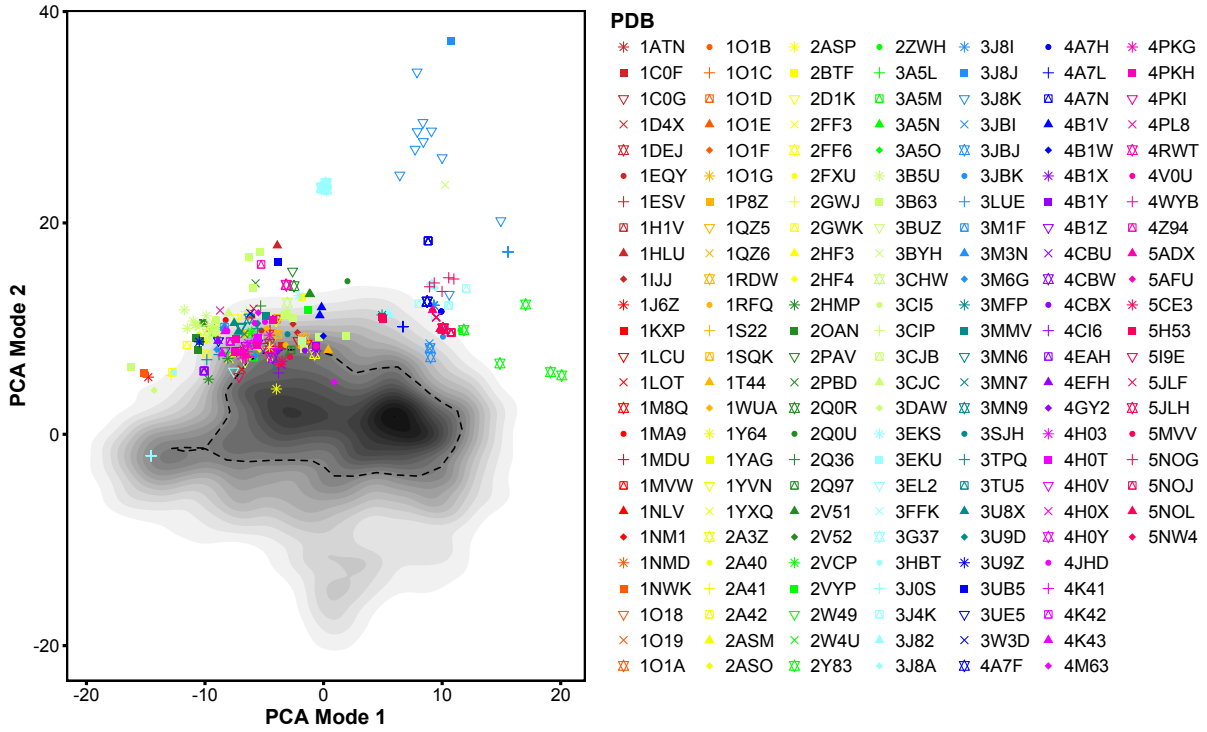


Figure 3.4 Crystal Structure Projection To observe the shift in PC space of our simulation structure compared to that of the various actin structures available in the PDB database, over 170 different PDBs, representing 535 unique structures were projected onto our basis set. The vast majority of structures lie outside our 50% density (black dashed line), with only two falling near either equilibrium structure (1Y64 and 4M63).

Further, if we measure the propeller-twist, phosphate clamps size (C_α separation of G15 and D157), and nucleotide cleft mouth size of the crystal structures, we find that the shift in PCA

space is largely due to the closing of the nucleotide binding cleft (Figure 3.5). In 2013, cysteine-crosslinking was performed using crosslinkers of various lengths, from 5.4 Å to 9.5 Å as well as a variable 6-22 Å crosslinker, to measure the cleft mouth size in yeast G-actin (Kudryashov & Reisler, 2013). For Mg ADP yeast G-actin, the 9.5 Å crosslinker was the most efficient. This matches well with the 9.73 ± 0.83 Å nucleotide cleft mouth size of our Mg ADP yeast G-actin simulation (data presented in Chapter 5). Our yeast simulation had a cleft mouth size that was 3.2 Å shorter than its associated crystal structure (1YAG: Vorobiev et al., 2003), this is consistent with the 3.6 Å shortening we observe for our muscle Mg ADP G-actin. Within the crystal complex, it appears that either the forces of the crystal or the modifications used to prevent polymerization, artificially open the nucleotide cleft mouth.

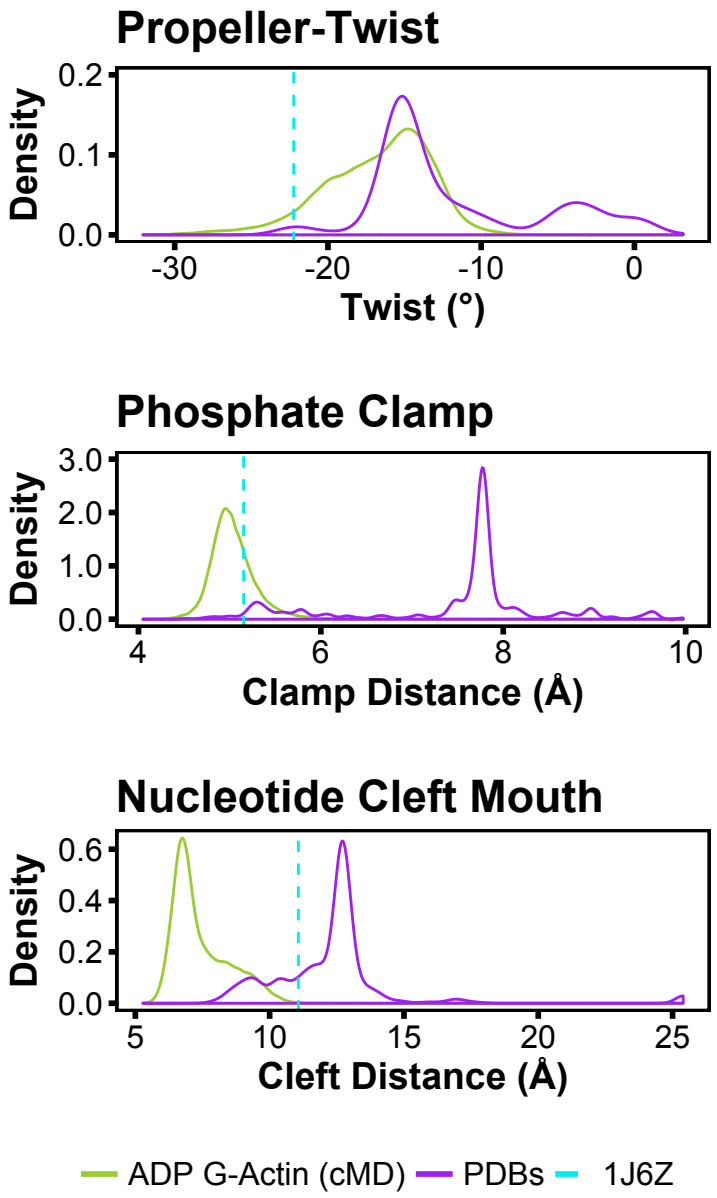


Figure 3.5 Propeller-Twist, Phosphate Clamp, and Nucleotide Cleft Mouth in ADP G-Actin vs PDB Structures A comparison of our ADP G-Actin simulation structures to that of the available PDB files (see Figure 3.4) shows that shift observed in PCA space has less to do with the propeller-twist angle and more to do with the nucleotide cleft mouth distance.

3.2.2 ADP G-actin secondary structure

As compared to the crystal structure of muscle ADP G-actin (Graceffa & Dominguez, 2003), we see essentially no change in the secondary structure apart from minor shifts or dynamical changes at the ends of α-helices or β-strands. The D-loop (residues 39-52) in subdomain 2 interacts

with DNaseI and has long been known to be a structurally dynamic region of the protein. Consistent with previous findings (Schwyter, Phillips, & Reisler, 1989; Strzelecka-Gołaszewska, Moraczewska, Khaitlina, & Mossakowska, 1993; Zheng et al., 2007; Dalhaimer et al., 2008; Durer et al., 2012; Kudryashov & Reisler, 2013), we see this region form a helix about 30% of the time in the ADP state (see Figure 3.6). Additionally, two helices at the C-terminus of the protein show reduced helical propensity, the FQQ-loop which forms a helix from 350-355 in the crystal forms at most 25% of the time, and a helix from 367-370 forms at most 67% of the time but is extended by an extra 2-3 residues in our simulation. Several other short 3_{10}-helices that form in the crystal have proved to be unstable in our simulation, mainly the helices from 172-175, 287-289 and 335-337.

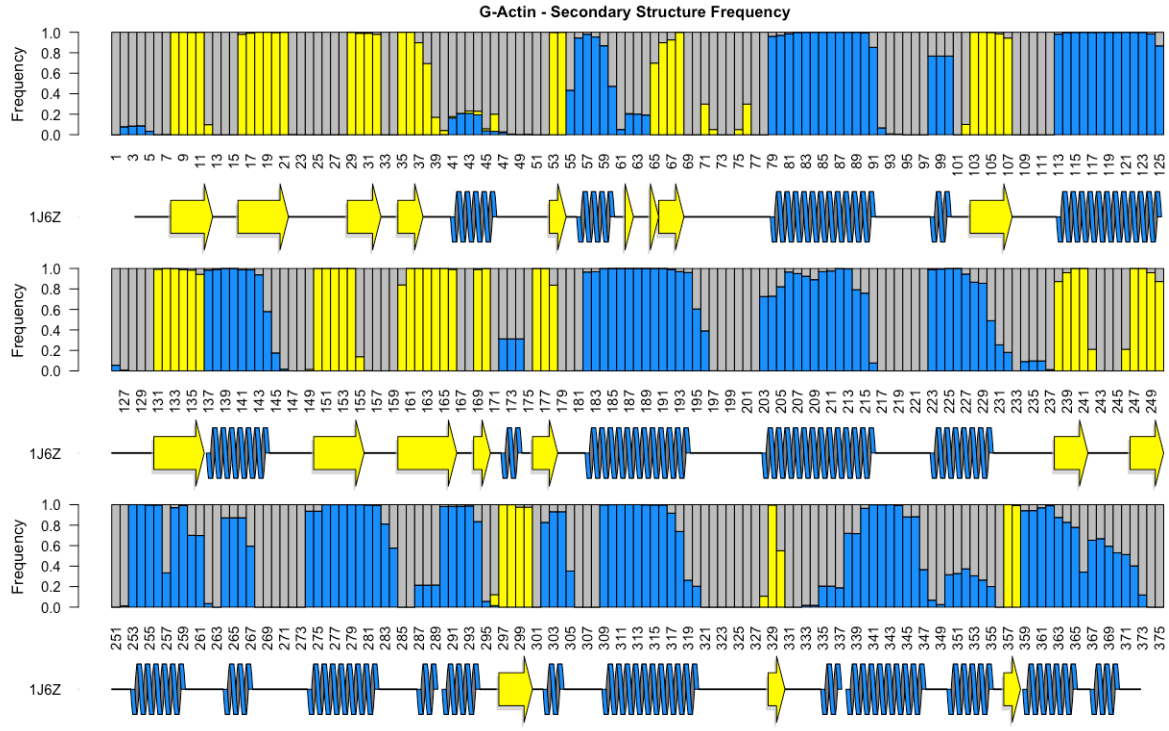


Figure 3.6 ADP G-Actin Secondary Structure A comparison of our cMD simulation ADP G-actin secondary structure to its corresponding crystal, 1J6Z (Otterbein et al., 2001).

3.2.3 Differences in Nucleotide States of G-Actin

It has long been known that there are functional differences between the ADP and ATP states of G-actin, not only in terms of polymerization but also from the interaction with various actin binding proteins. To determine the structural differences between ADP and ATP G-actin, we simulated the ATP G-actin structure (1NWK; Graceffa & Dominguez, 2003) and projected these results on to the same PCA space from the ADP G-actin simulations (Figure 3.7). Like the ADP simulation, the ATP monomer also shifts away from its starting crystal conformation, converging to the same two equilibrium points as our aMD ADP G-actin simulation, but with reversed intensity. Again, we sampled these states for an additional 1.2 μs using cMD. As with its ADP counterpart, the cMD simulations converged to a new more F-actin-like structure, and although the two G-actin structures have a similar range of conformations, there are some distinctions.

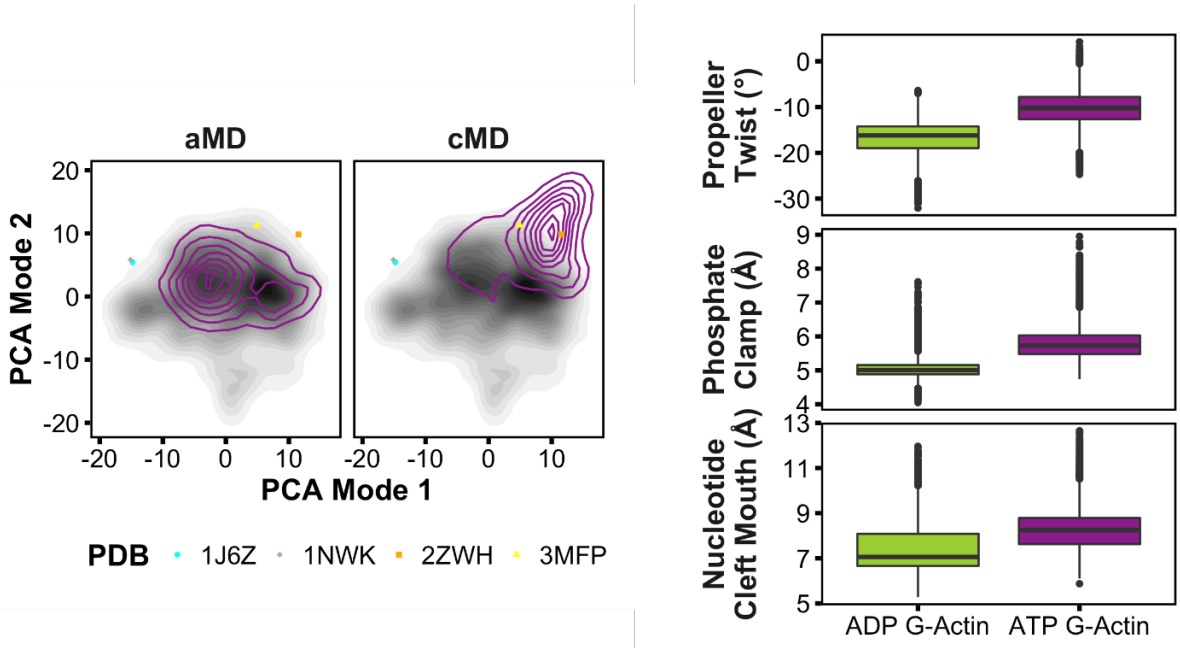


Figure 3.7 Tertiary Conformation of ATP G-Actin (A) The ATP G-actin molecular dynamics simulations were projected onto our ADP G-actin basis set. The aMD simulations converged to the same to equilibrium structures with reverse intensity. cMD simulations started from these points converged to a third more F-actin like state. (B) A comparison of the propeller-twist, phosphate clamp, and nucleotide cleft mouth distances show that ATP G-actin is flatter with a more open clamp and cleft than ADP G-actin.

Most notably, the ATP structure exhibits even less propeller-twist than the ADP form. This shift correlates with a 5.0° flattening of the subdomain 2-1-3-4 dihedral angle, from -16.7 ± 3.5° in ADP G-actin to just -11.7 ± 5.0° in the ATP state (Figure 3.7). This flattening is on the order of what Oda (Oda, Iwasa, Aihara, Maéda, & Narita, 2009; Oda & Maéda, 2010) and others (Fujii, Iwane, Yanagida, & Namba, 2010; Dominguez & Holmes, 2011) observed in the G- to F-actin transition, suggesting that ATP G-actin is in a more F-actin like conformation. Although we find the solution G-actin structures are already much flatter than their crystal counterparts, the ATP G-actin structure is even flatter and closer to the conformation of the F-actin protomer. Based solely on simple protein isomerization arguments, this is consistent with what we might expect since ATP G-actin polymerizes faster than ADP G-actin.

The shift in PCA space observed between ADP and ATP G-actin is also strongly correlated to opening of the nucleotide binding cleft. This opening was measured as the nucleotide cleft mouth size (C_α distance of Q59 and E207), as well as the phosphate clamp size (C_α distance of G15 and D157) (Dalhaimer et al., 2008). Both the phosphate clamp and nucleotide cleft mouth are about 1.3 Å wider in the ATP G-actin state than in the ADP state (Figure 3.7). Since there is an extra phosphate on the nucleotide, this is again not a surprising finding.

3.2.4 Analysis of Loop Dynamics

To this point we have focused on the overall tertiary structure of the protein and our analysis specifically removed unstructured region for the protein from consideration. To identify regions of G-actin that show differences between the ADP and ATP state, we performed targeted analysis just on the loop regions. In this case, the core of the protein was fit as described before, but PCA was done separately on each loop using the combined data sets for ADP and ATP G-actin. With this method of fitting and the combination of both nucleotide states, PCA will work to

maximize any conformation differences in these loop regions. Of the 24 loops that we analyzed, the majority showed no difference between ADP and ATP (Figure 3.8), but the section below highlights some key differences throughout the actin structure.

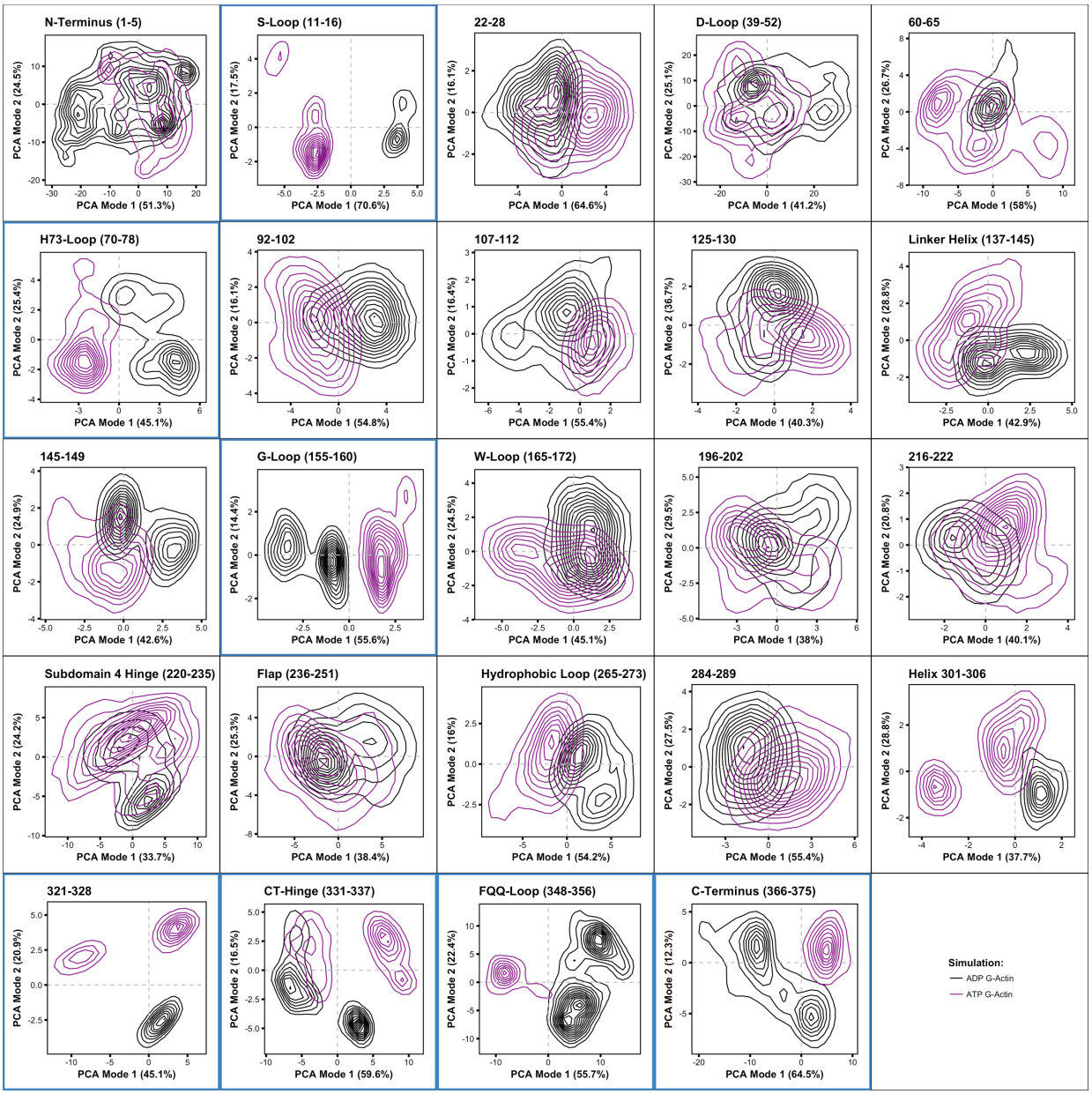


Figure 3.8 Loop PCA All of the regions that were analyzed using the loop PCA protocol. Of the loops shown, only a few were deemed significant: the S-Loop, H73-Loop, G-Loop, the loop from 321-238 (note this was not analyzed further as its motion was correlated with the CT-hinge), CT-Hinge, FQQ-Loop and the C-terminus.

3.2.4.1 S, G and H73 Loops

Within the nucleotide binding cleft, the S-, G-, and H73-loops (11-16, 155-160, 70-78) have all been shown to be state sensing (Graceffa & Dominguez, 2003; Otterbein et al., 2001; Zheng et al., 2007; Dominguez & Holmes, 2011). PCA on the combined basis set for these loops shows two distinct states for ADP and ATP G-actin (Figure 3.9). In the ATP G-actin state, the S- and G-loops open up, moving slightly away from each other allowing the H73-loop to come in and take up space previously occupied by those loops. This change is reflected in the hydrogen bonding pattern. In the ADP G-actin state, the S-loop is responsible for making hydrogen bonding interactions with the β-phosphate of ADP, when the nucleotide is switched for ATP, the majority of the phosphate contacts are switched to the G-loop. The exceptions are that the main chain of Ser14 which makes a stronger connection with ATP than it does ADP, and the side chains of Gly156 and Asp157 which make stronger connections in ADP than ATP. These two changes result from the rocking of the S- and G-loops with respect to the nucleotide. In addition to changes in conformation, the S- and H73-loops each have nucleotide dependent secondary structure (Figure 3.10) with the beta turn of the S-loop extending by one residue in the ATP state, and the H73-loop, which forms a short antiparallel β-sheet in the ADP state, is unstructured in the ATP state.

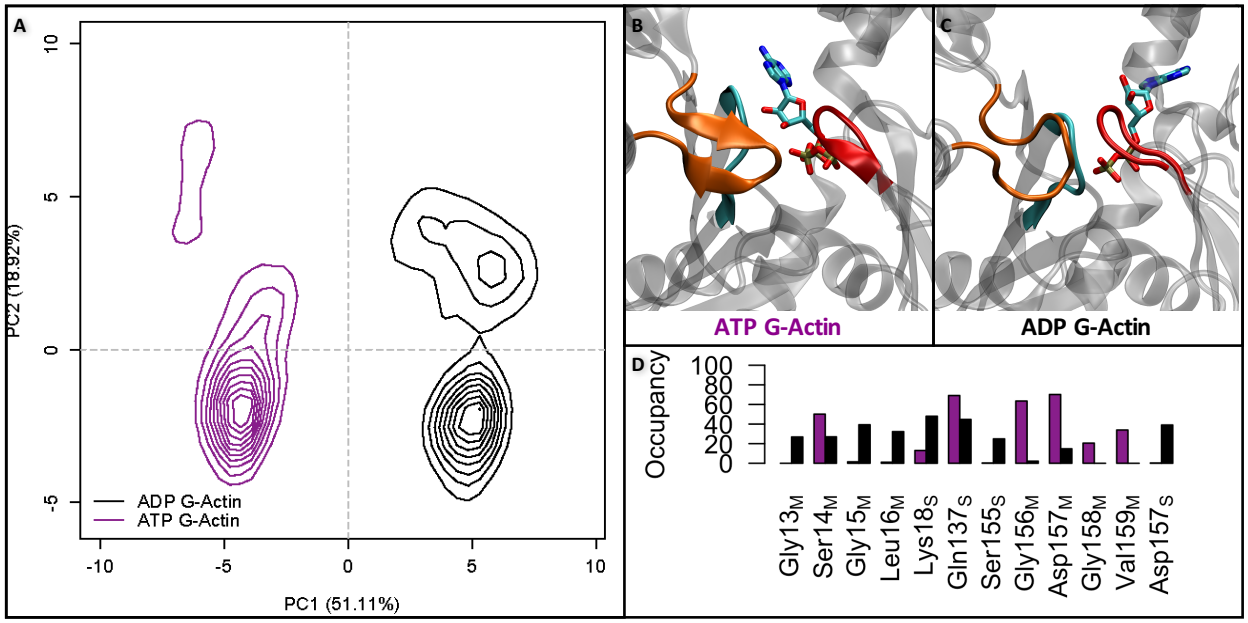


Figure 3.9 S, G, and H73 Loop Coordination (A) Loop PCA on the S-, G- and H73-loops (11-16, 155-160, 70-78) shows drastically different conformations between nucleotide states. (B,C) Loop dynamics represented by ATP and ADP G-actin respectively. In the ATP-state the S- and G-loops have more separation and the H73-loop forms a stable β-sheet. (D) Hydrogen bonds with the nucleotide that have a change in occupancy of at least 20 between nucleotide states.

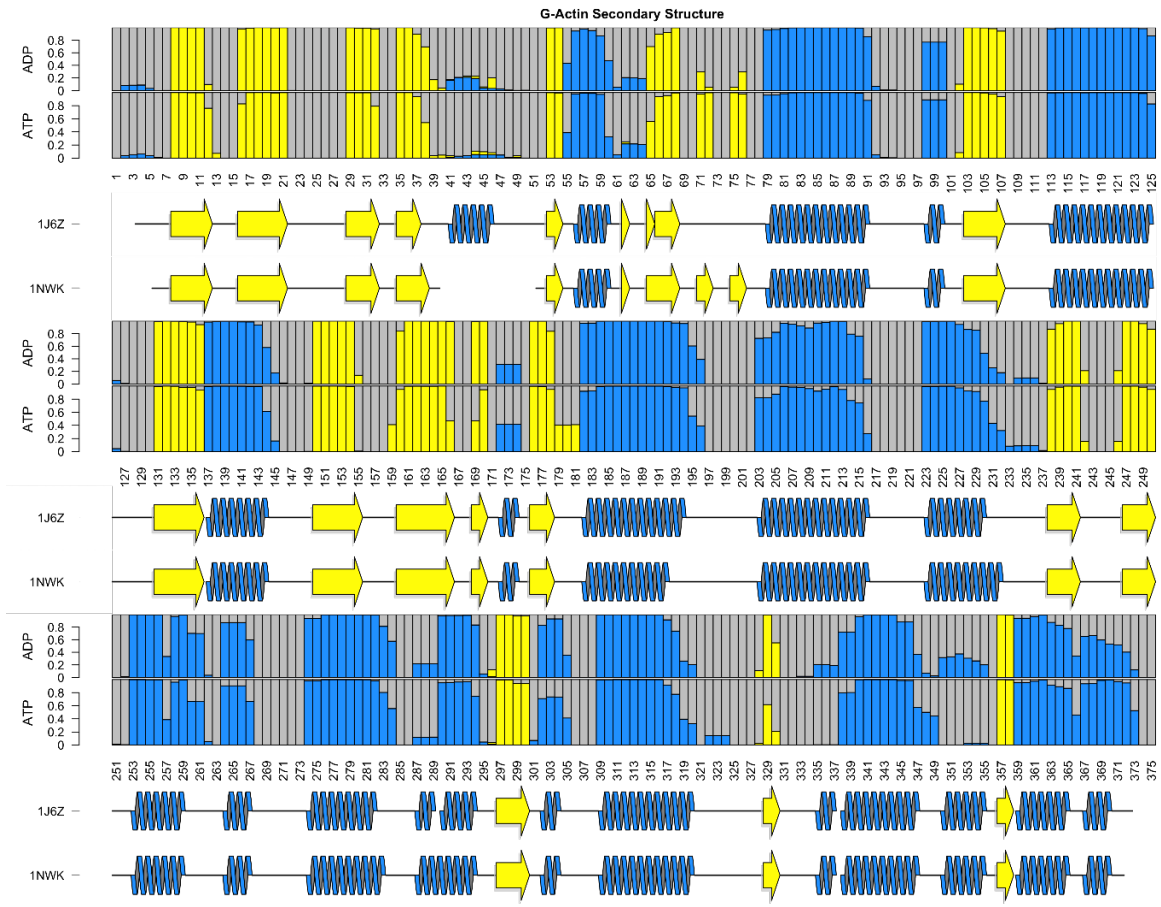


Figure 3.10 ADP vs ATP G-actin Secondary Structure A comparison of secondary structure between the ADP and ATP G-Actin nucleotide states shows a few key areas of difference. Regions that differ between the two states include the D-loop, H73-loop, W-loop, FQQ-loop and C-terminus.

3.2.4.2 C-Terminal Hinge

Connecting subdomains 1 and 3 are two hinge regions, one from Q137-S145 and a second from A331-Y337. The second, or C-terminal (CT) hinge, showed large RMSF deviations between states, with ATP G-Actin showing more variability (Figure 3.11). Our loop analysis showed a clear state dependence (Figure 3.12). The CT-hinge is able to move across the target binding cleft, as well as in and out from the protein from the surface. These motions are represented in the first two PC modes. The CT-hinge position was measured using the projection of C_α-atom of P333 onto the vector between C_α atoms of A144 and A310. The ADP state takes on two distinct positions, with the P333 spending time close to A144 and in the center of the cleft. When the nucleotide is

switched for ATP, a third state appears with P333 moving close to A310. With the increased flexibility of the CT-hinge in ATP G-Actin, the nucleotide can take on an altered conformation. Instead of stacking with E214, the adenosine ring flips and makes hydrogen bonding interactions with the backbone of K336. In the ADP state, K336 is in an upright conformation, with the side chain pointing towards the nucleotide binding cleft. Upon movement of the CT-hinge and nucleotide, K336 extends out away from the protein (Figure 3.12). The CT-hinge is at the binding surface for several actin binding proteins including cofilin, twinfilin and tropomyosin, each of which has an actin nucleotide preference. Additionally, when profilin is bound to actin in the open state, the CT-hinge takes on an extended conformation, more closely resembling what we observe in the ADP state (Chik, Lindberg, & Schutt, 1996).

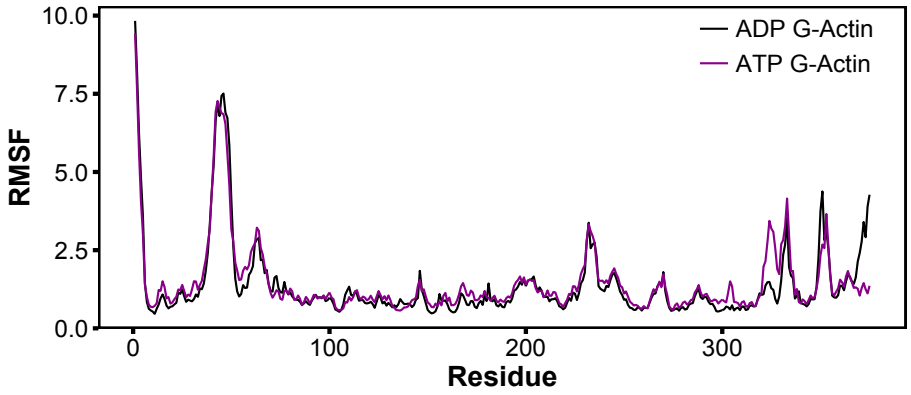


Figure 3.11 RMSF C_α RMSF for the cMD simulations of ADP and ATP G-actin.

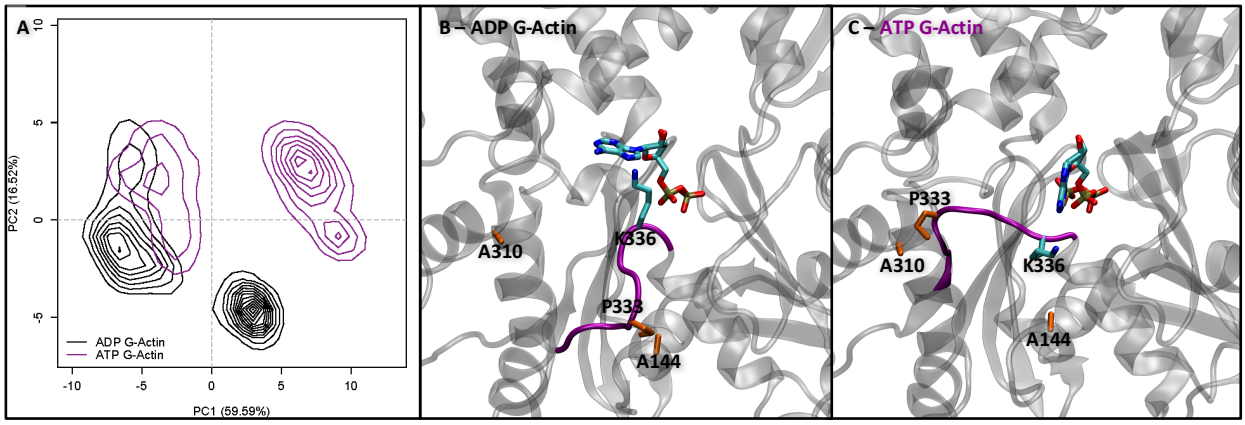


Figure 3.12 C-Terminal Hinge (A) Loop PCA on the CT-hinge (A331-Y337) shows differential dynamics between nucleotide states, where PC1 represents movement of the CT-hinge across the top of the target biding cleft, and PC2 represents the loop moving into and out of the protein. Movement across the target binding cleft can be viewed as the progression of P333 from A144 (B-low PC1-values), accessible in both nucleotide states, to A310 (C-high PC1-values), accessible only in the ATP state.

3.2.4.3 FQQ-Loop

Within the target binding cleft itself, a new loop was revealed that makes a drastic shift between nucleotide states. I refer to this as the FQQ-loop based on its sequence (S348-W356, SLST**FQQ**MW). This loop sits at the bottom of SD1 within the target binding cleft, directly across from the W-loop. In both the 1J6Z and 1NWK crystal structures, this loop forms a short 3_{10}-helix from S350 to M355. Our simulations showed this helix to be only metastable in the ADP G-Actin state, forming only about 25% of the time, and nonexistent in the ATP state (Figure 3.10). Loop analysis showed further differences between states (Figure 3.13). In the ADP state, the FQQ-loop extends out and away from the target binding cleft while in the ATP state, the loop extends towards the W-loop across the target binding cleft. To quantitatively measure the separation of the W and FQQ-loops, the C_α distance between loop residues were measured and the closest pair averaged. In the ADP state, the W- and FQQ-loops are separated by an average minimum of 8.7 ± 1.8 Å. In the ATP state, the loops are much closer, separated by only 4.0 ± 2.0 Å. The shift between the ADP and ATP state suggests that the loop could play a role in differentiating between actin binding proteins.

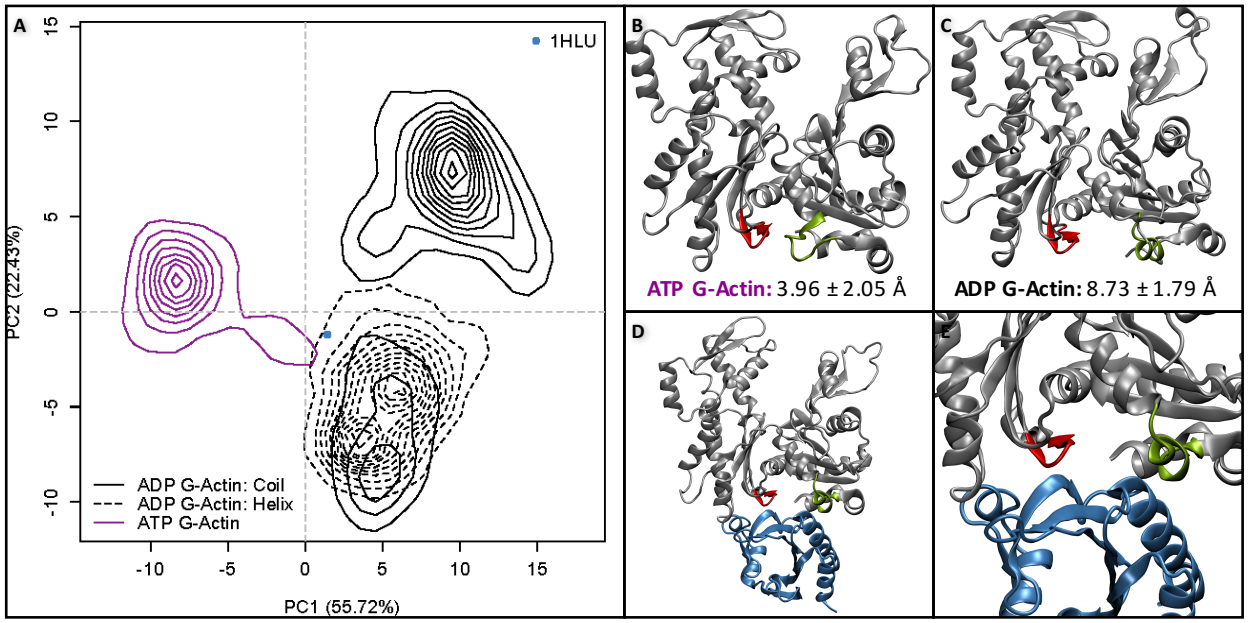


Figure 3.13 FQQ-Loop (A) Loop PCA on the FQQ-loop (S348-W356). PC1 shows the progression of the FQQ-loop across the target-binding cleft. In the ATP state (B) the FQQ-loop extends towards the W-loop partially closing the target-binding cleft. In the ADP state (C) the FQQ-loop moves away from the W-loop, opening the target-binding cleft. In this state, it can also take on a helical conformation (dashed lines in A). We believe that the FQQ-loop is key for determining the affinity of profilin (blue) for actin. (D-E) The crystal structure 1HLU (Chik et al., 1996) is β-actin bound to profilin; in this state actin's FQQ-loop is extended out away from the W-loop and is forming a helix, similar to the ADP G-actin state.

3.2.4.4 C-Terminus

Actin's crystal structures (1J6Z and 1NWK) form multiple helices at the C-terminus of the protein, one from 359 to 365 and another from 367 to 370 (Figure 3.10). In both the ADP and ATP states, the first helix is stable forming in more than 85% of the frames of both simulations. The second helix, 367-370, is extended by an additional three residues in both states but is less stable when ADP is bound, forming at most 67% of the time. When this helix is not formed the C-terminus can take on a stretched conformation, where F375 extends towards the target binding cleft (Figure 3.14). In either state, the C-terminus is able to form three hydrogen bonds with the pathogenic helix ($H371_S$-$E117_S$, $R116_S$-$F375_S$, $R116_S$-$F375_M$: S – sidechain hydrogen bond, M – mainchain hydrogen bond), these bonds form at least 20% more frequently in the ATP state. Both R116 and E117 have associated mutations, where the mutation of E117 to a Lys in β-cytoplasmic

actin causes Baraitser-Winter syndrome (Johnston et al., 2013; Rubenstein & Wen, 2014) and the mutation of R116 to Gln in α-smooth muscle actin causes TAAD aneurysms (Guo et al., 2007; Rubenstein & Wen, 2014). It is possible that these diseases may result in part from the disruption of interactions of the pathogenic helix with the actin's C-terminus.

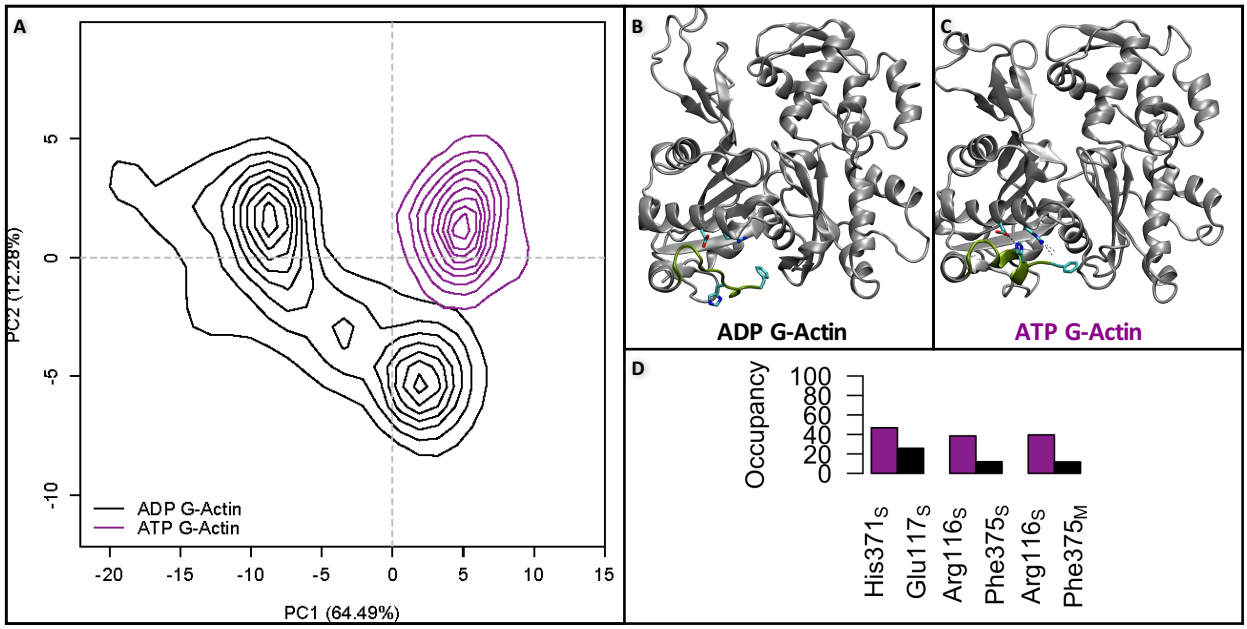


Figure 3.14 C-Terminus Actin's C-terminus (G366-F375) is another region with nucleotide dependent dynamics (A). In the ADP G-actin state (B), the C-terminus can either form a short helix or take on an extended conformation where F375 gets inserted into the target binding cleft. In the ATP state (C), the C-terminus forms the stable helix and makes hydrogen-bonding interactions with residues in actin's pathogenic helix (D).

3.2.4.5 D-Loop

Experimental evidence suggests that the D-loop (R39-S52) is more compact and can form a short helix in the ADP G-actin state (Durer et al., 2012; Kudryashov & Reisler, 2013; Schwyter et al., 1989; Strzelecka-Gołaszewska et al., 1993), but the evidence also shows that the secondary structure of the D-loop is highly unstable regardless of the bound nucleotide (Dalhaimer et al., 2008; Splettstoesser, Noé, Oda, & Smith, 2009; Durer et al., 2012). In our simulations, we find that the D-loop can form a short helix in the ADP G-actin state, but that it is unstable, forming at most 30% of the time. In the ATP G-actin state, the D-loop is unstructured for most of the

simulation, forming a helix at most 6% of the time and a β-strand at most 5% of the time. Despite the differences in secondary structure, PCA on loop analysis shows that the D-loop is not only highly dynamic, but that the dynamics are largely the same between states (Figure 3.15).

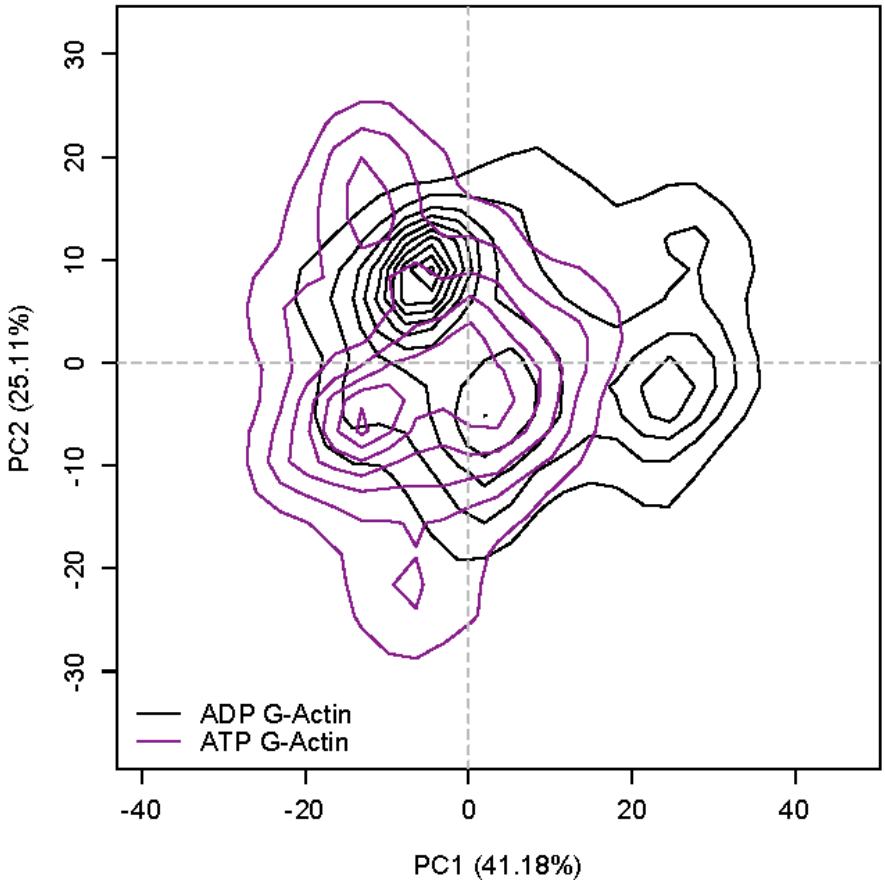


*Figure 3.*15 *D-loop* The D-loop is highly dynamic in both ADP and ATP G-actin. Loop PCA does not show significant nucleotide dependent dynamics.

3.3 Discussion

Consistent with our expectations, we see distinct differences between the ADP and ATP G-actin nucleotide binding cleft dynamics and nucleotide coordination. A simple comparison of the 1J6Z and 1NWK crystal structures reveals that the S-, G-, and H73-loops all move positions depending on which nucleotide is bound. Our simulations recapitulated these changes. We see that the S- and G-loops move closer in the ADP state, as measured by the phosphate clamp size. They also rotate with respect to each other (Figure 3.9), resulting in altered hydrogen bonding pattern

between the two nucleotides. As a result of the changes in the S- and G-loops, we see the H73-loop moving into the cleft and taking on a β-sheet structure in the ATP state.

In addition to having more hydrogen bonds, the adenosine ring in ADP forms stacking interactions with E214, essentially locking the ring in place for the duration of our simulation. While this conformation is seen in our ATP G-actin simulations it is not locked in place. The decreased coordination in the ATP state allows the ring to flip out and move into space previously occupied by the C-terminal hinge (CT-hinge). The CT-hinge (A331-Y337) is one of two short polypeptides that extend between subdomains 1 and 3. These loops also form as a barrier between the nucleotide and target-binding clefts. We believe that the CT-hinge, which makes direct contact with the nucleotide, helps communicate changes from within the nucleotide binding cleft to the target binding cleft. In addition to playing a role in communication between clefts, the CT-hinge is at the binding surface for several actin binding proteins including cofilin, twinfilin, and tropomyosin.

Within the target-binding cleft, previous studies had already implicated the W-loop as being an important nucleotide state sensor showing that it changes secondary structure based on bound nucleotide (Zheng et al., 2007). While the W-loop's change of structure may help determine the actin binding protein specificity for one nucleotide state or the other, it is not entirely sufficient. Here, in addition to the CT-hinge, I found two more regions that change conformation or secondary structure based on nucleotide state: the FQQ-loop (S348-W356) and the C-terminus (G366-F375).

FQQ-loop (S348-W356), which sits at the bottom of subdomain 1 across the target-binding cleft from the W-loop, underwent perhaps the most drastic change. In the ADP G-actin state, the FQQ-loop had some propensity to form a short 3_{10}-helix and extended out away from the target-binding cleft towards subdomain1. In the ATP G-actin state, the loop had little to no affinity for

the helical state and extended into the target-binding cleft, toward the W-loop, partially closing the cleft. The FQQ-loop has been shown to be at the binding surface for proteins such as dystrophin, profilin, and has the potential for myosin interactions (Fabbrizio, Bonet-Kerrache, Leger, & Mornet, 1993; Chik et al., 1996). Profilin facilitates the exchange of ADP for ATP and thus needs to recognize the ADP monomer. When bound to profilin, actin's W- and FQQ-loops are both structured and are in the open conformation (1HLU: Chik et al., 1996). This matches what is observed for the ADP G-actin state (Figure 3.13). It should also be noted that when profilin binds to actin, the CT-hinge takes on an extended conformation (Chik et al., 1996) matching what we observe for the ADP state.

The last loop at actin's barbed end that showed nucleotide state sensitivity was the C-terminus (G366-F375). While arguably not a part of the target-binding cleft, the C-terminus can insert the terminal Phe residue into the cleft and is at the protein surface making it an important binding site. In the ATP state, the C-terminus forms a stable helix (367-373). This conformation is also accessible when ADP is bound. However, in the ADP state the C-terminus can also take on an extended or stretched conformation. The distinct, nucleotide dependent conformations, plus the fact that the C-terminus is at the binding surface for several proteins, such as profilin and gelsolin (Crosbie et al., 1994; Chik et al., 1996; Porta & Borgstahl, 2012), indicates that it is important for nucleotide selectivity of actin binding proteins.

The concerted changes from the nucleotide-binding cleft down to the target-binding cleft, through the CT-hinge, all contribute to the nucleotide dependent affinity for actin binding proteins. It does not however, answer the question of why ATP G-actin polymerizes so much faster than its ADP counterpart (Pollard, 1986; Fujiwara et al., 2007). We will cover this in more detail in Chapter 4, but our initial simulations show that ATP G-actin takes on a flatter conformation more similar

to what is observed in F-actin. As ATP G-actin has a conformation similar to the F-actin protomer, it likely has a lower conformational energy barrier which explains its faster rates of polymerization.

3.4 References

Chik, J. K., Lindberg, U., & Schutt, C. E. (1996). The structure of an open state of beta-actin at 2.65 Å resolution. *J. Mol. Biol.*, *263*(4), 607-623.

Crosbie, R. H., Miller, C., Cheung, P., Goodnight, T., Muhlrad, A., & Reisler, E. (1994). Structural connectivity in actin: effect of C-terminal modifications on the properties of actin. *Biophysical journal*, *67*(5), 1957-1964.

Dalhaimer, P., Pollard, T. D., & Nolen, B. J. (2008). Nucleotide-Mediated Conformational Changes of Monomeric Actin and Arp3 Studied by Molecular Dynamics Simulations. *J. Mol. Biol.*, *376*(1), 166-183.

Dominguez, R. (2004). Actin-binding proteins--a unifying hypothesis. *Trends Biochem. Sci.*, *29*(11), 572-578.

Dominguez, R., & Holmes, K. C. (2011). Actin structure and function. *Annu. Rev. Biophys.*, *40*, 169-168.

dos Remedios, C. G., Chhabra, D., Kekic, M., Dedova, I. V., Tsubakihara, M., Berry, D. A. et al. (2003). Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol. Rev.*, *82*(2), 433-473.

Durer, Z. A., Kudryashov, D. S., Sawaya, M. R., Altenbach, C., Hubbell, W., & Reisler, E. (2012). Structural states and dynamics of the D-loop in actin. *Biophys J.*, *103*(5), 930-939.

Fabbrizio, E., Bonet-Kerrache, A., Leger, J. J., & Mornet, D. (1993). Actin-dystrophin interface. *Biochemistry*, *32*(39), 10457-10463.

Fujii, T., Iwane, A. H., Yanagida, T., & Namba, K. (2010). Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, *467*(7316), 724-728.

Fujiwara, I., Vavylonis, D., & Pollard, T. D. (2007). Polymerization kinetics of ADP- and ADP-Pi-actin determined by fluorescence microscopy. *Proc. Natl. Acad. Sci. USA*, *104*(21), 8827-8832.

Graceffa, P., & Dominguez, R. (2003). Crystal structure of monomeric actin in the ATP state. Structural basis of nucleotide-dependent actin dynamics. *J. Biol. Chem.*, *278*, 34172-34180.

Guo, D. C., Pannu, H., Tran-Fadulu, V., Papke, C. L., Yu, R. K., Avidan, N. et al. (2007). Mutations in smooth muscle alpha-actin (ACTA2) lead to thoracic aortic aneurysms and dissections. *Nature Genetics*, *39*(12), 1488-1493.

Johnston, J. J., Wen, K. K., Keppler-Noreuil, K., McKane, M., Maiers, J. L., Greiner, A. et al. (2013). Functional analysis of a de novo ACTB mutation in a patient with atypical Baraitser-Winter syndrome. *Human Mutation*, *34*(9), 1242-1249.

Kudryashov, D. S., Grintsevich, E. E., Rubenstein, P. A., & Reisler, E. (2010). A nucleotide state-sensing region on actin. *J. Biol. Chem.*, *285*(33), 25591-25601.

Kudryashov, D. S., & Reisler, E. (2013). ATP and ADP actin states. *Biopolymers*, *99*(4), 245-256.

Oda, T., & Maéda, Y. (2010). Multiple Conformations of F-actin. *Structure*, *18*(7), 761-767.

Oda, T., Iwasa, M., Aihara, T., Maéda, Y., & Narita, A. (2009). The nature of the globular- to fibrous-actin transition. *Nature*, *457*(7228), 441-445.

Otterbein, L. R., Graceffa, P., & Dominguez, R. (2001). The crystal structure of uncomplexed actin in the ADP state. *Science*, *293*, 708-711.

Pollard, T. D. (1986). Rate constants for the reactions of ATP- and ADP-actin with the ends of actin filaments. *J. Cell Biol.*, *103*, 2747-2754.

Porta, J. C., & Borgstahl, G. E. O. (2012). Structural basis for profilin-mediated actin nucleotide exchange. *Journal of Molecular Biology*, *418*(1-2), 103-116.

Rubenstein, P. A., & Wen, K. K. (2014). Insights into the effects of disease-causing mutations in human actins. *Cytoskeleton (Hoboken)*, *71*(4), 211-229.

Schwyter, D., Phillips, M., & Reisler, E. (1989). Subtilisin-cleaved actin: polymerization and interaction with myosin subfragment 1. *Biochemistry*, *28*(11), 5889-5896.

Splettstoesser, T., Noé, F., Oda, T., & Smith, J. C. (2009). Nucleotide-dependence of G-actin conformation from multiple molecular dynamics simulations and observation of a putatively polymerization-competent superclosed state. *Proteins*, *76*(2), 353-364.

Strzelecka-Gołaszewska, H., Moraczewska, J., Khaitlina, S. Y., & Mossakowska, M. (1993). Localization of the tightly bound divalent-cation-dependent and nucleotide-dependent conformation changes in G-actin using limited proteolytic digestion. *Eur. J. Biochem.*, *211*(3), 731-742.

Vorobiev, S., Strokopytov, B., Drubin, D. G., Frieden, C., Ono, S., Condeelis, J. et al. (2003). The structure of nonvertebrate actin: Implications for the ATP hydrolytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 5760-5765.

Zheng, X., Diraviyam, K., & Sept, D. (2007). Nucleotide effects on the structure and dynamics of actin. *Biophysical Journal*, *93*(4), 1277-1283.

Chapter 4

Filament Dynamics and the Effects of Polymerization on Actin

The proper regulation of actin filament (F-actin) dynamics is key for proper cellular and physiological function. Filament dynamics are regulated through a wide array of actin binding proteins – these proteins help regulate polymerization, depolymerization, branching, as well as the pool of unpolymerized G-actin monomers, and they influence the nucleotide state of actin itself. Actin filaments polymerize in a head-to-tail fashion, meaning that the filament is polar and the two filament ends are intrinsically different. The barbed end of the filament polymerizes more quickly, and the pointed end dissociates faster. As of right now, neither a F-actin crystal structure nor high resolution cryo-EM structures of either filament end are available, limiting our knowledge of important changes responsible for end-dependent dynamics. In this chapter, we use molecular dynamics simulations of two F-actin models to explore filament dynamics. We show that the filament's pointed end is significantly flatter than the remainder of the filament and is in a conformation not sampled by any of our other G or F-actin simulations. We also show that the barbed end of the filament takes on a conformation more similar to the ATP monomer – enhancing

ATP G-actin's ability to polymerize. Our observations in filament conformation help to answer many of the open questions in the field about F-actin dynamics (Pollard, 2017).

4.1 Introduction

Actin filaments are essential components of the eukaryotic cell and central to many processes including cell migration, cell division, and endocytosis. In order to accurately regulate and execute these functions, cells must be able to closely control filament dynamics. The ability of the cell to cycle and switch between periods of filament assembly and disassembly is accomplished through two primary means. The first is the large family of actin-associated proteins that function to nucleate, cap, branch or sever filaments, or otherwise regulate the pool of monomeric actin. The second means of control comes from the nucleotide state of the actin itself. Actin is an ATPase and the activity of actin depends on whether the bound nucleotide is ATP or ADP. This feature is important for both actin monomers and filaments, as well as the associated proteins that interact with them.

The actin filament, or F-actin, is formed by the polymerization of G-actin monomers. The actin subunits in the filament (protomers) are assembled in a head-to-tail fashion giving rise to a polar filament with distinct ends. It has long been known that the barbed end of the filament assembles faster than the pointed end, and also that ATP G-actin polymerizes faster the ADP G-actin at both ends of the filament (Pollard, 1986; Fujiwara et al., 2007) but the structural basis for these functional differences is unclear. Once polymerized, the ATP in the actin protomer is hydrolyzed into ADP-P_i, and ultimately the phosphate dissociates, leaving ADP. The kinetics of polymerization depend on the available pool of G-actin, but this rate is typically faster than the kinetics of hydrolysis and phosphate release, resulting in a filament with distinct regions of ATP, ADP-P_i and ADP. This feature is exploited by numerous proteins such as the Arp2/3 complex,

which preferentially binds to ATP portions of the filament, or ADF/cofilin which interacts more strongly to ADP F-actin.

In order to understand the molecular basis for these functional differences, we can turn to structural biology where we now have hundreds of crystal structures for actin and actin-like proteins. In the case of the actin filament, crystallization has not been possible, but techniques such as x-ray fiber diffraction (Oda, Iwasa, Aihara, Maéda, & Narita, 2009) and cryo-electron microscopy (Fujii, Iwane, Yanagida, & Namba, 2010; Galkin, Orlova, Schröder, & Egelman, 2010; Galkin, Orlova, Vos, Schröder, & Egelman, 2015; Von Der Ecken et al., 2015; Merino et al., 2018) have been successful in giving us filament structures with resolutions in the 3.3-7 Å range. These pieces of structural data have given us important insight into the structure-function relationship for actin, but simulations using these structures have provided another level of information by elucidating new details about dynamics, variance in secondary or tertiary structure, and other features. Much of the simulation work has been done through atomically-detailed molecular dynamics simulations, but a lot has also been accomplished using different levels of coarse-graining approaches (Chu & Voth, 2005; Chu & Voth, 2006; Pfaendtner, Lyman, Pollard, & Voth, 2010; Saunders et al., 2014).

In this chapter, we use all-atom molecular dynamics simulations of two F-actin filaments built from the Oda and the Fujii models to help elucidate the structural reasons for differing end dynamics, as well as the differences between the monomer and the protomer. Our simulations provide new insight into why ADP-actin binds slower than ATP-actin at the barbed ends, the slow association of subunits at actin's pointed end, as well as provides an explanation for why actin's affinity for phosphate is lower at the pointed end of the filament.

4.2 Results

4.2.1 Oda and Fujii Filament Models

Unlike its monomeric counterpart, no crystal structures exist for the F-actin filament requiring us to work with lower resolution F-actin structures derived from x-ray fiber diffraction and/or cryo-electron microscopy (Oda et al., 2009; Fujii et al., 2010; Galkin et al., 2010; Galkin et al., 2015; Von Der Ecken et al., 2015; Merino et al., 2018). We constructed 8-protomer canonical filaments using the models from both Oda *et al*. (2009) and Fujii *et al*. (2010). In a canonical filament, a rise of 27.6 Å and a twist of -166.4° exists between protomers (Figure 4.1). The filaments were simulated for a total of 650 and 550 ns respectively. Unlike many of the other F-actin simulations that have been performed, we did not use periodic boundary conditions to allow for both freely fluctuating ends and to prevent 'fixing' the filament twist.

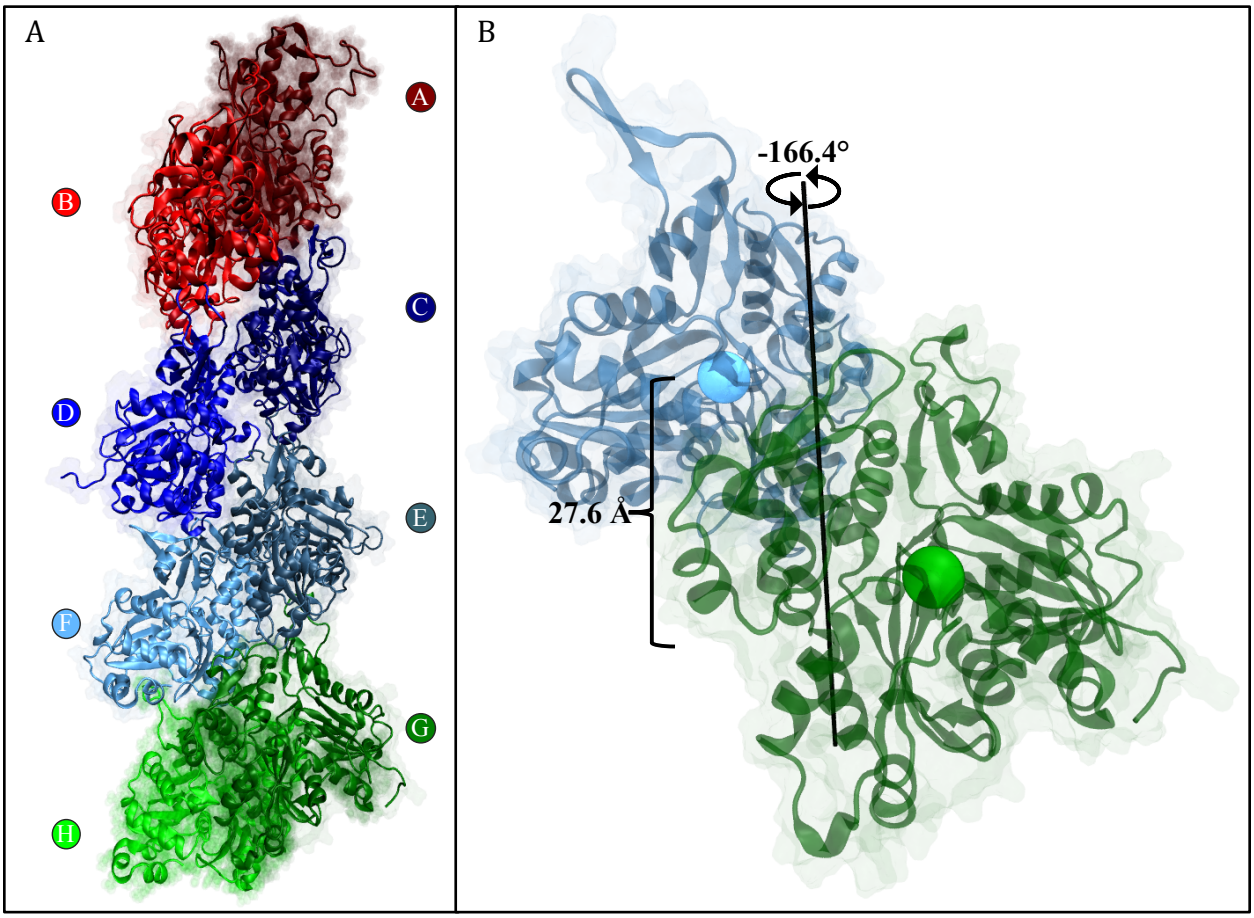


Figure 4.1 Filament Structure (A) For our simulations we built eight protomer long canonical filaments. Here, the pointed end is in red, the middle of the filament in blue, and the barbed end of the protomer is in green. (B) In a canonical filament, protomers are separated by a twist of -166.4° and a rise of 27.6 Å, spheres represent centers of mass.

As an initial analysis, we measured the twist and offset between protomers in the filaments using a modified Kabsch rotation matrix method. While this method does not perfectly capture the movement from one protomer to the next as it fails to capture any orientation changes, such as tilt, it does give an unbiased measurement of filament axis and twist. We found that our Oda F-actin simulation converged to canonical filament like parameters, with an average filament twist of -165.4 ± 0.8° and offset of 27.9 ± 0.1 Å (Figure 4.2 A). In contrast, the Fujii filament failed to converge, it was under twisted (-168.6 ± 0.5°) and had a short offset (26.9 ± 0.1 Å).

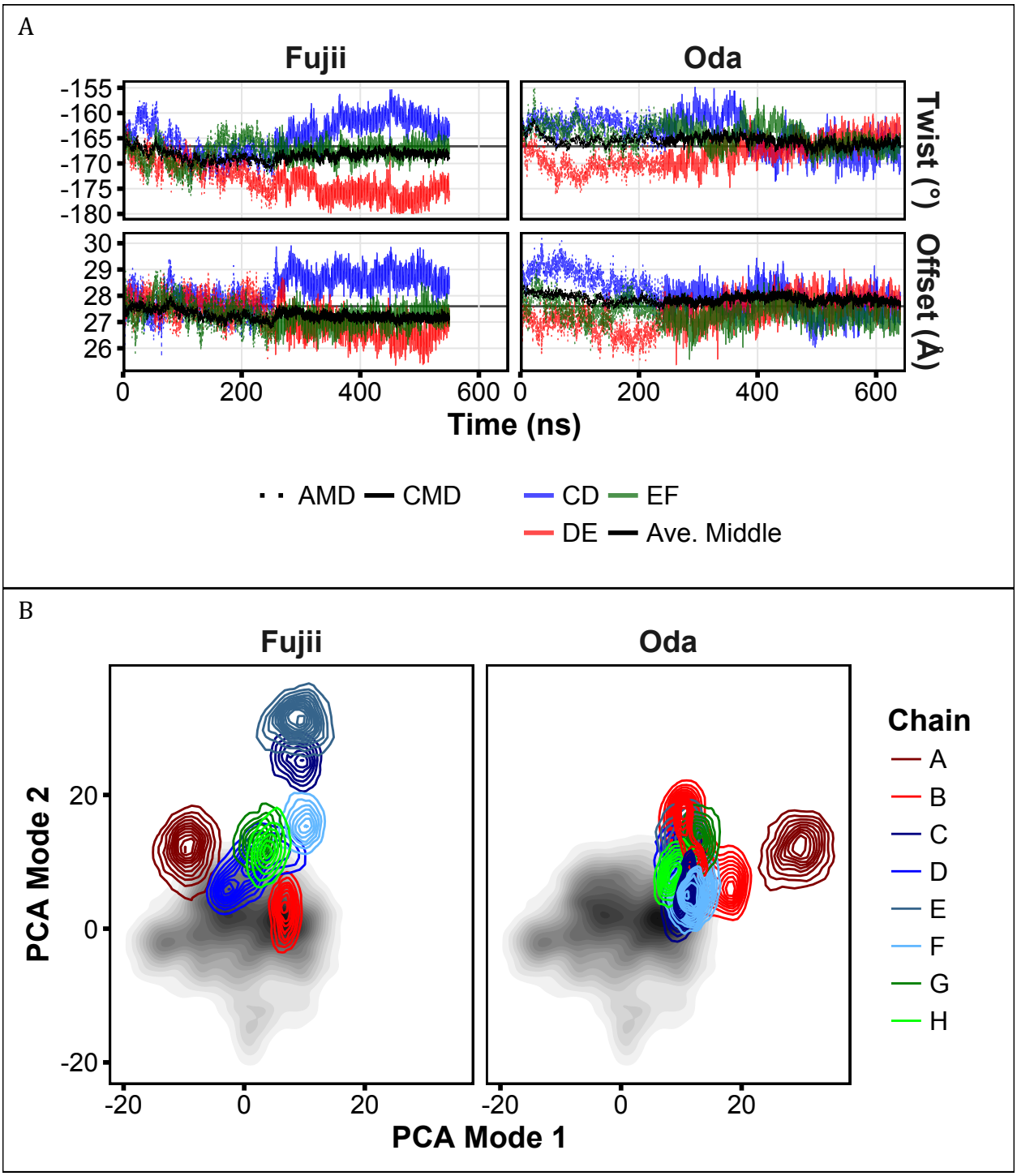


Figure 4.2 Filament Conformation (A) Protomer twist and offset between the middle four protomers of the Fujii and Oda F-actin models measured over the course of all simulations. (B) Projection of individual protomers from the middle of our Fujii and Oda filaments (cMD) onto our G-actin basis set. The Fujii protomers exhibit a high degree of structural variability, whereas the Oda protomers converge.

When we projected the conformation of our F-actin protomers from the middle of the filament on to our ADP G-actin PCA space, we see that the Oda protomer conformations are quite

homogeneous and remain fairly consistent with the initial starting point (Figure 4.2). The Fujii protomers also remain fairly consistent with the initial starting point but shows a high degree of structural variability. A comparison with the available F-Actin models (Figure 4.3) shows that while protomer conformation diverges in our Fujii filament, the divergent conformations are in line with Galkin's (2015) T1 and T2 filament models (PDBs 3J8J and 3J8K). In the tilted T1 and T2 states (solved at a resolution of ~12 Å) subunits have a more open cleft, similar to what was observed G-actin in the open state (Chik, Lindberg, & Schutt, 1996; Galkin et al., 2015). The protomers C and E within our Fujii model, that map to this state, have a phosphate clamp size of 8.2 ± 0.4 Å (± SD) and 9.9 ± 0.6 Å respectively, matching the definition of an open conformation (Dalhaimer, Pollard, & Nolen, 2008). It is also possible that the Fujii model's lack of convergence results from artifacts or distortions in the structure. During structural determination, conditions were chosen to create the straightest filaments possible. This straightening results from large forces on the filament (Egelman, 2012). It is possible that the large forces applied to straighten the filament prevented it from relaxing to a canonical filament structure over a reasonable amount of simulation time. As our Oda model had no such issues and we used this model for all subsequent analysis.

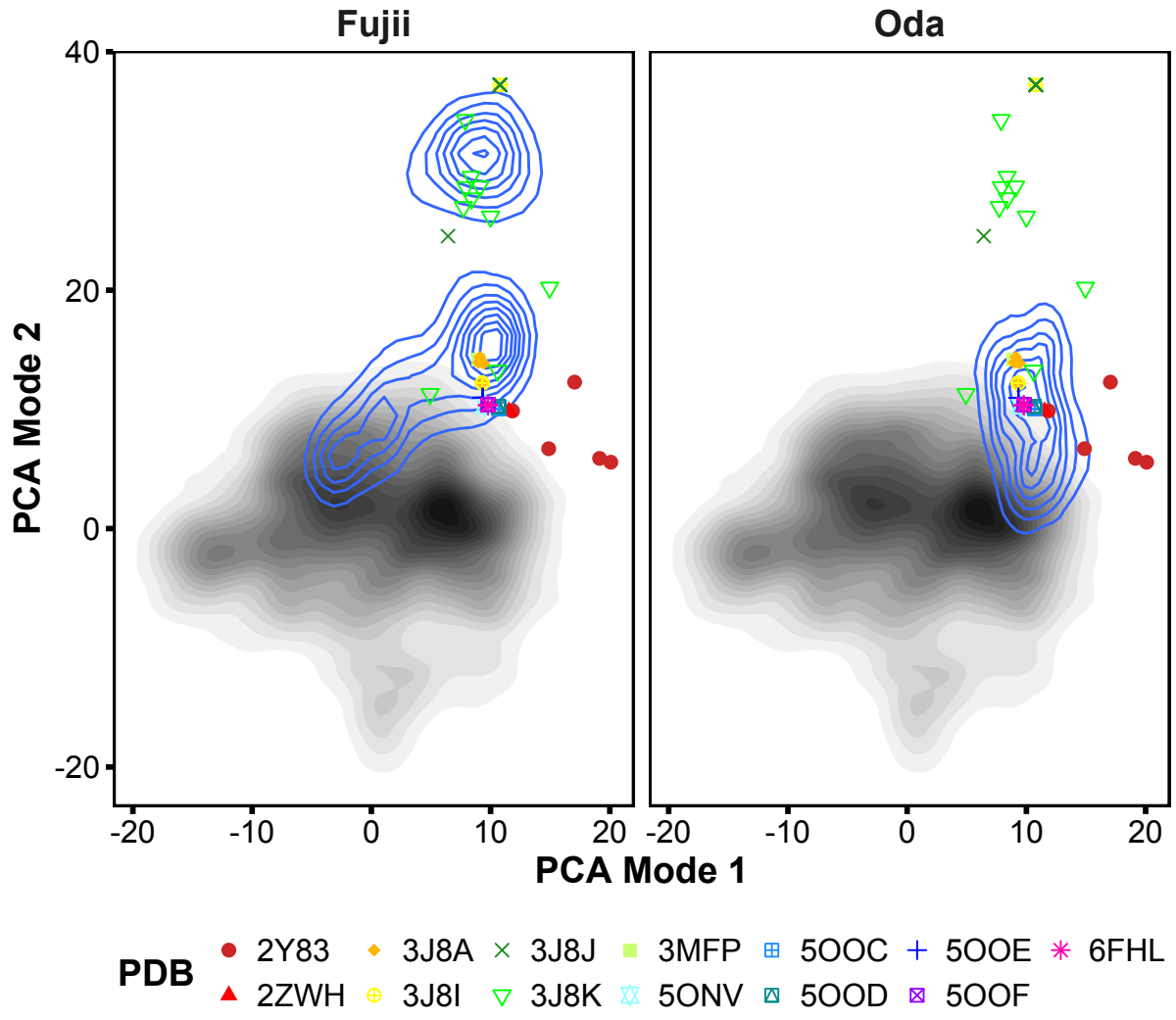


Figure 4.3 F-actin Simulations and F-Actin Models Our F-actin simulations show less deviations from their crystal starting points, especially our Oda filament model based off of PDB 2ZWH. It should be noted that while our Fujii model does drift from its starting point and show divergent protomer conformation, there is some overlap with Glakin's T1 and T2 filament models (PDBs 3J8J and 3J8K).

4.2.2 Barbed and Pointed End Differences

One of our primary goals in this work was to elucidate structural or dynamical differences between the two ends of the filament. The Voth lab has a number of publications on F-actin simulations (Chu & Voth, 2005; Chu & Voth, 2006; Pfaendtner et al., 2010; Saunders et al., 2014), but in all cases these simulations were done with periodic boundary conditions such that the filament appears like an infinite filament. This system has been useful for deriving many

mechanical features of the filament, however the periodic boundary conditions have the effect of fixing the mean twist of the filament and clearly cannot allow for freely fluctuating ends. Filament end structures are important, for understanding polymerization. To date, no high-resolution structures exist for the filament ends.

In terms of the overall filament geometry, we see that barbed and middle ends of the filament have very similar twist values, whereas the pointed end of the filament is divergent. The barbed end of the filament has an average twist of -165.3 ± 3.9° (± SD), which is statistically no different from the middle of the filament twist of -165.6 ± 1.2° (p-value of 0.22). The pointed end of the filament is slightly under-twisted with a value of -167.5 ± 3.0° (p-value of 3.95e-30). The twist between the pointed end penultimate and antepenultimate protomers shows the opposite trend and is over-twisted with a value of -162.7 ± 3.4°. Together, the two average to near that of the middle of the filament (-165.1 ± 1.5°). It is possible that the end dimer conformation could serve to block incoming protomers from addition. This hypothesis is similar to the argument proposed by Narita *et al.* (2011), they found that the terminal actin subunit was tilted over the penultimate subunit in such a way as to partially blocked the binding site for incoming subunits (Narita, Oda, & Maéda, 2011).

If we look at the protomers structure at the barbed and pointed ends, we again see differences from the middle of the filament (Figure 4.4). The propeller twist of the actin shows the largest difference. The four protomers in the middle of our filament have an average propeller twist of -10.8 ± 3.1°. The barbed end protomers are slightly flatter at -9.6 ± 2.3° while the pointed end is significantly flatter with a propeller twist of -2.4 ± 4.9°. Separating this out into individual subunits, we see that the terminal pointed end protomer had a twist of 2.1 ± 2.3° and the penultimate protomer had a twist of -6.9 ± 1.8°. While the penultimate subunit's twist is still flatter

than the remainder of the filament, it is 8.9° more twisted than the terminal subunit. This difference in propeller-twist values between the pointed end protomers is similar to the tilt observed in the cryo-EM structure of the pointed end (Narita et al., 2011).

If we look at the phosphate clamp and the nucleotide cleft mouth, the middle of the filament has values of 6.9 ± 1.5 Å and 9.3 ± 1.3 Å respectively. The barbed end protomers are more closed by both measures, with clamp distances of 6.1 ± 0.4 Å (p-value of 4.2e-88) and nucleotide cleft mouth distance of 8.4 ± 0.2 Å (p-value of 2.8e-138). The pointed end has a clamp distance that is more similar to the middle (6.7 ± 1.3 Å, p-value of 0.003), but the nucleotide cleft is much more open with an average distance of 10.8 ± 2.6 Å (p-value of 7.4e-54). Given that the nucleotide cleft is oriented toward the pointed end, this more open conformation might be expected since subdomains 2 and 4 are not constrained by additional protomers. The more open phosphate clamp at the pointed end is consistent with the finding that phosphate release is faster at the pointed end (Fujiwara et al., 2007).

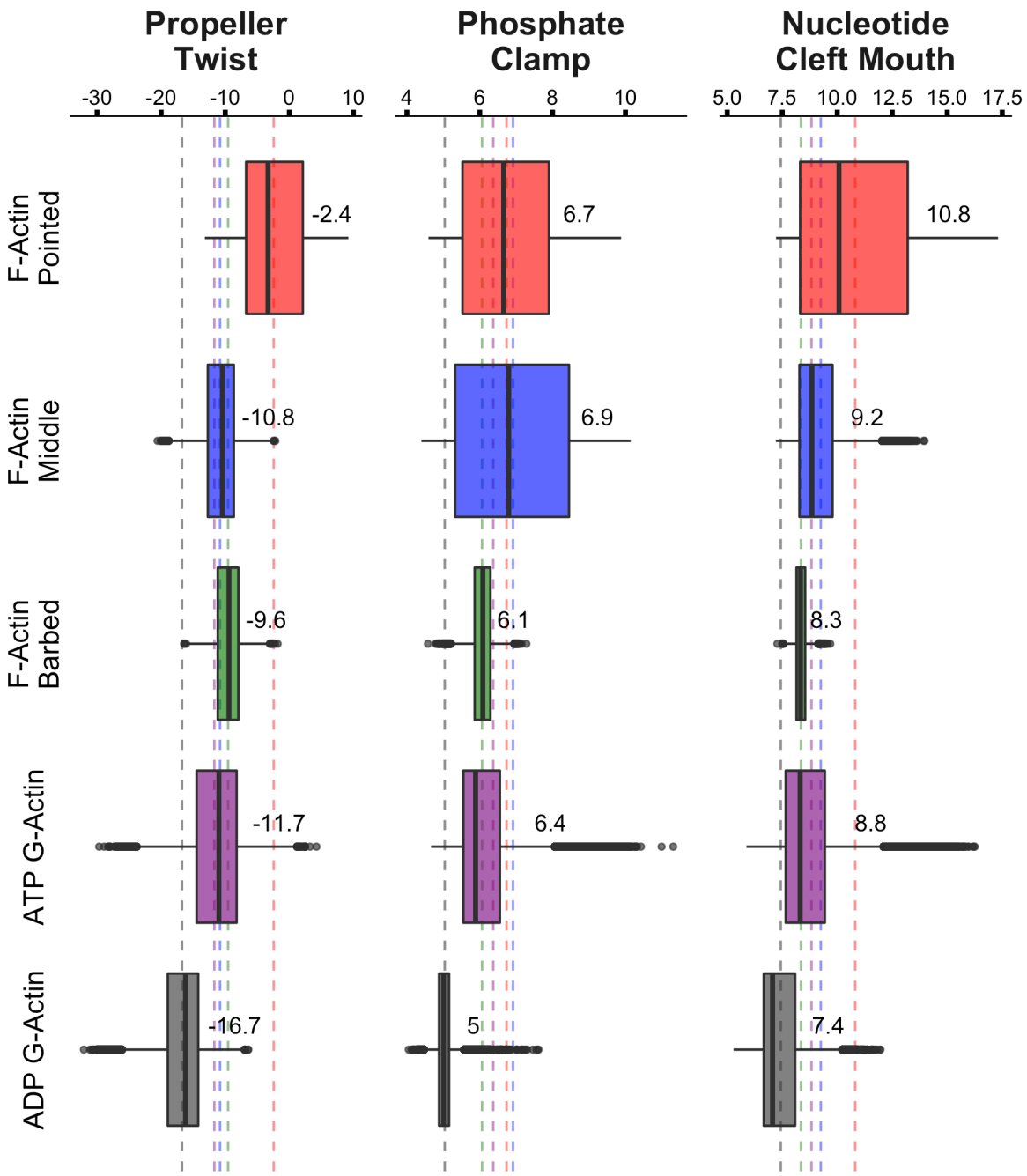


Figure 4.4 Actin's Propeller Twist, Phosphate Clamp, and Nucleotide Cleft Mouth Sizes Numbers on plot are the means of the distributions. For the pointed, middle and barbed ends of the filament represent, the means represent the averages of protomers A-B, C-F, and G-H respectively.

If we project the barbed and pointed end protomers on to our PCA space (Figure 4.5), we clearly see how the barbed end conformations are much closer to the G-actin structures than what

is seen for the pointed end. This has implications for the polymerization at the two ends of the filament, discussed below.

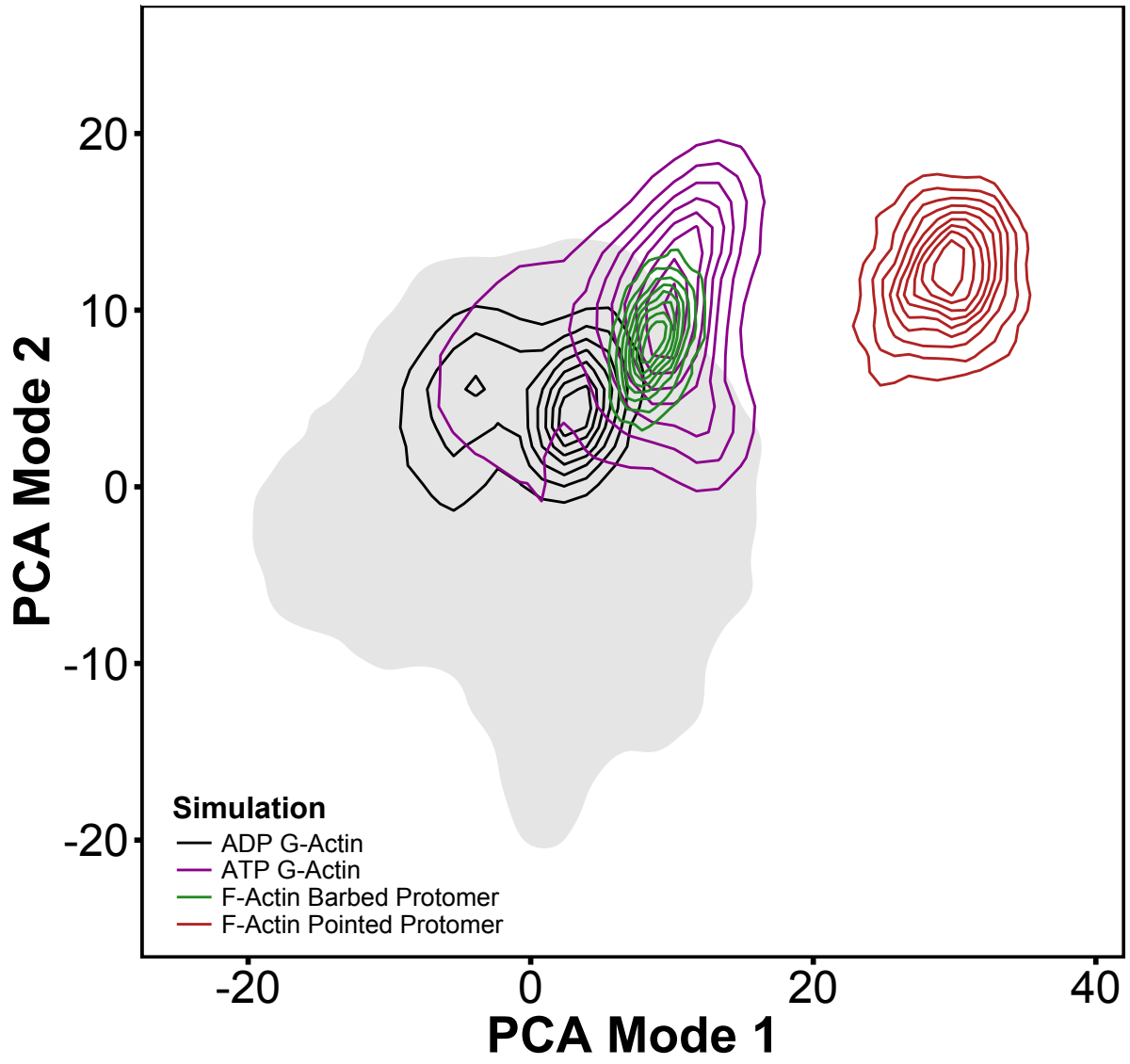


Figure 4.5 Projection of F-Actin's Barbed and Pointed Ends onto the G-actin PCA Basis Set F-actin's barbed end protomer (green) is shifted towards that of the G-actin monomers, specifically in the ATP state (magenta). The pointed end protomer (red) is shifted away from the remainder of the simulations.

4.2.3 Loop Dynamics

To examine the effects of the filament, the PCA on loops analysis was performed again with the inclusion of the middle four protomers of our filament simulation. For the vast majority

of loops, such as the FQQ and C-terminus, the ADP protomer looked like that of the ATP monomer (Figure 4.6).

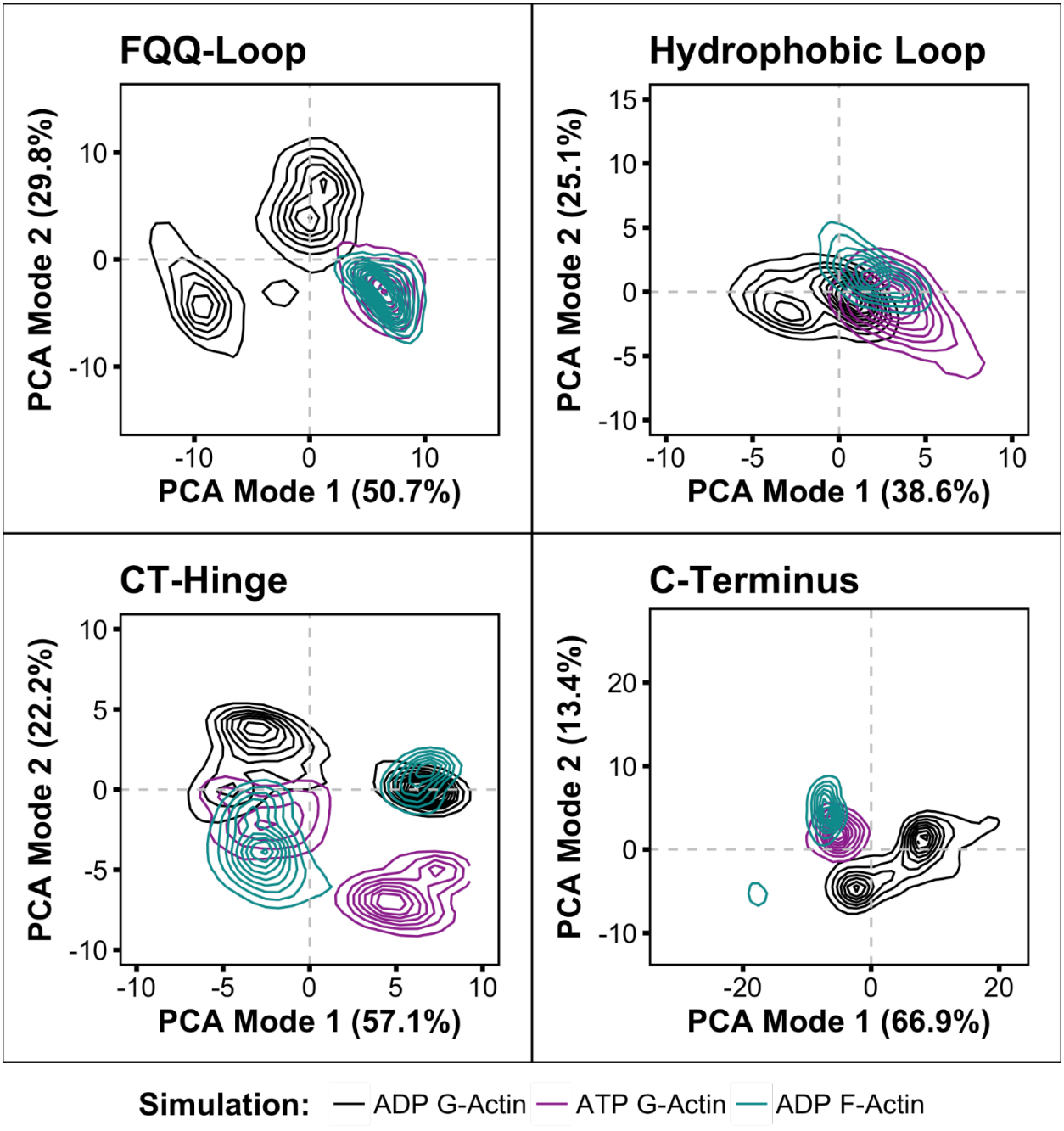


Figure 4.6 F-Actin PCA on Loops Shows Similarity to ATP G-Actin F-actins FQQ-loop (A), hydrophobic loop (B) and C-terminus (D) all take on conformations similar to that of the ATP G-actin monomer. F-Actin's CT-hinge resembles that of both the ATP and ADP G-actin protomers.

4.2.3.1 S, G, and H73 Loops

The S-, G- and H73-loops have protomer specific dynamics, with each protomer representing the middle of the filament having unique dynamics (Figure 4.7). Protomers D and E take on conformations that are similar to what we see in the ATP monomer, while protomer C and F are more divergent (see Figure 4.1A for protomer assignments). In protomer C, the H73 loop pulls out away from the monomer. It is possible that these altered loop dynamics play a role in the increased rate of nucleotide hydrolysis within the filament. The lack of convergence of these loops may also suggest that we have not had enough simulation time to allow these loops to converge to a common conformational space.

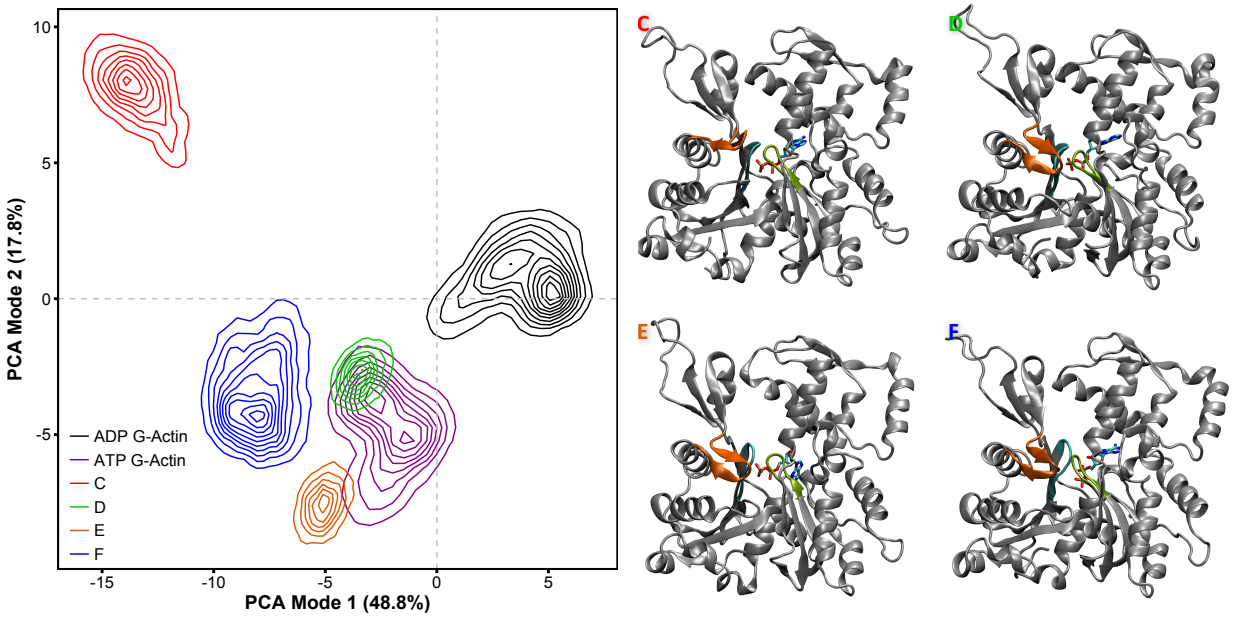


Figure 4.7 S, G, H73-Loop Filament Conformations The S, G, and H73-loops in the middle protomers of F-actin each take on a unique conformation.

4.2.3.2 Flap

The only other region that showed distinct protomer conformation was the flap in subdomain 4 (L236-G251). In the filament, the flap region is pulled up away from the rest of the monomer, with the loop region (L242-Q246) making longitudinal filament contacts (Figure 4.8).

Mutations to this region of the protein in muscle actin cause NEM3 or DFNA20/26 if found in γ-cytoplasmic actin (Agrawal et al., 2004; Morín & Bryan, F Goodyear, R Mencía, A Modamio-Høybjør, S del Castillo, I Cabalka, JM Richardson, G Moreno, F Rubenstein, PA Moreno-Pelayo, MA, 2009). Both kinds of mutation have been hypothesized to disrupt actin filament function leading either to growth defects or an accumulation of actin filaments, suggesting that this region plays an important role in both filament formation and filament stability.

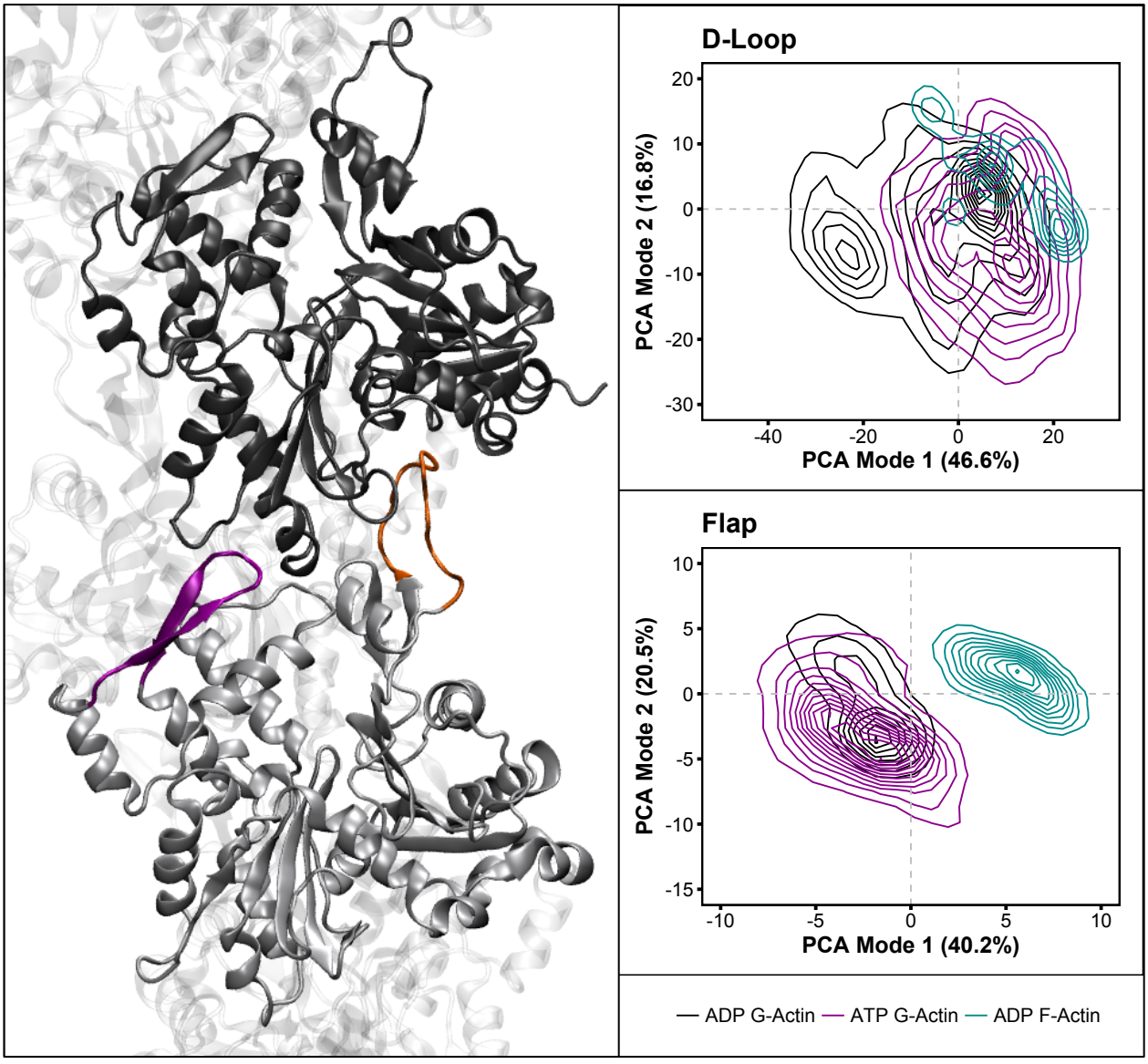


Figure 4.8 F-Actin flap and D-loop inter-protomer contacts Actin's flap and D-loop sit at the pointed end of the monomer and make lateral filament contacts within the filament. As such, the dynamics of the D-

loop, as seen in the loop PCA, are reduced compared to the monomer. While the flap does not have reduced dynamics, its dynamics are different from either monomer. The flap gets pulled up away from subdomain 4 towards the next lateral protomer.

4.2.3.3 Hydrophobic Loop

When the Holmes model was first proposed in 1990, it was postulated that the hydrophobic loop (also known as the hydrophobic plug or subdomain 3/4 loop - S265-G273) interacted with the target-binding cleft of a neighboring protomer and was thus key for filament stability (Holmes, Popp, Gebhard, & Kabsch, 1990). The idea was that the hydrophobic loop swung out away from protein and became locked in the target-binding cleft. This theory has been tested extensively over the intervening years. Shvetsov et al. (2002) showed that by locking the protein backbone of the loop you could inhibit polymerization. Orlova et al. (2004) further demonstrated that by crosslinking the loop within the target-binding cleft led to filament destabilization. These experiments seem to confirm the importance of the hydrophobic loop in filament stability. Indeed, in our simulations of the Oda F-Actin model, we see that there is a stable salt-bridge being formed between R39 in the D-loop and E270 in the H-loop. This is one of just a handful of interprotomer salt-brides formed within the filament.

While the evidence suggests that hydrophobic loop interactions are important for filament stabilization, the evidence does not suggest that there is a substantial conformational change of the loop upon polymerization. Interspin distances were similar for both G- and F-actin (Scoville et al., 2006). Additionally, EPR data shows that the loop is highly mobile in both the G- and F-actin states (Shvetsov et al., 2006). Our loop PCA data further supports that there is minimal change in the hydrophobic loop upon transition from G- to F-actin (Figure 4.6). EPR data did indicate that polymerization causes a substantial decrease in mobility for residues 267 and 269, as well as, to a lesser degree, a decrease in mobility for residues 265 and 266 (Scoville et al., 2006). In

concordance with this observation, we see a decreased RMSF of these residues in our F-Actin model compared to our G-Actin simulations (Figure 4.9).

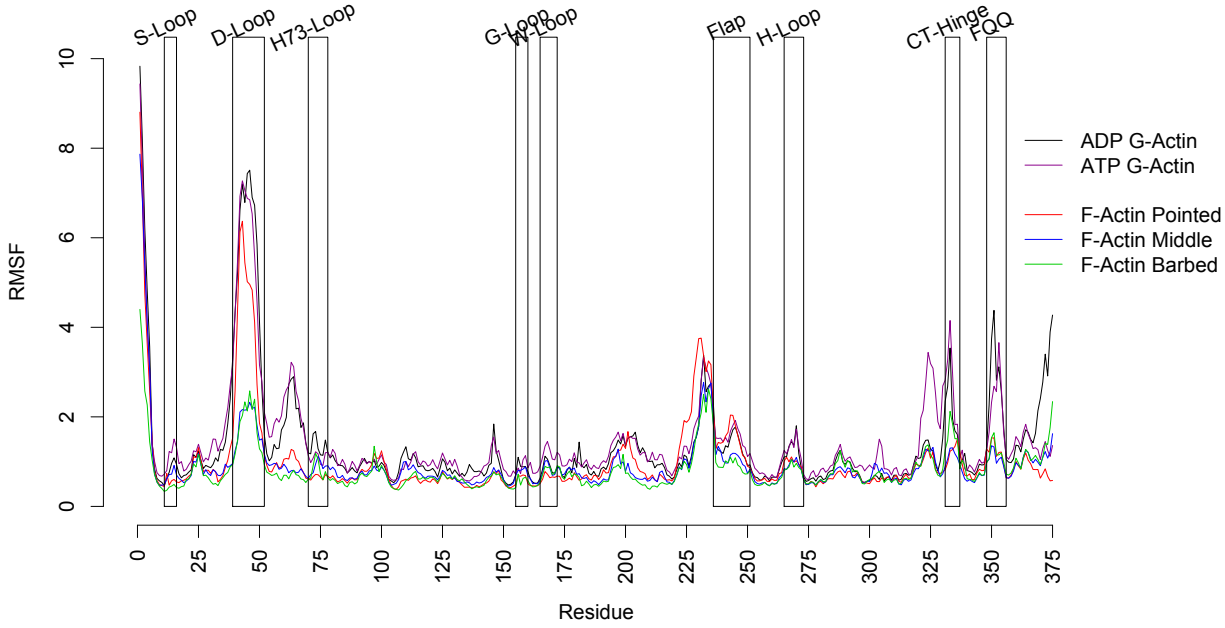


Figure 4.9 F-Actin RMSF by End The pointed end represents average of the last two protomers (A and B), the middle the average RMSF of the middle four protomers (C to F), and the barbed end the average RMSF of the terminal two protomers (G and H). Different fluctuations are present throughout the filament, with the pointed end showing more variability in subdomains 2 and 4.

4.2.3.4 D-Loop

Within the filament, the D-loop serves as an anchoring point between protomers. As a result, its dynamics are largely constrained compared to that of the monomer, this can be seen in our PCA on loops analysis (Figure 4.8), as well as with the decreased RMSF present in both the middle and barbed end D-loops (Figure 4.9). The D-loop remains unstructured in both the middle and barbed end filament protomers forming an extended coil structure which maximizes the lateral contacts with the target binding cleft of adjacent protomers. At the filaments pointed end, where the D-loop is free, we see increased dynamics, returning the RMSF to near that of the monomers (Figure 4.9), as well as the formation of either a short beta strand or helical structure (Figure 4.10).

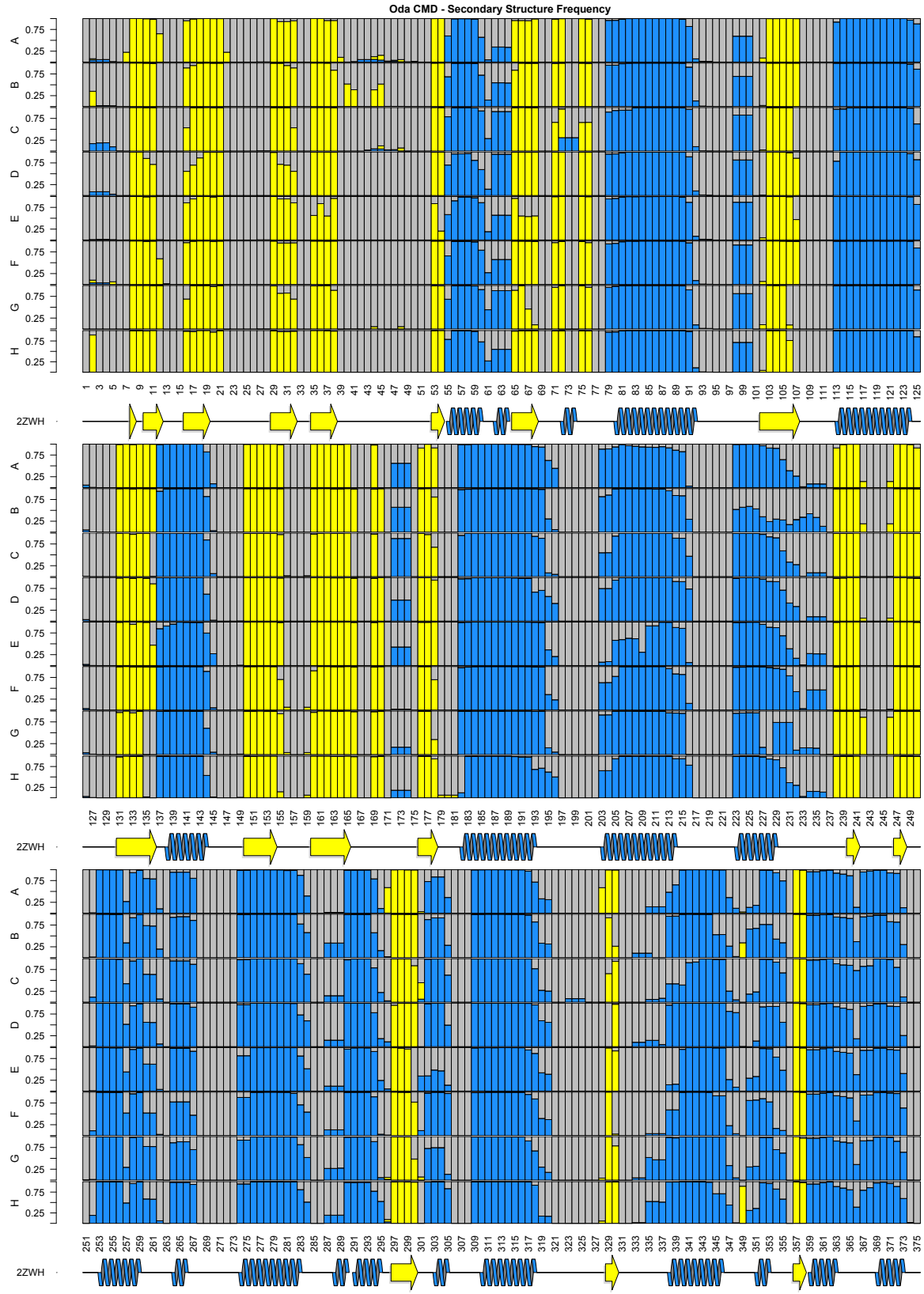


Figure 4.10 F-Actin Secondary Structure by Protomer While secondary structure is largely consistent throughout the filament, some protomers do exhibit variation. Here the pointed end is represented by

protomer A and the barbed end is represented by protomer H. The yellow represents β-strands, while the blue represents helices.

4.2.4 Monomeric Actin vs Barbed End - Energetics

Now that we have an understanding of how the filament ends differ as well as the differences between the nucleotide states of G-actin, we can put those pieces together to examine how this results in the kinetic and thermodynamic differences in barbed versus pointed end polymerization. In an idealized view, the polymerization process simply involves monomers adding to the end of the filament without any intermediate steps. However, as a consequence of the different conformations between the monomer and the protomers at the ends of the filament, G-actin must undergo a subsequent isomerization step after coming in contact with the end of the filament. The isomerization step has implications on both the thermodynamics and the kinetics, and since we have conformational differences in both the barbed and pointed ends, as well as differences between ADP and ATP G-actin, we expect a multiplicity of isomerization effects.

We have not explicitly modeled the process of a monomer interacting with the end of the filament, however our equilibrium distributions for the ADP and ATP monomers do show overlap with the barbed end protomers, and this allows us to estimate the penalty imposed by isomerization. Using inverse Boltzmann weighting, we can calculate the thermodynamic cost of an ADP or ATP monomer assuming the structure of the barbed end protomer (Figure 4.11). Since all other energetic contributions to the polymerization reaction should be the same (non-covalent bonds formed with the filament, removal of bound waters, etc.), this should represent the difference solely due to the nucleotide state of the monomer. For ADP-G-actin, we find the difference from the equilibrium monomer structure to the equilibrium barbed end conformation is 2.86 k_BT, however if we calculate the difference between randomly sampled points in both the monomer and barbed protomer distributions, the mean value (of points less than 10 k_BT) comes to

1.96 ± 1.69 k_BT (SD). The same sampling method for the ATP monomer produces a mean difference of -0.83 ± 1.04 k_BT. The difference between the values 2.78 k_BT and this should represent the difference in binding free energy between the two nucleotide states. The critical concentration for ATP G-actin at the barbed end is 0.12 μM while ADP G-actin has a value of 1.8 μM, (Pollard, 1986; Fujiwara et al., 2007) and this translates into a free energy difference of 2.7 k_BT, consistent with our findings. At the same time, we can think about the kinetics of this process, and using simple arguments from transition state theory, this addition isomerization barrier should decrease the kinetic rate constant by a factor of exp(2.78), or 16.12. In the case of the pointed end, the monomer and protomer distributions are much more separated (Figure 4.5) and we cannot perform this same inverse Boltzmann analysis, but this separation is entirely consistent with the decrease in both the free energy and association rate constants.

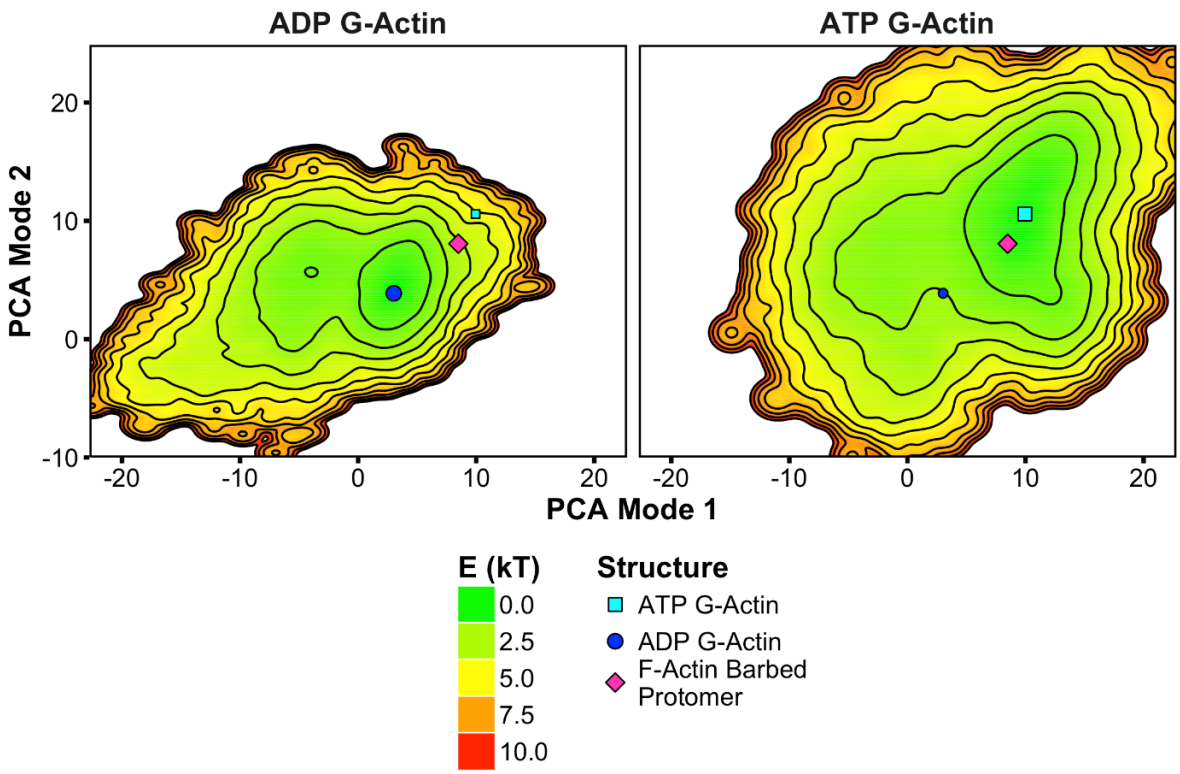


Figure 4.11 Energetics The inverse Boltzmann distributions of the PCA projections of the ADP and ATP G-actin. An energy barrier exists between the ADP monomer and the ADP F-actin barbed protomer, where there is no energy difference between the ATP monomer and the barbed protomer.

4.3 ATP Filament

In Chapter 3 I discussed the importance of G-actin's nucleotide state for determining its affinities for actin binding proteins. The same holds true for F-actin. ATP subunits add onto the barbed end of the filament and then are subsequently hydrolyzed to ADP-P_i, and finally the phosphate is released, leaving ADP filament. This leads to distinct filament regions of ATP, ADP-P_i and ADP subunits. Actin binding proteins take advantage of these properties and bind to different regions of the filament. To study how ATP affects filament dynamics, we took the end frame of our Oda filament simulation and converted all ADPs to ATP.

Filament simulations take time, and as of now we only have an initial 40 ns of aMD simulation to show. While this is not much, and our simulation has yet to converge, we can see

some trends emerge. First, in terms of protomer conformation there seems to be an opening up of the nucleotide cleft as seen by the shift in PCA space to higher PC2 values (Figure 4.12). Next, in terms of protomer arrangement, the filament is becoming more twisted. In the ADP Oda filament, the average filament twist was -165.4 ± 0.8°, in our ATP filament this has increased to -166.3 ± 0.6°.

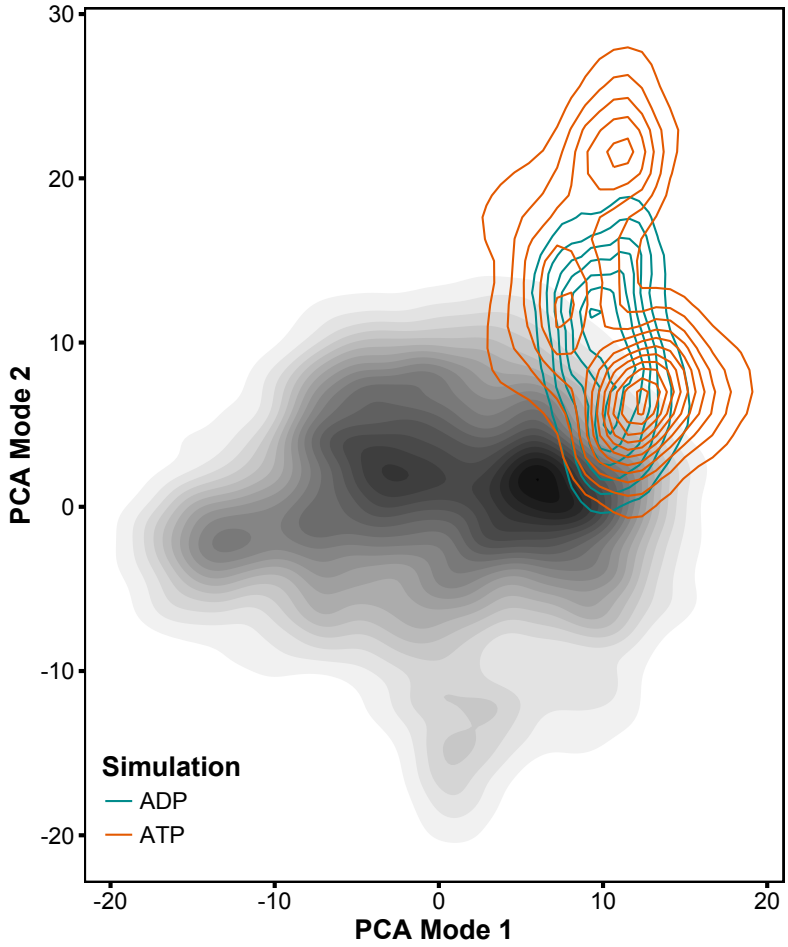


Figure 4.12 ATP F-Actin The middle four protomers of our initial 40 ns aMD simulation of ATP F-actin was projected onto our PCA space. There is some protomers are shifted to higher PC2 values indicating a more open nucleotide cleft.

4.4 Discussion

F-Actin filaments form in a head-to-tail fashion, leading to polar, directional filaments with intrinsic end differences. The barbed end of the filament, where subdomains 1 and 3 are exposed,

is where incoming actin monomers preferentially bind in a process called polymerization. The pointed-end, where subdomains 2 and 4 are exposed, has higher rates of subunit dissociation, or depolymerization. The question remains, however, as to the differences of the barbed and pointed ends that lead to their unique properties (Pollard, 2017). The structural reasons for these differences have been hard to pinpoint as no high-resolution structure exists for the filament ends.

The F-actin structure has been elusive and a crystal structure for the filament has yet to be captured leaving us to rely on filament models derived from x-ray fiber diffraction and cryo-electron microscopy. Two of the more well-known models are from Fujii *et al.* (2010) and Oda *et al*. (2009). These both represent structures from the middle of the filament. To probe end dynamics, we built 8 protomer long canonical filaments of each of these models and then simulated them using a combination of aMD and cMD techniques. We simulated our Fujii model for 550 ns and our Oda model for 650 ns. Unlike many of the F-actin simulations that have been done before, we did not use periodic boundary conditions to create infinite filaments, thus allowing for freely fluctuating filament ends.

Over the course of our simulation time, the Fujii model failed to converge (Figure 4.2), resulting from presence of filament artifacts or distortions caused by the extreme forces used to orient filaments during the cryo-electron microscopy process (Egelman, 2012). For this reason, the Fujii model was excluded from all further analysis. While the Oda model presented better convergence, many structural states were present within the filament. This is consistent with what Egelman and others have observed (Egelman, Francis, & Derosier, 1982; Galkin et al., 2015). In particular, the pointed end protomer diverged from the remainder of the filament (Figure 4.5), taking on a significantly flatter conformation with a more open nucleotide cleft mouth (Figure 4.4).

We believe that this dissimilar conformation imposes a high energy barrier limiting the addition of incoming subunits.

Unlike the pointed end protomer, the barbed end protomer converged to a conformation that was similar to monomeric G-actin. As our simulation PC spaces overlapped, we were able to use an inverse Boltzmann distribution to calculate the energy difference between the ADP and ATP monomers and the ADP barbed protomer. An energy difference of -0.83 ± 1.04 k_BT (SD) exists between the ATP monomer and the barbed end protomer, while a difference of 1.96 ± 1.69 k_BT exists between the ADP monomer and the barbed end protomer (Figure 4.11). Thus, the difference in binding free energies between the two monomer G-actin states is 2.78 k_BT. This value is consistent with the free energy difference of 2.7 k_BT derived from experiments (Fujiwara et al., 2007). Therefore, we would argue that the differences in polymerization kinetics between the ADP and ATP nucleotide states results from a difference in tertiary conformation presented in Chapter 3. Monomeric ATP G-actin is flatter than ADP G-actin and is in a conformation more similar to the ADP protomer (Figure 4.4).

In this chapter, we presented data resulting from our simulations of the F-actin filament, answering several of open questions within the actin community. Our simulations provided insight into the differing conformations of the barbed and pointed ends of the filament and how those conformations related to polymerization kinetics. We showed that the pointed end protomer is in a conformation that is dissimilar to the remainder of the filament and our monomeric G-actin simulations, imposing a conformational energy barrier inhibiting the addition of incoming subunits. Finally, we showed that the barbed end protomer, and actin protomers in general, are closer in conformation to the ATP monomer than the ADP monomer, essentially eliminating the conformational energy barrier for the addition of ATP monomers onto the end of the filament.

4.5 References

Agrawal, P. B., Strickland, C. D., Midgett, C., Morales, A., Newburger, D. E., Poulos, M. A. et al. (2004). Heterogeneity of nemaline myopathy cases with skeletal muscle α-actin gene mutations. *Annals of Neurology*, *56*(1), 86-96.

Chik, J. K., Lindberg, U., & Schutt, C. E. (1996). The structure of an open state of beta-actin at 2.65 Å resolution. *J. Mol. Biol.*, *263*(4), 607-623.

Chu, J., & Voth, G. A. (2005). Allostery of actin filaments: Molecular dynamics simulations and coarse-grained analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(37), 13111-13116.

Chu, J., & Voth, G. A. (2006). Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophysical journal*, *90*(5), 1572-1582.

Dalhaimer, P., Pollard, T. D., & Nolen, B. J. (2008). Nucleotide-Mediated Conformational Changes of Monomeric Actin and Arp3 Studied by Molecular Dynamics Simulations. *J. Mol. Biol.*, *376*(1), 166-183.

Egelman, E. H. (2012). *Structure and dynamic states of actin filaments* (4).

Egelman, E. H., Francis, N., & Derosier, D. J. (1982). F-actin is a helix with a random variable twist. *Nature*, *298*(5870), 131-135.

Fujii, T., Iwane, A. H., Yanagida, T., & Namba, K. (2010). Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, *467*(7316), 724-728.

Fujiwara, I., Vavylonis, D., & Pollard, T. D. (2007). Polymerization kinetics of ADP- and ADP-Pi-actin determined by fluorescence microscopy. *Proc. Natl. Acad. Sci. USA*, *104*(21), 8827-8832.

Galkin, V. E., Orlova, A., Schröder, G. F., & Egelman, E. H. (2010). Structural polymorphism in F-actin. *Nature Structural and Molecular Biology*, *17*(11), 1318-1323.

Galkin, V. E., Orlova, A., Vos, M. R., Schröder, G. F., & Egelman, E. H. (2015). Near-Atomic Resolution for One State of F-Actin. *Structure*, *23*(1), 173-182.

Holmes, K. C., Popp, D., Gebhard, W., & Kabsch, W. (1990). Atomic model of the actin filament. *Nature*, *347*(6288), 44-49.

Merino, F., Pospich, S., Funk, J., Wagner, T., Küllmer, F., Arndt, H. et al. (2018). Structural transitions of F-actin upon ATP hydrolysis at near-atomic resolution revealed by cryo-EM. *Nature Structural and Molecular Biology*, *25*(6), 528-537.

Morín, M., & Bryan, K. E. M.-M., F Goodyear, R Mencía, A Modamio-Høybjør, S del Castillo, I Cabalka, JM Richardson, G Moreno, F Rubenstein, PA Moreno-Pelayo, MA. (2009). In

vivo and in vitro effects of two novel gamma-actin (ACTG1) mutations that cause DFNA20/26 hearing impairment. *Human Molecular Genetics*, *18*(16), 3075-3089.

Narita, A., Oda, T., & Maéda, Y. (2011). Structural basis for the slow dynamics of the actin filament pointed end. *EMBO Journal*, *30*(7), 1230-1237.

Oda, T., Iwasa, M., Aihara, T., Maéda, Y., & Narita, A. (2009). The nature of the globular- to fibrous-actin transition. *Nature*, *457*(7228), 441-445.

Orlova, A., Shvetsov, A., Galkin, V. E., Kudryashov, D. S., Rubenstein, P. A., Egelman, E. H. et al. (2004). Actin-destabilizing factors disrupt filaments by means of a time reversal of polymerization. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(51), 17664-17668.

Pfaendtner, J., Lyman, E., Pollard, T. D., & Voth, G. A. (2010). Structure and dynamics of the actin filament. *Journal of Molecular Biology*, *396*(2), 252-263.

Pollard, T. D. (1986). Rate constants for the reactions of ATP- and ADP-actin with the ends of actin filaments. *J. Cell Biol.*, *103*, 2747-2754.

Pollard, T. D. (2017). *What we know and do not know about actin* (235).

Saunders, M. G., Tempkin, J., Weare, J., Dinner, A. R., Roux, B., & Voth, G. A. (2014). Nucleotide regulation of the structure and dynamics of G-actin. *Biophysical journal*, *106*(8), 1710-1720.

Scoville, D., Stamm, J. D., Toledo-Warshaviak, D., Altenbach, C., Phillips, M., Shvetsov, A. et al. (2006). Hydrophobic loop dynamics and actin filament stability. *Biochemistry*, *45*, 13576-13584.

Shvetsov, A., Musib, R., Phillips, M., Rubenstein, P. A., & Reisler, E. (2002). Locking the hydrophobic loop 262-274 to G-actin surface by a disulfide bridge prevents filament formation. *Biochemistry*, *41*(35), 10787-10793.

Shvetsov, A., Stamm, J. D., Phillips, M., Warshaviak, D., Altenbach, C., Rubenstein, P. A. et al. (2006). Conformational dynamics of loop 262-274 in G- and F-actin. *Biochemistry*, *45*(20), 6541-6549.

Von Der Ecken, J., Müller, M., Lehman, W., Manstein, D. J., Penczek, P. A., & Raunser, S. (2015). Structure of the F-actin-tropomyosin complex. *Nature*, *519*(7541), 114-117.

Chapter 5

Yeast G-Actin and the Allosteric Effects of Two Deafness-Causing Actin Mutations (DFNA20/26)

Point mutations in γ-cytoplasmic actin have been shown to result in autosomal dominant non-syndromic early-onset deafness. Two mutations at the same site, K118M and K118N, provide a unique opportunity to compare the effects of two dissimilar amino acid substitutions that produce a similar phenotype in humans. K118 resides in a helix that runs from K113 to T126, and mutations that alter the position, dynamics and/or biochemistry of this helix can result in a wide range of pathologies. Using a combination of computational and experimental studies[1], both using yeast actin, we find that these mutations at K118 result in changes in the structure and dynamics of the DNase-I loop, alterations in the structure of the H73 loop as well as the sidechain orientations of W79 and W86, changes in nucleotide exchange rates, and significant shifts in the twist of the actin monomer. Interestingly, for the case of K118N, the twist of the monomer is nearly identical to the F-actin protomer, and *in vitro* polymerization assays show that this mutation does result in

[1] This work was done in collaboration with Karina A. Kurth and Peter A. Rubenstein, both from the Department of Biochemistry at University of Iowa They were responsible for all experemental analysis presented in this chapter. Much of this work has been previously published (Jepsen, Kruth, Rubenstein, & Sept, 2016).

faster polymerization. Taken together, it is evident that mutations at this site give rise to a series of small changes that can be tolerated *in vivo*, but result in misregulation of actin assembly and dynamics.

5.1 Introduction

In Chapter 3, we discussed the importance of actin's nucleotide state in determining its affinity for interactions with actin-binding proteins. In Chapter 4, we discussed actin filament dynamics, including the unique dynamics present at filament ends, as well as how the nucleotide state of incoming subunits is important for polymerization. Actin's ability to function as part of the cytoskeleton is strongly dependent upon the strict regulation of filament polymerization and depolymerization; control of which relies both on nucleotide hydrolysis and interaction with actin binding partners. In this chapter, we will examine how disruption of actin's structure, by mutation of a lysine residue in the actin's pathogenic helix, blocks the proper function of actin leading to disease states. This work was done in collaboration with Karina A. Kurth and Peter A. Rubenstein, both from the Department of Biochemistry at University of Iowa, and much of it has been previously published (Jepsen, Kruth, Rubenstein, & Sept, 2016).

Point mutations in γ-cytoplasmic actin have been shown to cause autosomal dominant non-syndromic early-onset deafness, DFNA20/26 (Morín & Bryan, F Goodyear, R Mencía, A Modamio-Høybjør, S del Castillo, I Cabalka, JM Richardson, G Moreno, F Rubenstein, PA Moreno-Pelayo, MA, 2009; Rendtorff et al., 2006; Zhu et al., 2003). These mutations are found in γ-cytoplasmic actin, and are two of the eleven known DFNA20/26 actin mutations, and are unique in that they occur in the same residue. These mutations present in two different families, K118M in one family and K118N in a second family, providing an interesting opportunity to compare the effects of two dissimilar amino acid substitutions that produce a similar phenotype in humans. In

the first case, a charged but amphiphilic residue is replaced by a more hydrophobic species; in the second case, the mutation replaces the original residue with a smaller, neutral but still hydrophilic residue. The similarity in phenotype produced by such different amino acid substitutions presents a particularly interesting question regarding the role of K118 in actin function.

As outlined in our collaborators earlier work (Kruth & Rubenstein, 2012), K118 may play a previously unsuspected role in actin filament stability. K118 lies within a helix spanning from K113 to T126, extending from the inside to the outside of the actin filament, respectively (Figure 5.1). The residues of this helix are populated by mutations that lead to autosomal dominant actinopathies, and our collaborators recent work has demonstrated that the most interior residue of the helix, K113, is part of an inter-strand ionic bond involving E195 of an actin protomer in the opposing strand of the two-stranded actin helix (Lee et al., 2013). These results suggest that mutations within this helix, such as those at K118, may subsequently alter protomer-protomer interactions at this interface. Furthermore, K118 is located at or near the binding site for the Arp2/3 complex (Goley et al., 2010; Pfaendtner et al., 2012) and formins (Otomo et al., 2005; Thompson, Heimsath, Gauvin, Higgs, & Jon Kull, 2013), two important regulators of actin filament dynamics; the first mediates filament branching from a mother filament, and the second nucleates *de novo* filament formation. The co-localization of so many pathogenic mutations in one helix, coupled with the location of K118 at or near the binding sites of two important filament regulatory proteins, suggests that these mutations may be part of an allosteric regulatory system that controls actin filament stability.

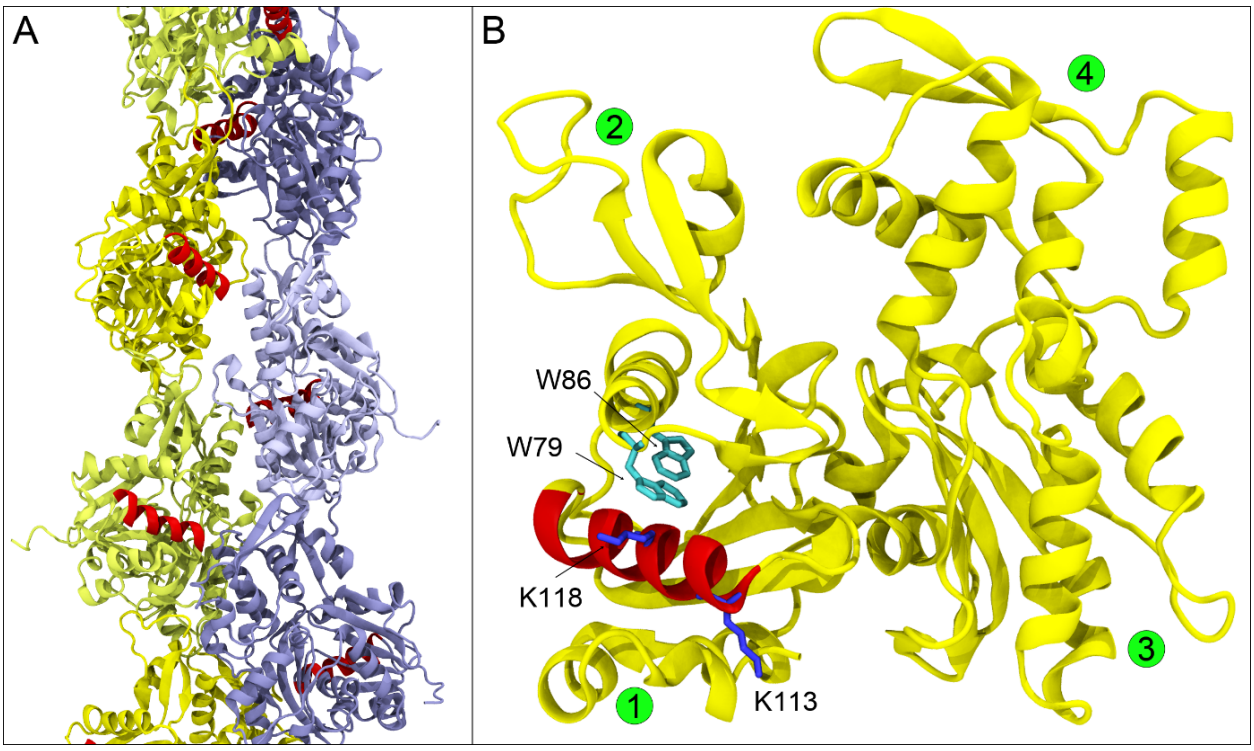


Figure 5.1 The Pathogenic Helix in F- and G-actin. The pathogenic helix containing K118 (red) is shown in context of (A) the F-Actin filament, and (B) monomeric G-Actin. K118 is positioned away from the actin-actin interface in the filament, while K113 extends towards this interface where it makes interstrand ionic interactions. Two tryptophans, W79 and W86, lie within the vicinity of K118 on a neighboring helix.

Using yeast actins, each carrying one of these two mutations, our collaborators previously demonstrated that the M and N mutations resulted in very different effects on actin behavior (Kruth & Rubenstein, 2012). By itself, K118M actin polymerized with the same kinetics as WT actin, whereas K118N polymerized twice as fast. This result was unexpected because the side chains for the amino acids at this site lie on the surface of the protein, away from the regions of G-actin that form protomer-protomer interfaces within the actin filament. Additional experiments polymerizing the K118M/N mutant actins in the presence of the Arp2/3 complex revealed that the mutations resulted in differential effects on Arp2/3-mediated branch formation. As measured by light scattering, polymerization of K118N actin showed almost no stimulation by the addition of Arp2/3 complex. The K118M actin, in contrast, did show some weak Arp2/3-dependent stimulation of polymerization, but it was dramatically reduced compared to WT. Our collaborators then

examined Arp2/3-mediated branching behavior in K118M actin using TIRF microscopy, which revealed that the branching frequency was drastically reduced. Furthermore, when branching did occur, the vast majority of branches formed near the pointed end of the filament instead of near the barbed end of the filament, where branching normally occurs.

Because of the potential importance of this residue in actin allosteric regulation and its association with human disease, we wished to gain further insight into structural consequences of these mutations that resulted in this altered polymerization behavior and its regulation by Arp2/3 complex. Using the mutant yeast actins previously generated, our collaborators employed a combination of spectroscopy methods, and protease digestion, while we used molecular modeling to elucidate a structural basis for the kinetics results they previously obtained. In the process, we have also gained further insight into the dynamics and behavior of the pathogenic regulatory helix in which K118 lies, and how these mutations influence the overall actin structure.

5.2 Results

5.2.1 Yeast vs Muscle G-Actin

The advantage of using yeast actin for *in vitro* studies lies in its similarity to α-muscle G-actin, the two share a high degree of sequence identity and similarity (86.7%, 93.3%) making it a good model system. The high degree of structural similarity should equate to similar structure and dynamics. However, knowledge of the similarities and differences between yeast and muscle actin will increase its usefulness as a model system for the *in vitro* study of actin related diseases. To accomplish this, we simulated ADP yeast G-actin (PDB 1YAG: Vorobiev et al., 2003) for 1.28 μs. Differing from our protocol for muscle actin, we started with a 300 ns cMD simulation followed by a 980 ns aMD simulation. Yeast actin is even more similar to γ-cytoplasmic actin

(89.1% identity and 95.2% similarity), but as we have not run any simulations of γ-cytoplasmic actin, we will continue to make comparisons to α-muscle actin.

To compare the dynamics of the two G-actins, we projected our yeast simulations onto our aMD ADP G-actin basis set presented in Chapter 3 (Figure 5.2). Despite the two simulations having very different starting points, they converged to the same equilibrium state. In terms of propeller-twist, phosphate-clamp and nucleotide cleft mouth sizes, yeast ADP G-actin is flatter and has a more open nucleotide cleft mouth compared to its muscle equivalent (-12.86 ± 2.81°, 6.42 ± 0.54 Å, and 9.73 ± 0.84 Å respectively). This same pattern, of yeast actin being flatter and more open than muscle actin, holds true when the two crystal structures (1YAG and 1J6Z) are compared.

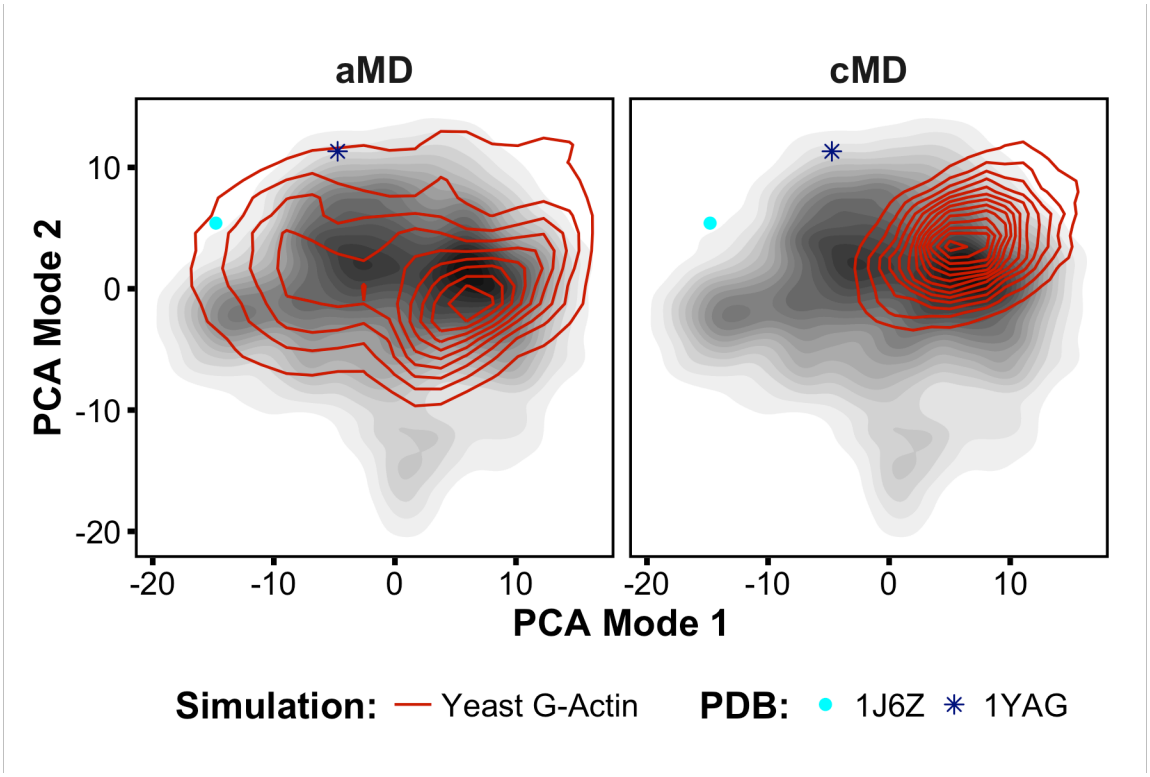


Figure 5.2 Yeast G-Actin on Muscle G-Actin PCA Space Despite the yeast and muscle ADP G-actin simulations having very different starting points (1YAG and 1J6Z respectively), the two simulations converge to nearly the same point.

Secondary structure is largely the unchanged in yeast actin (Figure 5.3). A few exceptions exist. First, within the nucleotide binding cleft the S-loop (D11-L16) is less structured (Figure 5.3) and the H73-loop (P70-N76) is more structured in yeast ADP G-actin. The difference in structure of the H73-loop likely arises from the lack of H73 methylation in yeast (Kalhor et al., 1999). Within the target binding cleft, the FQQ-loop (S350-M355) forms a helix at most 93.8% of the time.

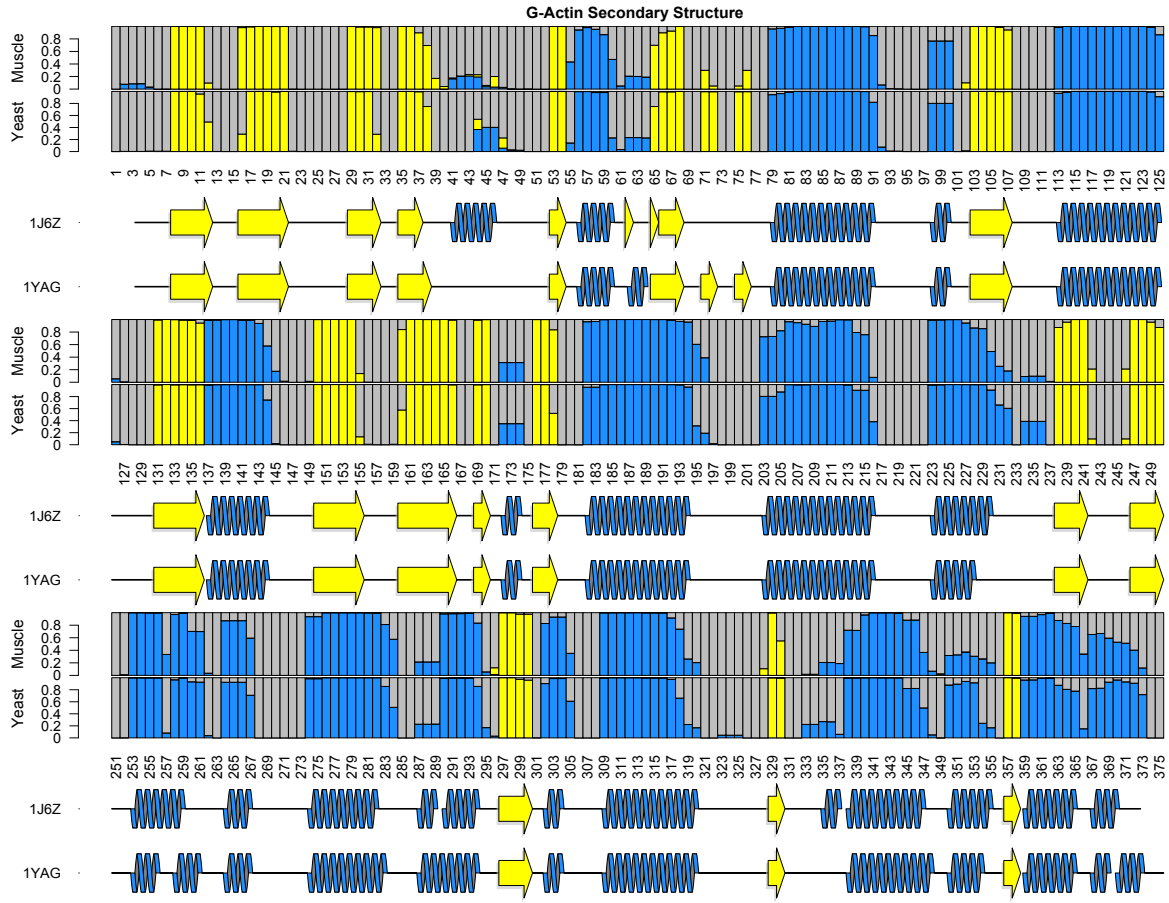


Figure 5.3 Yeast vs Muscle Secondary Structure A comparison of the secondary structure for our yeast and muscle ADP G-actin simulations, as well as their corresponding crystal structures, shows little difference. Some minor differences exist within the S, D, H73, and FQQ-loops.

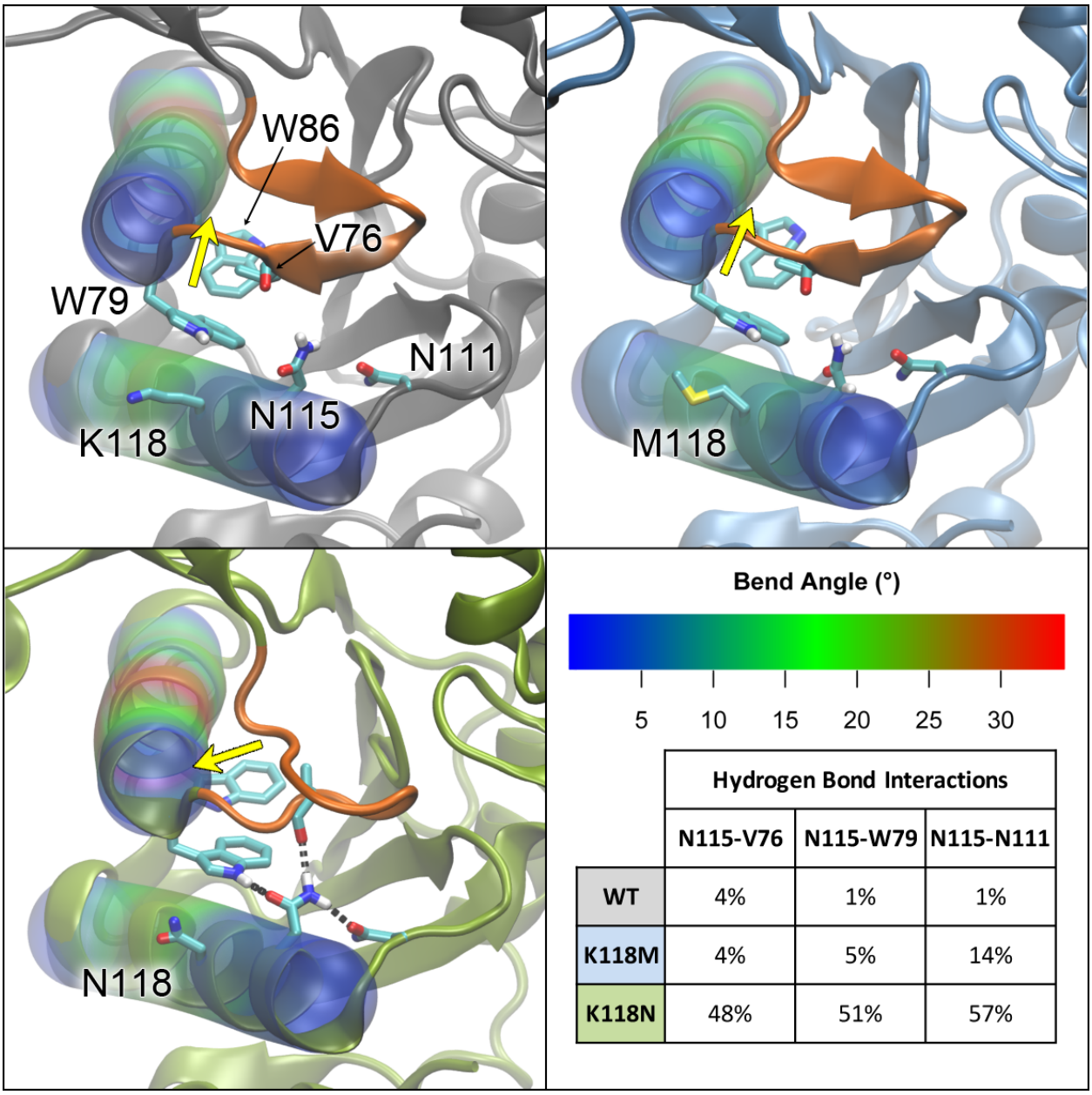
5.2.2 K118M/N

The previous observation that K118N actin polymerized approximately twice as fast as WT or K118M actin was puzzling because residue 118 lies on the surface of the actin monomer in

subdomain 1, extending out into the surrounding solvent and away from the nearest actin-actin interface in the filament. This result, coupled with the observation that the K118M mutation caused severe growth defects, suggested that the two mutant residues might generate allosteric and propagated conformational changes through the protein, altering both the structure of the monomer and the protomer contacts in F-actin.

5.2.2.1 Changes in the W79 and K118 Helices

In the actin monomer, the helix containing K118 is in close proximity to the helix containing tryptophans at positions 79 and 86. We find that while the local environments around W79 and W86 are very similar in the WT and K118M structures, the sidechain of W86 is flipped in the K118N structure (Figure 5.4). The flipped W86 sidechain in K118N is a result of localized conformational changes from the loss of interactions between K118 and W79. The aliphatic nature of the lysine results in the K118 sidechain being packed against the ring of W79 with the terminal amino group making contact with surrounding solvent. This interaction is lost upon mutation to asparagine, since it is both smaller and has less nonpolar surface area. Additionally, asparagine has a reduced helical propensity, weakening the K118 helix and allowing N115 to make new significant hydrogen bonds with W79, N111, and the backbone of V76 (Figure 5.4 and Table 5.1). These new interactions shift the K118 helix up towards the W79 helix and push W79 back into the protein, crowding W86 and causing its sidechain to flip. The shift of this helix also slightly shifts the bend of both the K118 and W79/W86 helices, becoming slightly straighter about K118 but more bent in the vicinity of W86.



	Hydrogen Bond Interactions		
	N115-V76	N115-W79	N115-N111
WT	4%	1%	1%
K118M	4%	5%	14%
K118N	48%	51%	57%

Figure 5.4 Localized structural changes around the K118 and W79 helices. For the WT, K118M and K118N structures we show the average conformation of the H73 loop (orange), orientation of the W86 indole ring (yellow arrow), bends of the W79 and W86 helices, and hydrogen bonding interactions. Helix bend angles are illustrated using the Bendix plugin in VMD (Dahl, Chavent, & Sansom, 2012). Amino acid distances are in Table 5.1.

Table 5.1 Distances between residues highlighted in Figure 5.4. The values shown are in Angstroms ± SD calculated over our MD simulations.

	$N115_{ND2}$ – $N111_{OD1}$	**$N115_{ND2}$ – $V76_{O}$**	**$W79_{NE1}$ – $N115_{OD1}$**	**$W79_{indole}$ – $W86_{indole}$**
WT	3.9 ± 0.6	4.4 ± 1.4	5.2 ± 0.7	8.1 ± 0.5
K118M	3.3 ± 0.4	4.0 ± 0.8	5.2 ± 1.2	7.6 ± 0.5
K118N	2.9 ± 0.2	3.0 ± 0.2	2.9 ± 0.2	7.1± 0.4

Since changes in tryptophan orientation and environment can affect their intrinsic fluorescence, our collaborators used fluorescence spectroscopy to measure the tryptophan fluorescence of the three different actins from 305 to 400 nm using an excitation wavelength of 300 nm. Typically, the λ_{max} for tryptophan fluorescence occurs in WT actin at 328 nm, but we see that both mutations affect the tryptophan fluorescence spectrum (Figure 5.5). The K118M mutation resulted in a quenching of the fluorescence intensity, but λ_{max} remained unchanged (328 nm). This result is probably not due to an induced protein conformation change but rather to the properties of the methionine itself, since the sulfur atom is known to cause quenching (Yuan, Weljie, & Vogel, 1998). In contrast, K118N actin had higher maximum fluorescence intensity as compared to WT, and there was a distinct and repeatable red shift in λ_{max} from 328 nm to 334 nm.

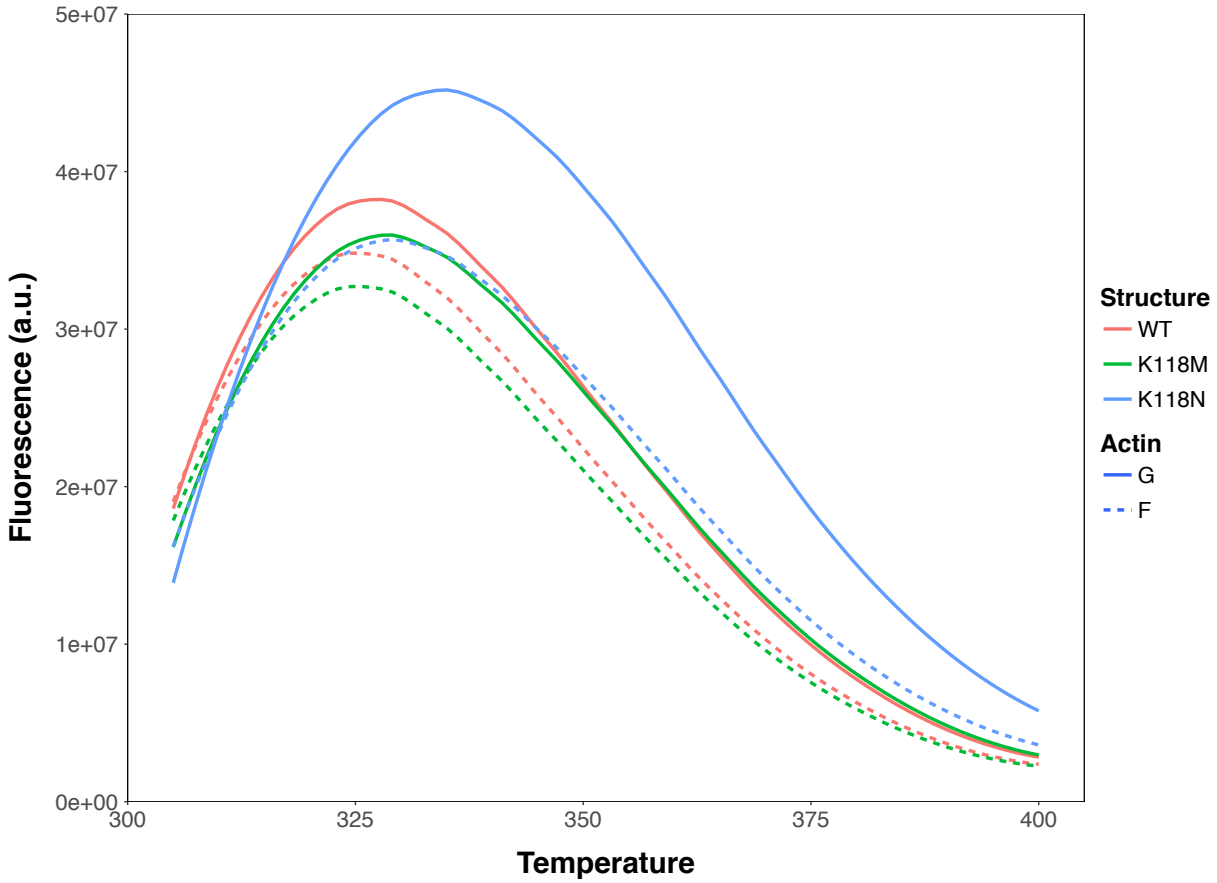


Figure 5.5 Intrinsic tryptophan fluorescence of G- and F-actin. Intrinsic tryptophan fluorescence was measured for both G- and F-actins to assess potential conformational changes in the protein. The above scans are representative scans that were acquired in triplicate from three separate actin preparations.

Doyle et al. previously showed that the fluorescence of W79 is quenched by non-radiative energy transfer to W86 in WT G-actin (Doyle, Hansen, & Reisler, 2001; Turoverov & Kuznetsova, 2003). Even though W79 has much greater solvent accessibility than any other tryptophan in WT and K118M G-actin structures, it does not contribute to overall fluorescence due to this quenching. Since the indole ring of W86 is flipped in K118N structure (Figure 5.4), W79 can now contribute more to the overall fluorescence spectrum, consistent with the red-shifted spectrum we observe.

Red shifts can also indicate an increase in solvent accessibility of one or more tryptophan residues (Vivian & Callis, 2001). Indeed, if we look at the total combined solvent-accessible surface area of tryptophans in our simulations, the distributions are not normal (Gaussian), and we see median areas of 12.8 and 14.6 $Å^2$ for WT and K118M, respectively (this does not include W79

due to quenching). However, in the case of K118N and the flipped W86 sidechain, the total tryptophan area should now include W79, which increases the median area to 26.0 $Å^2$ (Figure 5.6). Consistent with this finding, acrylamide quenching analysis, based on the Stern-Volmer equation (see Experimental Procedures in Jepsen et al., 2016), shows that K_{SV} values for WT and K118M G-actins were almost identical at 0.00227 and 0.00240, respectively. However, for K118N G-actin K_{SV} was 0.00359, nearly twice as large, indicating a much greater accessibility to acrylamide quencher (Figure 5.7).

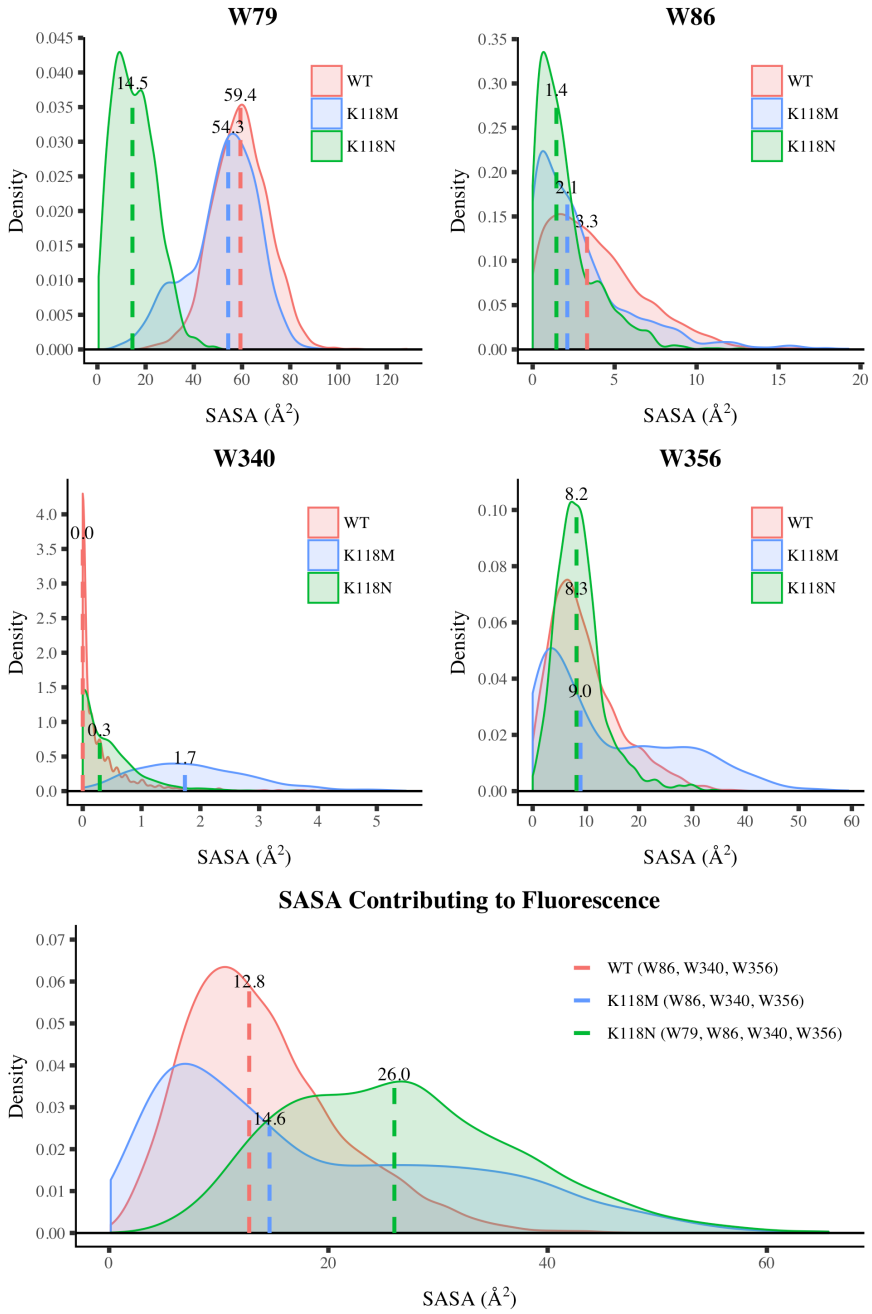


Figure 5.6 Solvent accessible surface area of tryptophans in G-actin. The distribution of SASAs (in Å^2) was calculated over the molecular simulation for each actin. The median values are indicated for each case by the dashed lines and numerical values. In the case of K118N, since the W86 sidechain is flipped, W79 should not be quenched and will contribute to the overall fluorescence.

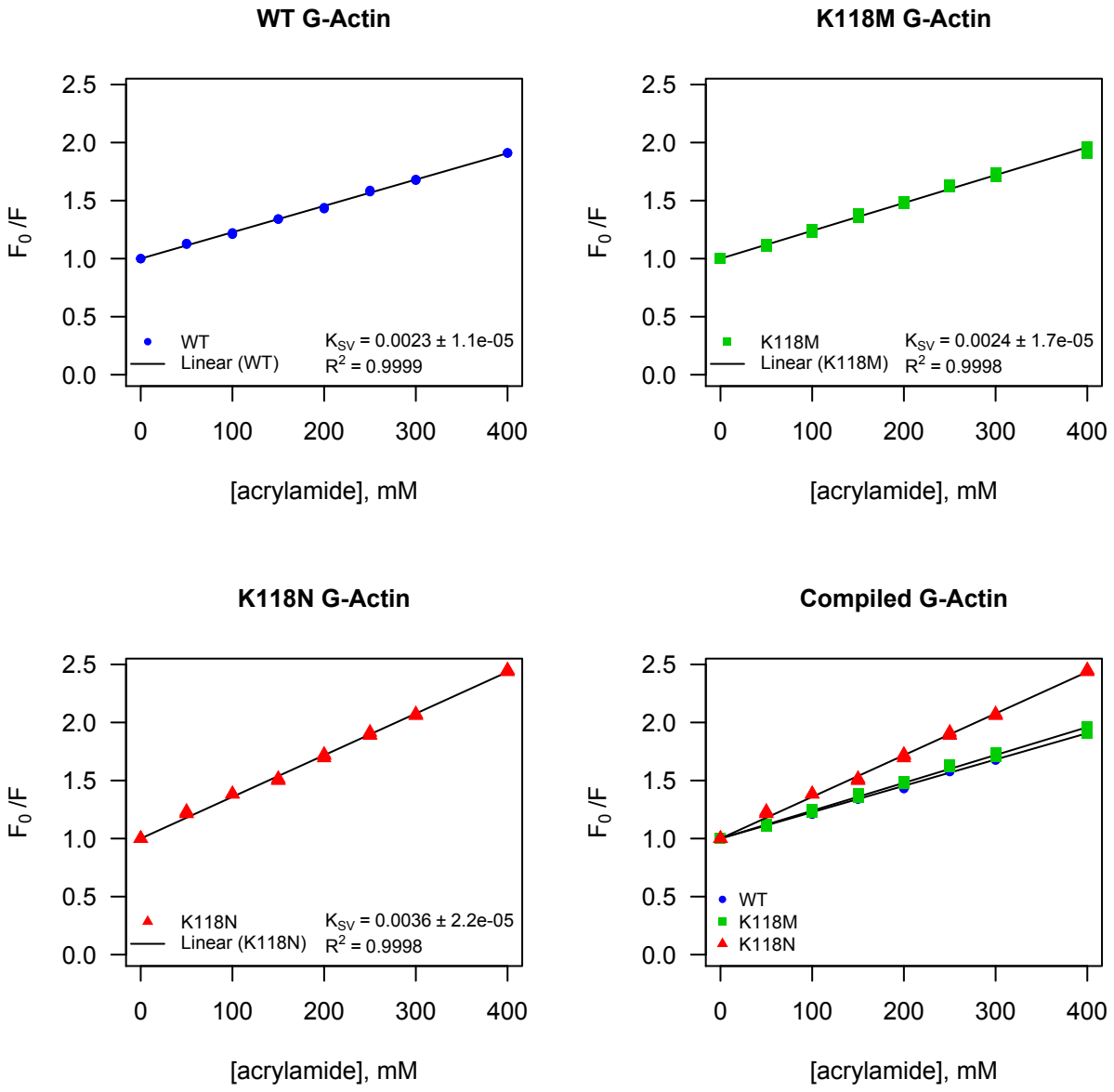


Figure 5.7 Quenching of intrinsic tryptophan fluorescence in G-actin Intrinsic tryptophan fluorescence quenching by the addition of increasing amounts of acrylamide. The ratio of the initial intensity without quencher, F_0, was divided by the intensity of fluorescence with the addition of acrylamide. The F_0/F ratio was plotted and fit to a linear regression with the y-intercept set to 1 according to the Stern-Volmer equation: $F_0/F=1+K_{SV}[Q]$, where [Q] is the concentration of quencher and K_{SV} represents the quenching constant. The above plots are representative of two assays with two different actin preparations.

5.2.2.2 Changes in the Actin Monomer

The localized changes we see in the W79 and K118 helices lead to larger scale changes in the protein. Using the results of our MD simulations, we examined the potential effects of the

K118M/N mutations on the overall structure of G-actin. As discussed in Chapter 3, the two major conformational changes of the actin monomer correspond to relative changes in the four subdomains – a propeller-twist and a scissor-like motion of the nucleotide binding cleft. The propeller-twist is characterized by the dihedral angle between the cores of subdomains 2-1-3-4. After extensive simulation, we found our WT yeast structure settled on an average twist of -12.9° (SD of 2.8°) (Figure 5.8). This conformation is slightly flatter than what is found in crystal structures of actin, which find a range of twists from -16 to -25° (Oda & Maéda, 2010). The K118M mutant adopts a more twisted structure, -28.2 ± 2.6° (SD), while the K118N conformation is significantly flatter, with a twist of -8.1 ± 1.8° (SD). This twist angle of K118N is consistent with what is found in the F-actin model structures (Fujii, Iwane, Yanagida, & Namba, 2010; Oda & Iwasa, T Maéda, Y Narita, A, 2009) and our F-actin simulations (see Chapter 4). If we compare the conformations of the two mutant actins to our muscle G-actin simulations, we see that the K118M mutant is shifted to lower PC1 values, away from that of the filament, representing a more twisted structure (Figure 5.8). K118N, on the other hand, is shifted toward where our F-actin simulations lie, this likely contributes to the observation that K118N polymerizes faster than either WT or K118M G-actin (Kruth & Rubenstein, 2012).

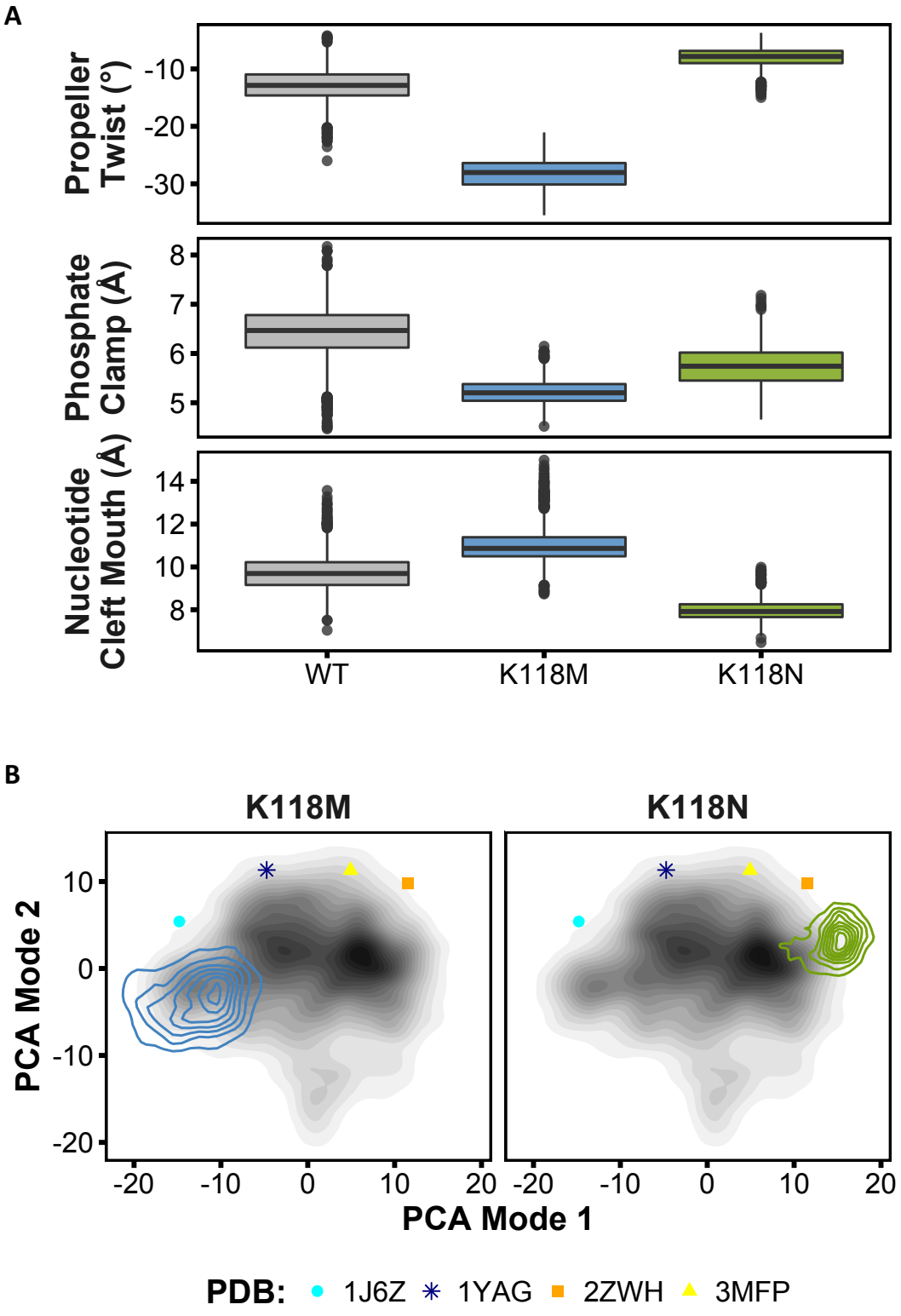


Figure 5.8 K118M/N Tertiary Structure (A) The propeller-twist, phosphate clamp and nucleotide cleft mouth sizes were measured for the three actin states. (B) Both the K118M and K118N mutants shift our PCA space away from WT yeast G-actin, with K118N being shifted closer to a F-actin conformation.

Next, we sought to ascertain if the effects of these two mutations might be due to a propagated conformational change in the actin. As a first step, our collaborators assessed the susceptibility of the mutant actins to different proteases whose cleavage sites were distant from the site of the mutations. Actin contains cleavage sites for trypsin, subtilisin and chymotrypsin near the DNase 1 loop (D-loop) in actin subdomain 2, and some mutations previously studied alter the conformation of these sites to the point that susceptibility to cleavage by these proteases is affected (Strzelecka-Gołaszewska, Moraczewska, Khaitlina, & Mossakowska, 1993; Yao, Grade, Wriggers, & Rubenstein, 1999). However, no changes compared to WT actin were observed when K118M and K118N actins were subjected to these proteases, suggesting that the mutations caused no significant conformational changes in this part of the protein (Figure 5.9 A). This observation is also consistent with our molecular dynamics simulations, which showed minor changes in the location of the helix within the D-loop, but not in such a manner as to inhibit cleavage (Figure 5.9 B). Although we observed changes in the arrangement of the four subdomains, we did not see any major structural changes that would likely affect proteolysis.

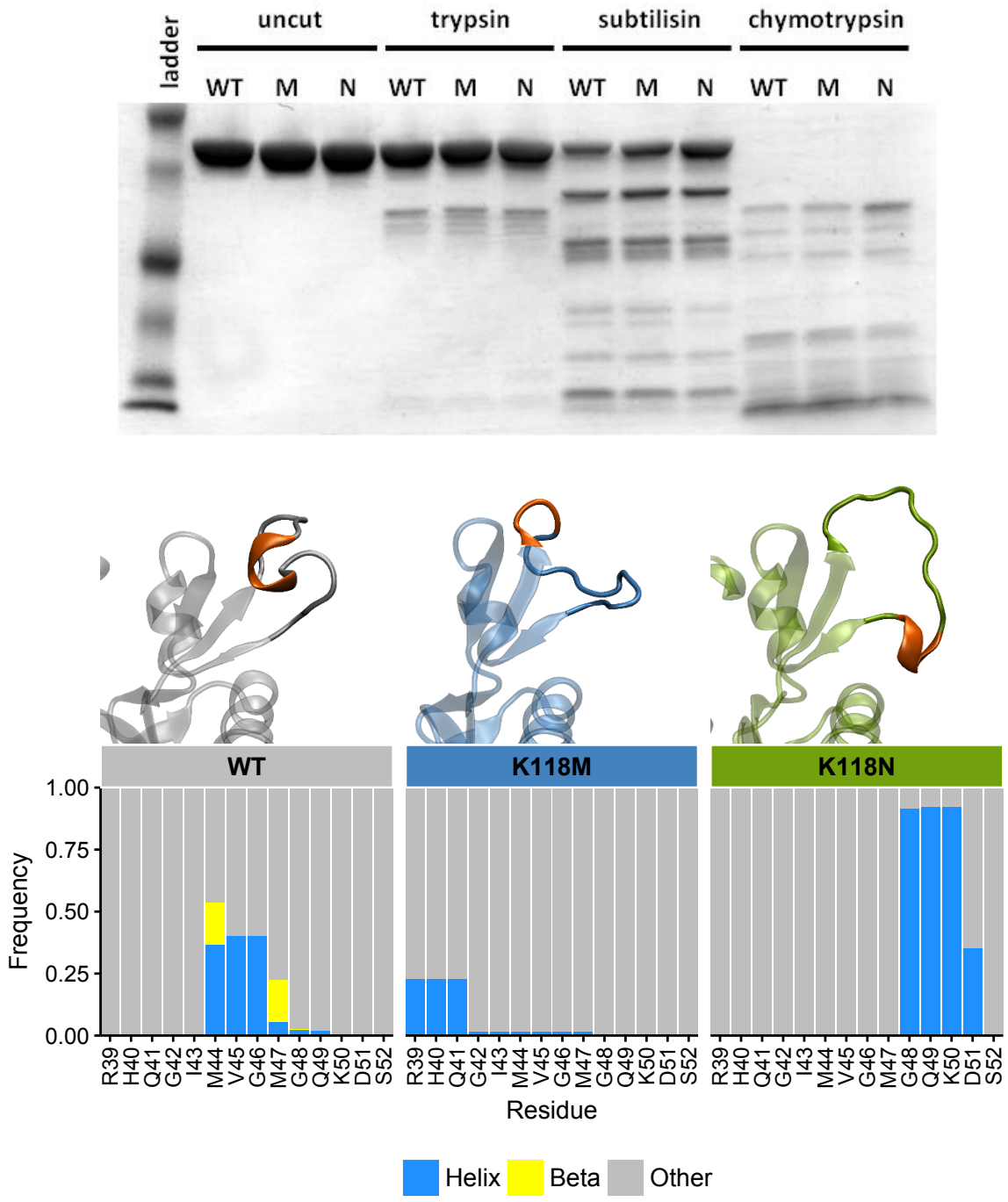


Figure 5.9 Subdomain 2 Conformation (A)Protease digestion of G-actin to determine subdomain-2 exposure. Purified monomeric WT and mutant yeast actins were incubated with trypsin, subtilisin, or chymotrypsin to test for protease accessibility in subdomain 2. Actin was digested at room temperature for 10, 15, or 20 minutes with trypsin, subtilisin, and chymotrypsin, respectively. If mutation of K118 alters the exposure of subdomain 2, we would likely see a change in the cleavage pattern produced upon protease digestion. We did not observe any significant changes in cleavage pattern, suggesting that mutation of K118 does not alter the exposure of subdomain 2 in the actin monomer. *(B) Conformation of D-Loop* Each actin was able to form a short helix within the D-loop, but the location and stability of this helix depended upon which actin it was found in.

We next wanted to detect structural changes. Our collaborators approached this via alterations in the CD spectra of the mutant proteins that, if they occurred, would indicate a mutation-based alteration in secondary structure. Using seven replicates over three different protein preparations we found no significant effect of either mutation on the G-actin CD spectra between 190 and 240 nm (Figure 5.10 A), although K118N actin did appear to have a slightly stronger signal at approximately 210 nm and 222 nm. To distinguish whether the difference was due to a concentration effect or the protein structure *per se*, they normalized the intensity of the K118N curve to that of WT actin and overlaid the two curves to observe potential differences in spectrum shape. The spectra were virtually identical (Figure 5.10 B), which reinforces the hypothesis that the protein secondary structures are not different. This was confirmed by our simulations, which showed only minimal changes in the secondary structure (Figure 5.11). As previously discussed, our simulations showed differences in the D-loop in all three actins, the H73-loop in the K118N mutant. Other small changes can be seen throughout the protein.

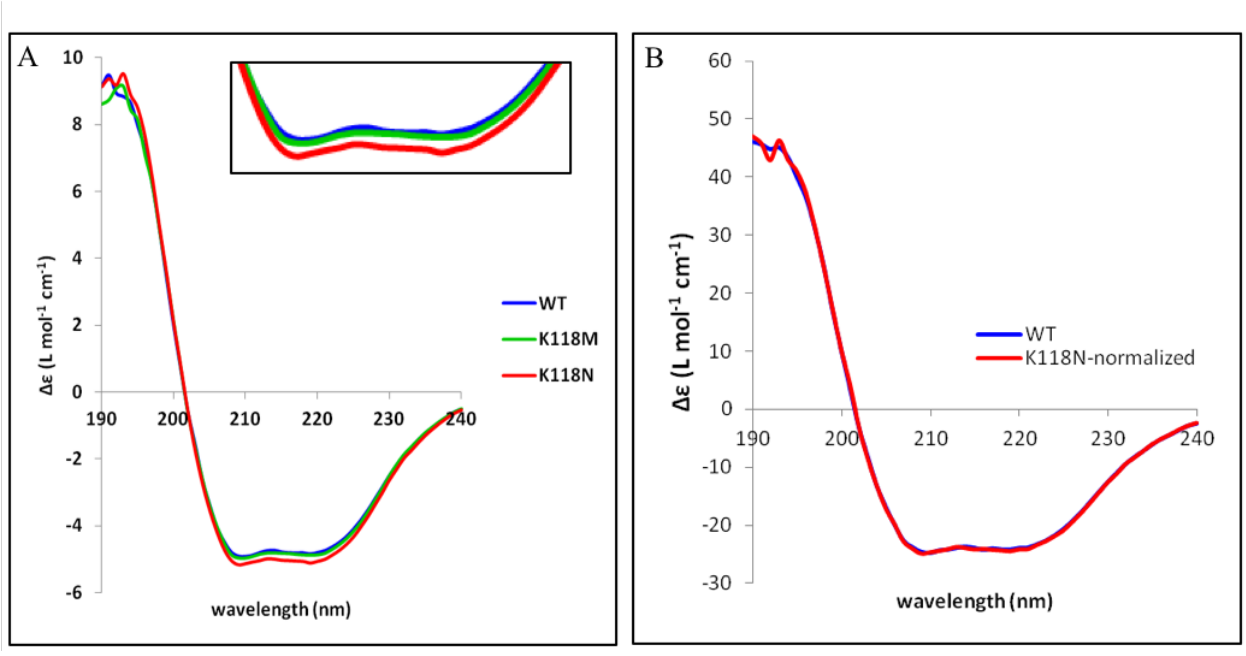


Figure 5.10 CD wavelength scans of monomeric actin Wavelength scans from 190 to 240 nm were performed using CD spectroscopy. (A) Plots represent the average molar ellipticity of purified WT, K118M, and K118N G-actins from seven runs over three different protein preparations. The K118N actin appeared

to display a consistent, small increase in α-helicity as evidenced by the slightly more negative Δε (B) The K118N wavelength scan was reduced in signal strength by 3.5%, according to the difference in stock concentration as determined by BCA assay, and overlaid with WT actin. The resulting curves are nearly identical.

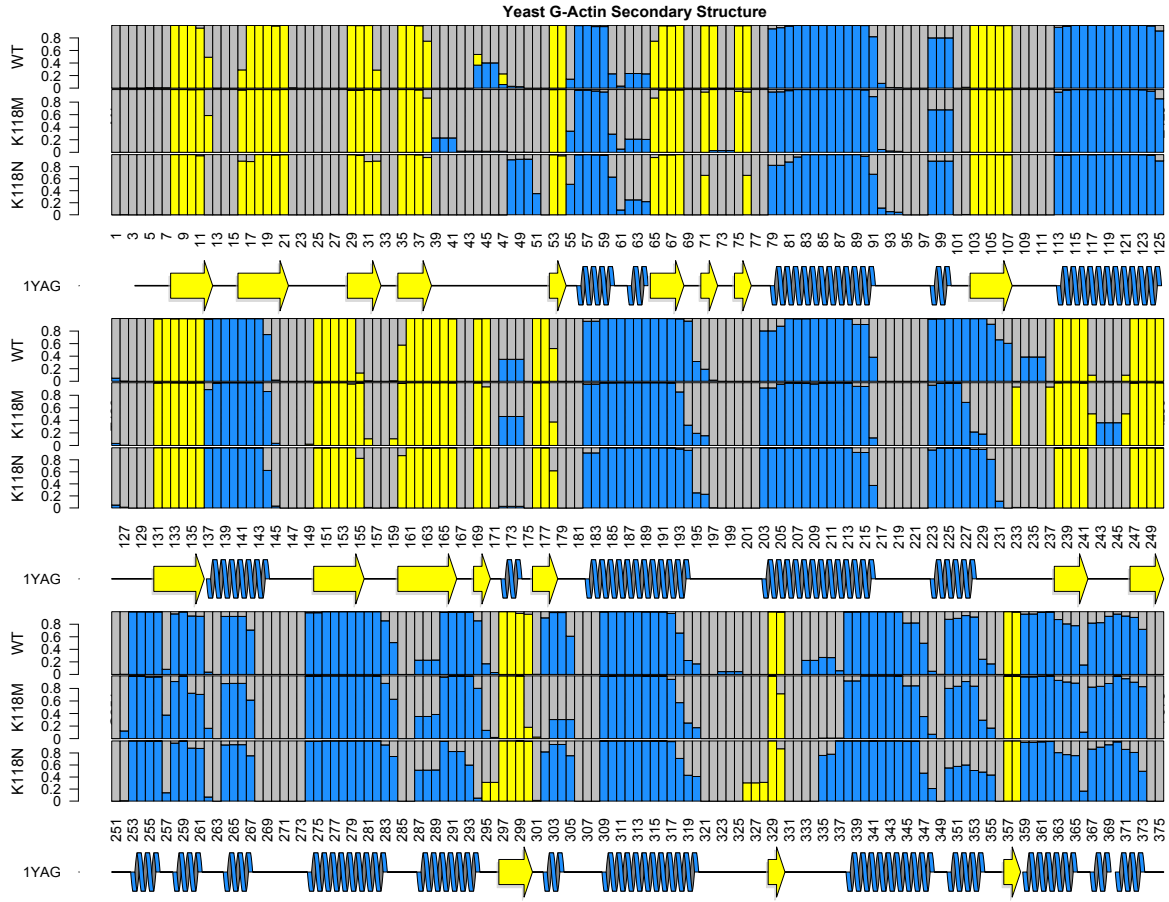


Figure 5.11 Secondary Structure of Yeast G-actin and the K118M/N Mutants A comparison of the three states of actin show minimal differences. The D-loop shows some differences across the board (see also Figure 5.9 B). The H73-loop is unstructured in the K118N mutant (see Figure 5.4**)**.

5.2.2.3 Changes in F-actin

We next wanted to determine how polymerization affected tryptophan fluorescence with the WT and K118M/N actins, as polymerization normally results in about a 15% decrease in intrinsic tryptophan fluorescence (Bryan & Rubenstein, 2005). Our collaborators previous results showed that the K118N mutation accelerated actin polymerization but provided no substantial information on the nature of the final filament. We therefore asked whether this polymerization-dependent fluorescence change also occurred in the mutant actins. We found the spectra of the WT

and K118M actins were almost identical (Figure 5.5), indicating that the chemical environment of the tryptophans after the conformational change from G- to F-actin is very similar for these two actins. In contrast, for K118N, polymerization caused a blue shift of the otherwise red-shifted G-actin fluorescence peak, returning the λ_{max} to near that of the WT and K118M proteins. The polymerization-dependent decrease in fluorescence was almost to the same level as seen with the other two actins. The tryptophan environment in K118N F-actin is thus very similar to that of the WT and K118M F-actins, suggesting that whatever conformational or environmental difference the asparagine mutation causes in G-actin, it reverts to almost normal in F-actin due to the constraints of the filament structure (Figure 5.12). From these data, it appears that the K118N mutation mainly affects G-actin conformation and the actin's ability to polymerize, but the final filament structure of K118N F-actin is likely similar to WT and K118M actins. A summary of the wavelength scan results is presented in Table 5.2.

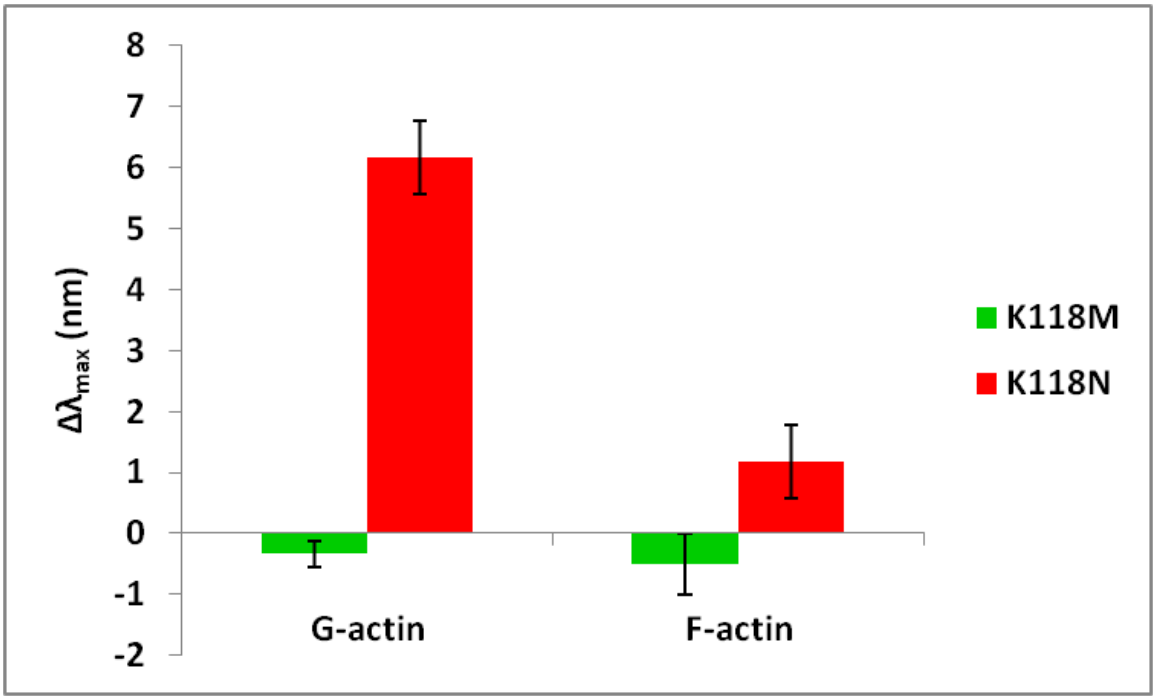


Figure 5.12 Change in λ_{max} of tryptophan wavelength scans The wavelength of λ_{max} was determined for WT, K118M, and K118N G- and F-actins, and the change in wavelength as compared to that of WT actin is shown above. Error bars represent standard error from six measurements over three separate actin preparations.

Table 5.2 Experimentally measure parameters for expressed actins. From left to right: peak wavelength for tryptophan fluorescence, quenching constants for G- and F-actin, halftime for nucleotide exchange, and rate constants for P_i release for both elongation and steady-state growth. All errors listed are standard errors of the mean. For the K_{SV} values, these are the standard errors resulting from linear regression.

	λ_{max} (nm) (n=6)	**G-actin K_{SV}**	**F-actin K_{SV}**	**$t_{1/2}$ of nucleotide exchange (s) (n=6)**	**$T_{1/2}$ (°C) (n=6)**	**Elongation P_i Release Rate (µM/s) (n=3)**	**Steady-State P_i Release Rate (µM/s) (n=3)**
WT	327.8 ± 0.12	0.00227 ± 0.00001	0.00219 ± 0.00003	37.1 ± 0.41	63.7 ± 0.08	0.0180 ± 0.0006	0.0024 ± 0.0001
K118M	327.5 ± 0.08	0.00240 ± 0.00002	0.00228 ± 0.00001	41.8 ± 0.86	61.2 ± 0.33	0.0203 ± 0.0014	0.0022 ± 0.0001
K118N	334.0 ± 0.16	0.00359 ± 0.00002	0.00282 ± 0.00003	39.4 ± 1.1	59.1 ± 0.20	0.0335 ± 0.0018	0.0031 ± 0.0003

The fluorescence quenching assays was repeated with F-actin samples of the three actins (Figure 5.13). Just as with G-actin, the K_{SV} for WT and K118M actins were nearly identical to each other at 0.00220 and 0.00228 respectively. The K_{SV} for K118N F-actin, however, was 0.00282, which is a significant (21%) decrease in quencher accessibility compared to K118N G-actin. These results are in agreement with our previous tryptophan fluorescence data, which indicate that the K118N tryptophans are in an environment more similar to WT upon polymerization. However, the K_{SV} for K118N F-actin is still 29% higher than for WT F-actin, and thus the environment of the tryptophans has not completely resumed a WT-like state. A summary of the results from the quenching assay is shown in Table 5.2.

While we did not perform any yeast F-actin simulations, we can look at our muscle G- and F-actin simulations for changes that occur upon polymerization that may affect tryptophan fluorescence and quenching. Limiting our analysis of F-actin to just the middle four protomers of our Oda filament model and to the last 100 ns of simulation time, we do see that the median solvent accessibility for tryptophan's decreases upon polymerization from 9.5 $Å^2$ in G-actin to 7.5 $Å^2$ in F-actin. However, our F-actin simulations display a wider range of tryptophan SASA values.

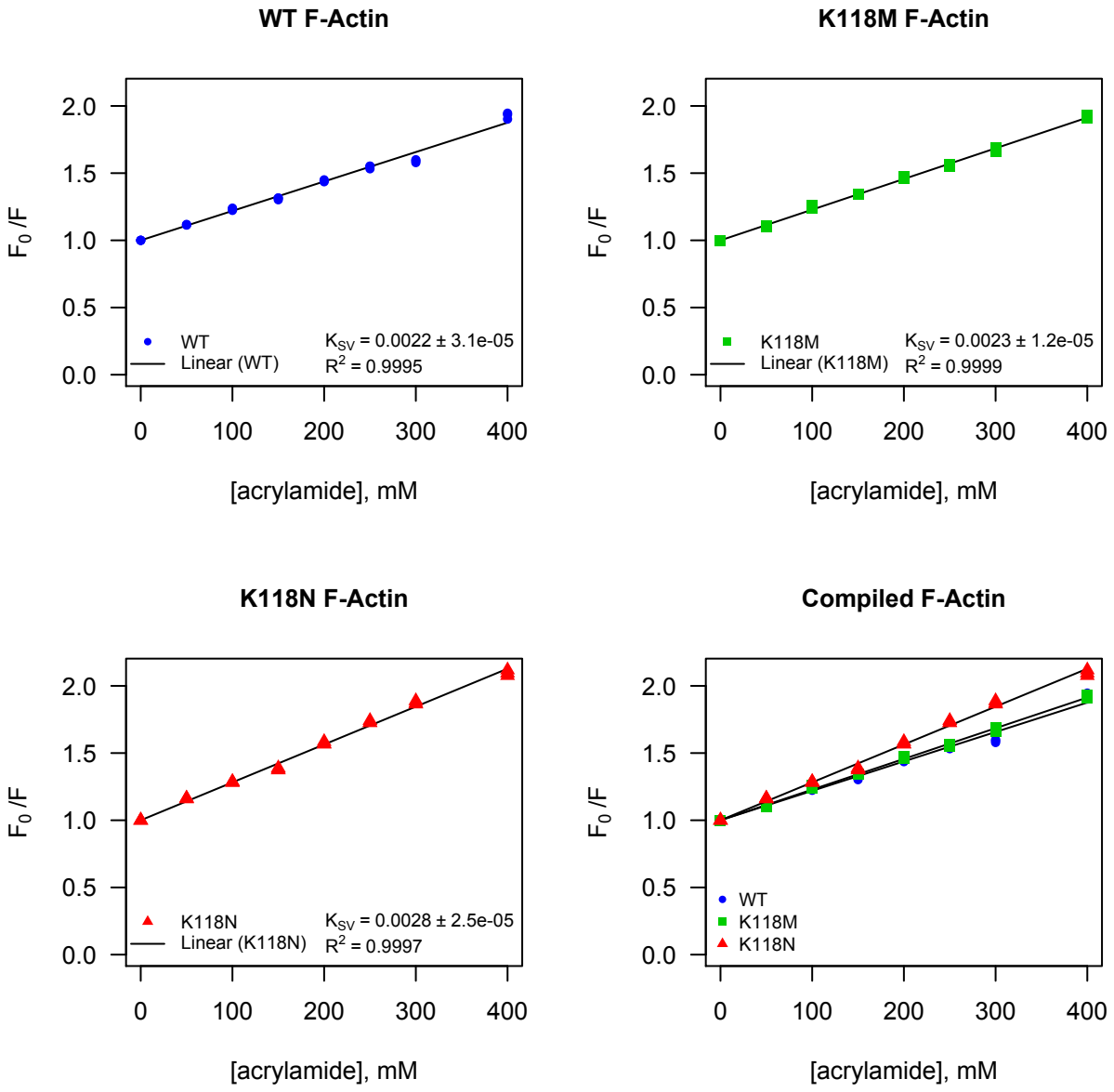


Figure 5.13 Quenching of intrinsic tryptophan fluorescence in F-actin Acrylamide quenching was repeated for F-actin samples. The above plots are representative of two assays with two different actin preparations.

5.2.2.4 Nucleotide Exchange

Because structural effects in actin are often propagated from the outside of the protein to the nucleotide binding cleft, and because the nucleotide dynamics of actin can influence polymerization, our collaborators measured the first-order rate of fluorescent etheno-ATP

exchange from WT, K118M, and K118N G-actin in the presence of a large excess of ATP. The $t_{1/2}$ values for WT and K118N were not statistically significant at about 37.1 and 39.4 s respectively (Table 5.2). The $t_{1/2}$ for K118M was slower at 41.8 s, and this was statistically significant from the WT ($p=0.002$) and seemed to indicate slight effects of the mutations on the dynamics of the ATP binding pocket. Although these shifts in $t_{1/2}$ are relatively small, the trend is consistent with our simulation results. The phosphate clamp is a structural measure of how tightly the nucleotide is coordinated in the binding site, and is characterized by the distance between the C_α atoms of D157 and G15. We find phosphate clamp distances (± SD) of 6.4 ± 0.5 Å and 5.8 ± 0.4 Å in WT and K118N G-actin respectively, but the coordination is tighter in K118M (5.2 ± 0.3 Å). If we calculate the Pearson correlation of the $t_{1/2}$ times and phosphate clamp distances we get a significant negative correlation (-0.99992, n=3, $p=0.0078$), suggesting that both mutants coordinate the nucleotide more tightly, resulting in longer exchange times.

5.2.2.5 Phosphate Release

It is possible that the mutations might cause alterations in the dynamics of the filament that could be ascertained by measuring the release of P_i from the nucleotide binding cleft. Our collaborators next measured the rate of free phosphate release during actin polymerization using a coupled enzymatic reaction method. Because there is a delay between ATP hydrolysis and release of the free phosphate in some actins, this assay does not provide results on the rate of ATP hydrolysis itself. Using cofilin-free G-actin, they reassessed phosphate release rates (Figure 5.14). Finding that for all three actins, the rate of phosphate release closely follows the polymerization curve during the initial elongation phase of polymerization, suggesting that the rate of phosphate release is proportionate to the polymerization rate. However, as polymerization levels off to the treadmilling phase where there is no net polymerization, the rate of phosphate release becomes

linear, according to the rate at which monomers cycle through the filament. Both WT and K118M show identical profiles, suggesting that the rate of phosphate release is unaffected by the K118M mutation. Although K118N actin shows a proportionate increase in phosphate release rate during elongation, it shows a small but reproducible increase in rate during treadmilling, suggesting that the K118N filament is more dynamic than its WT counterpart. Our simulations did not probe phosphate release directly, but the looser phosphate clamp in K118N would also be consistent with faster P_i release. These results are also summarized in Table 5.2.

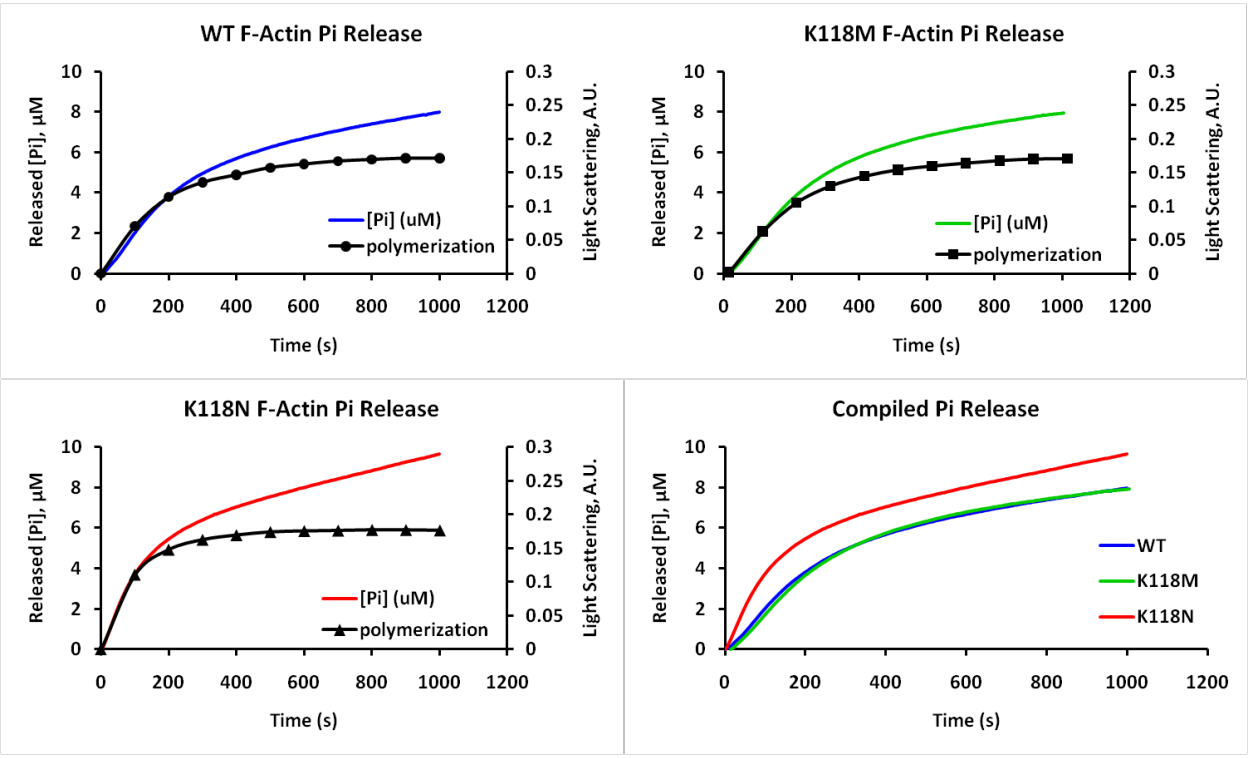


Figure 5.14 Phosphate release from actins Colored lines represent the concentration of unbound phosphate in µM, and black lines represent light scattering curves of actin polymerization. Phosphate measurements were recorded every 5 s for 1200 s total. For all three actins, the rate of phosphate release closely follows the polymerization curve during the initial elongation phase of polymerization. However, as polymerization levels off to the treadmilling phase, the rate of phosphate release becomes linear, according to the rate at which monomers cycle through the filament. Both WT and K118M actins are identical, whereas K118N actin shows a proportionate increase in phosphate release rate during elongation and shows a slight increase in rate during treadmilling.

5.3 Discussion

The fact that two very different mutations at residue 118 in actin lead to hereditary deafness provides a unique opportunity to gain insight into how this part of the actin structure relates to overall actin function. In the following discussion, we focus first on local and allosteric effects on the actin monomer, then the longer-range changes in the actin filament, and finally some comments on how these mutations may ultimately result in disease.

5.3.1 Effects on G-actin

The altered intrinsic tryptophan fluorescence spectra and decreased thermal stability suggest that both mutants cause localized changes in the actin structure. Our results show that the mutation to asparagine causes changes within the K118 helix, allowing N115 to form new hydrogen bonds with W79, N111, and the backbone of V76, resulting in the helix being shifted up toward W79. The movement of the K118 helix forces W79 into the space previously occupied by W86, causing the W86 sidechain to flip. These changes are consistent with the red-shifted tryptophan peak fluorescence observed for K118N actin and the increased susceptibility of K118N actin to acrylamide quenching. Since W86 has previously been shown to quench W79 by non-radiative energy transfer, the flipping of the W86 indole ring would increase the contribution of W79 to the observed fluorescence. Additionally, the polar group of the asparagine side chain would be closer to the tryptophan than is the lysine amino group, creating a more hydrophilic environment.

In the case of K118M, there is evidence that methionine can interact with tryptophans to contribute to protein stability (Valley et al., 2012), and thus an interaction between M118 and W79 could compensate for the loss of an interaction between the methylene carbons of K118 with W79. Such an interaction would also be consistent with an altered tryptophan electronic environment

that would explain the different fluorescence spectra of K118M. The fluorescence spectra of K118M showed no shift in peak fluorescence, but it did show a distinct quenching of this peak. This effect is likely due to the presence of the M118 sulfur, which has been shown to have a quenching effect on tryptophan fluorescence (Yuan et al., 1998).

The changes that we see within the K118 helix ultimately lead to changes in subdomain arrangement. The hydrogen bond between N115 and V76 in the K118N mutation results in a loss of secondary structure in the H73 loop (residues 70-78). As the H73 loop is part of a network of structural loops within actin that give structure to the nucleotide binding cleft (Zheng, Diraviyam, & Sept, 2007), the loss of secondary structure will alter the nucleotide site and should affect both the monomer twist and D-loop structure. Indeed, we observe a significant flattening of the monomer in the case of K118N and further twisting of the K118M structure, although the molecular basis for the changes in K118M are not as obvious. Interestingly, even with substantial subdomain rearrangement, we saw no significant change for either mutant in folding or overall secondary structure based on CD measurements, MD simulations, nucleotide exchange rates or susceptibility to limited protease digestion.

5.3.2 Effects in F-actin

Although we see differences in the monomer structure between WT and the K118M/N mutants, it appears as if these differences are lessened once actin is incorporated into the filament. Our fluorescence results show that putting the mutant monomer into the constraints of the filament structure significantly reverted the K118N protein to a more normal structure. Similarly, our measurement of phosphate release rates during polymerization showed that there does not appear to be a significant effect on nucleotide-binding dynamics as a result of the K118M/N mutations since phosphate release closely reflected the overall rate of polymerization in all cases. The rate of

phosphate release during the steady-state treadmilling phase, where monomer addition at one end of the filament is matched by monomer release at the other, was slightly faster for K118N than for WT actin, a feature that may result from the looser nucleotide coordination we observe in K118N simulations. For yeast actin, there is no lag in P_i release subsequent to ATP hydrolysis (Yao & Rubenstein, 2001). This finding, coupled with the fact that for all the actins studied here, P_i release and polymerization occurred at basically the same rate, implies that a slightly increased rate of P_i release during treadmilling would give a faster rate of monomer addition and P_i release. Such a result would be consistent with our hypothesis that the N mutation caused an increase in conformational flexibility that allowed the protein to more easily achieve a polymerization-competent state.

Our polymerization studies showed that while both the N and M mutations did not significantly alter the critical concentration for polymerization, the K118N mutation caused the actin to polymerize approximately two times faster than either the WT or K118M species. The maintenance of critical concentration suggests that the increased polymerization rate of the N mutant results from an increased ease of spontaneous filament nucleation, and this is consistent with our simulation results that show the K118N monomer structure is very close to that of the F-actin protomer. The cytoskeletal dysfunction associated with both mutations, coupled with the polymerization data, suggested that the mutations caused improper regulation of filament formation and turnover and that significant inherent polymerization defects were not the cause of the resulting altered cytoskeletal behavior.

Our collaborators previous work indicated that the K118 helix was at the strand-strand interface of the filament and connected the binding sites for Arp2/3 complex and formin with an interstrand ionic interaction that plays a role in filament stability and polymerization (Lee et al.,

2013). Our current work, especially with the K118N mutant, provides key insight into how altering a residue at a regulatory protein binding site can propagate a series of conformational changes to the actin monomer in terms of propeller twist, H73 loop, and D-loop alteration. All of these are important determinants of polymerization, and our results show that the mutation shifts the actin monomer into a conformation that polymerizes faster and is closer to the F-actin protomer structure. In essence, the mutation could function to deregulate polymerization by converting the protein to a molecule more likely to undergo spontaneous polymerization.

5.3.4 Potential Connection with Disease

A hallmark of proper cytoskeletal function is very tight regulation of actin polymerization and depolymerization. Patients with DFNA20/26 form a normally functioning auditory apparatus initially and gradually develop deafness as they age. The most widely accepted hypothesis for this phenotype suggests that the repair processes that maintain the structure over time, against various physical insults and normal protein turnover, fail with the mutant actins. This type of repair would represent a fine-tuning type of process, and if there is limited access to G-actin and the regulatory proteins that control it after the structure has been built, decreased regulation due to protein mutation could lead to the time-dependent disappearance of hearing that is described. The effects we have shown indicate that the K118 mutations have on yeast actin may be even more pronounced in the context of the auditory hair cell. The vast majority of the DNFA20/26 mutations are in γ-cytoplasmic actin. Our collaborators previous work (Bergeron, Zhu, Thiem, Friderici, & Rubenstein, 2010) demonstrated that this actin isoform polymerized more slowly than either yeast actin or β-cytoplasmic actin, the other actin isoform in the hair cell. Such an inherent bias would make spontaneous actin polymerization even less likely, underscoring the importance of strict regulation of filament nucleation. Apart from regulation of the filament, higher order regulation

of filament bundles could also be affected. The bundling protein espin is an attractive candidate in this regard since it is the major bundling protein within stereocilia and mutations in espin are also linked to autosomal recessive hearing loss (Donaudy et al., 2006). Previous work showed how mutations that weaken the cross-linking activity of espin have significant effects on the viscoelastic properties of actin bundles (Lieleg et al., 2009), and there is a possibility that K118 mutations could result in similar defects. Studies such as this will hopefully be the subject of future investigations.

5.4 References

Bergeron, S. E., Zhu, M., Thiem, S. M., Friderici, K. H., & Rubenstein, P. A. (2010). Ion-dependent polymerization differences between mammalian β- and γ-nonmuscle actin isoforms. *Journal of Biological Chemistry*, *285*(21), 16087-16095.

Bryan, K. E., & Rubenstein, P. A. (2005). An intermediate form of ADP-F-actin. *Journal of Biological Chemistry*, *280*(2), 1696-1703.

Dahl, A. C. E., Chavent, M., & Sansom, M. S. P. (2012). Bendix: Intuitive helix geometry analysis and abstraction. *Bioinformatics*, *28*(16), 2193-2194.

Donaudy, F., Zheng, L., Ficarella, R., Ballana, E., Carella, M., Melchionda, S. et al. (2006). Espin gene (ESPN) mutations associated with autosomal dominant hearing loss cause defects in microvillar elongation or organisation. *Journal of Medical Genetics*, *43*(2), 157-161.

Doyle, T. C., Hansen, J. E., & Reisler, E. (2001). Tryptophan fluorescence of yeast actin resolved via conserved mutations. *Biophysical Journal*, *80*(1), 427-434.

Fujii, T., Iwane, A. H., Yanagida, T., & Namba, K. (2010). Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, *467*(7316), 724-728.

Goley, E. D., Rammohan, A., Znameroski, E. A., Firat-Karalar, E. N., Sept, D., & Welch, M. D. (2010). An actin-filament-binding interface on the Arp2/3 complex is critical for nucleation and branch stability. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(18), 8159-8164.

Jepsen, L., Kruth, K. A., Rubenstein, P. A., & Sept, D. (2016). Two Deafness-Causing Actin Mutations (DFNA20/26) Have Allosteric Effects on the Actin Structure. *Biophysical journal*, *111*(2), 323-332.

Kalhor, H. R., Niewmierzycka, A., Faull, K. F., Yao, X., Grade, S., Clarke, S. et al. (1999). A highly conserved 3-methylhistidine modification is absent in yeast actin. *Archives of Biochemistry and Biophysics*, *370*(1), 105-111.

Kruth, K. A., & Rubenstein, P. A. (2012). Two deafness-causing (DFNA20/26) actin mutations affect Arp2/3-dependent actin regulation. *Journal of Biological Chemistry*, *287*(32), 27217-27226.

Lee, C.-Y., Lou, J., Wen, K.-K., Melissa, M., Eskin, S. G., Ono, S. et al. (2013). Actin depolymerization under force is governed by lysine 113:glutamic acid 195-mediated catch-slip bonds. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(13), 5022-5027.

Lieleg, O., Schmoller, K. M., Purdy Drew, K. R., Claessens, M. M. A. E., Semmrich, C., Zheng, L. et al. (2009). Structural and Viscoelastic Properties of Actin Networks Formed by Espin or Pathologically Relevant Espin Mutants. *ChemPhysChem*, *10*(16), 2813-2817.

Morín, M., & Bryan, K. E. M.-M., F Goodyear, R Mencía, A Modamio-Høybjør, S del Castillo, I Cabalka, JM Richardson, G Moreno, F Rubenstein, PA Moreno-Pelayo, MA. (2009). In vivo and in vitro effects of two novel gamma-actin (ACTG1) mutations that cause DFNA20/26 hearing impairment. *Human Molecular Genetics*, *18*(16), 3075-3089.

Oda, T., & Iwasa, M. A., T Maéda, Y Narita, A. (2009). The nature of the globular- to fibrous-actin transition. *Nature*, *457*, 441-445.

Oda, T., & Maéda, Y. (2010). Multiple Conformations of F-actin. *Structure*, *18*(7), 761-767.

Otomo, T., Tomchick, D. R., Otomo, C., Panchal, S. C., Machius, M., & Rosen, M. K. (2005). Structural basis of actin filament nucleation and processive capping by a formin homology 2 domain. *Nature*, *433*(7025), 488-494.

Pfaendtner, J., Volkmann, N., Hanein, D., Dalhaimer, P., Pollard, T. D., & Voth, G. A. (2012). Key structural features of the actin filament Arp2/3 complex branch junction revealed by molecular simulation. *Journal of Molecular Biology*, *416*(1), 148-161.

Rendtorff, N. D., Zhu, M., Fagerheim, T., Antal, T. L., Jones, M. P., Teslovich, T. M. et al. (2006). A novel missense mutation in ACTG1 causes dominant deafness in a Norwegian DFNA20/26 family, but ACTG1 mutations are not frequent among families with hereditary hearing impairment. *European Journal of Human Genetics*, *14*(10), 1097-1105.

Strzelecka-Gołaszewska, H., Moraczewska, J., Khaitlina, S. Y., & Mossakowska, M. (1993). Localization of the tightly bound divalent-cation-dependent and nucleotide-dependent conformation changes in G-actin using limited proteolytic digestion. *Eur. J. Biochem.*, *211*(3), 731-742.

Thompson, M. E., Heimsath, E. G., Gauvin, T. J., Higgs, H. N., & Jon Kull, F. (2013). FMNL3 FH2-actin structure gives insight into formin-mediated actin nucleation and elongation. *Nature Structural and Molecular Biology*, *20*(1), 111-118.

Turoverov, K. K., & Kuznetsova, I. M. (2003). Intrinsic Fluorescence of Actin. *Journal of Fluorescence*, *13*(1), 41-57.

Valley, C. C., Cembran, A., Perlmutter, J. D., Lewis, A. K., Labello, N. P., Gao, J. et al. (2012). The methionine-aromatic motif plays a unique role in stabilizing protein structure. *Journal of Biological Chemistry*, *287*(42), 34979-34991.

Vivian, J. T., & Callis, P. R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal*, *80*(5), 2093-2109.

Vorobiev, S., Strokopytov, B., Drubin, D. G., Frieden, C., Ono, S., Condeelis, J. et al. (2003). The structure of nonvertebrate actin: Implications for the ATP hydrolytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 5760-5765.

Yao, X., Grade, S., Wriggers, W., & Rubenstein, P. A. (1999). His73, often methylated, is an important structural determinant for actin. A mutagenic analysis of His73 of yeast actin. *Journal of Biological Chemistry*, *274*(52), 37443-37449.

Yao, X., & Rubenstein, P. A. (2001). F-actin-like ATPase Activity in a Polymerization-defective Mutant Yeast Actin (V266G/L267G). *Journal of Biological Chemistry*, *276*(27), 25598-25604.

Yuan, T., Weljie, A. M., & Vogel, H. J. (1998). Tryptophan fluorescence quenching by methionine and selenomethionine residues of calmodulin: Orientation of peptide and protein binding. *Biochemistry*, *37*(9), 3187-3195.

Zheng, X., Diraviyam, K., & Sept, D. (2007). Nucleotide effects on the structure and dynamics of actin. *Biophysical Journal*, *93*(4), 1277-1283.

Zhu, M., Yang, T., Wei, S., DeWan, A. T., Morell, R. J., Elfenbein, J. L. et al. (2003). Mutations in the γ-Actin Gene (ACTG1) Are Associated with Dominant Progressive Deafness (DFNA20/26). *American Journal of Human Genetics*, *73*(5), 1082-1091.

Chapter 6

Conclusions and Future Work

6.1 Monomeric G-Actin and the Effects of Nucleotide on Actin Structure and Dynamics

ATP-actin and ADP-actin are functionally different, with different affinities for actin binding proteins, as well as for polymerization. The structural reasons for these differences are not clear. Through the use of extensive molecular dynamics simulations, I sought to elucidate the changes to actin's structure and dynamics upon nucleotide hydrolysis. Actin has two primary clefts, the nucleotide binding cleft in the center of the protein, and the target-binding cleft at the bottom of the protein, between subdomains 1 and 3. The target-binding cleft is where the majority of actin binding proteins dock. As the two states have different affinities for binding partners, it follows that this cleft must exhibit some change in structure or dynamics upon nucleotide hydrolysis. There must also exist some allosteric connection between the two clefts.

To look for such allosteric connections, and to detect small scale changes within actin's loop regions, I developed a new variant of principal component analysis (PCA) to analyze the motion of loops. In this method, PCA analysis is performed on two or more structural states. The core of each actin was fit to a common structure, and then PCA was performed on small loop regions within the protein. Often these loop regions were not used in the fitting of the molecule.

This fitting allowed us to examine the small fluctuations of the loop within the context of the protein as a whole. By combining two or more states, the PCA seeks to maximize the differences between those states. As a whole, this new method allowed us to quickly examine all loop regions within actin for nucleotide state sensitivity.

As expected, the PCA on loops method showed that there was a concerted change of the S-, G-, and H73-loops within the nucleotide binding cleft. The S- and G-loops rocked with respect to each other, altering contacts with the nucleotide in the process. The S- and G-loops also moved away from each other in the ATP state. As a result, the H73-loop moved into the nucleotide binding cleft and took on a β-sheet structure in the ATP state. The results here recapitulated what many others had seen previously.

The loop PCA showed that actin's C-terminal hinge has nucleotide dependent dynamics. The C-terminal hinge is one of the two helices connecting subdomains 1 and 3 and forms a barrier between the nucleotide and target-binding clefts. In the ADP G-actins state, the C-terminal hinge has two accessible states. When the nucleotide is switched for ATP, the C-terminal hinge becomes more mobile and can take on a third conformation. K336 within the C-terminal hinge also makes hydrogen bonding contacts with the nucleotide. In the ADP state, K336 forms a backbone hydrogen bond with the nucleotide. When ATP is the nucleotide, K336 flips up parallel with the protein and makes side chain hydrogen bonds with the nucleotide. Due to both the location of the hinge within actin, as well as the contacts it makes with the nucleotide, we believe that the C-terminal hinge is the missing allosteric connection between the nucleotide and target-binding clefts.

Within the target-binding cleft, we show nucleotide state dependence of the W-loop, FQQ-loop, and C-terminus of the protein. As previously seen, we show that the W-loop is unstructured

in the ATP state and forms a β-sheet in ADP. But more interestingly, we identify the FQQ-loop which sits across the target-binding cleft from the W-loop at the bottom of subdomain 1. This loop, previously unidentified, moves by nearly 5 Å in the ATP state towards the W-loop to partially obscure the cleft. Further, in ADP G-actin, the FQQ-loop can form a short 3_{10}-helix. This loop at the binding site for several proteins, including profilin. The crystal structure of profilin bound actin, shows that the FQQ-loop is in its helical, open conformation. The final loop that shows substantial nucleotide dependent dynamics, was the C-terminus of the protein. While not technically within the target-binding cleft, it is at the bottom of subdomain 1 and can extend into the cleft. This extended conformation is unique to the ADP nucleotide state. The C-terminus has been shown to be at the binding surface for proteins such as profilin and gelsolin, both of which exhibit actin nucleotide dependent binding.

In summary, one of the two main goals of this work was to identify the basis for the nucleotide dependent conformational changes that lead to the differential affinities for actin binding proteins. Specifically looking for an allosteric pathway that communicates changes in the nucleotide-binding cleft down to the target-binding cleft. We believe that the C-terminal hinge is part of that allosteric pathway. Not only does it make contact with the nucleotide, but it also is the barrier separating the two clefts. We also showed that there were nucleotide dependent conformational changes with several loops within the target-binding cleft (W-loop, FQQ-loop and C-terminus), all of which make contacts with actin binding proteins. My second goal was to determine the nucleotide dependent changes within G-actin that lead to the two states having such different affinities for polymerization. I will discuss this more in the next section, but I show that ATP G-actin is much flatter than ADP G-actin. This flattening of the monomer puts it in a

conformation more similar to the F-actin protomer, providing a conformational argument for why ATP G-actin polymerizes so much faster than ADP G-actin.

6.1.1 Future Directions

In my work, I identified two previously unidentified loop regions that have substantially different conformations between nucleotide states, the C-terminal hinge and the FQQ-loop. The next step would be experimental validation of the different states of these loops.

All the above work was performed on α-skeletal muscle actin, just one of our six actin isoforms. In Chapter 5, I showed how a single point mutation in actin can drastically alter actin's dynamics. It would be interesting to study how the different actin isoforms, which can deviate by up to 7% from α-muscle actin, alter actin's dynamics.

6.2 Filament Dynamics and the Effects of Polymerization on Actin

G-actin comes together in a head-to-tail fashion to polymerize into F-actin filaments. Due to the nature of polymerization, the filament ends are intrinsically different. During filament treadmilling, there is net addition of protomers onto the barbed end of the filament and net dissociation of subunits from the pointed end. As of now, no high-resolution structures of the filament ends exist and much of the previous simulation work on the actin filament has been done on 'infinite' filaments using periodic boundary conditions. In Chapter 4, I performed molecular dynamics simulations on the F-actin filaments without periodic boundary conditions. I sought to determine the structural changes that occur during the monomer-to-protomer transition, as well as determine the differences between the filament ends that lead to their unique properties.

My first task was to examine the twist and offset between protomers within the filament. To do this, I developed a new method based on the Kabsch algorithm that allowed for an unbiased

determination of the filament axis and twist between protomers. The modified Kabsch algorithm represents a significant improvement over the old method which relied on the geometries of the centers-of-mass of two consecutive trimers to determine the filament axis. Using this method, we found that the middle and barbed ends of the filament had very similar twist values, but that the pointed end protomer was under twisted. We also showed, using PCA, that the middle and barbed ends of the filament displayed similar dynamics but that the pointed end, again, diverged taking on a much flatter conformation. Taken together, we believe that the differential conformation of the pointed end protomers may serve to block the addition of incoming actin subunits, as well as impose a high conformational energy barrier, inhibiting pointed end polymerization.

Comparing the monomer to the protomer, we found that the ADP F-actin protomer was flatter than its G-actin counterpart – more similar to what was observed for ATP G-actin. Using an inverse-Boltzmann distribution we were able to measure the free energy differences between the ADP and ATP G-actin monomers and the ADP F-actin barbed end protomer. We found that the difference in binding free energies between ADP and ATP G-actin was 2.78 k_BT, consistent with the free energy difference of 2.7 k_BT derived from experiments. This would suggest that the difference in polymerization kinetics results from the differences in tertiary conformations, presented in Chapter 3, between the G-actin states.

6.2.1 Future Directions

Despite our long simulation times, some of our results suggest that we still see a lack of complete convergence in our F-actin filaments (see sections 4.2.1 and 4.2.3.1), suggesting the need for further simulation, especially of the Fujii F-actin model. At the end of Chapter 4, I present our preliminary results from an ATP F-actin filament. I would suggest either extending this simulation, or building a new ATP F-actin filament from the newly available ATP F-actin model (Merino et

al., 2018). Just as with the monomer, the nucleotide state of the filament is key in determining which actin binding proteins interact with the filament, and is key for regulation of the filament dynamics within the cell.

6.3 Yeast G-Actin and the Allosteric Effects of Two Deafness-Causing Mutations (DFNA20/26)

As humans have six different actin isoforms, obtaining a pure sample of actin can be difficult. For this reason, researchers often turn to model systems to study actin mutations. Yeast actin has 86.7% sequence identity with α-muscle actin, making it a good model system. In Chapter 5, we performed extended molecular dynamics simulations of yeast G-actin in the ADP state. Despite yeast and muscle G-actin having differing sequences, and different crystallographic starting points, the two converged to the same state and displayed remarkably similar dynamics.

My second goal in Chapter 5, was to evaluate the effects of two dissimilar amino acid substitutions in actin's pathogenic helix that lead deafness in humans, K118M and K118N. To this end, we partnered with Karina A. Kurth and Peter A. Rubenstein from the Department of Biochemistry at University of Iowa, who performed biochemical experiments to complement our computational analysis.

K118 within the pathogenic helix lies at the binding surface for the Arp2/3 complex. Our collaborators had previously shown that both mutants led to altered interactions with Arp2/3. We believe that the mutation of lysine to the more hydrophobic methionine may inhibit the docking of actin binding proteins that interact through hydrogen bonds or ionic interactions in this region, such as the Arp2/3 complex. Our simulations of the K118N mutation show a significant alteration of both the pathogenic and W79 helices. The pathogenic helix in K118N becomes more bent. This allows N115 to form new hydrogen bonds with W79, N111 and the backbone of V76. This shifts

the pathogenic helix up, pushing W79 into the space that was previously occupied by W86. As a result, the side chain of W86 gets flipped. In WT actin, W79 fluorescence gets quenched by non-radiative energy transfer to W86. The flipping of W86 reduces this energy transfer, leading to the K118N actin having a red-shifted tryptophan fluorescent spectrum.

The new hydrogen bonding network also leads to loss of secondary structure of the H73-loop. The H73-loop has previously been shown to have implications in monomer twist. We think that the loss of secondary structure in the H73-loop leads to a flattening of the monomer in K118N. PCA analysis shows that the K118N takes on a flatter, more F-actin like conformation. Similar to the argument made in Chapters 3 and 4, we believe that putting K118N in a flatter more F-actin conformation explains why it polymerizes twice as fast as WT actin. By having a more similar conformation to the filament, there is a lower conformational energy barrier, facilitating polymerization.

Proper cytoskeletal function is dependent upon tight regulation of actin polymerization and depolymerization. Overall, we believe that the deafness caused by mutations to K118 results from a misregulation of polymerization and altered pattern of interaction with actin binding proteins.

6.3.1 Future Directions

Following up with the work done in Chapter 5, many future directions exist. Ten other mutations to actin's pathogenic helix have been identified, each with their own merit for further study. A logical next step would be to examine the N115T/S mutations. These mutations to α-muscle actin cause nemaline myopathies. The N115T mutation is also found in α-smooth muscle actin. In smooth muscle, N115T leads to TAAD aneurysms. Simulations of this mutant would have bearing on two different disease states. An alternative future direction would be the continued study of ADP G-actin in different actin systems, such as plasmodium or toxoplasma.

6.4 Significance

The last seven decades of actin research has gone a long way to reveal how actin functions within the cell, but some of the molecular details have remained elusive. Such as the structural differences between ADP and ATP G-actin that lead to their unique properties. Or why the pointed end of the actin filament polymerizes so much slower than the barbed end. My research helps to fill in these missing details. I show that ATP G-actin is in an F-actin like conformation, allowing it to polymerize faster than its ADP counterpart. I also found several new loop regions within actin that change their structure and dynamics depending on actin's nucleotide state, providing a structural basis for actin binding protein's nucleotide state-sensitivity. Within the filament, I show that the pointed end takes on a divergent conformation that prevents polymerization. Overall, my research reveals many of the structural details that influence polymerization kinetics.

6.5 Reference

Merino, F., Pospich, S., Funk, J., Wagner, T., Küllmer, F., Arndt, H. et al. (2018). Structural transitions of F-actin upon ATP hydrolysis at near-atomic resolution revealed by cryo-EM. *Nature Structural and Molecular Biology*, *25*(6), 528-537.

Appendix

Sample Analysis Scripts and Functions

A.1 R Functions

Note, the following functions require the bio3d R package and were bundled into a new R package called ljepsenThesis.

A.1.1 dssp.dcd.R

```
     #' dssp.dcd
     #'
     #' This is a modification of bio3d's dssp function. It
works similarily to \code{\link{[bio3d]dssp}}, but was modified
to work with my system and designed to only return secondary
structure information.
     #' @param pdb a structure object of class "pdb", obtained
from \code{\link{[bio3d]read.pdb}}
     #' @param dcd a trajectory object of class "xyz", obtained
from \code{\link{[bio3d]read.dcd}}
     #' @param exefile file path to 'DSSP' program
     #' @return returns a matrix where row corresponds to the
number of frames in 'dcd' and column corresponds to residue in
'pdb'. Entries indicate type of secondary structure element.
     #' \item{H}{alpha-helix}
     #' \item{B}{residue in isolated beta-bridge}
     #' \item{E}{extended strand, participates in beta ladder}
     #' \item{I}{pi-helix}
     #' \item{T}{hydrogen bonded turn}
     #' \item{S}{bend}
     #' \item{" "}{ unstructured loop or irregular coil}
```

```
     #' @export

     "dssp.dcd" <-
       function (pdb, dcd, exefile =
"~/Documents/Yeast_Actin/E117K2/analysis/PCA_without_loops/dssp"
, verbose=FALSE) {

         cl <- match.call()
         os1 <- .Platform$OS.type

         trim <- function(s) {
           s <- sub("^ +", "", s)
           s <- sub(" +$", "", s)
           s[(s == "")] <- NA
           s
         }
         split.line <- function(x, split = " ") {
           tmp <- unlist(strsplit(x, split = split))
           inds <- which(tmp != "")
           return(trim(tmp[inds]))
         }

         fin.sse = NULL

         nf = dim(dcd)[1]
         pb = txtProgressBar(1, nf, style = 3)

         for(frame in 1:nf){
           setTxtProgressBar(pb, frame)
           infile <- tempfile()
           outfile <- tempfile()
           write.pdb(pb, xyz = dcd[frame,], file = infile)

           if (os1 == "windows") {
             shell(paste(exefile, infile, outfile),
ignore.stderr = !verbose,
                   ignore.stdout = !verbose)
           }
           else {
             success <- system(paste(exefile, infile, outfile),
ignore.stderr = !verbose,
                               ignore.stdout = !verbose)
             if (success != 0)
               stop(paste("An error occurred while running
command\n '",
                          exefile, "'", sep = ""))
           }
```

```
            raw.lines <- readLines(outfile)
            unlink(c(infile, outfile))
            type <- substring(raw.lines, 1, 3)
            raw.lines <- raw.lines[-(1:which(type == "  #"))]
            aa <- substring(raw.lines, 14, 14)
            if (any(aa == "!")){
              raw.lines <- raw.lines[-which(aa == "!")]
            }

            sse <- substring(raw.lines, 17, 17)

            fin.sse = rbind(fin.sse, sse)

          } # End frame for loop
          cat('\n')
          return(fin.sse)

        } # End Function
```

A.1.2 get_sseFreq.R

```
#' Calculates Frequency of Secondary Structure Elements
#'
#' The goal of this function is to take the output from dssp.dcd
(sse) and calculte the frequency of element occurance for a
given frame across the trajectory.
#' @param sse matrix of secondary structure elements. Output of
\link{dssp.dcd}
#' @param combine logical, if TRUE will combine elements into
helix, beta, and other
#' @export

get_sseFreq = function(sse, combine = F){
  elementList = c(' ', 'S', 'T', 'E', 'B', 'H', 'G', 'I')

  nf = dim(sse)[1] # Number of Frames
  nr = dim(sse)[2] # Number of Residues

  if(combine){
    freq_helix = function(data){
      length(which(data %in% c('H', 'G', 'I')))/length(data)
    }
    freq_beta = function(data){
      length(which(data %in% c('E', 'B')))/length(data)
    }
    freq_other = function(data){
      length(which(data %in% c(' ', 'S', 'T')))/length(data)
```

```
    }

    freq = matrix(data = NA, nrow = 3, ncol = nr)
    rownames(freq) = c('Helix', 'Beta', 'Other')

    freq[1,] = apply(sse, 2, freq_helix)
    freq[2,] = apply(sse, 2, freq_beta)
    freq[3,] = apply(sse, 2, freq_other)

  } else {
    freq = matrix(data = NA, nrow = 8, ncol = nr)
    rownames(freq) = elementList

    for(x in 1:length(elementList)){
      tempFunction = function(data, element = elementList[x]){
        length(which(data == element))/length(data)
      }
      freq[x,] = apply(sse, 2, tempFunction)
    } # End element Loop
  }

  return(freq)
}
```

A.1.3 get_sse.R

```
#' Perform Secondary Structure Analysis
#'
#' This function is designed to perform secondary structure
analysis.
#' @param dcd a trajectory object of class "xyz", obtained from
\code{\link{[bio3d]read.dcd}}
#' @param pdb a structure object of class "pdb", obtained from
\code{\link{[bio3d]read.pdb}}
#' @param file logical (default = F), if TRUE indicates dcd and
pdb are paths to dcd and pdb files, and that they need to be
read in.
#' @return Returns a list of the following elements:
#' \item{sse}{secondary structure by residue by frame. Output of
\link{dssp.dcd}}
#' \item{sse.freq}{frequency of secondary structure elements.
Output of \link{get_sseFreq} with combine = F}
#' \item{sse.freq2}{frequency of secondary structure elements
combined into helix, beta, and other. Output of
\link{get_sseFreq} with combine = T}
#'@export
get_sse = function(dcd, pdb, file = F){
  if(file){
```

```
    dcd = read.dcd(dcd)
    pdb = read.pdb(pdb)
  }
  sse = dssp.dcd(pdb, dcd)
  sse.freq = get_sseFreq(sse, combine = F)
  sse.freq2 = get_sseFreq(sse, combine = T)

  results = list("sse" = sse, "sse.freq" = sse.freq, "sse.freq2"
= sse.freq2)
  return(results)
}
```

A.1.4 crystProject.R

The following collection of functions is used for the projection of PDB structures onto our G-Actin PCA Basis set.

A.1.4.1 check_cryst

```
#' Resolves object cryst into PDB object
#'
#' The goal of this function is to resolve the object cryst into
a PDB object.
#' @param cryst can be a pdbid with or without a chain
identifier, a path to a pdb file, or a pdb object.
#' @param type logical, if TRUE will return a list of the pdb
object and the original type for object cryst. (Default: TRUE)
#' @param verbose logical (Default: FALSE)
#' @return if type = TRUE, will return a list of a pdb object
and the original type of object cryst. If type = FALSE, will
just return a pdb object.
#' @examples
#' check_pdb("1j6z")
#' @note Will fail if given a pdb file with number of characters
less than 7. For example 'a.pdb'.
#' @export
check_cryst = function(cryst, type = T, verbose = F) {

  if(!is.pdb(cryst)){
    #          print('Cryst is not a pdb.')
    if(is.character(cryst) & nchar(cryst) > 6){
      if(verbose) print('Cryst is not a pdb. Checking if cryst
is a PDB file.')
      temp = try(read.pdb(cryst, verbose = F), silent = T)
      if(is.pdb(temp)){
        if(verbose) print("Cryst is a PDB file.")
```

```
        cryst = temp
        if(type) kind = 'file'
      } else {
        if(verbose) print("Cryst is not a PDB file.")
        stop("Please provide a PDB object, PDBID, or PBD File
for cryst.")
      }
    } else if(is.character(cryst)) {
      if(verbose) print('Cryst is not a pdb. Checking if cryst
is a PDBID.')
      temp = try(read.pdb(try(get.pdb(cryst, URLonly = T),silent
= T), verbose = F), silent = T)
      if(is.pdb(temp)){
        if(verbose) print('Cryst is a PDBID')
        cryst = temp
        if(type) kind = 'pdbid'
      } else {
        if(verbose) print("Cryst is not a PDBID. Checking if
cryst is a PDB file.")
        temp = try(read.pdb(cryst, verbose = F), silent = T)
        if(is.pdb(temp)){
          if(verbose) print("Cryst is a PDB file.")
          if(type) kind = 'file'
          cryst = temp
        } else {
          if(verbose) print("Cryst is not a PDB file.")
          stop("Please provide a PDB object, PDBID, or PBD File
for cryst.")
        }
      }
    } else {
      stop("Please provide a PDB object, PDBID, or PBD File for
cryst.")
    }

  } else if(type) {
    kind = 'pdb'
  }

  if(type) {
    return(list(pdb = cryst, type = kind))
  } else {
    return(cryst)
  }
}
```

A.1.4.2 cryst.process

```
#' Workhorse behind crystProject'
#'
#' Add documentation to this!
#' @export
cryst.process = function(cryst, pdb, xyz, pca, resno.xyz,
seq.xyz, pdb.seq, colinds, aln, crystName = 'cryst', muscle =
"~/Documents/muscle3.8.31_i86darwin64"  ){

  if(!(all(aln$ali[1,colinds] == seq.xyz))){
    stop("ERROR - Alignment colinds not equal to seq.xyz")
  }

  cryst.inds = suppressWarnings(pdb2aln.ind(aln, cryst, inds =
colinds, exefile = muscle, file = NULL))
  cryst.xyz = cryst.inds$b$xyz

  if(length(cryst.xyz) == length(pca$mean)){
    cryst.proj = project.pca(cryst$xyz[cryst.xyz], pca, fit = T,
fixed.inds = 1:length(pca$mean), mobile.inds =
1:length(cryst.xyz))
    aln = seqaln(seqbind(pdb.seq, pdbseq(cryst)), id = c("G-
Actin", crystName), exefile = muscle, outfile = NULL)
    ident = seqidentity(aln)[1,2]
    info = list(pdb = cryst, aln = aln, proj = cryst.proj, as =
cryst.inds$b, seqident = ident, work = T)
  } else {
    aln = seqaln(seqbind(pdb.seq, pdbseq(cryst)), id = c("G-
Actin", crystName), exefile = muscle, outfile = NULL)
    ident = seqidentity(aln)[1,2]

    info = list(pdb = cryst, aln = aln, proj = NULL, as = NULL,
seqident = ident, work = F )
  }
  return(info)
} # End cryst.process
```

A.1.4.3 crystProject

```
#' Projection of a PDB onto a PCA Basis Set
#'
```

```
#' This function creates a robust method for the projection of a
structure onto a PCA basis set. After obtaining the structure
specified by
#' cryst using \code{\link{check_cryst}}, it will align the
structure to the PDB corresponding to the pca basis set using
\code{\link[bio3d]{seqaln}}.
#' If the structure contains the required residues needed for
pca projection, it will be projected onto the basis set using
\code{\link[bio3d]{project.pca}}.
#' @param cryst The structure to be projected onto the pca basis
set. Can be a pdbid, a path to a pdb file, or a pdb object.
#' @param pdb An object of type pdb corresponding to the
structure used to create the pca basis set.
#' @param xyz The xyz indicies corresponding to indicies of the
pdb used to create the pca basis set.
#' @param pca The pca basis created using
\code{\link[bio3d]{pca.xyz}}.
#' @param chain A vector of chain identifiers. If provided,
cryst will be trimmed to those chains before proceeding.
(Default: NULL)
#' @param crystName A name given to the structure cryst to be
used in the alignment.
#' @param pdbName A name given to the pdb structure to be used
in the alignment.
#' @param split A logical indicating whether or not to split the
structure cryst into separate pdbs for each chain. (Default:
FALSE)
#' @param muscle A path to the muscle executable.
#' @param verbose A logical. (Default: TRUE)
#' @return Returns a list with the following attributes:
#' \item{pdb}{A pdb object or a list of pdbs with keys of chain
if split = T}
#' \item{aln}{An alignment between cryst and pdb}
#' \item{proj}{A vector or matrix of the projection of cryst
onto pca}
#' \item{as}{An atom selection or list of selections for the
indicies of cryst that are used in the projection}
#' \item{seqident}{The sequence identity of cryst compared to
pdb}
#' \item{work}{Logical indicating if projection worked, or a
vector of chains that projection worked for}
#' \item{call}{Call}
#' @export
crystProject = function(cryst, pdb, xyz, pca, chain = NULL,
crystName = 'cryst', pdbName = "G-Actin", split = F,
                        muscle =
"~/Documents/muscle3.8.31_i86darwin64", verbose = T  ){
```

```
  ## Example: crystProject('2W4U', pdb.gAct, ss.gAct.xyz,
gAct.pca, chain = c('E', 'F', 'G', 'H', 'I', 'J'), split = T)
  resno.xyz = pdb$atom[xyz2atom(xyz), "resno"]
  seq.xyz = pdbseq(pdb, inds = atom.select(pdb, resno =
resno.xyz, elety = "CA"))
  pdb.seq = pdbseq(pdb)

  colinds = which(names(pdb.seq) %in% resno.xyz)

  aln = seqaln(rbind(pdbName = pdb.seq, "DUP" = pdb.seq),
exefile = muscle, outfile = NULL)

  temp = check_cryst(cryst)
  if(temp$type == 'pdbid' & crystName == 'cryst'){
    crystName = cryst
  }
  cryst = temp$pdb
  if (verbose) cat('Retrieved Crystal\n')

  if(!is.null(chain))    {
    cryst = trim(cryst, inds = atom.select(cryst, chain =
chain))
    #print('Trimmed chain')
    if(length(chain)== 1 & split){
      split = F
      print('Set split to F')
    }
  }

  if(split){
    # print('in split')
    ## Need to process all chains with cryst.process
    chains = sort(unique(cryst$atom[,'chain']))

    pdb_list = list() # Only Save Working
    cryst_seqs = seqbind(pdb.seq) # Only Save Working
    working_chains = c()
    ids = c(pdbName)

    seqident = vector('numeric', length = length(chains))
    names(seqident) = chains

    cryst.proj = c()
    cryst.as = list()
    if (verbose) cat("Chains:",chains, "\n", sep = " ")
    for(c in chains){
      if (verbose) cat('\tProcessing:', c, '\n')
      id = paste(c(crystName, c), collapse = "_")
```

```
      temp_pdb = trim(cryst, inds = atom.select(cryst, chain =
c))
      temp.info = cryst.process(temp_pdb, pdb, xyz, pca,
resno.xyz, seq.xyz, pdb.seq, colinds, aln, crystName = id,
muscle)
      #cat(c,temp.info$seqident, '\n', sep = "\t")
      seqident[c] = temp.info$seqident
      if(temp.info$work){
        #cat(sprintf('Chain %s Works',c), '\n')
        working_chains = append(working_chains, c)
        pdb_list[[c]] = temp_pdb
        cryst_seqs = seqbind(cryst_seqs, pdbseq(temp_pdb))
        ids = c(ids, id)
        cryst.proj = rbind(cryst.proj, id = temp.info$proj)
        cryst.as[[c]] = temp.info$as
      }
    }

    rownames(cryst.proj) = working_chains
    if(!is.null(working_chains)){
      new_aln = seqaln(cryst_seqs, id = ids, exefile = muscle,
outfile = NULL)
      info = list(pdb = pdb_list, aln = new_aln, proj =
cryst.proj, as = cryst.as, seqident = seqident, work =
working_chains, call = match.call())
    } else {
      simpleWarning(sprintf('None of the chains in %s contain
all residues needed for projection onto pca.', crystName))
      info = NULL
    }
    #          info = list(pdb = pdb_list, aln = new_aln, proj =
cryst.proj, as = cryst.as, seqident = seqident, work =
working_chains, call = match.call())
    return(info)
  } else {
    temp.info = cryst.process(cryst, pdb, xyz, pca, resno.xyz,
seq.xyz, pdb.seq, colinds, aln, crystName, muscle)
    info = list(pdb = temp.info$pdb, aln = temp.info$aln, proj =
temp.info$proj, as = temp.info$as, seqident =
temp.info$seq.ident, work = temp.info$work, call = match.call())
    tryCatch(if(!(temp.info$work)) stop(sprintf('%s does not
contain all residues needed for projection onto pca.',
crystName)), finally = return(info))
  }
}
```

A.1.5 crystPropellerTwist.R

```
#' Propeller-Twist, Clamp and Cleft for a PDB
#'
#' @param cryst An object from \code{\link{crystProject}}
#' @param pdb The pdb object used for building cryst (pdb.gAct)
#' @param pdb.sub The residue numbers for subdomains 1-4 in pdb
used to calculate the propeller twist.
#' @param pdb.cleft The residue numbers (in pdb) for the defined
cleft mouth size.
#' @param pdb.cleft The residue numbers (in pdb) for the defined
phosphate clamp.
#' @param muscle Muscle alignment algorithm file location.
crystPropellerTwist = function(cryst, pdb, pdb.sub1 =
c(9:12,17:22,30:32,80:92,104:107,114:125,132:136,138:145,339:348
,358:365), pdb.sub2 = c(33,36:38,54:55,57:60,67:69), pdb.sub3 =
c(151:155,162:167,170:171,177:178,275:284,291:295,298:301,304:30
5,310:321,330:331), pdb.sub4 =
c(183:196,204:216,224:231,239:242,248:251,254:257,259:260),
pdb.cleft = c(60,208), pdb.clamp = c(16, 158), muscle =
"~/Documents/muscle3.8.31_i86darwin64"){

  pdb.seq = pdbseq(pdb)

  ### Need to Check how Cryst Saved ###
  w = cryst$work

  if(is.logical(w) && w){
    sub1 = pdb2aln.ind(cryst$aln, cryst$pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub1)$atom,"resno"]), exefile = muscle)$b
    sub2 = pdb2aln.ind(cryst$aln, cryst$pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub2)$atom,"resno"]), exefile = muscle)$b
    sub3 = pdb2aln.ind(cryst$aln, cryst$pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub3)$atom,"resno"]), exefile = muscle)$b
    sub4 = pdb2aln.ind(cryst$aln, cryst$pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub4)$atom,"resno"]), exefile = muscle)$b

    clamp.inds = pdb2aln.ind(cryst$aln, cryst$pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.clamp)$atom,"resno"]), exefile = muscle)$b

    cleft.inds = pdb2aln.ind(cryst$aln, cryst$pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.cleft)$atom,"resno"]), exefile = muscle)$b
```

```
    dihed = getDihedral(com(cryst$pdb, inds =
sub2),com(cryst$pdb, inds = sub1),com(cryst$pdb, inds =
sub3),com(cryst$pdb, inds = sub4))

    clamp =
dist.xyz(cryst$pdb$xyz[clamp.inds$xyz[1:3]],cryst$pdb$xyz[clamp.
inds$xyz[4:6]])
    cleft =
dist.xyz(cryst$pdb$xyz[cleft.inds$xyz[1:3]],cryst$pdb$xyz[cleft.
inds$xyz[4:6]])

    df = data.frame("Dihedral_2134" = dihed, "Clamp_Size" =
clamp, "Mouth_Size" = cleft)
    rownames(df)= unique(cryst$pdb$atom$chain)
    return(df)
  } else if(!is.logical(w)){
    dihed = vector('numeric',length = length(w))
    names(dihed)=w
    clamp = vector('numeric',length = length(w))
    names(clamp)=w
    cleft = vector('numeric',length = length(w))
    names(cleft)=w
    for(i in w){
      aln = cryst$aln
      cryst.pdb = cryst$pdb[[i]]

      sub1 = pdb2aln.ind(aln, cryst.pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub1)$atom,"resno"]), exefile = muscle)$b
      sub2 = pdb2aln.ind(aln, cryst.pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub2)$atom,"resno"]), exefile = muscle)$b
      sub3 = pdb2aln.ind(aln, cryst.pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub3)$atom,"resno"]), exefile = muscle)$b
      sub4 = pdb2aln.ind(aln, cryst.pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.sub4)$atom,"resno"]), exefile = muscle)$b

      clamp.inds = pdb2aln.ind(aln, cryst.pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.clamp)$atom,"resno"]), exefile = muscle)$b

      cleft.inds = pdb2aln.ind(aln, cryst.pdb, inds =
which(names(pdb.seq) %in% pdb$atom[atom.select(pdb, elety =
"CA", resno = pdb.cleft)$atom,"resno"]), exefile = muscle)$b
```

```
      dihed[i] = getDihedral(com(cryst.pdb, inds =
sub2),com(cryst.pdb, inds = sub1),com(cryst.pdb, inds =
sub3),com(cryst.pdb, inds = sub4))

      clamp[i] =
dist.xyz(cryst.pdb$xyz[clamp.inds$xyz[1:3]],cryst.pdb$xyz[clamp.
inds$xyz[4:6]])
      cleft[i] =
dist.xyz(cryst.pdb$xyz[cleft.inds$xyz[1:3]],cryst.pdb$xyz[cleft.
inds$xyz[4:6]])
    } # End chain for loop

    df = data.frame("Dihedral_2134" = dihed, "Clamp_Size" =
clamp, "Mouth_Size" = cleft)
    rownames(df) = w
    return(df)
  } # End if(!is.logical(w))
} # End Function
```

A.1.6 vecMath.R

A.1.6.1 veccross

```
#' Cross Product
#'
#' The cross product between to vectors of length 3.
#' @param a a vector of length 3
#' @param b a vector of length 3
#' @return
#' A numeric vector of length 3 equal to the cross product of
vectors a and b.
#' @export
#' @examples
#' veccross(c(1,6,3), c(6,9,2))
veccross = function(a,b){
  if(!all(c(length(a), length(b)) == 3)){
    warning("length of a and/or b not equal to 3")
  }
  # Cross product of vectors a and b
  return(c(a[2]*b[3]-a[3]*b[2],
           a[3]*b[1]-a[1]*b[3],
           a[1]*b[2]-a[2]*b[1]))

}
```

A.1.6.2 veclength

```
#' Vector Length
#'
#' Gives the length of the vector vec
#' @param vec a vector
#' @export
veclength = function(vec){
  # Get the length of the vector vec
  a = sqrt(sum(vec^2)) #(vec[1]^2 + vec[2]^2 + vec[3]^2)^0.5
  return(a)
}
```

A.1.6.3 unitvec

```
#' Unit Vector
#'
#' Returns a unit vector in the direction of vec.
#' @param vec a vector
#' @export
#' @examples
#' unitvec(c(1,6,3))
unitvec = function(vec){
  # Get a unit vector inthe direction of vec
  vlen = veclength(vec)
  return(vec/vlen)
}
```

A.1.6.4 my.angle

```
#' Angle
#'
#' Calculates the angle between 3 points.
#' @param p1 a vector of either length 3 or 9
#' @param p2 a vector of length 3
#' @param p3 a vector of length 3
#' @param degree a logical, default T, return value in degrees
#' @return
#' Returns the angle between P1-P2-P3
#' @examples
#' my.angle(c(1, 6, 3, 6, 9, 3, 7, 1, 2))
#' my.angle(c(1, 6, 3), c(6, 9, 3), c(7, 1, 2))
#' my.angle(c(1, 6, 3), c(6, 9, 3), c(7, 1, 2), degree = F)
```

```
my.angle = function(p1, p2 = NULL, p3 = NULL, degree = T){
  if(is.null(p2) || is.null(p3)){
    p2 = p1[4:6]
    p3 = p1[7:9]
    p1 = p1[1:3]
  }

  v1 = c(unitvec(p2 - p1))
  v2 = c(unitvec(p2 - p3))

  a = atan2(veclength(veccross(v1,v2)), v1 %*% v2)
  if(degree) a = a*180/pi
  return(a)
}
```

A.1.6.5 Tr

```
#' Trace
#'
#' Returns the trace of a matrix.
#' @param x a matrix
#' @return
#' Returns the trace of the matrix x, aka the sum of the
diagonal.
#' @export
#' @examples
#' Tr(matrix(1:9, nrow = 3))
Tr = function(x){
  return(sum(diag(x)))
}
```

A.1.6.6 projABC

```
#' Vector Projection
#'
#' The vector projection of AC onto AB
#' @param a a point in 3 space, or 3 points (vector length 3 or
9)
#' @param b a vector of length 3
#' @param c a vector of length 3
#' @export
projABC = function(a, b = NULL, c = NULL){
  if(length(a) == 9){
    b = a[4:6]
    c = a[7:9]
```

```
    a = a[1:3]
  }
  ## Projection of AC onto vector AB
  ## a, b, c are all pionts

  ab = b-a
  ac = c-a

  p = ((ac%*%ab)/(ab%*%ab))*ab

  return(p)
}
```

A.1.6.7 getDihedral

```
#' Dihedral Angle
#'
#' Calculates the dihedral angle between four points: com1-com2-
com3-com4
#' @param com1 vector of xyz coordinates for first point
#' @param com2 vector of xyz coordinates for second point
#' @param com3 vector of xyz coordinates for third point
#' @param com4 vector of xyz coordinates for fourth point
#' @param atan2 logical (default = TRUE), indicates if use of
atan2 should be used for angle calculation. If false, will use
acos instead
#' @param degree logical (default = TRUE). If true will return
angle in degrees. If false, will return angle in radians.
#'
#' @export
getDihedral = function(com1, com2, com3, com4, atan2 = TRUE,
degree = T){
  b1 = com2 - com1
  b2 = com3 - com2
  b3 = com4 - com3

  n1 = unitvec(veccross(b1, b2))
  n2 = unitvec(veccross(b2, b3))

  if(atan2){
    m1 = veccross(n2, unitvec(b2))

    x = n1 %*% n2 #  [vecdot $n1 $n2]
    y = m1 %*% n1 #set y  [vecdot $m1 $n1]

    a = atan2(y,x)
```

```
    ifelse(degree, return(a*180/pi), return(a))
  } else {
    c = acos(n1 %*% n2)

    ifelse(degree, return(c*180/pi), return(c))
  }
}
```

A.1.7 twist_offset.R

This collection of functions is used to analyze filament twist and offset

A.1.7.1 getU

```
#' getU
#'
#' This function finds the fit between two sets of coordinates.
#' (1) Moves both objects to the origin
#' (2) Gets the covariance of the two objects \code{\link{cov}}
#' (3) Singular value decomposition of the covariance matrix
\code{\link{svd}}
#' (4)
#' @param xyz1 mobile xyz indicies
#' @param xyz2 fixed xyz indicies
#' @param verbose a logical
#' @export
getU = function(xyz1, xyz2, verbose = F){
  # Mobile: xyz1
  # Fixed:  xyz2
  xyz1 = as.xyz(xyz1)
  xyz2 = as.xyz(xyz2)

  P = matrix(moveby(xyz1, com(xyz1)), ncol = 3, byrow = T)
  Q = matrix(moveby(xyz2, com(xyz2)), ncol = 3, byrow = T)

  A = t(P) %*% Q#cov(P,Q)
  A.svd = svd(A)
  A.svd$D = diag(A.svd$d) # A = A.svd$u %*% A.svd$D %*%
t(A.svd$v)

  d = sign(det(A.svd$v %*% t(A.svd$u)))
  U = A.svd$v %*% diag(c(1,1,d)) %*% t(A.svd$u)
  if(verbose){cat(sprintf("{%s}\n",paste(apply(U,1,paste,
collapse = " "), collapse = "} {")))}

  return(U)
```

```
}
```

A.1.7.2 moveby

```
#' moveby
#'
#' This function moves coordinates xyz by vector v
#' @param xyz xyz coordinates
#' @param v a vector
#' @param ... other parameters passed to \link{sweep}
#' @export
moveby = function(xyz, v, ...){
  # v is a vector to subtract from xyz
  return(as.xyz(as.vector(sweep(matrix(xyz,nrow = 3), 1, v,
...))))
}
```

A.1.7.3 move

```
#' move
#'
#' This function applies a 3x3 rotation matrix to a set of
indicies.
#' @param xyz a set of XYZ coordinates
#' @param U a matrix representing a rotation about an axis
#' @export
move = function(xyz, U){
  #U is a rotation matrix
  xx = matrix(xyz,nrow =3, )
  xx = U %*% xx
  return(as.xyz(as.vector(xx)))
}
```

A.1.7.4 getEigen

```
#' getEigen
#' @export
getEigen = function(t, realOnly = T){
  e = eigen(t)
  x = which(Im(e$values) == 0)
  if(length(x) > 1){
    e$values = e$values[x]
    e$vectors = e$vectors[,x]

    x = which.min(abs(Re(e$values) - 1))

    e$values = e$values[x]
```

```
    e$vectors = e$vectors[,x]
    if(realOnly){
      e$values = Re(e$values)
      e$vectors = Re(e$vectors)
    }
    return(e)
  } else if (length(x) == 1) {
    e$values = e$values[x]
    e$vectors = e$vectors[,x]
    if(realOnly){
      e$values = Re(e$values)
      e$vectors = Re(e$vectors)
    }
    return(e)
  }
}
```

A.1.7.5 decompU

```
#' decompU
#'
#' Decomposes rotation matrix found using \code{\link{getU}}
into an angle and an axis.
#' @param U rotation matrix found using \code{\link{getU}}
#' @param degree logical. If TRUE, returns the rotation angle in
degrees (May be of opposite sign...)
#' @export
decompU = function(U, degree = T){
  u = getEigen(U)$vectors
  x = 0.5*(Tr(U)-1)
  theta = atan2(sqrt(1-x*x), x)

  if(degree){theta = theta*180/pi}
  return(list(theta = theta, axis = u))
}
```

A.1.7.6 fit

```
#' fit
#'
#' Designed to calculate the fit between two objects in terms of
rotation about an axis, and offset along the axis.
#' xyzM and xyzF represent a the coordinates of a dimer pair
centered on the origin.
#' First calculates the fit matrix U using \code{\link{getU}}
#' @param xyzM coordinates of mobile protomer where dimer has
been moved to the origin
```

```
#' @param xyzF coordinates of the fixed protomer where dimer has
been moved to the origin
#' @export
fit = function(xyzM, xyzF){
  # Mobile: xyzM
  # Fixed:  xyzF

  U = getU(xyzM,xyzF)
  e = decompU(U)

  xyzM.bar = com(as.xyz(xyzM))
  xyzF.bar = com(as.xyz(xyzF))
  xyzM.proj = ((xyzM.bar %*% e$axis)/(e$axis %*% e$axis)) %*%
e$axis
  xyzF.proj = ((xyzF.bar %*% e$axis)/(e$axis %*% e$axis)) %*%
e$axis

  trans = xyzF.proj-xyzM.proj ##e$axis*c(com(as.xyz(xyzF)) -
com(as.xyz(xyzM)))
  offset = veclength(trans)
  twist = e$theta
  m = rbind(cbind(U,c(trans)),c(0,0,0,1))
  VMD = sprintf("[list {%s}]",paste(apply(m,1,paste, collapse =
" "), collapse = "} {"))
  #cat(VMD, '\n')

  return(list(twist = e$theta, axis = e$axis, offset = offset, U
= U, t = trans, vmd = VMD ))
}
```

A.1.7.7 twist_offset

```
#' twist_offset
#'
#' This function is used to calculate the twist and offset
between two protomers in F-Actin. This is done in a series of
steps:
#' (1) Moves the COM of the dimer to the origin
#' (2) Calculates fit using the \code{\link{fit}} function
#' (3)
#' @param fixed numeric vector of xyz coordinates, can have more
than one coordinate set.
#' @param mobile numeric vector of xyz coordinates, can have
more than one coordinate set.
#' @param fixed.inds a vector of indices that selects the
elements of fixed upon which fitting should be based.
#' @param mobile.inds a vector of indices that selects the
elements of mobile upon which fitting should be based.
```

```
#' @param getU logical, if TRUE it will return the elements of
the rotation matrix U
#' @param verbose logical.
#' @return Returns a matrix of the following attributes
#' \item{theta}{rotation angle in degrees}
#' \item{offset}{offset distance between fixed and mobile}
#' \item{Cx Cy Cz}{center of the dimer}
#' \item{Vx Vy Vz}{vector of rotation and translation.}
#' \item{Pre-rmsd}{rmsd between fixed and mobile before fitting}
#' \item{Fit-rmsd}{rmsd between fixed and mobile after fitting
using just rotation and translation}
#' \item{Best-rmsd}{best fit rmsd between the fixed and mobile.}
#' \item{U}{Elements of the rotation matrix U}
#' @export
twist_offset = function(fixed, mobile, fixed.inds = NULL,
mobile.inds = NULL, getU = F, verbose = T){
  ## inds are from atom.select(...)$xyz

  if(!is.xyz(fixed)){fixed = as.xyz(fixed)}
  if(!is.xyz(mobile)){mobile = as.xyz(mobile)}
  if(nrow(fixed) != nrow(mobile)) stop("mismatched number of
frames")

  if(is.null(fixed.inds)){ fixed.inds = 1:ncol(fixed)}
  if(is.null(mobile.inds)){ mobile.inds = 1:ncol(mobile)}

  nf = nrow(fixed)
  if(nf < 2 & verbose) {
    verbose = F
  } # Do this because for txtProgressBar, max > min


  if(verbose){
    cat(sprintf("%d Frames Total\n", nf))
    #cat("Frame", "Twist", "Offset", "Fit RMSD", "\n", sep =
"\t")
    pb = txtProgressBar(1, nf, style = 3)
  }
  if(getU){
    out = matrix(NA, nrow = nf, ncol = 20)
    colnames(out) = c('theta', 'offset','Cx', 'Cy','Cz', 'Vx',
'Vy', 'Vz', 'Pre-rmsd', 'Fit-rmsd', 'Best-rmsd', sprintf("U%d",
1:9))
  } else {
    out = matrix(NA, nrow = nf, ncol = 11)
    colnames(out) = c('theta', 'offset','Cx', 'Cy','Cz', 'Vx',
'Vy', 'Vz', 'Pre-rmsd', 'Fit-rmsd', 'Best-rmsd')
  }
```

```
  for(f in 1:nf){
    ## See wikipedia for Kabsch Algorithm and Bio3Ds rot.lsq

    ## Moves the center of xyzFM to the origin
    xyzM = mobile[f,]
    xyzF = fixed[f,]
    p.rmsd = rmsd(xyzF, xyzM, fixed.inds, mobile.inds, fit = F)
# Pre-fit rmsd
    b.rmsd = rmsd(xyzF, xyzM, fixed.inds, mobile.inds, fit = T)
# Best possible rmsd
    #cat("Pre-RMSD:", p.rmsd, '\n', sep = '\t')
    #cat("Best RMSD:", b.rmsd, '\n', sep = '\t')
    xyzFM.com = com(as.xyz(c(xyzF,xyzM))) # Dimer COM
    xyzM = moveby(xyzM, xyzFM.com)
    xyzF = moveby(xyzF, xyzFM.com)

    ## Get Fit
    MF.fit = fit(xyzM[mobile.inds], xyzF[fixed.inds])
    xyzM.fit = moveby(move(xyzM, MF.fit$U), MF.fit$t, "+") #
Apply Rotation then translation

    ## Get final RMSD
    f.rmsd = rmsd(xyzF, xyzM.fit, fixed.inds, mobile.inds, fit =
F) # Post-fit rmsd
    #cat("Fit RMSD:", f.rmsd, '\n', sep = '\t')
    if(verbose){
      setTxtProgressBar(pb, f)
      #cat(f,MF.fit$twist, MF.fit$offset, f.rmsd, "\n", sep =
"\t")
    }
    if(getU){
      out[f,] = c(MF.fit$twist, MF.fit$offset, xyzFM.com,
MF.fit$axis, p.rmsd, f.rmsd, b.rmsd, as.vector(MF.fit$U))
    } else {
      out[f,] = c(MF.fit$twist, MF.fit$offset, xyzFM.com,
MF.fit$axis, p.rmsd, f.rmsd, b.rmsd)

    }
  } # end of for loop
  if(verbose) cat('\n')
  return(as.data.frame(out))
}
```

A.1.7.8 makeU

```
#' makeU
#'
```

```
#' This function is designed to give you the rotation about an
axis and translation upon said axis
#' @param axis a unit vector <u,v,w>
#' @param theta rotation angle in radians
#' @param offset offset distance (distance to translate upon
axis)
#' @param trans a logical indicating weather or not to return a
4x4 translation matrix (if F will return 3x3 rotation matrix)
#' @param vmd a logical indicating if the vmd matrix function
should be printed
#' @export
makeU = function(axis, theta, offset = 0, trans = T, vmd = T){
  u = axis[1]
  v = axis[2]
  w = axis[3]

  U=matrix(
    c(c(u^2 + (v^2 + w^2)*cos(theta), u*v*(1-cos(theta)) -
w*sin(theta), u*w*(1-cos(theta))+v*sin(theta), u*offset),
      c(u*v*(1-cos(theta))+w*sin(theta),v*v + (u^2 +
w^2)*cos(theta) , u*v*(1-cos(theta))-u*sin(theta), v*offset),
      c(u*w*(1-cos(theta))-v*sin(theta), v*w*(1-
cos(theta))+u*sin(theta), w^2+(u^2+v^2)*cos(theta), w*offset),
      c(0,0,0,1)), nrow = 4, ncol = 4, byrow = T)
  if(trans){
    if(vmd){
      cat(sprintf("{{%s}}", paste(apply(U,1,paste, collapse = "
"), collapse = "} {")))
      cat("\n")
    }
    return(U)

  } else {
    if(vmd){
      cat(sprintf("{{%s}}", paste(apply(U[1:3,1:3],1,paste,
collapse = " "), collapse = "} {")))
    }
    return(U[1:3,1:3])
  }
}
```

A.1.7.9 u_tcl

```
#' u_tcl
#'
#' This function is designed to take in a 3x3 matrix (or a
vector of 9 elements) and output a matrix for VMD.
```

```
#' More specifically, as written it will take in the 3x3 U
matrix and convert it to a 4x4 matrix.
#' @param U A 3x3 numerix matrix or a vetor of 9 elements
#' @param dim4 Default TRUE. As of yet, setting this to FALSE
will not allow the function to finish.
#' @export
u_tcl = function(U, dim4 = T){
  ## Function to print the U matrix into a form that can be
copied and pasted into VMD
  ## If dim4 == T, it will make the 3x3 U matrix into a 4x4
matrix

  if(dim4) {
    #dummy = matrix(0, nrow = 4, ncol = 4)
    #dummy[4,4] = 1

    if(all(dim(U) == c(3,3))){
      #dummy[1:3,1:3] = U
      #U = dummy
      U = rbind(cbind(U,c(0,0,0)),c(0,0,0,1))
    } else {
      #dummy[1:3,1:3] = matrix(U, nrow = 3, ncol = 3, byrow = T)
      #U = dummy
      U = matrix(U, nrow = 3, ncol = 3, byrow = T)
      U = rbind(cbind(U,c(0,0,0)),c(0,0,0,1))
    }

    ## Now that we have a 4x4 matrix we need to print it
    #temp = "{%s} {%s} {%s} {%s}"

    #tcl = sprintf(temp, apply(U,1,paste,collapse = " "))
    tcl = sprintf("{{%s}}", paste(apply(U,1,paste,collapse = "
"),collapse = "} {"))
    cat(tcl)
    #cat("\n")
    #return(tcl)
  }
}
```

A.2 Example R Analysis Scripts

A.2.1 Analysis of aMD G-Actin Trajectory – Secondary Structure Analysis and Building PCA Basis Set

```
library(ljepsenThesis)
library(bio3d)
```

```
library(MASS)

#####################
### Read in Files ###
#####################
pdb.gAct = read.pdb("~/Documents/G_Actin/amd/result.pdb") # PDB
File for ADP G-Actin
gAct.amd =
read.dcd("~/Documents/G_Actin/amd/run1t10_10t110_fit.dcd") #
Trajectory file aMD simulation of ADP G-actin

###################################
### Secondary Structure Analysis ###
###################################

gAct.ss = get_sse(dcd = gAct.amd, pdb = pdb.gAct)

# Get Protein Core
ss.gAct = which(gAct.ss[['sse.freq2']] [3,] < 0.20) # Residues
that are part of helical or beta structures for at least 80% of
the frames of the simulation
ss.gAct.xyz = atom.select(pdb.gAct, resno = ss.gAct + 1, elety =
"CA")$xyz # Note: +1 to get residue numbers. In PDB the first
residue is the N-terminal Acetylation.

##############################
### Building PCA Basis Set ###
##############################

# Fit Trajectory
gAct.fit.amd = fit.xyz(fixed = gAct.amd[1,], mobile = gAct.amd,
fixed.inds = ss.gAct.xyz, mobile.inds = ss.gAct.xyz) # Fitting
to first frame of trajectory using the core of the protein
rm(gAct.amd)

# Do PCA
gAct.pca = pca.xyz(gAct.fit.amd[,ss.gAct.xyz])

# Get Projection
gAct.pca.proj = project.pca(gAct.fit.amd[,ss.gAct.xyz],
gAct.pca) # Note: This is the same as gAct.pca$z
gAct.pca.proj.density = kde2d(gAct.pca.proj[,1],
gAct.pca.proj[,2])

#################
### Save Data ###
#################
```

```
# Save Secondary Structure
save(gAct.ss, file = "gAct_ss.RData")

# Save Fitted Trajectory
save(gAct.fit.amd, file = "gAct_fit_amd.RData")

# Save PCA Data
save(pdb.gAct, ss.gAct, ss.gAct.xyz, gAct.pca, gAct.pca.proj,
gAct.pca.proj.density, file = "gAct_pcaBasis.RData")
```

A.2.2 Projection of cMD ADP and ATP G-actin Simulation onto PCA Basis Set

```
library(ljepsenThesis)
library(bio3d)
library(MASS)

###################################
### Read in Files and Load Data ###
###################################

# Load PCA Basis
load("gAct_pcaBasis.RData")

# Load cMD ADP G-actin
pdb.gAct_cmd =
read.pdb("~/Documents/G_Actin/cmd/data/result_mg_model.pdb")
gAct_cmd.cmd =
read.dcd("~/Documents/G_Actin/cmd/data/gAct_cmd_allSims.dcd")

# Load cMD ATP G-actin
pdb.gAct_atp_cmd =
read.pdb("~/Documents/ATP_Actin/G_Actin/cmd/data/result_mg_model
.pdb")
gAct_atp_cmd.cmd =
read.dcd("~/Documents/ATP_Actin/G_Actin/cmd/data/gAct_atp_cmd_al
lSims.dcd")

########################
### Fit Trajectories ###
########################

# Select 'Core'
ss.gAct_cmd.xyz = atom.select(pdb.gAct_cmd, resno = ss.gAct + 1,
elety = "CA")$xyz
ss.gAct_atp_cmd.xyz = atom.select(pdb.gAct_atp_cmd, resno =
ss.gAct + 1, elety = "CA")$xyz

# Fit to PCA Mean
```

```
gAct_cmd.fit.cmd = fit.xyz(fixed = gAct.pca$mean, mobile =
gAct_cmd.cmd, mobile.inds = ss.gAct_cmd.xyz)
gAct_atp_cmd.fit.cmd = fit.xyz(fixed = gAct.pca$mean, mobile =
gAct_atp_cmd.cmd, mobile.inds = ss.gAct_atp_cmd.xyz)
rm(gAct_cmd.cmd, gAct_atp_cmd.cmd)

######################
### PCA Projection ###
######################

# Project
gAct_cmd.pca.proj =
project.pca(gAct_cmd.fit.cmd[,ss.gAct_cmd.xyz], gAct.pca)
gAct_atp_cmd.pca.proj =
project.pca(gAct_atp_cmd.fit.cmd[,ss.gAct_atp_cmd.xyz],
gAct.pca)

# Density
gAct_cmd.pca.proj.density = kde2d(gAct_cmd.pca.proj[,1],
gAct_cmd.pca.proj[,2])
gAct_atp_cmd.pca.proj.density = kde2d(gAct_atp_cmd.pca.proj[,1],
gAct_atp_cmd.pca.proj[,2])

#################
### Save Data ###
#################

# Fit Trajectories
save(pdb.gAct_cmd, gAct_cmd.fit.cmd, file =
"gAct_cMD_fit.RData")
save(pdb.gAct_atp_cmd, gAct_atp_cmd.fit.cmd, file =
"gAct_atp_cMD_fit.RData")

# PCA Projections
save(pdb.gAct_cmd, ss.gAct_cmd.xyz, gAct_cmd.pca.proj,
gAct_cmd.pca.proj.density, file =
"gAct_cMD_gActBasis_pcaProject.RData")
save(pdb.gAct_atp_cmd, ss.gAct_atp_cmd.xyz,
gAct_atp_cmd.pca.proj, gAct_atp_cmd.pca.proj.density, file =
"gAct_atp_cMD_gActBasis_pcaProject.RData")
```

A.2.3 PCA on Loops – ADP vs ATP G-actin (cMD)

```
library(bio3d)
library(MASS)

#################
### Load Data ###
```

```
#################
load("gAct_cMD_fit.RData")
load("gAct_atp_cMD_fit.RData")

#############
### Loops ###
#############

loops = list("N-term" = c(1:5),
             "S-Loop" = c(11:16),
             "D-Loop" = c(39:52),
             "H73-Loop" = c(70:78),
             "LinkerHelix" = c(137:145),
             "G-Loop" = c(155:160),
             "W-Loop" = c(165:172),
             "Sub4Hinge" = c(220:235),
             "Flap" = c(236:251),
             "HydrophobicLoop" = c(265:273),
             "CT-hinge" = c(331:337),
             "C-term" = c(349:375),
             "S_G-Loop" = c(11:16,155:160),
             "S_G_H3-Loops"= c(11:16,70:78,155:160),
             "22-28" = c(22:28),
             "60-65" = c(60:65),
             "92-102"=c(92:102),
             "107-112" = c(107:112),
             "125-130" = c(125:130),
             "145-149" = c(145:149),
             "196-202" = c(196:202),
             "216-222" = c(216:222),
             "284-289" = c(284:289),
             "321-328" = c(321:328),
             "348-356" = c(348:356),
             "h301-306" = c(301:306),
             "W-loop_348-356" = c(165:172, 348:356),
             "W-loop_348-356_370-375" = c(165:172, 348:356,
370:375))
## extra unnamed loops added using elements not included in
first PCA. Loops had to be more than 4 residues long. Loops
meeting this criteria but that were either an extension or a
subset of the named loops were excluded.

##############
### Do PCA ###
##############
pca.loops = list()
proj_atp = list()
proj_adp = list()
```

```
pdb.loops = list()

for(i in names(loops)){
  # Break out XYZ
  as.adp = atom.select(pdb.gAct_cmd, resno = loops[[i]]+1, elety
= "CA")
  as.atp = atom.select(pdb.gAct_atp_cmd, resno = loops[[i]]+1,
elety = "CA")
  pca.loops[[i]] =
pca.xyz(rbind(gAct_cmd.fit.cmd[,as.adp$xyz],gAct_atp_cmd.fit.cmd
[,as.atp$xyz]))
  proj_adp[[i]] = project.pca(gAct_cmd.fit.cmd[,as.adp$xyz],
pca.loops[[i]])
  proj_atp[[i]] = project.pca(gAct_atp_cmd.fit.cmd[,as.atp$xyz],
pca.loops[[i]])

  pdb.loops[[i]] = trim.pdb(pdb.gAct_cmd, inds = as.adp)

  #quartz(antialias = F, title = i, bg = 'white')
  #pca_trace(pca.loops[[i]], proj.add = list("ADP" =
proj_adp[[i]], "ATP" = proj_atp[[i]]),
  #          proj.cols = c("cyan4", "darkmagenta"), main = i,
xlim = c(-10,10), ylim = c(-10,10))
  #quartz.save(sprintf("loopsPCA_%s.png",i), antialias = F, type
= 'png')
  #dev.off()
}

save(pca.loops, proj_atp, proj_adp, pdb.loops, loops,
pdb.gAct_cmd, pdb.gAct_atp_cmd, file =
"gAct_atp_adp_cmd_loops_pca.RData")

##############
### PC Trj ###
##############
for(i in names(loops)){
  mktrj.pca(pca.loops[[i]], pc = 1, pdb = pdb.loops[[i]], file =
sprintf("loops_pc1_%s.pdb", i), rock = F)
  mktrj.pca(pca.loops[[i]], pc = 2, pdb = pdb.loops[[i]], file =
sprintf("loops_pc2_%s.pdb", i), rock = F)
}
```

A.2.4 Projection of PDB Structures onto G-Actin PCA Basis

```
library(bio3d)
library(reshape2)
library(MASS)
library(ljepsenThesis)
```

```
###############
## Load Data ##
###############
load("gAct_pcaBasis.RData")

# Use Bio3D Wevserver to search against PDB 1J6Z, saved Results
bio3d.1j6z = read.csv("Bio3D-web.csv") # Used bio3d webserver to
do search, saved results
rownames(bio3d.1j6z) = bio3d.1j6z$ID

# Fix ID Column (remove '● ')
bio3d.1j6z$ID = colsplit(as.character(bio3d.1j6z$ID), '.{2}',
names = c("JUNK","ID"))$ID

# Set BitScore cutoff
bitCutoff = 665
hits = colsplit(bio3d.1j6z$ID[bio3d.1j6z$BitScore>bitCutoff],
'_', names = c("PDBID", "Chain"))
idList = list()
for(id in sort(unique(hits$PDBID))){
  idList[[id]] = hits[which(hits$PDBID == id),2]
}

#################
## Process IDs ##
#################

cryst.info = list()
proj_tbl = c()
working_ids = c()
for(id in sort(names(idList))){
  cat("###", id, "###", '\n')
  cryst.info[[id]] = crystProject(id, pdb.gAct, ss.gAct.xyz,
gAct.pca, chain = idList[[id]], split = T)
  if(!is.null(cryst.info[[id]])){
    y = cryst.info[[id]]$proj
    if(!is.null(y)){
      x = apply(cbind(rep(id, length(rownames(y))),
rownames(y)), 1, paste, collapse = "_")
      rownames(y) = x
      working_ids = append(working_ids, x)
      proj_tbl = rbind(proj_tbl, y)

      rm(x)
    }
    rm(y)
  }
```

```
  cat('\n')
}

### Get Named Projection Table ###

# Function:
extract_proj = function(i, crystInfo = cryst.info){
  proj_mat = c()
  for(j in i){
    p = crystInfo[[j]]$proj
    if(!is.null(p)){
      if(nrow(p) == 1){
        if(is.logical(crystInfo[[j]]$work)){
          n = sprintf("%s_%s",j,
unique(crystInfo[[j]]$pdb$atom$chain))
          rownames(p) = n
        } else {
          n = sprintf("%s_%s",j,rownames(p))
          rownames(p) = n
        }
      } else {
        n = sprintf("%s_%s",j,rownames(p))
        rownames(p) = n
      } # End else
      proj_mat = rbind(proj_mat,p)
    } # End !is.null(p)
  } # End For loop
  return(proj_mat)
}

# Get Table:
proj_tbl = extract_proj(names(cryst.info))

working_ids = rownames(proj_tbl)

### Get Working ID Table ###
idList.working = list()
x = colsplit(working_ids, '_', names = c("PDBID", "Chain"))
for(id in sort(unique(x$PDBID))){
  idList.working[[id]] = x[which(x$PDBID == id),2]
}
rm(id, x)


### Get Sequence Identity List ###
cryst.seqidentity = vector('numeric', length(working_ids))
names(cryst.seqidentity) = working_ids
for(id in names(idList.working)){
```

```
  x = cryst.info[[id]]$seqident[idList.working[[id]]]
  y = apply(cbind(rep(id, length(x)), names(x)), 1, paste,
collapse = "_")
  cryst.seqidentity[y] = x
}
rm(x,y,id)

cp.ident = colorRampPalette(c("blue", "red"))

############
### Save ###
############
bio3d.1j6z.working = bio3d.1j6z[working_ids,]
write.csv(bio3d.1j6z.working, file = "bio3d_pdb_info.csv",
row.names = T)
save(cryst.info,cryst.seqidentity,working_ids,idList.working,bio
3d.1j6z,pdb.gAct,bio3d.1j6z.working,proj_tbl, file =
"crystal_projection.RData")
```

A.2.5 Measurement of PDB Files Propeller-Twist, Phosphate-Clamp and Nucleotide Cleft Mouth Size

```
library(bio3d)
library(ljepsenThesis)

#################
### Load Data ###
#################

load("crystal_projection.RData")

################################
### Calculate Propeller-Twist ###
################################

df = c()
for(i in names(idList.working)){
  cat(i,'\n')
  temp = crystPropellerTwist(cryst.info[[i]],pdb.gAct)
  rownames(temp)=sprintf("%s_%s",i,rownames(temp))

  df = rbind(df,temp)
}
df

cryst.ptwist.df = df
save(cryst.ptwist.df, file = "cryst_ptwist_df.RData")
```

A.2.6 Filament Twist and Offset

```
library(bio3d)
library(ljepsenThesis)

################################
### Measure Twist and Offset ###
################################

# Note: Each chain within the filament was saved individually
using VMD. This significantly speeds up calculation time by
limiting the amount of data that is loaded in R at one time.

fActO_cmd.twist_offset = list()
a.xyz = read.dcd("chainA_0t400.dcd")
b.xyz = read.dcd("chainB_0t400.dcd")
fActO_cmd.twist_offset[["AB"]] = twist_offset(a.xyz,b.xyz)
rm(a.xyz)
c.xyz = read.dcd("chainC_0t400.dcd")
fActO_cmd.twist_offset[["BC"]] = twist_offset(b.xyz,c.xyz)
rm(b.xyz)
d.xyz = read.dcd("chainD_0t400.dcd")
fActO_cmd.twist_offset[["CD"]] = twist_offset(c.xyz,d.xyz)
rm(c.xyz)
e.xyz = read.dcd("chainE_0t400.dcd")
fActO_cmd.twist_offset[["DE"]] = twist_offset(d.xyz,e.xyz)
rm(d.xyz)
f.xyz = read.dcd("chainF_0t400.dcd")
fActO_cmd.twist_offset[["EF"]] = twist_offset(e.xyz,f.xyz)
rm(e.xyz)
g.xyz = read.dcd("chainG_0t400.dcd")
fActO_cmd.twist_offset[["FG"]] = twist_offset(f.xyz,g.xyz)
rm(f.xyz)
h.xyz = read.dcd("chainH_0t400.dcd")
fActO_cmd.twist_offset[["GH"]] = twist_offset(g.xyz,h.xyz)
rm(g.xyz,h.xyz)

############################
### Get Twist Data Frame ###
############################

fActO_cmd.twist = data.frame("AB" =
fActO_cmd.twist_offset[["AB"]]$theta,
                             "BC" =
fActO_cmd.twist_offset[["BC"]]$theta,
```

```
                                "CD" =
fActO_cmd.twist_offset[["CD"]]$theta,
                                "DE" =
fActO_cmd.twist_offset[["DE"]]$theta,
                                "EF" =
fActO_cmd.twist_offset[["EF"]]$theta,
                                "FG" =
fActO_cmd.twist_offset[["FG"]]$theta,
                                "GH" =
fActO_cmd.twist_offset[["GH"]]$theta)
fActO_cmd.twist$meanP =
apply(fActO_cmd.twist[,c("AB","BC")],1,mean) # Average Pointed
End by Frame
fActO_cmd.twist$meanM =
apply(fActO_cmd.twist[,c("CD","DE","EF")],1,mean) # Average
Middle by Frame
fActO_cmd.twist$meanB =
apply(fActO_cmd.twist[,c("FG","GH")],1,mean) # Average Barbed
End by Frame
fActO_cmd.twist$mean = apply(fActO_cmd.twist,1,mean) # Average
Filament by Frame

##############################
### Get Offset Data Frame ###
##############################

fActO_cmd.offset = data.frame("AB" =
fActO_cmd.twist_offset[["AB"]]$offset,
                                "BC" =
fActO_cmd.twist_offset[["BC"]]$offset,
                                "CD" =
fActO_cmd.twist_offset[["CD"]]$offset,
                                "DE" =
fActO_cmd.twist_offset[["DE"]]$offset,
                                "EF" =
fActO_cmd.twist_offset[["EF"]]$offset,
                                "FG" =
fActO_cmd.twist_offset[["FG"]]$offset,
                                "GH" =
fActO_cmd.twist_offset[["GH"]]$offset)
fActO_cmd.offset$meanP =
apply(fActO_cmd.offset[,c("AB","BC")],1,mean) # Average Pointed
End by Frame
fActO_cmd.offset$meanM =
apply(fActO_cmd.offset[,c("CD","DE","EF")],1,mean) # Average
Middle by Frame
```

```
fActO_cmd.offset$meanB =
apply(fActO_cmd.offset[,c("FG","GH")],1,mean) # Average Barbed
End by Frame
fActO_cmd.offset$mean = apply(fActO_cmd.offset,1,mean) # Average
Filament by Frame

#################
### Save Data ###
#################
rm(list = ls()[grep("xyz", ls())])
save(fActO_cmd.twist_offset, fActO_cmd.twist, fActO_cmd.offset,
file = "fActO_cmd_prod0t400_twistOffset.RData")
```

A.3 VMD Scripts

The following are TCL scripts to be run in VMD for analysis.

A.3.1 Vector Math Functions

```
package require math::constants

proc ::getDistance { point1 point2 } {
     set vec1 [vecsub $point2 $point1]
     return [veclength $vec1]
}
# Returns the distance between point1 and point2

proc ::getDistance_sel {sel1 sel2} {
     set cent1 [measure center $sel1]
     set cent2 [measure center $sel2]
     #set vec [vecsub $cent1 $cent2]
     set vlen [getDistance $cent1 $cent2]
     #set vlen [veclength $vec]

     return $vlen
}
# dist measures the distance between the centers of mass of the
two selections

proc ::getDihedral { point1 point2 point3 point4 {degree False}
{atan2 True}} {
     #http://math.stackexchange.com/questions/47059/how-do-i-
calculate-a-dihedral-angle-given-cartesian-coordinates
     ::math::constants::constants pi
```

```
     if ($atan2) {
          set b1 [vecsub $point2 $point1]
          set b2 [vecsub $point3 $point2]
          set b3 [vecsub $point4 $point3]

          set n1 [vecnorm [veccross $b1 $b2]]
          set n2 [vecnorm [veccross $b2 $b3]]

          #set m1 [veccross $n1 [vecnorm $b2]]
          set m1 [veccross $n2 [vecnorm $b2]]
          set x  [vecdot $n1 $n2]
          #set y  [vecdot $m1 $n2]
          set y  [vecdot $m1 $n1]

          if ($degree) {
               return [expr [expr 180/$pi]*atan2($y, $x)]
          } else {
               return [expr atan2($y, $x)]
          }

     } else {
          set vec1 [vecsub $point2 $point1]
          set vec2 [vecsub $point3 $point2]
          set vec3 [vecsub $point4 $point3]

          set norm1 [vecnorm [veccross $vec1 $vec2]]
          set norm2 [vecnorm [veccross $vec2 $vec3]]
          if ($degree) {
               return [expr [expr 180/$pi]*acos([vecdot $norm1
$norm2])]
          } else {
               return [expr acos([vecdot $norm1 $norm2])]
          }
     }

}
# Returns the point1-point2-point3-point4 dihedral angle

proc ::getAngle { point1 point2 point3 } {
     set pi [expr 4.0*atan(1.0)]
     set rd [ expr 180.0/$pi] # radians to degrees

     set vec21 [vecsub $point1 $point2]
     set vec23 [vecsub $point3 $point2]

     set norm21 [vecnorm $vec21]
     set norm23 [vecnorm $vec23]
```

```
     return [expr $rd*acos([vecdot $norm21 $norm23])]
}
# Returns the angle between point1-point2-point3

proc ::normvec {pt1 pt2 pt3} {
     set pt1pt2 [vecsub $pt2 $pt1]
     set pt1pt3 [vecsub $pt3 $pt1]

     set n [vecnorm [veccross $pt1pt2 $pt1pt3]]

     return $n
}
# normal vector to a plane defined by 3 points

proc ::vecangle { vec1 vec2 {degree True} } {
     ### This function returns the angle between two vectors

     # Get Unit Vectors
     set nvec1 [vecnorm $vec1]
     set nvec2 [vecnorm $vec2]

     if $degree {
          set ang [expr acos([vecdot $nvec1 $nvec2]) * [expr
{180.0/(4.0*atan(1.0))}]]
          # Note: [expr {180.0/(4.0*atan(1.0))}] = 180/Pi -->
conversion from radians to degrees
     } else {
          set ang [expr acos([vecdot $nvec1 $nvec2])]
     }

     return $ang
}
# angle between two vectors in degrees, or radians if degree ==
False

proc ::vecangle_atan2 { vec1 vec2 {degree True} } {
     ### This function returns the angle between two vectors

     if $degree {
          set ang [expr atan2([veclength [veccross $vec1
$vec2]],[vecdot $vec1 $vec2]) * [expr {180.0/(4.0*atan(1.0))}]]
          # Note: [expr {180.0/(4.0*atan(1.0))}] = 180/Pi -->
conversion from radians to degrees
     } else {
          set ang [expr atan2([veclength [veccross $vec1
$vec2]],[vecdot $vec1 $vec2])]
     }
```

```
     return $ang
}
# angle between two vectors in degrees, or radians if degree ==
False


proc ::midpt {pt1 pt2} {
     set mid [vecscale [vecadd $pt1 $pt2] 0.5]
     return $mid
}

proc ::getVec { pt1 pt2 {scale 1.0} {unit False} } {
     set vec [vecsub $pt2 $pt1]

     set vec2 [vecscale $vec $scale]

     if $unit {
          return [vecnorm $vec]
     } else {
          return $vec2
     }
}
# returns vector between to points scaled by scale, or if unit
== True, a unit vector
# in the direction of the vector between points 1 and 2

proc ::projvP {lcom lvec point} {

     set lambda [expr [vecdot $lvec [vecsub $point $lcom]] /
[veclength2 $lvec]]
     set b [vecadd $lcom [vecscale $lvec $lambda]]

     return $b
}
## Projects point onto vector defined by lvec that passes
through lcom
## Note b lies on lvec
```

A.3.1 Propeller-Twist, Phosphate Clamp, and Nucleotide Cleft Mouth Size

```
source ~/Desktop/tcl_Scripts/vector_math_functions.tcl

proc ::propeller_twist { {molid top} {fName
"dihedral_clamp_cleft.txt"} {chain X} {degree True} {atan2 True}
{crystal False}} {
     ### If you don't want to save the results, set fName to
False
```

```
     # Get Number of Frames
     set nf [molinfo $molid get numframes]

     # Open Save File
     if {[string is boolean $fName] && ! $fName} {
          set save False
          puts "Frame \t Dihedral_2134 \t Clamp_Size \t
Mouth_Size"
     } else {
          set save True
          if {[string is boolean $fName]} {set fName
"dihedral_clamp_cleft.txt"}
          set file [open $fName w]
          puts $file "Frame \t Dihedral_2134 \t Clamp_Size \t
Mouth_Size"
     }

     ##########################
     ### Get Measurements  ###
     ##########################
     if {$crystal} {
          set com1 [atomselect $molid "resid 8 to 11 16 to 21 29
to 31 79 to 91 103 to 106 113 to 124 131 to 135 137 to 144 338
to 347 357 to 364 and chain $chain" ]
          set cg2  [atomselect $molid "resid 32 35 to 37 53 to
54 56 to 59 66 to 68 and chain $chain" ]
          set com3 [atomselect $molid "resid 150 to 154 161 to
166 169 to 170 176 to 177 274 to 283 290 to 294 297 to 300 303
to 304 309 to 320 329 to 330 and chain $chain" ]
          set cg4  [atomselect $molid "resid 182 to 195 203 to
215 223 to 230 238 to 241 247 to 250 253 to 256 258 to 259 and
chain $chain" ]

          set d157 [atomselect $molid "resid 157 and name CA and
chain $chain and resname ASP" ]
          set g15  [atomselect $molid "resid 15 and name CA and
chain $chain and resname GLY"  ]

          set e207 [atomselect $molid "resid 207 and name CA and
chain $chain and resname GLU" ]
          set q59  [atomselect $molid "resid 59 and name CA and
chain $chain and resname GLN"  ]
     } else {
          set com1 [atomselect $molid "resid 9 to 12 17 to 22 30
to 32 80 to 92 104 to 107 114 to 125 132 to 136 138 to 145 339
to 348 358 to 365 and chain $chain" ]
          set cg2  [atomselect $molid "resid 33 36 to 38 54 to
55 57 to 60 67 to 69 and chain $chain" ]
```

```
          set com3 [atomselect $molid "resid 151 to 155 162 to
167 170 to 171 177 to 178 275 to 284 291 to 295 298 to 301 304
to 305 310 to 321 330 to 331 and chain $chain" ]
          set cg4  [atomselect $molid "resid 183 to 196 204 to
216 224 to 231 239 to 242 248 to 251 254 to 257 259 to 260 and
chain $chain" ]

          set d157 [atomselect $molid "resid 158 and name CA and
chain $chain and resname ASP" ]
          set g15  [atomselect $molid "resid 16 and name CA and
chain $chain and resname GLY"  ]

          set e207 [atomselect $molid "resid 208 and name CA and
chain $chain and resname GLU" ]
          set q59  [atomselect $molid "resid 60 and name CA and
chain $chain and resname GLN"  ]
     }
     for {set x 0} {$x < $nf} {incr x} {
          ### Measuring 2-1-3-4 Dihedral ##
          $com1 frame $x
          $cg2  frame $x
          $com3 frame $x
          $cg4  frame $x

          set dihedral [getDihedral [measure center $cg2]
[measure center $com1] [measure center $com3] [measure center
$cg4] $degree $atan2]

          ### Measuring Clamp Size ###
          $d157 frame $x
          $g15  frame $x

          set clamp [vecdist [measure center $d157] [measure
center $g15]]

          ### Measuring Cleft Size ###
          $e207 frame $x
          $q59  frame $x

          set cleft [vecdist [measure center $q59] [measure
center $e207]]

          ### Write Out Results ###i
          if {$save} {
               puts $file "$x \t $dihedral \t $clamp \t $cleft"
          } else {
               puts "$x \t $dihedral \t $clamp \t $cleft"
          }
```

```
        }
        if {$save} {
             close $file
        }
}
```