In presenting this dissertation/thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish, this thesis/dissertation may be granted by the professor under whose direction it was written when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. In the absence of the professor, the dean of the Graduate School may grant permission. It is understood that any copying from, or publication of, this thesis/dissertation which involves potential financial gain will not be allowed without written permission.

_____

Dina Nicole Greene

Studies on the Activation Mechanism and Identification of a Substrate for Giant Kinases in *C. elegans* Muscle

By
Dina Greene
Doctor of Philosophy
Program in Biochemistry, Cell, and Developmental Biology
Graduate Division of Biological and Biomedical Sciences

_____
Guy Benian
Adviser

_____
Keith Wilkinson
Committee Member

_____
Anita Corbett
Committee Member

_____
David Lynn
Committee Member

_____
David Pallas
Committee Member

_____
David Dunlap
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

_____
Date

Studies on the Activation Mechanism and Identification of a Substrate for Giant Kinases in *C. elegans* Muscle

By

Dina Greene
B.S., University of Florida, 2003

Adviser: Guy M. Benian, M.D.

An abstract submitted to the Faculty of the Graduate
School of Emory University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Program in Biochemistry, Cell, and Developmental Biology
Graduate Division of Biological and Biomedical Sciences

2008

ABSTRACT

The muscles of virtually all animals contain giant (>700,000 Da) polypeptides that consist primarily of multiple copies of immunoglobulin (Ig) and fibronectin type III domains, and one or even two protein kinase domains. In various muscles these proteins have several roles. For example, directing assembly of the sarcomere and providing passive elasticity (titin of vertebrates), maintenance of the "catch" state (twitchin in mollusks), and ability of insect flight muscle to beat at high frequencies (projectin of insects). We are studying two such proteins in *C. elegans*, twitchin and TTN-1.

Human titin kinase has been implicated as an initiating catalyst in a signaling pathway that ultimately results in muscle cell growth. The enzyme is negatively regulated by intramolecular interactions occurring between the kinase catalytic core and a downstream autoinhibitory region. The precise mechanism(s) resulting in the conformational changes that relieve the kinase of this autoinhibition are unknown. Force-probe molecular dynamics simulations suggest that human titin kinase may act as a force sensor. This study predicts that the small forces that are generated with each contraction/relaxation cycle are sufficient to remove the autoinhibitory region thereby activating the enzyme. We experimentally tested this force activation hypothesis using atomic force microscopy to analyze the kinase and flanking domains of *C. elegans* TTN-1 (a titin-like protein) and twitchin. Our results show that these kinase domains have a remarkably high mechanical stability. Further, in response to applied force, these kinase domains unfold in a stepwise manner, first an unwinding of the autoinhibitory region, followed by a two-step unfolding of the catalytic core. These data directly support the hypothesis that the titin and titin-like kinase domains function as effective force sensors.

In an ongoing effort to identify binding partners and substrates for the protein kinase domains of these giants, we have discovered an excellent candidate for the TTN-1 protein kinase. The interacting partner is UIG-1, previously defined as an UNC-112 binding partner with Cdc42 GEF activity located in the dense bodies/I-bands of *C. elegans* striated muscle. An intragenic deletion of *uig-1* displays disorganized myofibrils. Using the yeast 2-hybrid method we have determined which portions of each protein are required for this interaction, and we have confirmed this interaction using an *in vitro* binding assay. By immunofluorescence microscopy, UIG-1 partially co-localizes with TTN-1. Further, TTN-1 kinase phosphorylates UIG-1 *in vitro* in regions outside the DH and PH domains. We speculate that phosphorylation of UIG-1 by TTN-1 regulates either (1) the interaction of UIG-1 with UNC-112, (2) UIG-1's localization to dense bodies, or (3) the GEF activity of UIG-1. Whichever is the case, these studies have identified the first substrate for a muscle giant kinase in *C. elegans* and revealed a function for one of these kinases in sarcomere assembly.

Studies on the Activation Mechanism and Identification of a Substrate for Giant Kinases in *C. elegans* Muscle

By

Dina Greene
B.S., University of Florida, 2003

Adviser: Guy M. Benian, M.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Program in Biochemistry, Cell, and Developmental Biology
Graduate Division of Biological and Biomedical Sciences

2008

ACKNOWLEDGEMENTS

Graduate school has been filled with triumph and failure. I knew when I entered that graduate school was going to be hard, but I did not fully understand the magnitude to which I would grow and develop, as both a scientist and an individual. In fact, I don't think any amount of warning could have braced me for the process, for my entire experience. Although, in the end, I feel like the road was traveled independently, fueled by self-determination, I never would have made it without the advice, companionship, and nourishment of others.

When I began at Emory, the faculty were kind of scary. Even an extrovert like me felt uncomfortable in conversations. I slowly silenced my hesitations and began to embrace their personalities. I realized that these "seniors" would not only push my intellect to its maximum, but they would also offer advise, teach me to think better, and allow me to embrace my independence. Guy Benian has been the epitome of this faculty. Guy believed in me when I didn't even believe in myself. He has taught me not only to be a better scientist, but also to be a better person. Although Guy has been monumental, many other faculty have simultaneously inspired and supported me through this process: Keith Wilkinson, always the best mediator with the best "final answers"; Anita Corbett, always willing to talk to me about anything, who gives me advice as both a friend and a mentor; Win Sale, for the best hugs and one liners; David Pallas for me to say "Hi Dr. Pallas" to; and the many others that exchanged smiles and words with me throughout the years.

I am often asked, "Why did you choose Emory"? I usually answer this question in some sort of roundabout way, but what I should really say is "Marie". Marie Cross has

been the best colleague I ever could have hoped for. She has sacrificed many tissues, and possible confrontations with her adviser, to make me smile and to make sure that I was OK. The combination of Marie with the rest of my classmates has been a support I never expected from an academic setting. I owe tons of thanks to Rob, Seth, Branch, Emma, Avanti, and Lori. And I can't forget my lovely ladies, Jocelyn Lee and Laura McLane, who have been two of the best pairs of ears to listen and arms to hug.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Introduction

Muscle structure: an overview of myofibrils

A fundamental characteristic of animals is muscle-driven motility. Thus, muscle is an abundant and fascinating tissue found in all animals that allows for voluntary movement, and also involuntary movement required for pumping of the heart, blood vessel tone, gut peristalsis, excretion, reproduction, and respiration. There are three main types of muscle: skeletal, cardiac, and smooth (Murphy 1976). Skeletal muscle is fundamental to the overall structural anatomy of an animal. It is strong, quick, and allows for voluntary movement. Cardiac muscle constitutes the main component of the heart and has evolved for strong, quick, involuntary contractions of the heart in order to pump blood, and hence oxygen, throughout the animal. Smooth muscle is found in the walls of all hollow, vital internal organs (except the heart). It is weak, slow, and involuntary; its action allows for the movement of nutrients and fluids throughout the animal. An additional useful categorization of muscle is to distinguish "striated" from "smooth" muscle, based on the appearance by light microscopy of regular alternating bands of dark and light material. Both skeletal and cardiac muscles have this appearance and thus are considered "striated". Smooth muscle lacks this appearance and thus, is not striated. The remainder of this thesis will use the general term "muscle" to denote striated muscle, as it is the focus of our laboratory. Regardless of the muscle type, its contraction, that involves the interaction of myosin and actin proteins, provides the physiological force that promotes the biological activity (Davies 1963).

To understand the architecture of skeletal muscle (such as the biceps muscle), let's begin with its appearance by eye (Figure 1.1). The individual muscle is surrounded by dense connective tissue, which is continuous with a tendon that provides the

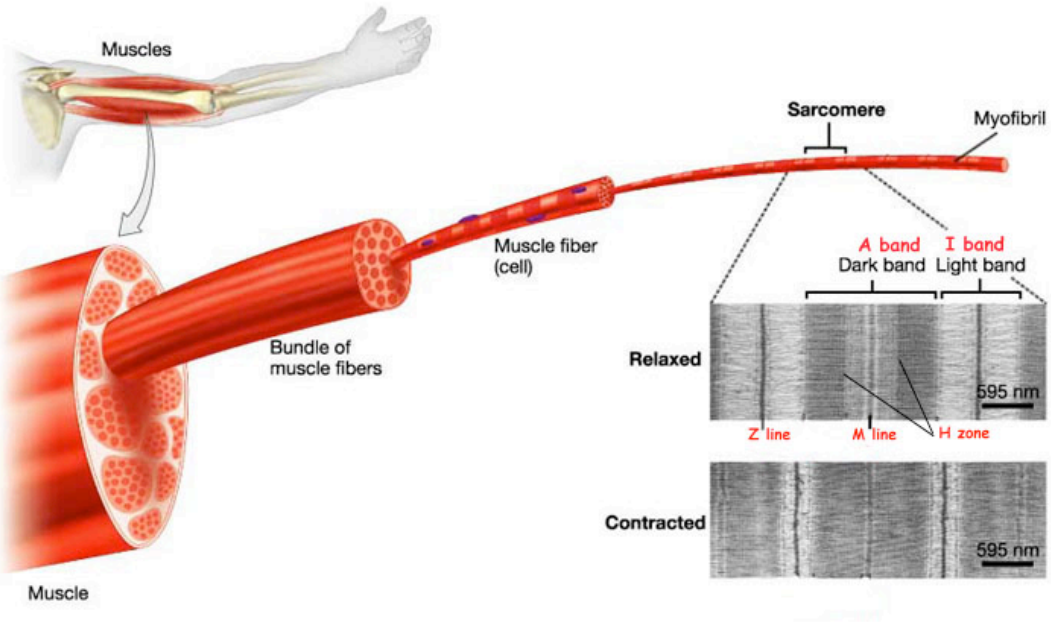


Figure 1.1- The hierarchy of muscle tissue. A single vertebrate muscle provides the forces required to generate a specific movement of the body. The muscle is composed of fasiculi, or bundles of muscle cells, which are called muscle fibers. Within the muscle fibers (the cells) are cylinders known as myofibrils. The myofibril further breaks down into the individual contractile unit, the sarcomere. Sarcomeres are anatomically defined by their iterative patterning, visualized here in electron micrographs. The Z-line marks the boundaries between individual sarcomeres, and also serves to anchor the actin thin filaments. The bright regions, or I-bands, change length during contraction are made up of thin filaments, excluding the region where the thin filaments and myosin thick filaments overlap. The dark, central portion is termed the A-band and is the location of the thick filaments; the outer A-band also contains the regions where thick and thin filaments overlap. The M-line is a structure in which the myosin thick filaments are cross-linked, and it is located in the center of the sarcomere. Figure courtesy of http://fig.cox.miami.edu/~cmallery/150/neuro/sf43x15a.jpg.

attachment to bone. This connective tissue further divides the muscle into bundles of fasicles, each containing multiple "muscle fibers", the somewhat confusing term for the muscle cells. These muscle cells are very long (up to 30 cm) with multiple peripherally localized nuclei resulting from the fusion of myoblasts during development. Each muscle cell, if cut in cross-section can be seen to be packed full of the contractile units of muscle, the "myofibrils", long cylinders that are oriented parallel to the long axis of the cell. In between the myofibrils reside the mitochondria and each myofibril is surrounded by a hollow tube containing the sarcoplasmic reticulum/T-tubule system (SR/T-tubules) (Ogata and Yamasaki 1990). The SR is a specialized endoplasmic reticulum and is the storage and release depot for calcium (Costantin 1975; Costantin 1975). Calcium is released from the SR in response to action potentials initiated at the cell surface and conveyed deep into the cell by the T-tubules. The SR/T-tubule system permits the nearly synchronized contraction of all the myofibrils in the cell. The myofibrils are composed of a long series of smaller cylinders of uniform size stacked end to end; these are the "sarcomeres", and these are the fundamental repeating units of the myofibril, and of muscle contraction, and in human muscle are typically 2.5 μm long (Spiro 1956) (Figure 1.2).

The sarcomere contains a set of three overlapping filaments: myosin containing thick filaments, actin containing thin filaments, and titin filaments (Figure 1.2, reviewed in (Lange, Ehler et al. 2006)). The sarcomere is actually composed of several hundred proteins; new protein components are being discovered each year; and each protein can have multiple, even dynamic localizations. The uniform makeup of the sarcomere allows for distinct anatomical definitions within the muscle cell. The striations that are easily

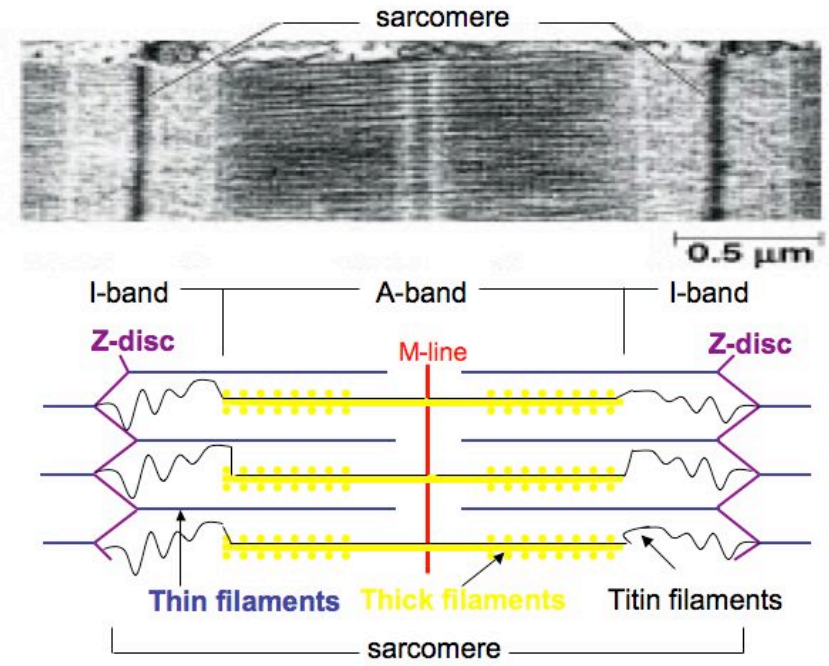


Figure 1.2 Filament composition of the sarcomere. The patterning of myofibrils is a function of the filament and protein composition. There are three types of overlapping filaments that provide the foundation for the sarcomere: thin filaments, thick filaments, and titin filaments. Thin filaments are primarily actin polymers, thick filaments are primarily assemblies of conventional myosin, and titin filaments consist of a few molecules of the giant protein called titin. Besides providing the structural architecture, the filaments are invaluable to the function of muscle. The interaction of the myosin heads projecting from the surface of thick filaments, with the thin filaments is fundamental to the shortening and lengthening of the sarcomere during contraction. Titin filaments provide the passive elasticity and a spring-like mechanic to the muscles (Linke, Ivemeyer et al. 1998). During muscle contraction, neither thick nor thin filaments change their lengths, but the I-band portion of titin filaments will straighten-out from its coiled position.

visualized when muscle is analyzed under a light microscope are functions of the filament and protein positioning. With the usual fixation and staining techniques (e.g. hematoxylin/eosin staining), the striations, or alternating light and dark bands, are called the I-bands and A-bands, respectively. Actually, the terms I-bands and A-bands (abbreviations for isotropic and anisotropic) arise from their appearance without fixation under polarized light (the I-bands will be dark and the A-bands light). In a stained specimen, the molecular anatomy of the sarcomere also includes two sectional markers, visualized as distinct dark lines, one bisecting the I-band called the Z-disc, and one that bisects the A-band called the M-line. The Z-disc defines the boundaries between the individual sarcomeres.

Healthy muscle tissue will have a regular pattern throughout the muscle cell, corresponding to the individual sarcomeres that function in unison. In abnormal or diseased muscle tissue, these structures will typically be distorted, corresponding to a disruption in one or more of the filament types. Thin filaments are actin polymers (F-actin) consisting of two strands of actin monomers that intertwine into a helix. One turn of the helix contains 7 actin monomers on one strand or a total of 14 actin monomers and spans a distance of 37 nm. Thin filaments are not just F-actin; many other proteins are associated, the most prominent of which are the troponins and tropomyosin. One set of troponins (troponins I, C and T) and one rod-shaped coiled-coil dimer of tropomyosin are associated with each turn of the F-actin helix. The F-actin of thin filaments is polarized, based on the asymmetry of the actin monomers, with the barbed (or plus) end attached to the Z-disc and the pointed (or minus) end located in the middle of the A-band. As mentioned below, the interaction between actin in thin filaments and myosin heads is

regulated by calcium. In striated muscle, this regulation occurs through the tropomyosin/troponin complex.

The major component of thick filaments is myosin, an asymmetric molecule that appears under EM as 2 globular heads attached to a rod. Myosin is a hexamer consisting of two myosin heavy chains and two copies each of two types of myosin light chains. The N-terminal half of each myosin heavy chain forms the globular head. The C-terminal halves of both myosin heavy chains form an α-helical coiled-coil dimer, which is the rod portion of the myosin molecule. Individual thick filaments from mammalian skeletal muscle can be imaged by EM and appear as bipolar filaments (approx. 1.6 μm long), with "heads" consisting of the globular portions of myosin projecting from the surface, except in the middle of the filament, the "bare zone". The rod portions of myosin molecules have the ability to self-assemble into the shaft of the thick filament, mostly in a side-to-side parallel fashion, except in the middle of the thick filament, where the packing is anti-parallel. Consideration of simply the primary amino acid sequence and the α-helical secondary structure of the myosin heavy chains can explain both the formation of the coiled-coil dimer and how these myosin rods assemble in a parallel fashion into the thick filament shaft (McLachlan and Karn 1982). Indeed, this modeling can also explain the evolutionarily conserved and uniform stagger between myosin heads on the surface of the thick filaments (shortest distance between neighboring heads is 14.3 nm). However, neither consideration of the myosin heavy chain sequences nor any other data can explain how the anti-parallel assembly occurs, how anti-parallel switches to parallel assembly along the thick filament shaft, nor how the observed uniform length of thick filaments is achieved and maintained. The globular heads of myosin can bind to both actin and ATP,

in a mutually exclusive manner. The ATPase activity of myosin provides the energy necessary to exert the mechanical force of contraction. The cyclical interaction between a myosin globular head and actin in the thin filaments explains muscle contraction. A combination of mechanical measurement, crystal structures of the myosin head bound with various nucleotides and spectroscopic observations of contracting muscle have led to a likely mechanism by which chemical energy from ATP is converted into force: the dissociation of phosphate from the myosin head is associated with a large conformation change in the myosin head and a swinging of the myosin neck (which acts as a lever arm).

Titin filaments are composed of a few molecules of the giant protein titin (Maruyama 1976; Wang, McClure et al. 1979). In vertebrates, single titin polypeptides span half of a sarcomere, from Z-disc to M-line. The N-terminus of titin is tightly anchored at the Z-disk; the C-terminus of titin is anchored at the M-line. The A-band portion is associated with the thick filament shaft, and helps assemble the A-band by associating with myosin and a large number other proteins. The I-band portion of titin is largely "free" and acts as an elastic or rubber band: it provides passive elasticity so that sarcomeres can maintain a set resting length. I will discuss more about this fascinating molecule later in the Introduction and elsewhere.

Muscle evolved to suit the voluntary and involuntary movements required by animals. The mechanical forces that allow such movement to occur are derived from muscle contraction, or the rapid shortening and relaxing of the sarcomere. The molecular mechanisms governing muscle contraction have been nicely characterized and involve the myosin-dependent pulling of thin filaments towards the mid-line of the sarcomere.

Myosin activation occurs in one of two calcium-dependent ways that are tissue specific. In striated muscle the calcium acts through the troponin/tropomyosin portion of the thin filament. Troponin and tropomyosin wind the length of the thin filament in an orientation that, at rest, blocks the myosin-binding sites of actin monomers within F-actin. Elevated cellular concentration of calcium permits the association of calcium with troponin C, which causes conformational changes in the protein that somewhat expose the myosin-binding sites. As a result, myosin binds, further displacing tropomyosin, and securing the interaction between the myosin head and the thin filament. In smooth muscle and non-muscle cells calcium activation of the myosin-actin interaction occurs through calcium binding to calmodulin, and calcium/calmodulin binding to myosin light chain kinase (MLCK), which is then converted from an autoinhibited to an active kinase. MLCK then phosphorylates the regulatory myosin light chain, which is situated on the myosin neck and this in some still mysterious way activates the myosin head (probably by decreasing the space between myosin head and thin filaments). In either case, the structural basis for the conversion of free energy into force is provided by conformational changes in the myosin head when the inorganic phosphate (the product of ATP hydrolysis) is released. More specifically, phosphate release (and the conformational changes that occur as a result) lead to a swinging of the neck region of myosin, which results in a pull on the thin filament, generally called the “power stroke”. Synchronicity of power strokes between and within individual sarcomeres is the key to rapid contraction.

The general features of the sarcomere are essential to further our understanding of the micro functions that allow for the assembly and maintenance of this molecular

machine. The decades of work that precede this thesis have provided an invaluable foundation for developing an even more detailed comprehension of muscle biology. Such knowledge will continue to prove useful in both the basic and applied scientific fields.

Caenorhabditis elegans: an excellent model system for studying muscle

Caenorhabditis elegans is a simple, multicellular free-living nematode species that grows to about 1 mm in length in the adult stage, and contains an invariant number of somatic cells (Brenner 1974). In many ways, the simplicity of *C. elegans* renders the organism microbe-like, and thus, the nematode is a terrific model system for researching intracellular/multicellular processes. The fixed number, predetermined fate, and light microscope visibility of the individual cells in *C. elegans* adds to its virtue, facilitating genetic analysis and elucidating mechanisms involved in animal development/anatomy (Wood 1988). More specifically, the organism allows for single cell resolution in an intact animal. Equally important to the utility of the nematode, especially for genetic analysis, is the short lifespan (~3 days) (Hosono, Mitsui et al. 1982), large brood size (~250 progeny by self-fertilization and over 1000 progeny by cross-fertilization with a male) (Hodgkin 1983), relatively small genome (100 Mb on 6 chromosomes) (Hodgkin, Plasterk et al. 1995), ability to freeze and revive genotypic strains, and the ease of molecularly cloning genes first identified through mutants with phenotypes (Jorgensen and Mango 2002). In addition, *C. elegans* exists as two sexes, the hermaphrodite and male (Hodgkin 1983). The hermaphrodite sex produces both eggs and sperm and can therefore reproduce by self-fertilization; the male sex only produces sperm and therefore

must mate with hermaphrodites to produce cross-fertilized progeny. The hermaphrodite allows for the passage of homozygous mutations without additional mating. The hermaphrodite also allows for the propagation of animals that are deficient in motility (due to muscular or neuronal defects). These animals would otherwise be unable to mate, as movement is critical for the sexual union of two organisms. *C. elegans* can be successfully employed to study germ cells, spermatogenesis, the nervous system, signal transduction, sex determination, apoptosis, cell migration, chemotaxis, evolution, and more. Our lab uses *C. elegans* to dissect muscle biology: the assembly and maintenance of the sarcomere, and the causes and effects of muscle contraction.

Similar to higher animals, *C. elegans* has muscle tissue functionally dispersed throughout the organism (Waterston, Epstein et al. 1974). The muscles of the pharynx allows for the grinding and pumping of bacteria into the gut (for nutrients), while the enteric muscles assist in the completion of digestion and excretion. The vulval muscles provide the pushing/contractile forces for egg laying and the body wall muscles of *C. elegans* are directly responsible for locomotion. The spindle shaped body wall muscle cells line four quadrants that run the length of the nematode, arranged in interlocking pairs (Moerman and Williams 2006) (Figure 1.3). The muscle cells basally juxtapose the hypodermis and laterally line their neighbor muscle cells. The intimate linking of body wall muscle to the hypodermis and outer covering of the worm (the cuticle) allows for synchronized contraction and movement of the animal. The myofibrils exist in a narrow

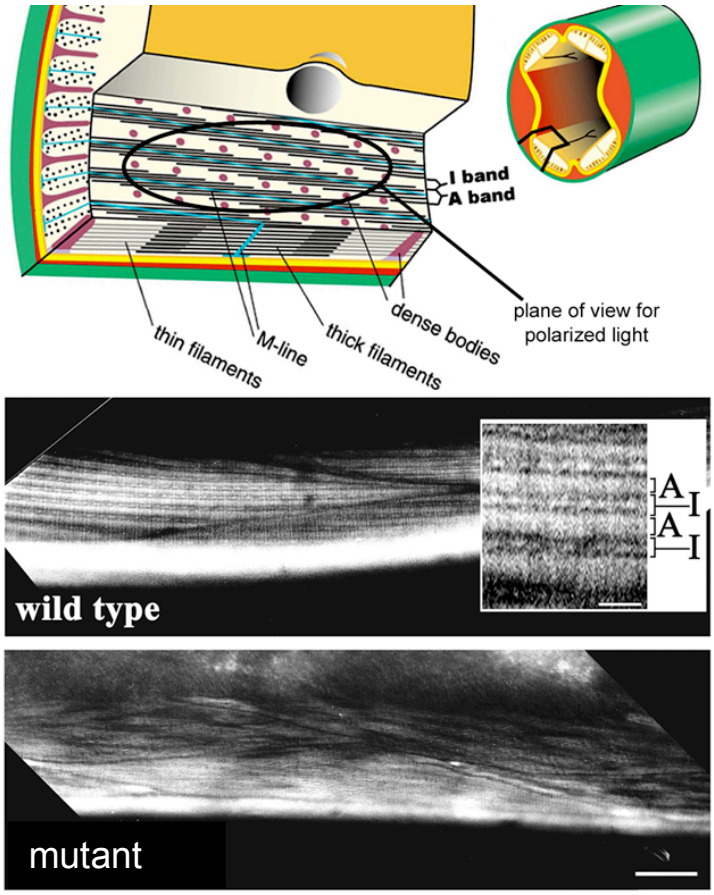


Figure 1.3 Utility of polarized light microscopy to evaluate *C. elegans* muscle structure. The top panel is an illustration showing the position of the four muscle quadrants (right) and a portion of a single muscle cell emphasizing the myofilament lattice. The oval indicates the plane of view seen by polarized light or by immunofluorescent microscopy. The bottom two panels show polarized light microscopy of adult body wall muscle from wild type and from a typical mutant animal. Each shows portions of 3–4 spindle shaped muscle cells. The middle panel inset shows an enlarged view from wild type muscle. Wild type muscle shows a high degree of organization with birefringent A-bands alternating with dark I-bands, each containing a row of dense bodies. In contrast, the mutant displays disorganized, irregular myofibrils.

zone approx. 1.5 μm tall, along one side of the muscle cell. All the myofibrils are attached to the muscle cell membrane and ultimately the basal lamina and hypodermis at each M-line and dense body (the anatomical equivalent of the Z-disc). Thus, the M-line and dense bodies serve the same function as their equivalent structures in vertebrate muscle (cross linking of thick filaments and attachment of thin filaments, respectively), and in addition, function as costameres in vertebrate muscle. Costameres are the focal adhesion-like structures that attach the peripherally located myofibrils to the muscle cell membrane. Indeed, as we are learning more about the composition of nematode M-lines and dense bodies, this equivalence to costameres and focal adhesions is becoming even more apparent.

Both the transparency of *C. elegans* and the oblique striations patterned through the muscle cells allow for the rapid analysis of sarcomere integrity using a polarized light microscope (Zengel and Epstein 1980) (Figure 1.3). Thus, live adult animals can be placed directly on a slide to evaluate muscle tissue. These characteristics, accompanied with the ease of generating and analyzing mutants (Jansen, Hazendonk et al. 1997) and RNAi knockdowns (Fire, Xu et al. 1998), are extremely powerful tools for dissecting the molecular contributions of individual proteins to sarcomere assembly and maintenance. Early mutagenesis screens identified two major phenotypic classes of mutants that affect myofibril assembly or structure: the “Uncs” (uncoordinated) and the “Pats” (paralyzed at two-fold stage) (Moerman and Williams 2006). The Uncs are uncoordinated adults that include mutations in about 130 different genes, 40 of which specifically affect muscle, the other 90 specifically affecting the nervous system. The Uncs can be identified using a dissecting light microscope at low power to find animals that are slow moving or

paralyzed or move in a way different from the typically smooth, sinusoidal movement of wild type animals. For example, worms that are lacking or otherwise mutant for the giant muscle protein encoded by the *unc-22* gene have a characteristic "twitching" phenotype after about the L4 larval stage, and thus the protein was called "twitchin" (Benian, Kiff et al. 1989). Pat mutations are more severe, as the worms cease development within the embryo at the 2-fold stage and do not move within the eggshell (Williams and Waterston 1994). There are about 18 genes that when mutant result in the Pat phenotype. A few genes have loss-of-function Unc and null Pat phenotypes; examples include *unc-97* (the mammalian homolog encodes PINCH; (Hobert, Moerman et al. 1999)) and *unc-45* (encodes a conserved myosin chaperone; (Barral, Bauer et al. 1998)). A model for myofibril assembly has been proposed in which assembly is initiated or directed by signals first laid down in the extra cellular matrix (ECM) and muscle cell membrane (Williams and Waterston 1994; Moerman and Williams 2006). This is supported by the observation that the Pat mutants showing the greatest degree of disorganization define genes that encode components of the ECM (e.g. UNC-52 (perlecan) and muscle cell membrane (e.g. PAT-3 (β-integrin)). This model is further supported by immunolocalization studies revealing the timing and localization of these proteins during embryonic development (Hresko, Williams et al. 1994).

Although *C. elegans* proves a useful model system for elucidating muscle biology, there are muscular differences when directly compared to vertebrates. Vertebrate muscle cells are multinucleated, resulting from the fusion of myoblasts during development (Loomis, Wahrmann et al. 1973). *C. elegans* muscle cells are mononucleated, as development proceeds directly from a single precursor cell into the

muscle cell (Gossett, Hecht et al. 1982). The vertebrate myofibrils fill the entire cell, while in *C. elegans* the myofibril is restricted to a narrow zone beneath the cell membrane along one side of the body wall muscle cells (Moerman and Williams 2006). In aggreement with this restriction, all of the myofibrils within a *C. elegans* muscle cell are tightly anchored to the cell membrane through the M-lines and dense bodies. In contrast, in vertebrate striated muscle cells, the myofibrils fill the entire cross sectional area of the cell and only the outermost myofibrils (located at the periphery of the cell) are anchored to the muscle cell membrane via costameres. A final difference is in the geometry of the muscle sarcomeres. In both vertebrates and invertebrates, the Z-discs of the sarcomere are perpendicular to the longitudinal axis of the muscle cell, while the filaments are parallel. In vertebrates, the Z-discs stack directly on top of each other. In contrast, the sarcomeres within *C. elegans* slightly stagger on top of each other, positioning the sarcomeres at a 5-7° angle relative to the longitudinal axis of the muscle, resulting in oblique striations (Mackenzie and Epstein 1980). In addition, invertebrate sarcomeres are usually much longer than those in vertebrates. For example, *C. elegans* adult body wall muscle thick filaments are 9.7 μm long, as compared to the typical thick filament in vertebrate striated muscle of 1.6 μm. These longer thick filaments result in many more cross bridges and thus probably higher force production per thick filament / thin filament pair. The physiological implications of these differences are unknown, but it is hypothesized that the angular positioning of the sarcomeres and the longer thick filaments in the nematode may allow for the transmission of greater forces during contraction (Burr and Gans 1998).

Titin and titin-related proteins

Vertebrate titin

Early anatomical observations of muscle cells revealed that filaments provide the foundation for the architectural substructure. At first, it was thought that there were two types of filaments, both fairly rigid and homogenous: thin filaments and thick filaments (Huxley 1957). Later, a third type of filament was identified and named titin (Maruyama 1976; Wang, McClure et al. 1979; Maruyama, Kimura et al. 1981). Titin differs from the other filaments because it is a single, giant, polypeptide. Typically ~6 molecules of titin make a single filament, in contrast to the several hundred molecules of actin and myosin that form thin and thick filaments, respectively (Liversage, Holmes et al. 2001). Titin is translated from an 83kb transcript into an ~40,000 amino acid protein (Labeit and Kolmerer 1995, Bang, 2001 #168). The first electron micrographs that visually identified titin predicted modular repeats of similar, if not identical, domains (Maruyama, Kimura et al. 1984; Trinick, Knight et al. 1984, Wang, 1984 #183). The protein appeared to have a "pearls on a string" appearance, the pearls ranging from 3-4nm (figure 1.4). Sequence analysis confirmed that 90% of titin is made up of 100 residue repeats, corresponding to Ig or Fn domains, which, in fact, measure ~3-4nm in diameter (Labeit and Kolmerer 1995). Titin spans ~1.2μm, the length of half of a sarcomere (Nave, Furst et al. 1989). Elegantly positioned, titin filaments are thought to provide balanced force between the two halves of the sarcomere.

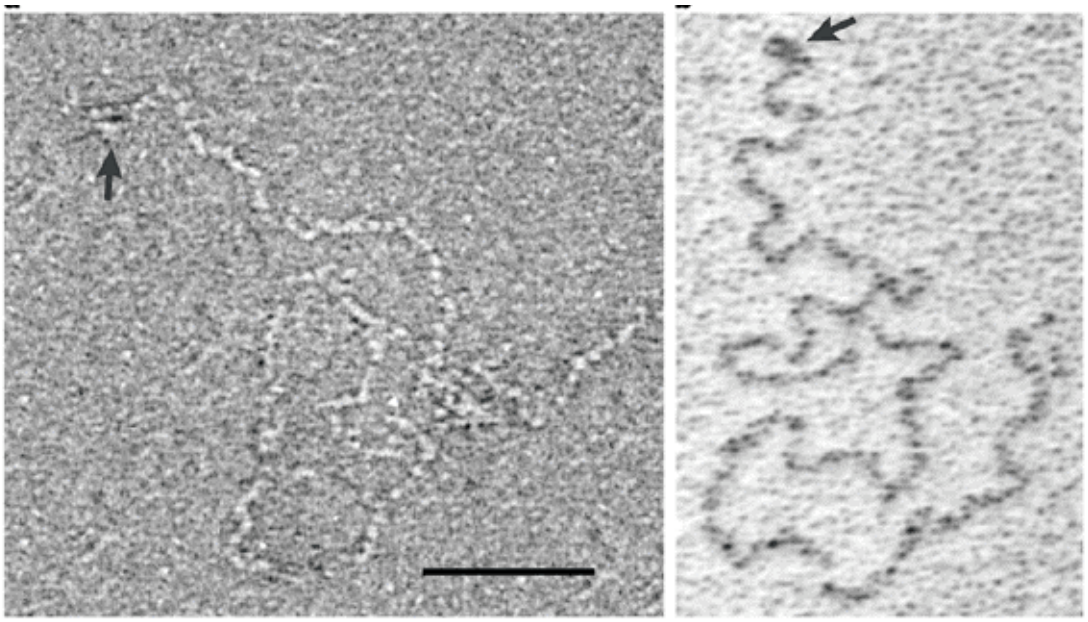


Figure 1.4 Electron micrograph of the giant titin protein. The flexible, filamentous titin molecules reach over 1 μm in length and 4 nm in diameter. The staining reveals the bead-like substructure, which sequence analysis and crystal structures indicate consist of individual Ig and Fn3 domains. Arrows point to the head of titin, which is located at the C-terminus (M-line portion) of titin. The head also contains two M-line proteins that remain tightly associated with titin during purification (Nave, Furst et al. 1989). Figure used with permission from (Tskhovrebova and Trinick 2003). Titin filaments were isolated from rabbit skeletal muscle. The left panel was graphed on a carbon substrate after negative staining with uranyl acetate. The right panel was dried on mica and rotary shadowed by platinum.

The N-terminus is docked at the Z-disc stretching titin two through the I band and A band regions, ultimately to the M-line, where the C-terminus attaches (Figure 1.2) (Nave, Furst et al. 1989; Whiting, Wardale et al. 1989, Zou, 2006 #69). The A band region of titin is a regular arrangement of domains that bind to other proteins of the thick filaments (Labeit and Kolmerer 1995; Sebestyen, Fritz et al. 1996). This region includes the single catalytic domain, a kinase near the C-terminus. Within the I-band titin filaments are extensible, providing the intrinsic elasticity of vertebrate striated muscle cells, which is crucial for maintaining structural integrity during rest and during the application/release of contractile forces (Granzier and Labeit 2007). This spring-like quality is primarily provided by an elastic region, the PEVK region, named because 70% of its sequence consists of proline, glutamate, valine and lysine. Although initial sequence analysis did not reveal any repeats, several years later it was discovered that the PEVK region is mostly made up of tandem repeats of 28 residues (Labeit and Kolmerer 1995; Linke, Stockmeier et al. 1998) often interrupted by glutamate-rich sequences (Freiburg, Trombitas et al. 2000; Greaser 2001; Gutierrez-Cruz, Van Heerden et al. 2001). An NMR structure of one of these 28 residue modules reveals it to consist of alternating polyproline II helices and random coils (Ma, Kan et al. 2001). Titin is a complex macromolecule with multiple functions, it is critical for sarcomeric assembly, therefore making it likely responsible for the ordered assembly of 100's of protein subunits into the organized sarcomere (Lange, Ehler et al. 2006). Thus, titin has its own organized proteome, creating a giant communicating network, integrating signals from various directions within the muscle cell.

Many interactions between the structural regions of the giant proteins related to titin and various sarcomeric proteins have been identified. However, identifying substrates for the catalytic kinase domain(s) of the giants has proven difficult. A recombinant twitchin kinase from *Aplysia* phosphorylates regulatory myosin light chain *in vitro* when incubated with an actomyosin preparation or purified myosin (Heierhorst, Probst et al. 1995). However, when a similar experiment was performed with nematode twitchin kinase, such phosphorylation was not observed (J. Heierhorst and G.M. Benian, personal communication). Vertebrate titin kinase is the only other giant muscle protein kinase that has known substrates. None of these substrates can account for all the predicted roles of the enzyme, begging for scientific progress in understanding the physiological implications of the kinase and its homologs. In mature muscle, titin has a single known substrate, nbr1(Lange, Xiang et al. 2005). Titin-phosphorylated nbr1 transmits a signal through p62, leading to a subsequent increase in nuclear SRF. Nuclear SRF increases gene expression, including expression of myofibril proteins. The implications of the substrate interaction between nbr1 and titin are significant, as mutations in human patients that disrupt the interaction result in a hereditary muscle disease. Further, the results likely explain why exercise leads to increased myofibrils, while disuse leads to atrophy. In developing muscle, the sarcomere is not yet assembled, and therefore titin is hypothesized to adopt a different conformation with different interactions. There is one identified substrate for titin in developing muscle, telethonin (also known as Tcap) (Mayans, van der Ven et al. 1998). We will discuss the implications of titin activation and substrate interactions during development later in the introduction.

Vertebrate titin is an extremely diverse molecule. Within a single organism, the dozens of titin isoforms are vast and complex, including variability in isoform expression even within individual skeletal muscles (i.e.- psoas vs. soleus muscles) (Granzier, Helmes et al. 2000). More obvious variability occurs between skeletal and cardiac muscles or when comparing isoform expression in different vertebrates (Granzier and Labeit 2007). The multiple isoforms add to the complexity of titin, and also muddy the boundaries of classifying a titin homolog.

Titin-like proteins in C. elegans

C. elegans has 3 titin-like proteins: TTN-1, twitchin, and UNC-89 (Ferrara, Flaherty et al. 2005). This classification is based on the enormous size of these polypeptides (>700,000 Da), the presence of multiple Ig and Fn3 domains in conserved patterns, and the presence of one or two related protein kinase domains.

Twitchin (encoded by the *unc-22* gene) (Figure 1.5) was the first intracellular protein discovered that belongs to the Ig superfamily (Benian, Kiff et al. 1989, Benian, 1993 #54, Moerman, 1988 #11). The ~800kDa protein is primarily comprised of 31 Fn3 domains, 30 Ig domains, and a single protein kinase domain. Upon its discovery, twitchin was hypothesized to be an important regulator of muscle contraction, mainly due to the motility phenotype of the *unc-22* mutant and MLCK-like kinase domain. Later, its role in inhibiting the rate of relaxation was identified in studies of twitchin from *Aplysia* and *Mytilus* (Probst, Cropper et al. 1994; Siegman, Funabara et al. 1998). Located at the outer portion of the muscle A-band (Moerman, Benian et al. 1988), twitchin could possibly serve a similar role as the A-band portion of vertebrate titin in thick filament positioning

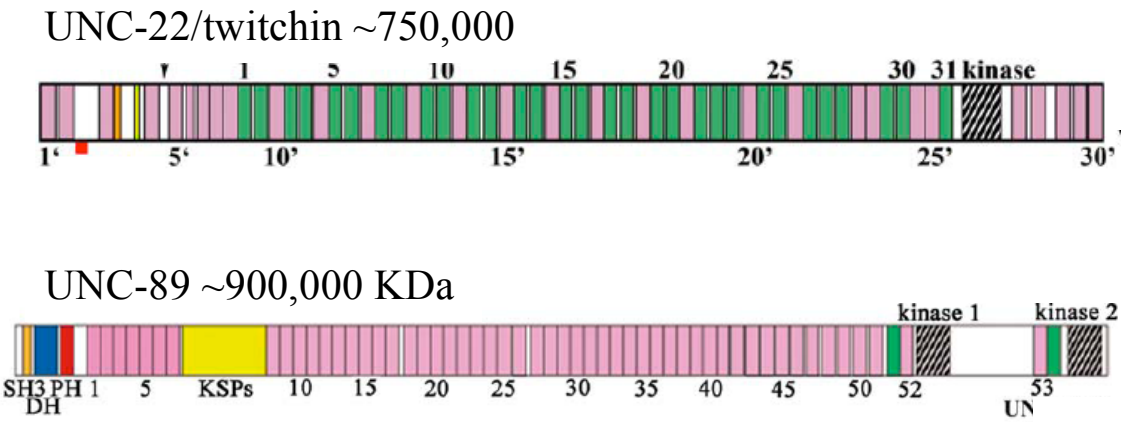


Figure 1.5- Domain organization of twitchin and the largest isoform of UNC-89. The giant proteins are composed of repeating Ig (pink boxes, Ig domains of twitchin are notated with a ‘) and Fn3 (green boxes) domains. The C-terminal kinase domains, characteristic of titin-like proteins, are striped black and white. In addition, twitchin, located in the A-band region, has stretches of amino acids predicted to be intrinsically unstructured (red tab), and a glycine-rich region, which might act as a hinge (denoted by black arrow). In the M-line, UNC-89 has several additional domains located near its N-terminus (SH3, DH and PH). The yellow region marked “KSPs” contains ~60 copies of this sequence motif and is likely to be phosphorylated. Note that UNC-89 has two protein kinase domains, only one of which (kinase 2) is predicted to be catalytically active.

and sarcomere integrity. Indeed, both twitchin and the A-band portion of vertebrate titin are composed almost exclusively of alternating patterns of Fn3 and Ig domains (usually Fn3-Fn3-Ig, or Fn3-Fn3-Fn3-Ig) (Benian, Kiff et al. 1989; Benian, L'Hernault et al. 1993; Labeit and Kolmerer 1995). Moreover, the pattern of organization of 16 of these Fn3 and Ig domains surrounding the protein kinase domain is conserved between twitchin and vertebrate titin. However, whereas the C-terminal half of vertebrate titin spans a vertebrate thick filament (Nave, Furst et al. 1989), this is not possible for the smaller twitchin and the much larger nematode thick filament (up to 10 μm in adult muscle) (Benian, Tang et al. 1996). In the case of twitchin, to cover the length of the thick filament either twitchin must form a polymer or be positioned periodically along the thick filament (with gaps). *Aplysia* twitchin has been shown, *in vitro*, to phosphorylate regulatory myosin light chains (Heierhorst, Probst et al. 1995). There are no other endogenous substrates identified for the kinase domain of twitchin, however *C. elegans* twitchin has catalytic activity *in vitro*, both in autophosphorylation and towards peptides derived from vertebrate myosin light chains (Lei, Tang et al. 1994; Heierhorst, Tang et al. 1996).

UNC-89 (Figure 1.5) is the invertebrate homolog of mammalian obscurin, but can be considered titin-like because of the large numbers of Ig domains, its massive size (up to 900,000 Da), presence of one or two protein kinase domains, and the importance of the macromolecule in correct sarcomere formation (Benian, Tinley et al. 1996; Small, Gernert et al. 2004). Localized to the M-line region of the A-band, worm strains harboring mutant alleles of *unc-89* have a disorganized myofilament lattice and, for most alleles, no M-lines (Ferrara, Flaherty et al. 2005). *unc-89* is a complex gene, and through

the use of three promoters and alternative splicing, it generates at least six major polypeptides ranging in size from 156,000 to 900,000 Da. Like the other titin-like giant proteins, it is thought that UNC-89 participates in transmitting cellular signals via phosphorylation. In contrast to the other *C. elegans* titin-like proteins that have a single protein kinase domain, some isoforms of UNC-89 have two. Although neither kinase domain has thus far demonstrated catalytic activity, modeling suggests that one of the domains is active, while the other is probably a psuedokinase (Small, Gernert et al. 2004). Additionally, UNC-89 harbors SH3, DH, and PH domains. DH domains typically provide Rho-GEF activity, and evidence from our lab suggests that the DH domain of UNC-89 is no exception (Quadota et al, manuscript in revision). The data reveals an interaction between UNC-89 and Rho-1 that promotes the exchange of GDP for GTP.

The final titin-like protein in *C. elegans*, TTN-1 (Figure 1.6), is more similar to vertebrate titin than twitchin or UNC-89 (Flaherty, Gernert et al. 2002). Significantly larger, TTN-1 is 2.2MDa consisting of 56 Ig domains, 11 Fn3 domains, 5 regions of short sequence repeats, and a single protein kinase domain. TTN-1 localizes to the I-band region of the sarcomere. It is thought to span 2μm, the length of half of an I-band, from the outer edges of the dense body to the A-band boundary. Recently, TTN-1 specific antibodies were developed and upon western blot analysis of worm extract these antibodies recognize a ~2MDa protein, providing visual evidence for the massive polypeptide in *C. elegans* (Ma et al, manuscript in preparation). Besides the large size and localization of TTN-1, the most striking feature that provides the foundation for the similarity to the I-band portion of vertebrate titin is the presence of several regions that are predicted to be elastic. Most noticeably is the PEVT region, which is similar in

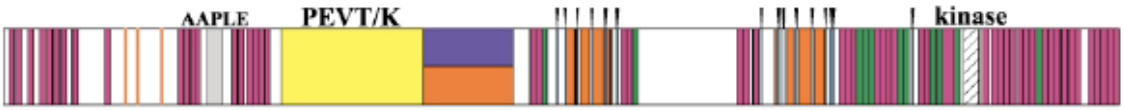


Figure 1.6- Domain organization of TTN-1. With a molecular mass of 2.2 MDa, TTN-1 is the largest known protein in *C. elegans*. Similar to the other titin-like proteins, TTN-1 has multiple Ig (pink boxes) and Fn3 (green boxes) domains, as well as a single catalytic protein kinase domain (black and white striped box). The giant protein localizes to the I-band region and has multiple regions predicted to be coiled coil (the smaller of the repeats depicted as orange boxes), with the largest stretching for 1500 amino acids (orange and blue box). The 250 residue AAPLE repeat region (grey box) contains tandem copies of a 14 residue repeat. The PEVT/K region (yellow) has 56 tandem copies of a 39 residue repeat, interrupted by 5 copies of a 50 residue repeat, totaling 2400 amino acids. Thin blue lines with projections mark a 30 amino acid dispersed repeat. Recent CD studies are consistent with the idea that the AAPLE, PEVT/K, and dispersed repeats consist of a combination of polyproline II helix and random coil, and thus all three types of repeats might be molecular springs analogous to the PEVK region of vertebrate titin.

sequence (59% P, E, V and T), but not homologous to, the PEVK region of titin (Flaherty, Gernert et al. 2002). Like vertebrate titin PEVK, the TTN-1 PEVT is composed of tandem repeats (in this case 39 residues long) interrupted by glutamate-rich regions. Moreover, CD data are consistent with these repeats containing a combination of polyproline II helix and random coil, just like the vertebrate titin PEVK repeats (Ma et al, manuscript in preparation). Furthermore, there is an additional region that is composed of tandem repeats, and a dispersed repeat, which again, by CD experiments are suggested to have similar secondary structures as the vertebrate PEVK. Thus, we predict that TTN-1 is responsible for providing the worm sarcomere with passive elasticity, an invaluable feature of muscle stability.

Like twitchin and titin, TTN-1 also has a single protein kinase domain located at the C-terminus of the large molecule (Flaherty, Gernert et al. 2002). We will discuss the biochemical similarities between these protein kinases in detail in later sections. *In vitro* phosphotransferase activity has been demonstrated for TTN-1 kinase against model peptides derived from vertebrate myosin light chains. However, there is no further evidence to support that MLCK is a substrate *in vivo*. This thesis provides evidence for the first substrate interaction of TTN-1 kinase.

Structure/function relationships and regulation of giant protein kinases

C. elegans twitchin and mammalian titin: crystal structures and autoinhibition

Structural data provides insight towards the biological properties of a macromolecule. Specifically, crystal structures reveal a snapshot of a protein's conformation, generally accepted to be similar to the native conformation the protein

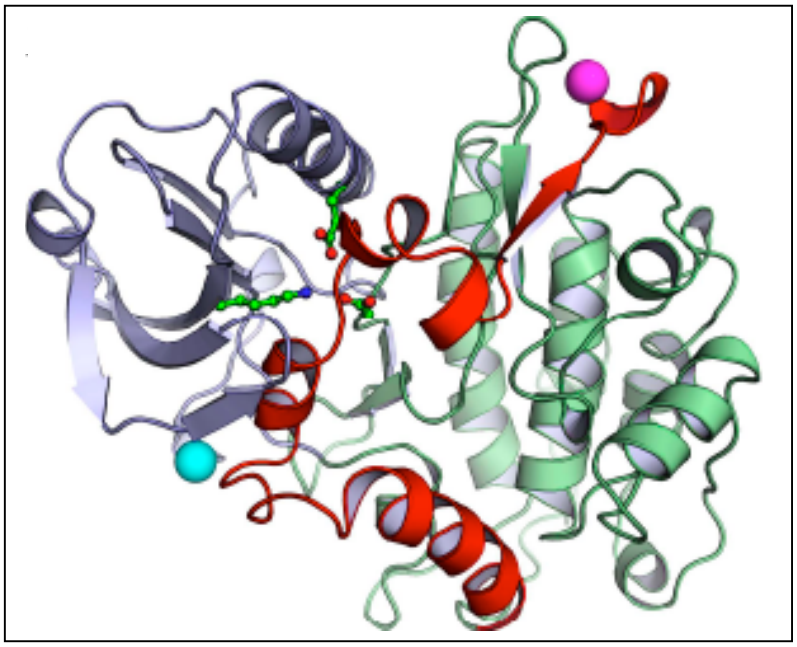


Figure 1.7- Crystal structure of twitchin kinase. Twitchin kinase is composed of 3 subdomains: an alpha-helical rich large lobe (green), a small lobe of mainly beta sheets (grey), and the autoregulatory tail (red) (structure taken from (Kobe, Heierhorst et al. 1996)). The autoregulatory tail is situated between the 2 lobes making extensive contact with the active site. To orient the active site conserved residues that help neutralize the ATP binding pocket are shown in a ball and stick interpretation. The N terminal alpha carbon is marked with a cyan sphere; the C terminal alpha carbon is marked with a magenta sphere.

assumes. The crystal structures of three of the giant protein kinases from *C. elegans* twitchin (Figure 1.7), *Aplysia* twitchin, and vertebrate titin (Hu, Lei et al. 1994; Kobe, Heierhorst et al. 1996; Mayans, van der Ven et al. 1998) have been solved. These structures reveal an intrinsic mechanism of protein regulation, comparable to the MLCK family.

Similar to most protein kinases, twitchin and titin adopt a bi-lobal topology, with the smaller, N-terminal lobe folding into anti-parallel β sheets and the larger, C-terminal lobe mainly forming α helices (Knighton, Pearson et al. 1992; Hu, Lei et al. 1994; Taylor and Radzio-Andzelm 1994; Kobe, Heierhorst et al. 1996; Mayans, van der Ven et al. 1998) (Figure 1.7). The lobes are connected through a glycine hinge region that determines the degree of separation and rotation between the substructures (Taylor and Radzio-Andzelm 1994). Many protein kinases have two conformations, open and closed (Cox, Radzio-Andzelm et al. 1994; Huse and Kuriyan 2002). However, both twitchin and titin always seem to occupy an "open" state, stemming from the intrasteric mechanism of regulation (Kobe, Heierhorst et al. 1996; Mayans, van der Ven et al. 1998). Twitchin and titin each have an autoregulatory tail that makes extensive contact with the active site, mimicking a substrate interaction and inhibiting catalysis. In twitchin, the traversion of the regulatory tail through the entire substrate-binding groove and partially through the ATP binding site makes 47 hydrogen bonds and 351 van der Waals contacts (Hu, Lei et al. 1994; Kobe, Heierhorst et al. 1996). With such intimate contacts, it is hard to imagine conformational changes that support the removal of the autoinhibitory tail without disruption to the active site and catalytic integrity of the enzyme.

Almost all protein kinases share the core topology of ~270 amino acids (Hanks and Hunter 1995). When active, the structures of different kinases tend to be similar in both the ATP and protein substrate recognition regions. However, in general, the inactive structures have more variability between kinases. The intrasteric regulation mechanism discussed above for twitchin and titin kinase is also found in the MLCK family (Knighton, Pearson et al. 1992). MLCKs have a regulatory tail located just downstream of the catalytic core and following the regulatory tail is a calmodulin binding site. Although there are no solved crystal structures of MLCKs, the structure of calmodulin kinase 1 (CaMK1) has been solved (Goldberg, Nairn et al. 1996). CamK1 is homologous to the MLCKs and contains the same regulatory regions following the catalytic core. When visually comparing twitchin, titin, and CaMK1 a distinct difference is seen in the C-terminal tail (Goldberg, Nairn et al. 1996; Kobe, Heierhorst et al. 1996; Mayans, van der Ven et al. 1998). In all 3 structures, the regulatory tail makes extensive contact with the protein substrate binding groove, but in CaMK1 the short tail that binds calmodulin wraps around the N-terminal lobe, causing severe distortion. In contrast, the same region of twitchin and titin occupies part of the ATP binding site. Thus, it is tempting to imagine that, although quite similar, the enzymes have different mechanisms of regulation.

Activation of MLCKs, twitchin and titin

Phosphorylation via protein kinases is one of the most common forms of regulation in all cells types. Phosphorylation can occur quickly and reversibly in a substrate specific manner, making it an extremely powerful on/off switch. Specificity

and spatial regulation is mainly provided by substrate/active site conformations and nanoscale localization, however the temporal regulation often requires modulation of the kinase activity. Thus, in addition to providing a means of cellular regulation, protein kinases themselves must be controlled. As discussed earlier, the MLCKs and the giant protein kinases have an intrinsic regulatory tail that makes intimate contact with the substrate-binding site (Knighton, Pearson et al. 1992; Goldberg, Nairn et al. 1996). The MLCKs are relieved of this inhibition by the binding of calcium/calmodulin to the autoregulatory domain. In contrast, the giant protein kinases have alternative mechanisms for activation (Mayans, van der Ven et al. 1998; Grater, Shen et al. 2005; Lange, Xiang et al. 2005; Greene, Garcia et al. 2008).

For the MLCK proteins, activation has been fairly well characterized and occurs in a step-wise manner (Goldberg, Nairn et al. 1996; Mirzoeva, Weigand et al. 1999). First, an influx of intracellular calcium allows binding of calcium to calmodulin, causing calmodulin to assume a different conformation. The calcium/calmodulin complex binds to the regulatory region of the MLCK causing the regulatory domain to collapse. The conformational changes in the MLCKs adopted upon calcium/calmodulin binding expose the catalytic cleft and allow catalysis to occur. The binding of the protein substrate to the active site of the MLCK results in the closure of the catalytic cleft and reorientation of calmodulin. Because of the structural and sequence similarities between the MLCKs and the giant protein kinases, it was initially assumed that they would be activated by a similar mechanism. However, although twitchin kinase is partially activated by the binding of S100 (a protein similar to calmodulin) (Heierhorst, Kobe et al. 1996; Heierhorst, Tang et al. 1996; Heierhorst, Mann et al. 1997) and titin kinase activation is

enhanced by calmodulin binding during development (Mayans, van der Ven et al. 1998), neither event can account for the complete activation of the giant protein kinases (discussed further below).

Human titin kinase has two known substrates, telethonin and nbr1 (Mayans, van der Ven et al. 1998; Lange, Xiang et al. 2005). The substrate interaction between titin kinase and telethonin was elegantly examined using structural, cell biological, and biochemical techniques (Mayans, van der Ven et al. 1998). Titin interacts with telethonin during development, before titin has assumed its location traversing the sarcomere. In mature muscle, the catalytic kinase domain of titin is positioned near the M-line, while telethonin is restricted to the Z-discs. However, in developing muscle the localizations are different and immunofluorescence detects a spatial similarity between the C-terminus of titin and telethonin. Mayans et al. (1998) noticed a tyrosine residue deeply buried in the active site at rest. The tyrosine residue exists in the P + 1 loop (substrate binding pocket) and, when titin kinase is at rest, binds to a catalytic aspartate. By making a phosphomimetic mutation (to glutamate) in the suspected tyrosine residue, they showed a ten-fold increase in the specific activity of titin for telethonin. Further, if the phosphomimetic kinase was incubated with calcium/calmodulin the specific activity increased an additional 100-fold. The calcium/calmodulin had no effect on the wildtype titin kinase. Most interesting, the researchers also showed that when purified wildtype kinase is incubated with extracts from differentiating myocytes, phosphorylation is detected. However, if they incubated wildtype kinase with extracts from mature muscle, or mutated the tyrosine to glutamate, phosphorylation was not detected. These results support the importance of the tyrosine residue in activation of the kinase in developing

muscle, but also suggest that mature muscle has an alternative activation mechanism for titin kinase. It is noteworthy that at the comparable position of this tyrosine in titin kinase, there is an alanine in all the twitchin kinases, and a cysteine in TTN-1 kinase.

The activation mechanism of *Aplysia* and *C. elegans* twitchin kinase has been carefully studied, but many interesting questions still remain. In the mid 90's, Heierhorst et al showed that S100A2, but not S100B, binds and activates *Aplysia* twitchin kinase by ~10-fold (Heierhorst, Kobe et al. 1996). The addition of zinc to the reaction increases the activation by an order of magnitude. *C. elegans* twitchin kinase, which is only 38% different in its amino acid sequence when aligned with *Aplysia* twitchin kinase, was also shown to be activated ~10-fold by S100A2. However, zinc was shown to completely inhibit the *C. elegans* twitchin kinase reaction. The high similarities in the sequences and topologies of the twitchin kinases suggest functional conservation, but research has shown that they differ in substrate specificity, enzymatic activity, inhibitor sensitivity, and calcium/calmodulin affinity (Heierhorst, Kobe et al. 1996; Heierhorst, Tang et al. 1996). In general, *Aplysia* twitchin kinase is much more biochemically similar to a MLCK than to *C. elegans* twitchin kinase. The activation of *Aplysia* twitchin kinase by S100A2 is an inviting hypothesis, and supported nicely by *in vitro* experiments using S100A2 from bovine brain. Unfortunately, S100 proteins have yet to be identified in an invertebrate, suggesting that there is either an alternative calcium binding protein that confers activation, or an unknown mechanism of activation for twitchin kinase.

Crystal structures of the kinase domains of the giant proteins clearly reveal the major conformational changes that the enzyme must undergo to expose the active site (Figure 1.7). How does the active site escape the plethora of contacts that the regulatory

tail imposes? One hypothesis is that the giant protein kinases are activated by force; the forces generated from the contraction/relaxation cycles of muscle activity result in a "pulling" of the regulatory tail away from the active site, allowing substrate entry and catalysis to occur (Grater, Shen et al. 2005; Lange, Xiang et al. 2005). Thus, the giant protein kinase is speculated to behave as a force sensor. Physical characteristics of titin and titin-like proteins that support this idea include: the elasticity of the macromolecules, the positioning through the sarcomere, and the mechanical deformation of the M-line during stretch/contraction. This hypothesis was elegantly analyzed using force probe molecular dynamic simulations (FPMD) (Grater, Shen et al. 2005), and further supported with cell biological and our own single molecule studies (Greene, Garcia et al. 2008) (please see chapter 2).

One way to study complicated unfolding patterns is *in silico* using molecular dynamic simulations (Fersht and Daggett 2002; Oberhauser and Carrion-Vazquez 2008). Studying titin kinase, given the enormous size of full-length titin, is quite a feat, even using recombinant portions of the macromolecule. To add complexity, measuring the effect of force on an enzyme is quite difficult in practice. Thus, the use of computers, simulations, and modeling are an important asset to understanding certain biological phenomena. FPMB is a type of computer simulation that provides models to explain experimentally measured forces or provide estimates on the forces required to unfold specific macromolecules, unbinding reaction pathways, and interatomic interactions (Heymann and Grubmuller 2001). Using FPMB, Grater et al (2005) applied a harmonic pulling potential to the termini of titin kinase, mimicking the tension that titin is expected to experience during stretch (Figure 1.8). These simulations showed that not only can the

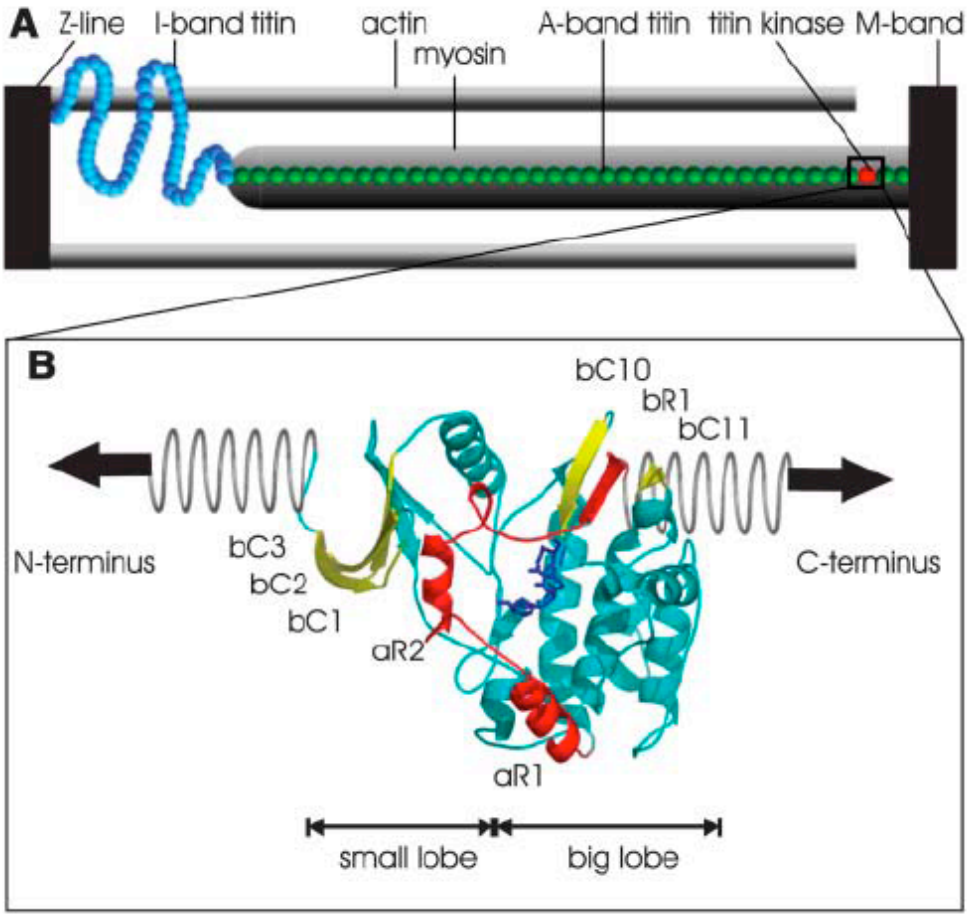


Figure 1.8- Positioning of titin for simulated analysis of force extension. A. Positioning of titin within the sarcomere. The kinase domain (red ball) is hypothesized to be positioned perfectly to behave as a force sensor. B. To study the force induced activation hypothesis, Grater et al (2005) simulated tension with harmonic pulling from end to end (the potential of the pulling symbolized as springs). The catalytic site is shown in blue, the regulatory tail in red, and the terminal β sheets in yellow. The abbreviations help orient the important secondary structures, "a" for α helix, "b" for β sheet, "C" for catalytic, and "R" for regulatory. Figure used with permission from (Grater, Shen et al. 2005).

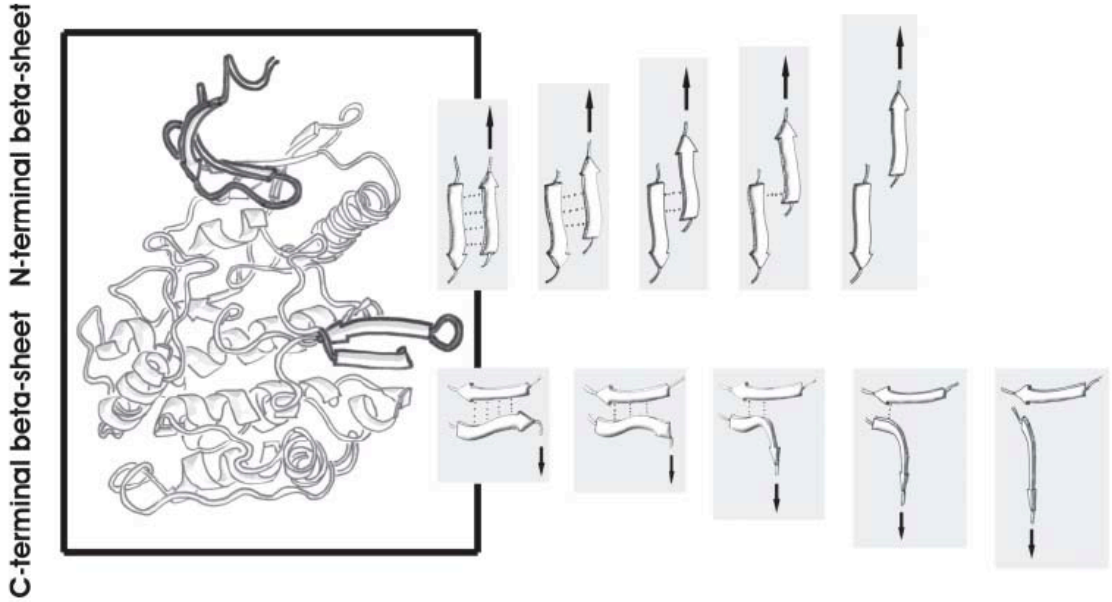


Figure 1.9- Hydrogen bond rupture patterns of force induced unfolding of titin kinase. The simulations predict that the secondary structures of titin kinase are positioned perfectly to respond to forces. When exposed to low, constant, force, the C-terminal β sheet will rupture first, after about 7.5 ns, beginning to relieve the active site of inhibition in order for substrate binding to occur. In contrast, the N- terminal β strands will slowly slide along the longitudinal axis, forming new hydrogen bonds as the original ones rupture, stabilizing the underbelly of the enzyme, and in turn allowing for the structural integrity to be maintained. Rupture of the N-terminal β strands takes 17ns, ~ 3 times longer than the C-terminal strands. Figure used with permission from (Grater, Shen et al. 2005).

kinase withstand expected forces, but that the kinase domain is also positioned in perfect orientation within the "molecular spring" of the titin molecule, making the kinase an ideal force sensor.They conclude that the strain on the autoinhibitory domain leads to an ordered sequence of conformational changes that open the catalytic cleft, while maintaining the structural integrity of the enzyme. This remarkable stability is a function of the orientation of the molecule with respect to the pulling force (Figure 1.9); the pulling geometry of activation is such that the beta sheets most responsible for the force resistance are located parallel to the pulling force, while the beta sheets responsible for exposing the active site are oriented perpendicular to the force. This simple, yet elegant, difference in the orientation of beta-sheets sets the stage for the hypothesis that of the kinase will release the autoinhibitory region from the catalytic core, while maintaining the structural integrity of the active site. The molecular dynamics simulations support a model in which human titin kinase will sense force and pass along the message by phosphorylation of its substrate.

Lange et al (2005) provided additional, cell biological, data to support the force activation mechanism. This intricate work dissected both the substrate interaction between titin and nbr1, and the downstream implications, which includes increasing SRF in the nucleus with consequent increases in SRF-driven gene expression. The group showed that titin kinase interacts with nbr1 through a mechanically inducible conformation. To mimick the mechanically induced intermediate conformation, titin kinase plus the N-terminal 10 amino acids of the regulatory tail was used as bait for a yeast-2-hybrid screen. Nbr1 interacts with this segment of titin kinase, but not with the kinase that has the full regulatory tail, or no regulatory tail, suggesting that the interaction

requires a semi-opened active site and the early portion of the regulatory tail. Using various tissue culture and *in vivo* experiments, the work showed that the mechanical signals (i.e.- mechanically active or mechanical arrest by various agents or denervation) have drastic effects on the localization of MURF2. In mechanically arrested cells, MURF2 concentrates in the nucleus and prevents SRF from activating transcription. In contrast, overexpression of titin kinase has the opposite effect of mechanical arrest, decreasing the overall cellular amount of MURF2 and increasing the expression of SRF-driven genes. Further, an arginine to tryptophan mutation in the fourth amino acid of the regulatory tail that abolishes the interaction between titin and nbr1, and mislocalizes many of the proteins involved in the signaling pathway, causes a severe human myopathy that often results in death.

The simulation studies of titin kinase, accompanied by the implications of mechanical signals on cellular physiology, support a force activation hypothesis. One major experiment that begs to be completed is the direct observation of the giant kinase sensing force. Using single molecule studies, we believe that this data can be obtained. We have begun to test the force activation hypothesis using atomic force microscopy, of which the details are discussed below (and further in chapter 2).

Molecular Force Spectroscopy

The study of single molecules using Molecular Force Spectroscopy

Biological systems are highly complex, even when starting at the atomic level. As the system increases in size, the layers of complexity build, and thus biologists have quite a challenge deciphering cellular mechanisms. One way to understand the individual

components of a molecular pathway is to study the macromolecule of interest in isolation, purified. To simplify even further, the purified molecules can be studied one at a time, using single molecule techniques (Oberhauser and Carrion-Vazquez 2008). Single molecule studies allow biophysical and structural characterization of macromolecules by "feeling" and manipulation of the native conformation (Altmann and Lenne 2002). This approach is of most interest to proteins involved in exerting mechanical forces. It is easy to often overlook the vast cellular processes that require mechanical forces, but these forces are critical for mechanisms as basic as DNA replication and as complex as the building of cells into organs. Thus, by probing bond dynamics in real time, under psuedophysiological conditions, the mechanical thresholds of a single protein can be studied.

The elegant design of the molecular force spectroscope (MFS) allows biologists and physicists to gain an understanding of the mechanical properties of a protein (Altmann and Lenne 2002). The instrument uses a piezo electric positioner, a cantilever tip, and a laser/split diode detector positioned to measure deflection of the cantilever (Figure 1.10). A slide (the material of the slide can vary, but is usually glass) is mounted on the piezo electric positioner, and a dilute protein solution is adsorbed onto the slide. Piezo electric materials have the ability to smoothly expand at the nanometer scale upon application of electric voltage. The piezo electric positioner moves at a fixed, calibrated speed and gently brings the protein solution into contact with the cantilever tip. The attachment process is random, and so, by chance, a fraction of the events will result in the correct alignment of a single protein molecule attached to both the cantilever tip and the glass slide. Regardless of the attachment, the glass slide will move away from the

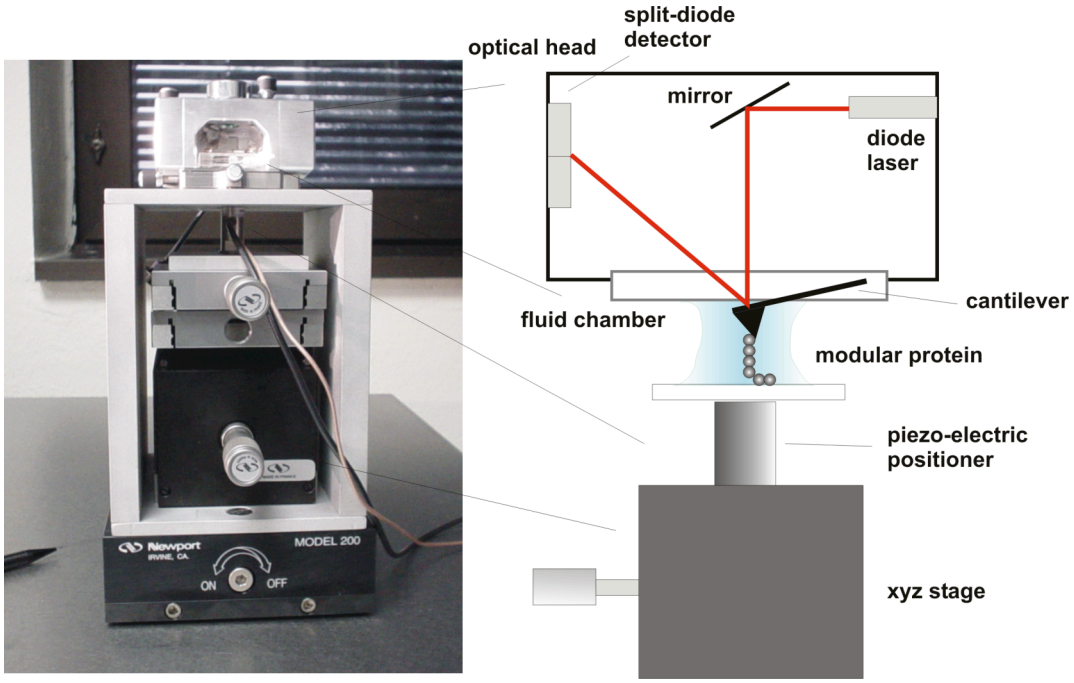


Figure 1.10- Components of an MFS.

cantilever at a constant speed. In the case of a "correct" attachment, as the glass slide moves away from the tip, the protein begins to stretch (extension) and the distance traveled is proportional to the movement of the piezo. Once the protein is stretched out to its secondary structure (or individual domains) tension builds and the molecule becomes taut, pulling the cantilever towards the piezo, increasing the deflection. As the piezo continues to move away from the cantilever, the protein can no longer withstand the pull, and the non-covalent bonds rupture. The rupture causes a relaxation of the cantilever and hence, the deflection also relaxes, abruptly moving the laser back to the starting position. In practice, proteins or fragments of proteins containing several domains are probed. The domain with weakest total intramolecular bonds unwinds first. The piezo continues to move away and the unfolded domain is extended into its primary structure. When the next domain is reached, the molecule will once again to become taut, and then this domain ruptures, relaxing the cantilever. Eventually, all of the domains have ruptured and the protein is in an extended structure. The final bit of force that the instrument will detect is the detachment of the protein from either surface (the strength of these non-specific bonds are much stronger than the protein domains due to the large entropic contributions). The pull generates a force-extension curve; the force required to extend the macromolecule into its primary structure. This process is visually described in Figure 1.11.

Molecular Force Spectroscopy and titin

Titin (or at least the domains of titin) has become a model system for single molecule experiments (Linke and Grutzner 2008). First, the intricate domain structure

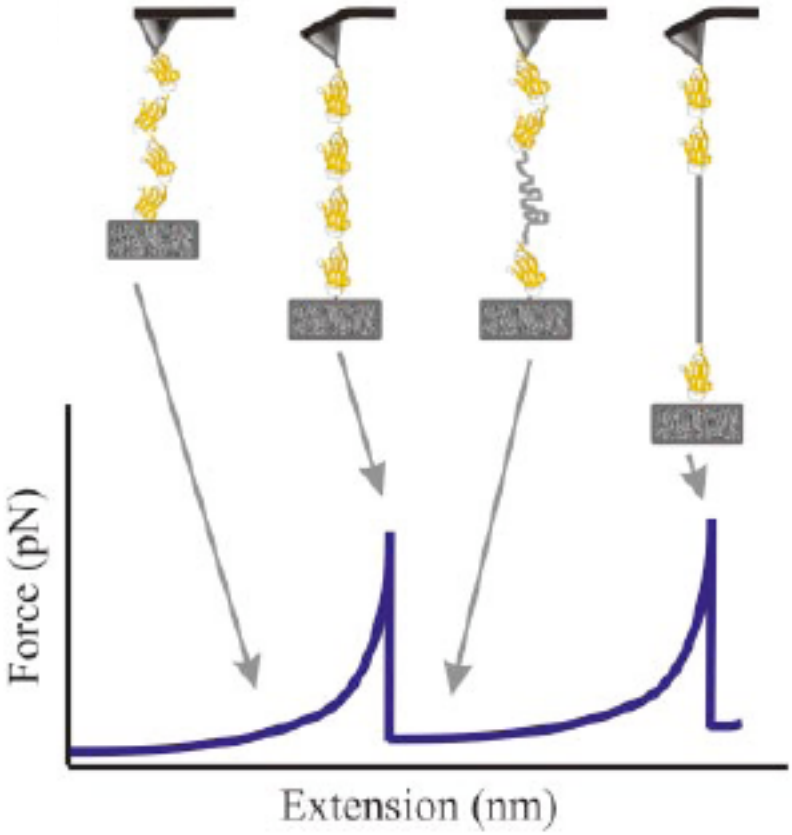


Figure 1.11- Force extension curves model the unfolding of domains within a single protein. Cartoon depicting the unfolding events that occur when force is applied to a protein with multiple domains. The sequential unfolding of the domains results in a saw-tooth pattern, and provides insight into the mechanical properties of the protein of interest. Figure used with permission from (Oberhauser and Carrion-Vazquez 2008).

that provides both passive elasticity and structural integrity makes it an excellent candidate for understanding the different mechanical properties within a single polypeptide. In addition, the linear pulling geometry of an MFS or optical trap adequately mimics the end-to-end pulling expected to occur physiologically in the sarcomere. The main implications of these studies contribute to understanding of protein unfolding/folding and understanding the intrinsic elasticity of mechanical proteins.

A central problem in biology is describing the complicated, yet usually spontaneous, path that leads a macromolecule to its native structure. MFS experiments have been successful in studying the unfolding/folding pathway of titin Ig domains, showing that refolding occurs spontaneously, without the aid of chaperones (Rief, Gautel et al. 1997; Rief, Gautel et al. 1998). The protein (typically tandem Ig domains) is attached to the cantilever tip, and as the piezo moves away from the tip, the protein is stretched. Before detachment of the protein from the surface, the stretch is relaxed and the protein is given a time interval to refold. After the time interval, the protein is again stretched, and the number of refolded domains can be measured. The measurements provide the rate of refolding, which can be calculated by the percent of domains that have refolded over the time interval. The proper, spontaneous refolding of the Ig domains suggests that, in vivo, the Ig domains may behave as shock absorbers; if the macromolecule experiences great force the Ig domains can "take the hit" without compromising the integrity of the sarcomere. Further, the Ig domains can refold in seconds even when experiencing 25-30 pN of force, suggesting that the sarcomere does not have to be fully relaxed for spontaneous refolding to occur.

Adjusting the passive stiffness of muscle tissue is of great interest to cardiologists, as such modulation occurs naturally, in both fetal/perinatal development and in heart disease. A contributing factor to this physical parameter is the expression of specific titin isoforms (Fukuda, Wu et al. 2005; Lim and Sawyer 2005). The main regions thought to provide elasticity to the titin filament are the PEVK and N2 domains, both located in the I band region of the sarcomere (Linke, Ivemeyer et al. 1998; Linke, Stockmeier et al. 1998; Li, Oberhauser et al. 2001; Leake, Grutzner et al. 2006; Granzier and Labeit 2007). The length of these regions is variable, depending on the type, age, and health of the muscle tissue (Granzier, Labeit et al. 2003). For example, the PEVK region can vary from 163 residues in titin isoforms of stiff cardiac muscle to 2174 residues in some compliant skeletal muscles. To better understand the elastic properties of PEVK and N2 the unique sequences were engineered with flanking tandem Ig domains (Li, Oberhauser et al. 2001; Watanabe, Nair et al. 2002; Leake, Grutzner et al. 2006). The Ig domains provide a mechanical "fingerprint" as their stretch-unfolding pattern is well established, and thus allows for the researcher to identify the stretched, single molecules. These data revealed that the elastic regions extend at low forces, where the probability of Ig unfolding is still very low. Thus, the elasticity of a titin filament is conferred by modular unique sequences, but the design of the giant macromolecule has many features that contribute to its physiological functions.

These data and others have led to a model for the behavior of the titin filament when exposed to tension or stretch (Linke and Grutzner 2008). At rest, there is slack in the sarcomere, and titin assumes a tertiary/quaternary structure that includes interdomain interactions. At low stretch forces the titin molecule begins to lengthen, with the distal

and proximal Ig regions straightening first. With continued stretch, the unique sequences extend (these include the N2 and PEVK regions). The specific isoform, which will vary in i) how many tandem Ig domains are present, ii) how long the PEVK region is, and iii) the type of N2 region present, will determine the stretch force resistance of the macromolecule. Lastly, and only under extreme tension, the Ig and Fn domains unfold.

Single Molecule Techniques to Support the Force Activation Mechanism of Giant Protein Kinases

The activation mechanism of titin and the other giant protein kinases has proven a difficult biological problem to solve. As described earlier, unlike their MLCK relatives, the activation of the giant kinases can only be partially explained through the binding of calcium binding proteins like calmodulin or S100. We, and others, hypothesize that the giant protein kinases are activated by force; the forces generated from the contraction/relaxation cycles of muscle activity are necessary and sufficient to unleash the enzymes of their intrasteric regulatory tail, exposing the catalytic site and permitting catalysis (Grater, Shen et al. 2005; Lange, Xiang et al. 2005; Greene, Garcia et al. 2008). Until recently, this hypothesis has not been supported experimentally.

MFS is a powerful tool for understanding the mechanical stability of a protein (Oberhauser and Carrion-Vazquez 2008). Typically, MFS experiments have focused primarily on cytoskeletal and adhesion proteins, exploiting the strength that these proteins have understandably evolved to possess. We decided to test the mechanical resistance of the kinase domains of *C. elegans* twitchin and TTN-1 kinase using MFS (Greene, Garcia et al. 2008). In the native polypeptides, the kinase domains are flanked by Fn and Ig

domains. Ig and Fn domains are similar ~100 amino acid structures, composed of 2 βsheets that form a sandwich (Kobe, Heierhorst et al. 1996; von Castelmur, Marino et al. 2008). Various MFS studies have characterized the force-induced unfolding of Ig and Fn domains (Rief, Gautel et al. 1997; Rief, Gautel et al. 1998; Rief, Gautel et al. 2000; Oberhauser, Hansma et al. 2001; Li, Linke et al. 2002; Li and Fernandez 2003, Oberhauser, 2002 #148; Bullard, Garcia et al. 2006), assigning them a molecular "fingerprint" that is easily identified by force-extension analysis. We therefore used a recombinant portion of the giant kinases containing the Fn, kinase, and Ig domains for MFS analysis (Greene, Garcia et al. 2008). Our results were the first to characterize the force-induced unfolding of kinase domains. The data revealed that, indeed, the kinases are mechanically resistant (relative to proteins that are not expected to withstand applied forces, such as barnase (Best, Li et al. 2001) and apo-calmodulin (Hertadi and Ikai 2002)) and that the kinases unfold in a step-wise manner, with the smaller lobe rupturing first, followed by the larger lobe. These results indicate that giant kinases are designed to withstand forces. Further, although below the range of detection for the MFS, but supported by our simulation studies, we believe that the regulatory tail dislodges at a much lower force than required to unfold the small lobe. The implications of these results are quite exciting: the kinase may act as a force sensor! For details of these results, please refer to chapter 2.

The present resolution (> ~20pN) and lack of spatial manipulation of the MFS prohibits controlled pulling of a single kinase molecule at forces low enough to be physiologically similar to what we expect the domain to experience during muscle contraction. Thus, the experiment that directly tests if the application of low forces on

the kinase results in accelerated substrate binding or turnover must be done using alternative instruments. We predict that using molecular tweezers to study the effects of small forces on catalysis will prove fruitful. Intricate protein engineering, a strong understanding of fluorescence, and experience with the biophysical instrument are all crucial factors in completing the desired task. In collaboration with several other labs, we will assemble this team of scientists and answer the important question: can the giant protein kinases sense force? (please see chapter 4 for further discussion).

Identification of substrates for protein kinases

Kinase/Substrate Interactions

The importance of protein kinases in the regulation and maintenance of healthy cellular pathways is now well established. Unfortunately, one major roadblock in understanding the mechanisms that contribute to the functions of protein kinases is identifying substrates. Many sophisticated techniques have been developed to focus on a single type of kinase, exploiting their individual properties to select for cellular substrates. Such assays include using phospho-motif antibodies and phospho-motif-binding proteins to identify the substrate targets *in vivo*. Other, more generic, assays include biochemical approaches: beginning with cellular extract and finding a "band on a gel" that is phosphorylated upon incubation with a purified or recombinant kinase and using classical purification techniques to eventually identify a single protein substrate; co-immunoprecipitations; in gel kinase assays; or affinity chromatography. Genetic approaches can also be useful, but these can cause complications in elucidating direct interactions versus effects of upstream or downstream signaling (for more detail on the

techniques listed above, please see the following reviews (Berwick and Tavare 2004; Elphick, Lee et al. 2007)). We chose to use the yeast-2-hybrid approach, which combines both genetics and biochemistry to screen for proteins that interact with TTN-1 kinase.

Yeast-2-Hybrid as a Method to Identify Kinase Substrates

The yeast-2-hybrid system is a genetic/biochemical technique that typically allows for the identification of protein-protein interactions (Fields and Song 1989). Invented in1989, yeast-2-hybrid screens have become an indispensable laboratory tool for helping to explore known interactions and to discover new protein-protein interactions. The system is conceptually and practically quite simple, exploiting the ease of genetic manipulation in the single celled eukaryote, *Saccharomyces cerevisiae*. As the name implies, the premise of the screen is two hybrid proteins. In one version of the two hybrid, a protein of interest is fused to the GAL4 DNA binding domain (the bait), and a second protein is fused to the GAL4 activating domain (the prey). If the bait and prey proteins interact, the GAL4 transcriptional activating protein will translocate to the nucleus and activate transcription of a reporter gene containing an upstream GAL4 binding site. If the proteins do not interact, the GAL4 activation and DNA binding domains will not be brought into proximity and the reporter genes will not be turned on. The hybrid proteins are expressed on two separate plasmids, allowing for i) the screening of libraries and ii) the quick isolation of the clone corresponding to the unknown interacting protein.

Although the researchers intuition might insist that yeast-2-hybrid interactions cannot identify a substrate interaction (due to the often transient nature of

enzyme/substrate binding) previous studies have identified enzyme substrates using this approach. In fact, one of the first studies to use the yeast-2-hybrid method to identify protein-protein interactions identified a kinase/substrate interaction, promoting the utility of the assays as applicable to identifying many types of intracellular contacts (Yang, Hubbard et al. 1992). Although it is not always fruitful in identifying positive interactions, and it is even more difficult to identify a substrate interaction, some very prominent kinase/substrate interactions have been discovered using the method. For example, the interaction between vertebrate titin kinase and nbr1 was first identified using the yeast-2-hybrid (Lange, Xiang et al. 2005). As discussed earlier, this interaction yielded scientific insight into both the applied and basic aspects of titin, by elucidating the cause of a hereditary human myopathy and also providing support for the force-activation hypothesis of giant protein kinases.

Using the yeast-2-hybrid approach we screened for proteins that interact with TTN-1 kinase. We were successful in identifying a kinase/substrate interaction between TTN-1 kinase and the Cdc42 guanine exchange factor (GEF) UIG-1. These results provide the foundation for the identification of the first substrate of any of the giant protein kinases in *C. elegans*. Because of the giant kinases' high sequence and structural homology to the MLCKs, it was thought that their substrates might be similar. Our data, instead, links the giant protein kinase TTN-1 to GTPase signaling, and thus possibly providing a direct link between TTN-1 activation and the integrity of the cytoskeletal architecture (please see chapter 3 for further details).

Rho Family of GTPases

Biological Significance

The building and maintenance of cells, either individually or into tissues, requires many forms of regulation. We have already discussed the implications of phosphorylation (via kinases) in the transmission of various signals that contribute to cellular homeostasis, and we will now introduce GTPase signaling, which constitutes another fundamental regulatory system in all eukaryotic cells (Bos, Rehmann et al. 2007). The signals transmitted by GTPases influence gene expression, differentiation, cell division, motility, vesicle transport, cytoskeletal integrity, nuclear transport, and various other essential cellular processes (Bos, Rehmann et al. 2007, Bernards, 2007 #179). There are currently five general groups of GTPases responsible for these activities: heterotrimeric G proteins (membrane bound, typically sense environmental ligands such as hormones or other membrane diffusible small molecules), initiation/elongation factors (translation), signal recognition particle/signal receptor (ER translocation), tubulins and cytoskeletal motors, and the Ras superfamily of monomeric, small G proteins. All of the GTPases have a common catalytic core, capable of binding and hydrolyzing GTP. GTP hydrolysis triggers conformational changes and the specific conformation of the GTPase directly affects cellular events. The Ras superfamily further breaks down into the Ras, Rho, Rab, Arf, and Ran subfamilies (Wennerberg, Rossman et al. 2005). The members of each subfamily share functional relationships (rather than sequence homology) from vesicle transport to nuclear import. This thesis will focus mainly on the Rho subfamily of Ras superfamily, which includes Cdc42 and is primarily responsible for cell shape, cytoskeletal dynamics, and cell migration. Although the GTPase families are segregated

for simplicity purposes, many of the basic features described below are common among the GTPase phylogeny.

GTPases On/Off switch: GEFs and GAPs

Contrary to what the name implies, GTPases are not very efficient at hydrolyzing GTP (k_{cat}~0.0002-0.025s^{-1}(Higashijima, Ferguson et al. 1987; Huang, Surka et al. 2006), compared to an average kinase k_{cat}~ 50s^{-1} (Jetten, Gubler et al. 1994; Zhang, Zhang et al. 2006; Lee, Mwongela et al. 2008)). Regulatory accessories provide the basis for the precise regulation of GTP binding and hydrolysis (Bos, Rehmann et al. 2007). Equally important to the accurate regulation of GTPases is the cellular concentration of GTP, which is much greater (~10-fold) than the GDP concentration. At rest, the GTPase is GDP bound, and considered inactive or "off"; during active signaling, the GTPase is GTP bound, and considered "on" (Figure 1.12). The GTPase cycles between the on/off states depending on the needs of the cell. The exchange and hydrolysis of GDP/GTP is controlled by two lineages of multidomain proteins, guanine exchange factors (GEFs) and guanine activating proteins (GAPs) (Figure 1.12).

GTPases have a very high affinity (low nM/high pM) for the guanine nucleotide, regardless of whether there are two or three phosphates (Manne, Yamazaki et al. 1984). Thus, the k_{off} or rate of dissociation is very slow, taking hours to release the nucleotide spontaneously. GEFs accelerate the release of the nucleotide from the GTPase by orders of magnitude (Bernards and Settleman 2007). In general, the GEF acts as the sensor of biological signals, and it transmits this signal by modulating its catalytic activity (i.e.- GDP release) for the specific GTPase. GEFs are highly specific, and their domain

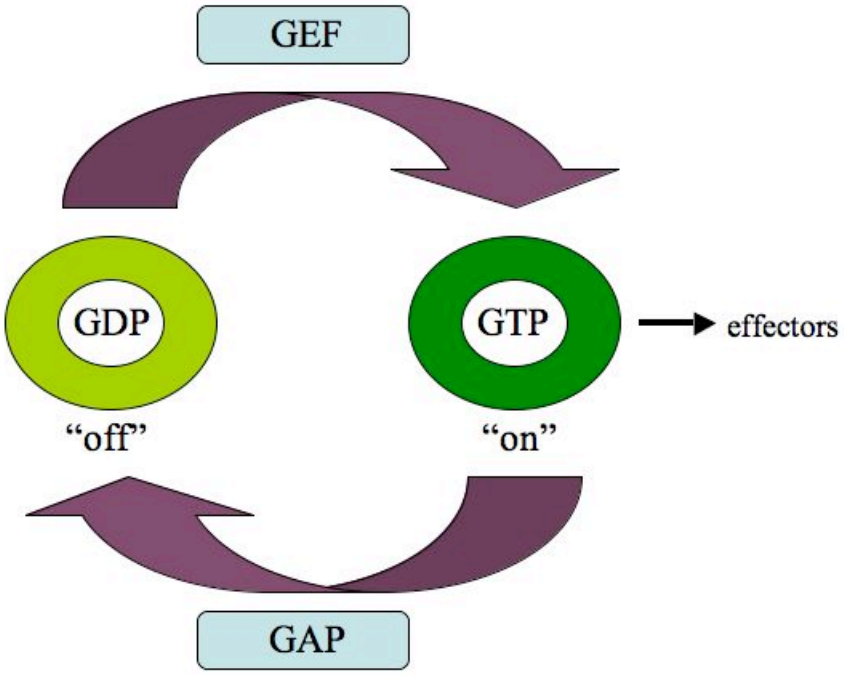


Figure 1.12- Activation cycle of GTPases. GTPases exist in two main states, GTP bound (dark green circle) or GDP bound (light green circle). When the GTPase is bound to GDP, it cannot interact with effector molecules and is therefore considered to be off. To exchange the GDP for GTP, a guanine exchange factor (GEF) promotes the disassociation of GDP, and by virtue of the high concentration of GTP in the cell, the GTPase re-associates with the tri-phosphate form of the guanine nucleotide. A GTPase with GTP bound signals to effectors, deeming it active, or on. To return to the dormant state, a GTPase activating protein (GAP) facilitates the hydrolysis of GTP.

structure varies depending on the GTPase family they regulate. For the Rho family of GTPases, the GEFs tend to have a tandem arrangement of dbl (DH)/plekstrin homology (PH) domains (Bos, Rehmann et al. 2007). GEFs are not usually promiscuous, they recognize and regulate one or few GTPase, and will have no catalytic effect on the others (Nassar, Cancelas et al. 2006; Garcia-Mata and Burridge 2007). Of course, there are some examples of GEFs that have multiple GTPase substrates (Olson 1996), but the inferred specificity can also vary depending on if the study was done *in vitro* (which tends to promote non-specific activation) or *in vivo*. GEF binding causes conformational changes in the phosphate region of the nucleotide-binding site, in turn decreasing the affinity of the GTPase for the nucleotide (Vetter and Wittinghofer 2001). Once the GDP is released, rebinding is almost exclusively limited to GTP, as a result of the relative intracellular concentration of the nucleotides. Nucleotide re-binding results in the release of the GEF, and an active GTPase.

Turning the GTPase off requires hydrolysis of the GTP to GDP. As mentioned earlier, the k_{cat} for this reaction is very slow. GAPs assist the GTPase in overcoming this catalytic limitation by providing spatial and temporal control over GTP hydrolysis (Bernards and Settleman 2004; Bernards and Settleman 2005). Similar to GEFs, GAPs have minimal sequence and/or structural similarity in their primary or tertiary structures, respectively. GAPs also tend to be highly specific and recognize only one member of a GTPase subfamily. GAPs influence catalysis in two major ways. First, they stabilize the GTPase active site in a conformation that is most complementary to the pentacoordinate transition state of the γ phosphate during GTP hydrolysis (Sprang 1997). Second, they insert a catalytic group *in trans*. The catalytic group is typically a basic residue that

activates water for S_N2 nucleophilic attack. The GTP is hydrolyzed by direct in-line transfer of the γ phosphate from GTP to the activated water.

Regulation of Guanine Exchange Factors

GEF regulation is complex and despite the advances that have been made thus far, it is becoming increasingly clear that developing an in-depth understanding of GEF regulation will require focused attention on the individual members of the family (Schmidt and Hall 2002). A general rule of thumb is that the "default" state for a particular GEF is inactive, and that biochemical signals provide the cues that result in an active conformation (Bernards and Settleman 2007; Bos, Rehmann et al. 2007). Most (possibly all) GEFs are autoinhibited at rest and require activation to relieve the active site of this inhibition. Further, in addition to their catalytic domain, GEFs contain multiple functional domains that can contribute to their regulation. Although there is clear diversity between the regulation of different GEFs, the general principles of activation include: protein/protein interaction, protein/lipid interactions, second messenger binding, and post-translational modifications. Mechanistically, these cellular events lead to GEF activation by releasing the intramolecular inhibitory sequences (either providing allosteric changes in the active site or physically removing an intrasteric domain that blocks the active site) and/or translocation to an alternative subcellular location.

Negative regulation of GEFs also exists, but is currently less understood relative to activation (Schmidt and Hall 2002). One recurring theme is targeted GEF degradation via the ubiquitin proteasome pathway (Pham and Rotin 2001; Kamynina, Kauppinen et

al. 2007; Yamaguchi, Ohara et al. 2008). However, other examples show that monoubiquination can stimulate GEF activity (Davies, Topp et al. 2003). Yet another inconsistent theme occurs with GEF phosphorylation. Typically, phosphorylation will positively influence GEF activity (Crespo, Schuebel et al. 1997; Han, Das et al. 1997; Salojin, Zhang et al. 1999; Kato, Kaziro et al. 2000; Yamauchi, Chan et al. 2005; Zhang and McCormick 2007; Fujishiro, Tanimura et al. 2008), but it has also been shown to have a negative or no effect on GEF activity (Albright, Giddings et al. 1993; Birkenfeld, Nalbant et al. 2007; Chahdi and Sorokin 2008). The seeming contradictions and intricate mechanisms in the regulation of GEFs inspire us to understand the substrate interaction between TTN-1 and the Cdc42 GEF UIG-1. Our studies will continue to pave the way towards a better understanding of the complex regulation of GEFs, and hence, the regulation of essential cellular pathways.

Summary

Ultimately, our lab strives to understand the mechanisms involved in building and maintaining the sarcomere. Using *C. elegans* as a model system, this thesis provides both biophysical and biochemical data to further our conceptual understanding of the conserved giant muscle protein kinases. Molecular force spectroscopy allowed us to investigate the force-induced unfolding of the giant protein kinases, twitchin and TTN-1. These data provided some of the first experimental evidence to support that the giant protein kinases may act as force sensors. Yeast-2-hybrid screening identified an interaction between TTN-1 kinase and the Cdc42 GEF UIG-1; further investigation indicated that UIG-1 is a substrate for the kinase. This interaction represents the first

substrate identified for any of the giant protein kinases in *C. elegans*, and adds to the very short list of known substrates for the giant kinases in any animal. Further, the interaction between UIG-1 and TTN-1 may bridge the activation of TTN-1 to cytoskeletal rearrangements, providing a novel role for this giant protein. Together, this work contributes to the global understanding of muscle biology at the molecular level.

CHAPTER TWO

Single-Molecule Force Spectroscopy Reveals a Stepwise Unfolding of *C. elegans* Giant Protein Kinase Domains

(Most of the contents of this chapter were published in Greene et al. Biophysical Journal. 95(3):1360-70)

INTRODUCTION

All animals have complex mechanisms to regulate their overall muscle mass. The functional unit of a muscle cell is the sarcomere, a miniature machine comprised of overlapping, interacting filaments. During muscle activity, the sarcomeres act in unison, undergoing rapid contraction/relaxation, transmitting work throughout the muscle. The three filaments that give the sarcomere its integrity are thin filaments, thick filaments, and titin filaments. These filaments, together with their multi-protein attachment structures, the M-lines and Z-disks, must undergo assembly and disassembly during myofibril growth and maintenance. Although progress has been made in understanding this intricate system, the precise mechanisms that control these assembly/disassembly processes remain central questions in muscle biology. Data presented by Lange et al. (2005) is consistent with a model in which signaling from mammalian titin kinase leads to the building of myofibril (Lange, Xiang et al. 2005). When in its active conformation, titin kinase transmits a phosphorylation cascade that ultimately leads to the expression of genes involved in myofibril assembly. Specifically, titin kinase phosphorylates nbr1, resulting in less MuRF2 and more SRF in nuclei; SRF directly increases myofibrillar gene expression. However, recently, it has been shown that deletion of MuRF2 in mice has no affect on the cellular distribution of SRF (Witt, Witt et al. 2008). Moreover, both knockout and conditional mutants deficient for the titin kinase function have been studied (Gotthardt, Hammer et al. 2003). In these murine models, normal myofibrils are assembled, but either appear to have a defect in proper maintenance or have impaired

contractile regulation. Thus, the physiological functions for titin kinase are still unknown. What we wish to elucidate is how the normally autoinhibited titin kinase is activated.

In *C. elegans* there are two titin-like proteins, twitchin and TTN-1. These giant muscle proteins, like those found in other species, are comprised primarily of multiple copies of immunoglobulin (Ig) and fibronectin type-III (Fn) domains (Ferrara, Flaherty et al. 2005). Twitchin (~800kDa), named for its characteristic twitching mutant phenotype, has 30 Ig domains, 31 Fn domains, and a single kinase domain (Benian, Kiff et al. 1989; Benian, L'Hernault et al. 1993). Upon its discovery, twitchin was hypothesized to be an important regulator of muscle contraction, mainly due to its motility phenotype and MLCK-like kinase domain. Later, its role in inhibiting the rate of relaxation was identified in studies of twitchin from *Aplysia* and *Mytilus* (Probst, Cropper et al. 1994; Siegman, Funabara et al. 1998). TTN-1 is significantly larger than twitchin, with a molecular weight of 2.2 MDa. In addition to 56 Ig domains and 11 Fn domains, TTN-1 has several regions predicted to be coiled-coil, and two regions consisting of tandem repeats (Flaherty, Gernert et al. 2002). The largest of these tandem repeat regions, called PEVT, is similar in amino acid composition and tandem repeat structure to the main elastic element of vertebrate titin called the PEVK region. Thus, this region of nematode TTN-1 is hypothesized to be elastic. Similar to mammalian titin kinase, TTN-1 and twitchin both have a single kinase and regulatory domain located near the C-terminal end of the giant molecule (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002). In the primary sequence, the inhibitory/regulatory region is located just downstream of the catalytic core (Figure 2.1A & B) and in 3-D space (Figure 2.2A & B) the domain is wedged in between the two subdomains of the catalytic core, making extensive contact

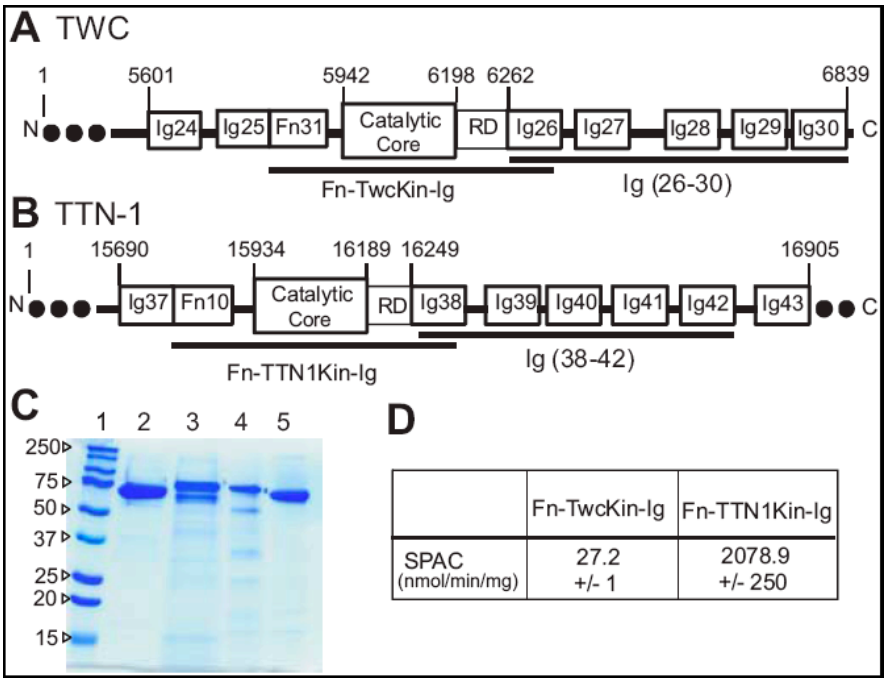


Figure 2.1- Expression, purification, and enzyme activity of recombinant kinase domains and tandem Ig domain segments from *C. elegans* giant proteins. A, B) Twitchin and TTN-1 recombinant protein constructs for MFS experiments. The kinase constructs have an autoregulated kinase domain flanked by Fn and Ig domains. The Ig constructs consist of 5 tandem Ig domains that immediately follow the kinase domain. The numbers denote the position of the amino acid in the full-length polypeptide. C) 12% SDS PAGE stained with coomasie brilliant blue shows that the proteins were isolated to > 95% purity. In each lane, 2 micrograms of protein was loaded. The lanes are as follows: 1) molecular weight standard 2) Fn-TwcKin-Ig, 3) Fn-TTN1Kin-Ig, 4) Twc Ig 26-30, 5) TTN-1 Ig 38-42. D) The purified TTN-1 and twitchin kinases retain phosphotransferase activity in vitro. Using ^{32}P-ATP and a model peptide as substrates the specific activity values (SPAC) obtained are similar to those previously published (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002). Enzyme activities were determined in triplicate on freshly purified protein.

with the active site, blocking substrate entry (Hu, Lei et al. 1994; Kobe, Heierhorst et al. 1996). The kinases with the highest homology, the vertebrate smooth muscle and non-muscle myosin light chain kinases (MLCKs), are also autoinhibited by a sequence just C-terminal of the catalytic core. In the case of the MLCKs, binding of the autoinhibitory sequence to Ca^{+2}/calmodulin causes a conformational change sufficient to allow access to its substrate (Wilmann, Gautel et al. 2000). Although the giant kinases can also bind to Ca^{+2}/calmodulin, this alone is not sufficient to activate the enzymes (Heierhorst, Kobe et al. 1996). In developing muscle, the combination of Ca^{+2}/calmodulin and phosphorylation of a key tyrosine residue can activate vertebrate titin kinase (Mayans, van der Ven et al. 1998). However, the required tyrosine phosphorylating activity is not found in mature muscle. Furthermore, this tyrosine is not conserved: in the analogous position, in twitchin it is alanine (Mayans, van der Ven et al. 1998), and in TTN-1 it is cysteine (Flaherty, Gernert et al. 2002). Thus, to date, the precise mechanism(s) resulting in the conformational changes that relieve the giant kinases of their autoinhibition remain a mystery.

One hypothesis is that the giant kinases may act as force sensors; that the forces generated during the contraction/relaxation cycles of muscle activity are sufficient to unleash the regulatory domain from the catalytic core (Grater, Shen et al. 2005). Thus, after experiencing a certain threshold of force, the kinase would become active. Grater et al. (2005) used molecular dynamics simulations to pull the human titin kinase from its amino and carboxy termini to simulate the strain the molecule would undergo during muscle activity (Grater, Shen et al. 2005). The authors found that not only can the kinase withstand expected forces, but that the kinase domain is also positioned in perfect

orientation within the "molecular spring" of the titin molecule, making the kinase an ideal force sensor. They conclude that the strain on the autoinhibitory domain leads to an ordered sequence of conformational changes that open the catalytic cleft, while maintaining the structural integrity of the enzyme. This remarkable stability is a function of the orientation of the molecule with respect to the pulling force; the pulling geometry of activation is such that the beta sheets most responsible for the force resistance are located parallel to the pulling force, but the beta-sheets responsible for exposing the active site are oriented perpendicular to the force. This simple, yet critical, difference in the orientation of beta-sheets sets the stage for the ability of the kinase to simultaneously remove the autoinhibitory region from the catalytic core, while maintaining the structural integrity of the active site. The molecular dynamics simulations support a model in which human titin kinase will sense force and pass along the message by phosphorylation of its substrate.

In order to begin to test this hypothesis experimentally, we recombinantly expressed twitchin and TTN-1 kinases, the human titin kinase homologs, from *C. elegans*. Using single-molecule molecular force spectroscopy (MFS) we analyzed the mechanical strength of the kinase and their flanking Ig/Fn domains, along with a tandem repeat of five Ig domains that immediately follow the kinase domains in the primary structure. Our results show that these kinase domains have remarkably high mechanical stability. The kinase domain unfolds at a force range of ~30-150pN, a value that is similar to the unfolding range of nematode Ig and Fn domains (40-180pN). Further, in contrast to the Ig/Fn domains, which unfold in an all-or-none highly cooperative fashion, the unfolding of the kinase is sequential, first an unwinding of the autoinhibitory region,

followed by the biphasic rupture of the catalytic core. These data provide support for the hypothesis that the kinase domains of the giant muscle proteins function as effective force sensors.

RESULTS

3D structures of C. elegans twitchin and TTN-1 kinase domains

Crystal structures for both *C. elegans* and *Aplysia* twitchin kinases (Hu, Lei et al. 1994; Kobe, Heierhorst et al. 1996) and human titin kinase (Mayans, van der Ven et al. 1998) have been previously solved. However, the structure of TTN-1 kinase is unknown. In the interest of developing a better understanding of how the structural similarities among the enzymes might relate to their activation mechanism(s), we developed a homology model of TTN-1 kinase catalytic core and autoinhibitory region (Figure 2.2B). Overall, the predicted structure is very similar to the *C. elegans* twitchin kinase. The catalytic region has 2 lobes, a smaller, beta rich lobe, and a larger alpha helical lobe (Figure 2.2B, grey and light green, respectively). The regulatory tail (in red) wraps between the 2 lobes, nesting itself into the active site. This inhibited conformation would be maintained as the native structure during catalytic arrest. When the muscle cells require the catalytic activity of the giant protein kinases, the regulatory domain must be removed for optimal catalysis. Closer examination of how the regulatory domain interacts with the catalytic core reveals secondary structure elements that are believed to be the fundamental basis for the force activation hypothesis (Grater, Shen et al. 2005).

The molecular dynamic simulations by Grater et al. (2005) predicted the N and C terminal beta sheets (Figure 2.2B, dark blue and dark green plus red beta strand,

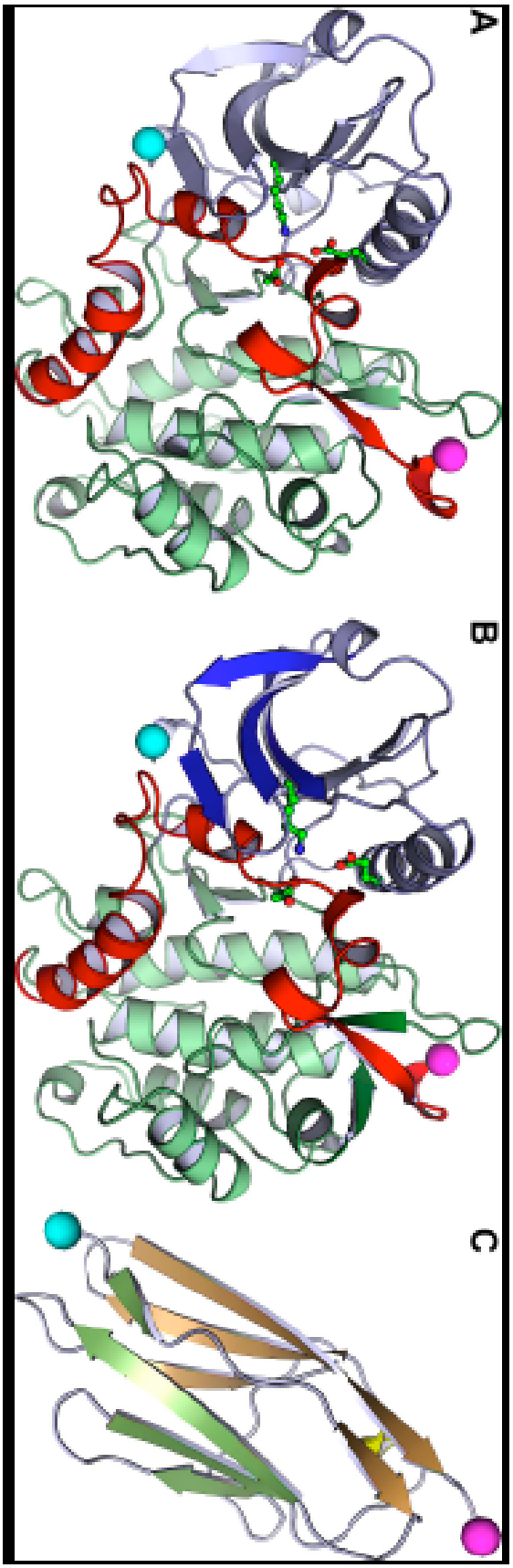

A
B
C

Figure 2.2- 3D structures of *C. elegans* twitchin Ig and kinase domains and homology model for TTN-1 kinase. A) Twitchin kinase is composed of 3 subdomains: an alpha-helical rich large lobe (green), a small lobe of mainly beta sheets (grey), and the autoregulatory tail (red) (structure taken from (Kobe, Heierhorst et al. 1996)). The autoregulatory tail is situated between the 2 lobes making extensive contact with the active site. The active site has a set of conserved residues that help neutralize the ATP binding pocket. These are shown in a ball and stick interpretation. B) 3D homology model of TTN-1 kinase. The N-terminal beta sheet (strands betaC1-C3) predicted to be parallel to the pulling force is colored dark blue. This sheet is thought to stabilize the kinase upon activation by force. The C-terminal beta sheet is predicted to be perpendicular to the pulling force and is composed of betaC10 (colored dark green), betaR1 (colored red), and betaC11 (colored dark green) strands. The autoinhibitory domain is colored red and includes the betaR1 strand and alphaR1-R2 helices. Shown in ball and stick representation are key residues from the N-terminal beta strands interacting directly with the ATP binding pocket of the active site. C) Twitchin Ig domain 26, which immediately follows the twitchin kinase domain (structure taken from (Kobe, Heierhorst et al. 1996)). The two beta sheets, characteristic of an Ig fold are depicted in green and brown. All three structures have their N terminal alpha carbons marked with cyan spheres and their C terminal alpha carbons marked with magenta spheres.

respectively) to be the primary mechanical elements responsible for the force resistance (Grater, Shen et al. 2005). Similar to the known crystal structures, the homology model of TTN-1 kinase maintains the alignment of these important substructures. The C-terminal beta sheet consists of 3 beta strands: betaC10, betaC11, and betaR1 ("C" for catalytic; "R" for regulatory) (Figure 2B; dark green plus red beta strand). Grater et al. (2005) showed that during activation, the C-terminal beta sheets are positioned perpendicular to the pulling force and are expected to undergo the initial rupture that leads to the activation of the kinase. When comparing the crystal structures to the model, there is a general conservation along betaC11 visualized by comparing the hydrogen-bonding pattern and the side chain interactions. betaR1 is sandwiched between betaC10 and betaC11 and continues to reveal the consistent bonding pattern between the enzymes.

According to the force activation model, the N-terminal beta sheet (betaC1-C3, Figure 2.2B, dark blue) is the region that lies parallel to the pulling force to maintain the active site integrity during force activation (Grater, Shen et al. 2005). This sheet is strikingly similar in all four enzymes. Almost all of the backbone interactions are maintained. The side chain interactions, which further stabilize the beta sheets, are also consistent between the kinases. A conserved lysine residue in the N-terminal betaC3 sheet (K82 KOB, K5971 KOA, K57 TTN1, K53 TKI) interacts with the regulatory tail alphaR2 via van der Waals interactions between the side chain and isoleucine/leucine and valine. This lysine also interacts directly with the active site through electrostatic interactions with an aspartic acid and glutamic acid (Hu, Lei et al. 1994; Kobe, Heierhorst et al. 1996). Linked intimately to the N-terminal beta strands, the alphaR2 helix is a tightly packed 3_{10} helix buried beneath the betaC1-C3 sheets, the substrate

binding site, and the portion of the protein that lies directly upstream of the helix (the alphaR1 helix and the linker between alphaR1 and alphaR2 helices). The face of the helix that is in contact with the beta strands is notably similar in each of the enzymes; a string of hydrophobic interactions from the non-polar side chains of the beta strands contacts the non-polar face of the helix (with the exception of one serine present only in TTN-1 and *C. elegans* twitchin). On the alternate side of the helix, the residues interact with both the substrate-binding site and the longer of the regulatory helices, alphaR1. Many of the interactions are conserved between all four kinases, allowing the alphaR2 helix to retain the overall stability and topology of the enzymes.

Force-extension relationships of C. elegans TTN-1 and twitchin Ig domains

For titin-like proteins, the organization is such that a set of five tandem Ig domains immediately follows the kinase catalytic core and its regulatory sequence (Figure 2.1A & B). If the kinase were to act as a force sensor, we would expect these domains to withstand forces greater than the kinase itself. To test this hypothesis we used single-molecule MFS techniques to analyze the mechanical properties of recombinant proteins containing 5 tandem Ig domains from twitchin and TTN-1 (Figure 2.1A-C). Random segments of these proteins were picked up by the MFS tip and then stretched with a pulling speed of ~0.5 nm/ms. The resulting force-extension curves showed a sawtooth-like pattern, characteristic of the unfolding of Ig domains (Figure 2.3A & D) (Li, Linke et al. 2002; Bullard, Garcia et al. 2006). To analyze the spacing between peaks in the sawtooth patterns we used the worm-like chain (WLC) model for polymer elasticity, which predicts the entropic restoring force (F) generated upon the extension (x)

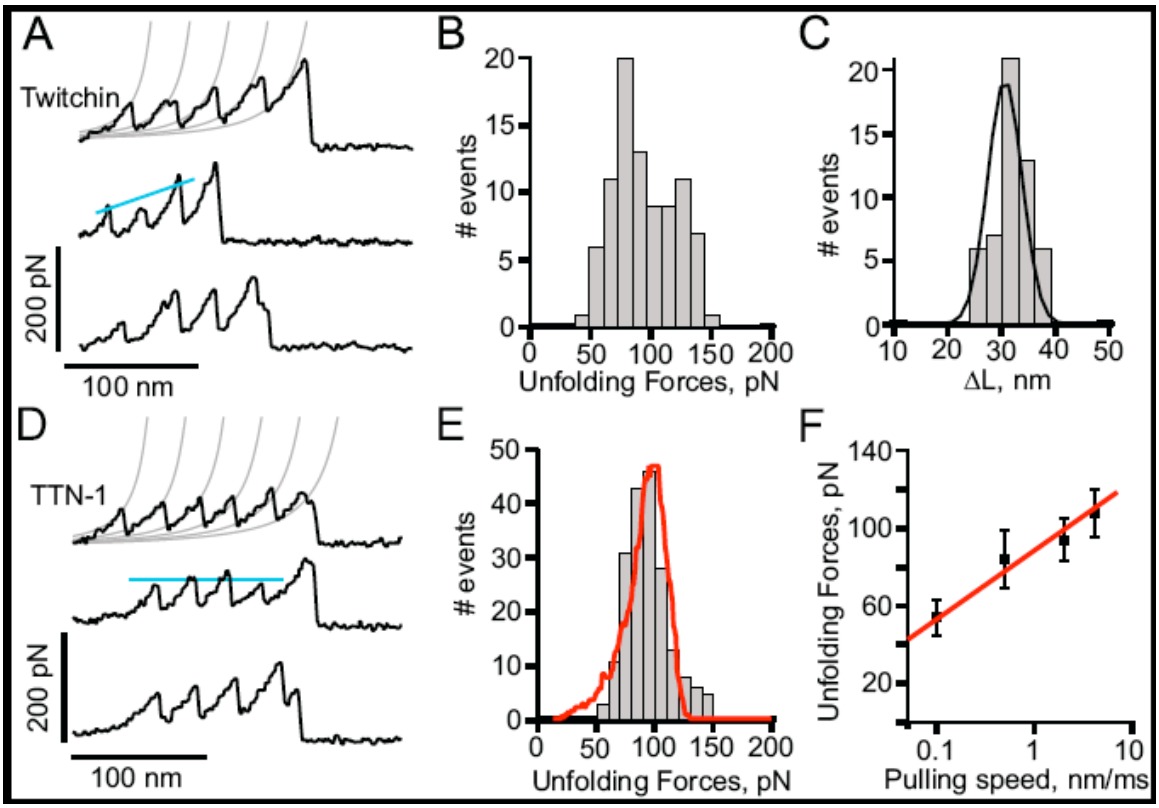


Figure 2.3- Force-extension relationships of TTN-1 and twitchin Ig domains.

A, D) Several examples of force-extension curves obtained after stretching twitchin (A) and TTN-1 (D) Ig domains. The grey lines were generated with the WLC equation using a persistence length of 0.4 nm and contour length increments, ΔL, of 30 nm. B, E) Unfolding force histograms for twitchin and TTN-1 domains. The mean force peak values are 93 ± 25 pN, (n=88 peaks) and 85 ± 22 pN (n=193 peaks), respectively. In E) the red line corresponds to a Monte-Carlo simulation of TTN-1 Ig using $k_u^o = 5x10^{-2}\ s^{-1}$ and $\Delta x_u = 0.35$ nm at a pulling speed of 500 nm/s. C) Histogram of contour length increments observed upon unfolding of twitchin Ig domains shows one main peak centered at ~30nm (Gaussian fit: 30.6 ± 3.2 nm). F) The unfolding forces of TTN-1 Ig domains depend on pulling speed. The experimental data (black symbols) can be well described by Monte-Carlo simulations (red line) using $k_u^o = 5x10^{-2}\ s^{-1}$, $\Delta x_u = 0.35$ nm.

of a polymer ((Bustamante, Marko et al. 1994; Marko 1995); see methods). The thin lines in Fig. 2.3A and 2.3D correspond to fits of the WLC equation to the curve that precedes each force peak. We found that the separation between force peaks for both proteins is ~30 nm (30.6 nm $\pm$ 3.2 for the twitchin protein; n= 53 force peaks; Fig. 2.3C). This value corresponds very well with the expected increase in contour length of a 95 amino acid (aa) Ig domain: 95 aa x 0.35 nm (length of an aa) – 3 nm (size of folded Twc Ig domain 26; (Kobe, Heierhorst et al. 1996) and Figure 2.2C) = 30.3 nm.

Unfolding force histograms show that the Ig domains from twitchin and TTN1 unfold at a similar range of forces and similar average unfolding force values (~90 pN and ~85 pN, respectively; Fig 2.3B & E). The unfolding forces of these Ig domains from twitchin and TTN-1 fall within the range or are somewhat weaker than those reported for Ig domains from vertebrate titin (Li, Linke et al. 2002) and insect projectin (Bullard, Garcia et al. 2006). It is noteworthy that the recordings from the 5 Ig domains of twitchin show an ascending pattern of peak heights suggesting a hierarchy in mechanical stabilities for these Ig domains (Figure 2.3A). In addition, we found that mechanical unfolding of TTN1 Ig domains is fully reversible and that they refold to their native states with a rate at zero force of ~2 s^{-1} (data not shown).

Our results show that the TTN-1 and twitchin Ig domains unfold at ~90 pN at a fixed pulling speed of 0.5 nm/ms. However, during normal muscle contraction cycles these domains may experience a wide range of stretching speeds. Since the mechanical stability may not be the same at different pulling speeds, as shown for titin domains (Rief, Gautel et al. 1997; Carrion-Vazquez, Oberhauser et al. 1999) we studied the rate-dependency of the stability of Ig domains from TTN-1 and twitchin. Figure 2.3F shows

the relationship between the unfolding force and the pulling speed (0.1-5nm/ms) for TTN-1 Ig domains. A tenfold decrease in pulling speed decreases the unfolding forces by only 20 pN indicating that these domains are mechanically stable over a wide range of pulling speed. The continuous lines correspond to the result of Monte Carlo simulations of two-state unfolding of this sequence at the corresponding range of pulling rates (Rief, Gautel et al. 1997; Oberhauser, Marszalek et al. 1998; Carrion-Vazquez, Oberhauser et al. 1999; Best, Fowler et al. 2002). We found that, by using a combination of the rate constant at zero force, k_u^o, of $5x10^{-2}$ s^{-1} and the unfolding distance Dx_u between the folded state and transition state of 0.35nm, we can adequately describe the unfolding force histogram (Figure 2.3E) as well as the speed dependency of unfolding forces (Figure 2.3F). These values are comparable to those used for several other mechanically stable beta-strand rich domains (Oberhauser, Marszalek et al. 1998; Bullard, Linke et al. 2002; Li, Linke et al. 2002; Sharma, Perisic et al. 2007). For example human titin Ig domains have Dx_u values ranging from 0.25nm (I27) to 0.35nm (I1) but it can be as high as 0.4nm in fibronectin beta-sandwich domains. These results are consistent with the hypothesis that these Ig domains resist denaturation under low and moderate stretching forces.

Force-extension relationships of C. elegans giant kinases and flanking domains

The protein kinase domains of twitchin and TTN-1 are autoinhibited, containing an endogenous regulatory sequence situated between the two subdomains, making extensive contact with residues essential for ATP binding, substrate recognition and catalysis. Unlike their closest homologs, the MLCKs, binding to Ca^{+2}/calmodulin does

not relieve autoinhibition. It has been hypothesized that the giant kinases may act as force sensors (Grater, Shen et al. 2005); that the forces generated from the contraction/relaxation cycles of muscle activity are sufficient to unleash the regulatory domain from the catalytic core and activate the kinase. The relaxation of a stretching force would restore the inhibited conformation of the kinase.

We used single-molecule MFS to analyze the mechanical strength of the kinase and their flanking Ig/Fn domains. To be sure that the proteins were in their native conformations, the kinase activity of the proteins was measured. Although the proteins are autoinhibited, they have modest activity in vitro (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002). We found that our recombinant kinase constructs have specific activities comparable to those previously published (Figure 2.1D) (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002).

For the force measurements, dilute solutions of the recombinant kinase proteins were non-specifically attached to a glass coverslip. Random segments of the proteins were picked up by the MFS tip and then stretched with a pulling speed of 0.5 nm/ms. Figure 4A shows examples of force–extension curves obtained after stretching single Fn-Twc kinase-Ig molecules. We typically observed recordings with multiple force peaks before the last detachment peak. In the Fn-Twc kinase-Ig construct the Ig and Fn domains have ~95 aa and should therefore contribute to an increase in contour length, DL, of ~30nm (Figure 2.3). Hence, we attribute the two force peaks before the detachment peak as the unfolding of Fn and Ig domains and the initial two force peaks to the sequential unfolding of the kinase domain, which has a contour length of ~95 nm. Figure 2.4B shows a histogram of increases in contour length increments observed upon unfolding of

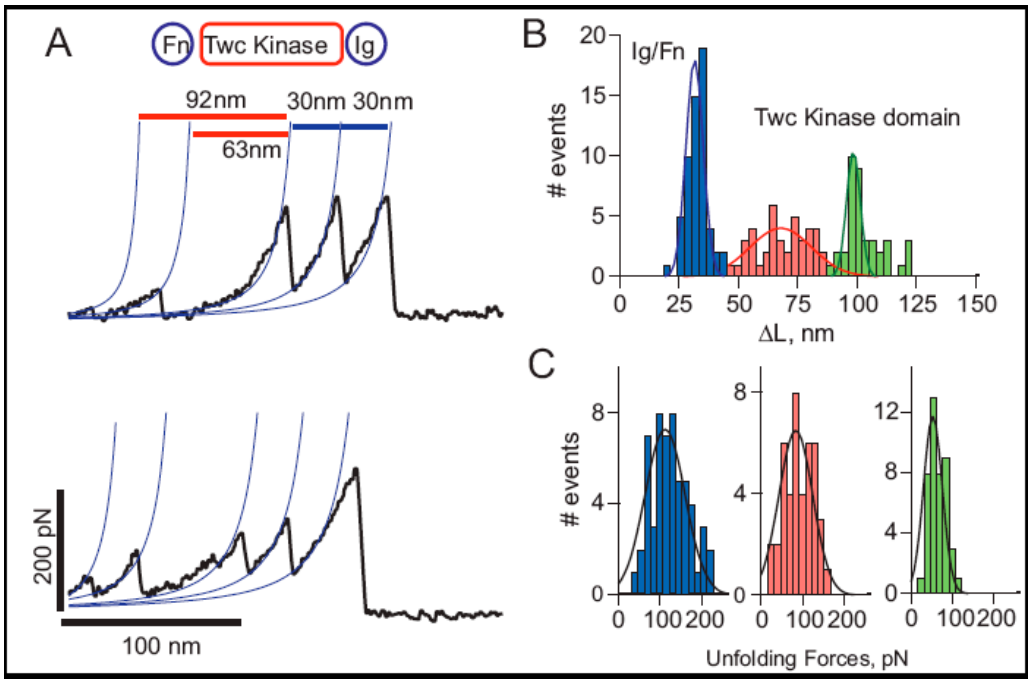


Figure 2.4- Mechanical properties of *C. elegans* twitchin kinase. A) Two examples of force-extension curves obtained for the Fn-Twc kinase-Ig construct. The two small force peaks correspond to the stepwise unfolding of the Twc kinase domain and the last two peaks to the unfolding of the flanking Ig/Fn domains. B) Histogram of increases in contour length increments observed upon unfolding, ΔL, of Fn-Twc kinase-Ig. There are peaks at ~30 nm, 65 nm and 95 nm (Gaussian fits: 31±5 nm, 67±18 nm and 97±10 nm, *n*=142), which correspond to the unfolding of Ig/Fn domains (blue bars), and kinase domain (red and green bars). C) Unfolding force distributions for Ig/Fn domains (blue bars) and kinase domain (red and green bars); the respective force peaks are at 111±67 pN (n=57), 83±57 pN (n=41) and 52±11 pN (n=43).

Fn-Twc kinase-Ig. There are peaks at ~30 nm, 65 nm and 95 nm, which correspond to the unfolding of Ig/Fn domains (blue bars), and kinase domain (red and green bars), respectively. The corresponding unfolding forces are 111±67 pN (n=57), 83±57 pN (n=41) and 52±11 pN (n=43) (Figure 2.4C). The small peak has a contour length of ~30nm, a length that would be expected to fit the unwinding of the ~100 aa small lobe. The second peak gives a contour length of ~65nm, fitting nicely to the unwinding of the ~190 aa large lobe. The regulatory domain (which includes the betaR1 C-terminal strand) probably unfolds at very low forces (<10pN), forces indistinguishable from noise on the MFS. Hence, we interpret the data shown in Figure 4 as the stepwise unfolding of the two lobes of the kinase. The smaller lobe, composed mainly of beta sheets will unfold first, leaving the larger, alpha helical lobe intact. This hypothesized order of rupture, and lack of any unfolding peak from the regulatory domain, is supported by our SMD simulations (Figure 2.6).

The unfolding pattern of TTN-1 Fn-Kinase-Ig parallels that of twitchin Fn-Kinase-Ig. The complete unfolding of TTN-1 kinase has four force peaks (Figure 2.5A) corresponding first to the biphasic rupture of the catalytic core, followed by the unfolding of Ig and Fn domains. There are peaks at ~30nm, 65nm and 95nm (Figure 2.5B). To further characterize the mechanical unfolding of the kinase domains, we analyzed the unfolding kinetics by doing experiments at different pulling speeds. Figure 2.5C shows a plot of the average unfolding force versus the pulling rate for TTN-1 Ig/Fn domains and the kinase domain. For the kinase, we analyzed the first force peak that gives a DL of ~30nm, which interpret as the unfolding of the small, beta-sheet rich lobe. The continuous lines correspond to the result of Monte Carlo simulations of two-state

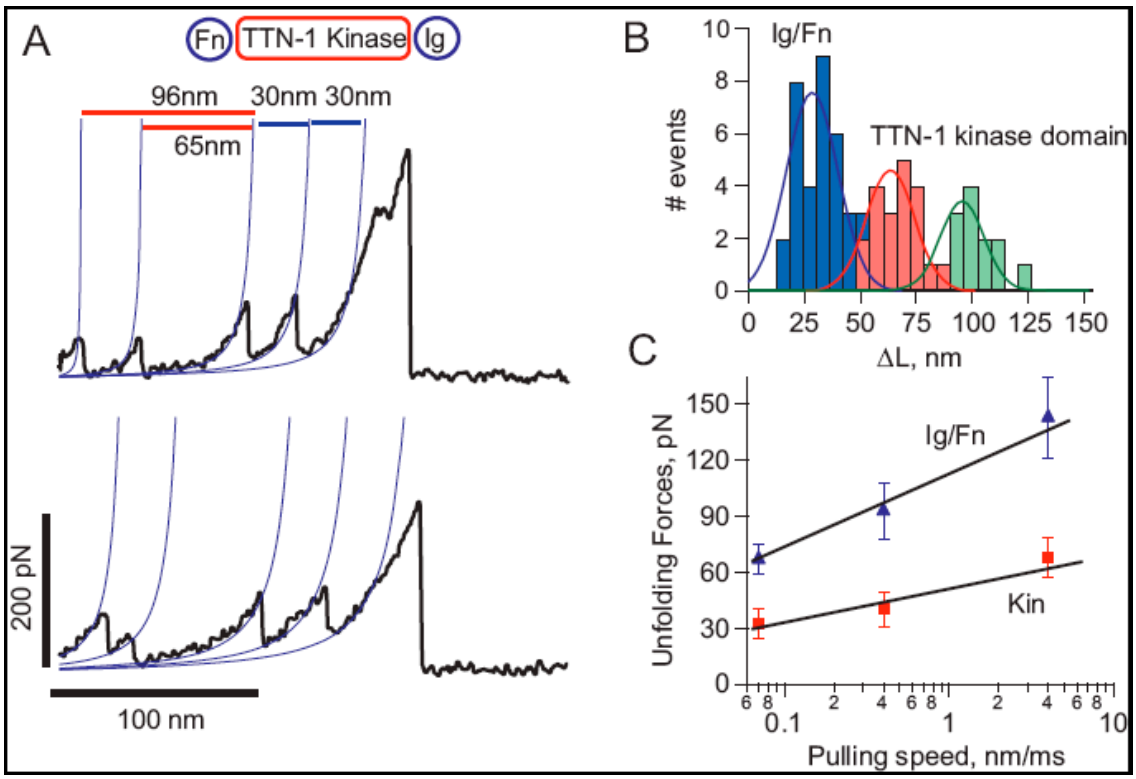


Figure 2.5- Mechanical properties of *C. elegans* TTN-1 kinase. A) Two examples of force-extension curves obtained for the Fn-TTN1 kinase-Ig construct. B) Histogram of increases in contour length increments observed upon unfolding, ΔL, of the Fn-TTN-1 kinase-Ig construct (16 molecules). There are peaks at ~30nm, 65nm and 95nm (n=35). (Gaussian fits: 29±16 nm, 63±15 nm and 95±13 nm). C) Plot of the average unfolding force versus the pulling rate for TTN-1 Ig/Fn domains and the kinase domain. For the kinase, we analyzed the first force peak with a ΔL of ~95nm. The continuous lines correspond to the result of Monte Carlo simulations of two-state unfolding at the corresponding pulling rates. The parameters used for the Monte-Carlo simulation are: $k_u^o = 1.4 \times 10^{-2}\ s^{-1}$, $\Delta x_u = 0.35$nm for the Ig/Fn domains and $k_u^o = 4 \times 10^{-2}\ s^{-1}$, $\Delta x_u = 0.6$nm for the TTN-1 kinase.

unfolding at the corresponding pulling rates. The parameters used for the Monte-Carlo simulation are: $k_u^o = 1.4x10^{-2}\ s^{-1}$, $Dx_u = 0.35nm$ for the Ig/Fn domains and $k_u^o = 4x10^{-2}\ s^{-1}$, $Dx_u = 0.6nm$ for the TTN-1 kinase. The unfolding rate constant at zero force, k_u^o, and unfolding distance between the folded and the transition state, Dx_u, for the Ig and Fn beta-sandwich domains are similar to that for TTN-1 Ig domains 38-42 (Figure 2.3F). However, the transition state distance Dx_u for the kinase alpha helical lobe is considerable larger than in the beta-sandwich domains (0.6nm vs. 0.35nm), suggesting different transition state structures. From these data, we can conclude that the TTN-1 and twitchin kinase domains show a significant mechanical resistance and they unfold at forces similar to those for Ig/Fn beta-sandwich domains (30-150pN). These results further support the force activation hypothesis for the giant protein kinases.

DISCUSSION

The kinase domain of human titin has been postulated to act as a force sensor (Grater, Shen et al. 2005; Lange, Xiang et al. 2005). The kinase domain is thought to be autoinhibited at rest, but during muscle activity, catalysis by the kinase results in a phosphorylation cascade that ultimately causes the expression of genes important in myofibril maintenance and growth. Human titin spans half of the sarcomere, from the M-line to the Z-disk (Furst, Osborn et al. 1988). The segment of titin in the A-band ending at the M-line is fixed, whereas the portion of titin in the I-band varies its length in response to the state of muscle contraction (Tskhovrebova, Trinick et al. 1997). The kinase domain, which is located at the periphery of the M-line (Obermann, Gautel et al. 1996), is in an ideal position and orientation to sense the mechanical strain that occurs during the

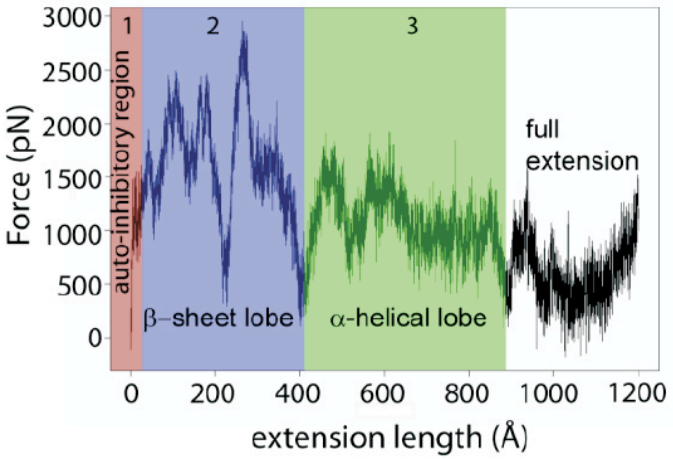
contraction/relaxation cycle of muscle activity. In *C. elegans* striated muscle, the proteins of greatest similarity to human titin are TTN-1 located in the I-bands (Flaherty, Gernert et al. 2002), and twitchin located at the non-M-line portions of A-bands (Moerman, Benian et al. 1988). Given the high degree of similarity of the kinase catalytic cores and the conservation of organization of surrounding Ig and Fn domains, it is reasonable to hypothesize that force activation might also occur in TTN-1 and twitchin.

If the kinase domains were to act as force sensors, we would expect them to withstand stretching forces. Here we used single-molecule force spectroscopy to test the mechanical properties of *C. elegans* TTN-1 and twitchin kinases. The proteins were recombinantly expressed and shown to retain activity in their inhibited form. We found that the mechanical stabilities of the kinase domains are compatible with their postulated function. The kinases unfold in two clearly resolvable steps at forces of ~50pN and ~80pN. These unfolding forces are lower than most beta-strand rich domains, range ~80-250pN (Rief, Gautel et al. 1997; Li, Linke et al. 2002; Oberhauser, Badilla-Fernandez et al. 2002; Carrion-Vazquez, Li et al. 2003; Bullard, Garcia et al. 2006; Sharma, Perisic et al. 2007) but similar to alpha-helix rich proteins, range ~30-100pN (Best, Li et al. 2001; Ainavarapu, Li et al. 2005; Li, Wetzel et al. 2006).

Our results show that the TTN-1 and twitchin kinase domains unfold in a biphasic, stepwise fashion, indicated by the presence of two peaks in the force-extension recordings. The first force peak probably represents the unwinding of the smaller lobe of the catalytic core, corresponding to the rupture of beta sheets. Following the initial rupture is a second peak, probably representing the rupture of the larger, alpha helical lobe of the catalytic core. The small peak has a contour length DL of ~30nm, a length

that is consistent with the unwinding of the ~100 aa small lobe. The second peak gives a contour length of ~65nm, which may correspond to the unwinding of the ~190 aa large lobe. Unexpectedly, these results contrast data showing that a-helices are less mechanically stable than beta sheets (Oberhauser and Carrion-Vazquez 2008). Finally, the regulatory domain most likely unfolds at very low forces (<10pN), which is below the resolution of our MFS.

To further understand the molecular origin for the mechanical unfolding of the *C. elegans* twitchin kinase we carried out steered molecular dynamics (SMD) simulations (Figure 2.6). Molecular dynamics simulations have been extensively used to examine the mechanical unfolding of a wide variety of proteins and in general, there is a good correlation with single-molecule data (Lu, Isralewitz et al. 1998; Lu and Schulten 1999; Best, Li et al. 2001; Sotomayor and Schulten 2007). To validate the accuracy of our *in silico* experiments we performed SMD simulations on other protein domains, such as titin I27, ubiquitin and synaptotagmin C2A. The I27, ubiquitin and C2A results are very similar to those published previously (Rief, Gautel et al. 1997; Lu, Isralewitz et al. 1998; Phillips, Rosemary et al. 2005) (data not shown). The magnitude of the forces observed in the SMD simulations does not directly correspond to those measured with MFS. This is partially because the pulling speeds are several orders of magnitude different. However, the simulations are qualitatively consistent with the MFS results. For example, similar to our MFS results twitchin Ig26 unfolds at lower forces than I27. In addition, as shown by MFS data (Carrion-Vazquez, Oberhauser et al. 2000) the C2A domain unfolds at much lower forces (~50pN) than I27 or ubiquitin (~200pN) and does not show an initial force-extension burst.


3000
2500
2000
1500
1000
500
0
Force (pN)
1
2
3
auto-inhibitory region
β–sheet lobe
α-helical lobe
full
extension
0
200
400
600
800
1000
1200
extension length (Å)

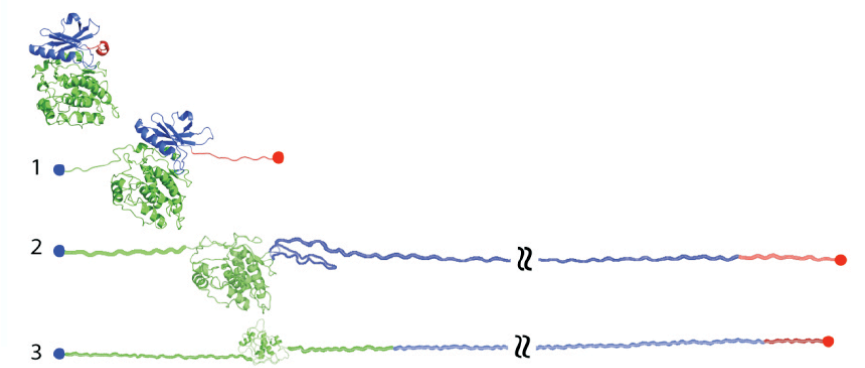

1
2
3

Figure 2.6- Constant velocity steered molecular dynamics simulation of the mechanical unfolding of twitchin kinase. Left) Force-extension curve obtained from SMD simulations by stretching the twitchin kinase domain (1KOA) between its C terminus and its N terminus at a pulling speed of 0.5Å/ps. The total simulation time was 2.4ns using 33,428 total atoms including 18 Na^+ and 9,305 water molecules. The fixed atom was Tyr 5915 and the SMD atom Arg 6261. Right) 4 snapshots of twitchin kinase stretched from its termini taken at no extension (rest), after 65Å (1), 340Å (2) and 639Å (3) of extension. At rest, the kinase domain is in a closed conformation. The active site is occupied by the autoinhibitory region (red), which makes extensive contact with the catalytic site, blocking substrate binding. 1) At low forces the regulatory tail will unravel reversibly and expose the active site to its substrates. 2) At high forces the kinase begins to unfold and the integrity of the active site is disrupted. The small lobe (blue), made mainly of beta sheets, unravels first followed by the unfolding of the alpha helical rich large lobe (green).

Figure 2.6 shows constant velocity SMD simulation of the mechanical unfolding of the twitchin kinase domain. The force-extension curve was obtained from SMD simulations by stretching the *C. elegans* twitchin kinase domain (1KOA) between its C terminus and its N terminus at a pulling speed of 0.5Å/ps. On the right four snapshots of twitchin kinase are shown taken at no extension (rest), after 6.5nm (1), 34nm (2) and 64nm (3) of extension. At rest, the kinase domain is in a closed conformation. The active site is occupied by the autoinhibitory region (red), which makes extensive contact with the catalytic site, blocking substrate binding. At low forces, the regulatory tail (which includes the C-terminal beta strand BR1 and alphaR1-R2 helices) will unravel reversibly and expose the active site to its substrates (snapshot 1; red region in the force-extension plot). According to this model the C-terminal beta sheet (strands betaC10, C11 + betaR1; dark green and red in Fig 2.2B) should unfold together with the alphaR1 and R2 helices. Because their different mechanical topologies (zipper vs. shear), the C-terminal beta sheet should unfold at much lower forces than the N-terminal beta-sheet. As shown by Grater et al (2005), we also observe that after breakage of the C-terminal beta-sheet and unraveling of the regulatory domain, the tertiary structure of the kinase catalytic site remains intact. At this point, the kinase is in an open, active conformation and downstream signaling can occur. During muscle activity, low forces may result from the repeated contraction/relaxation of the sarcomeres. At high forces, the kinase begins to unfold and the active site is disrupted (snapshot 2; blue region in the force-extension plot). The small lobe (blue), made mainly of beta sheets, unravels first followed by the unfolding of the alpha helical rich lobe (green; snapshot 3; green region in the force-extension plot). Eventually, with sustained high forces, the enzyme will completely

unravel (white region in the force-extension plot). We also performed constant velocity SMD simulations of the mechanical unfolding of the model for TTN-1 kinase (data not shown). The unfolding trajectory for TTN-1 kinase is nearly identical to that of the Twitchin kinase indicating a similar stepwise mechanical unfolding pathway, as seen in the MFS data (Figure 2.5).

Grater et al. suggested a similar model from results of molecular dynamics simulations (Grater, Shen et al. 2005). However, in their studies they predicted that only human titin, but not *Aplysia* twitchin, would be likely to function as a force sensor. The simulations show that when force is applied to human titin the termini of the protein shift away from the kinase, the autoinhibitory region detaches from the active site, and that this release is accompanied by a shift in the two lobes of the catalytic core that results in an active site that is accessible to substrates. For *Aplysia* twitchin it was predicted that although applied force would remove the inhibitory region, the lobes of the catalytic core remain static and the enzyme would still be inactive. Our data argues that, in *C. elegans,* both TTN-1 and twitchin kinases meet the requirements of an enzyme likely to be activated by mechanical forces.

The most common feature of all the giant titin-like muscle proteins is the presence of multiple copies of Ig and Fn domains, either as tandem Ig domains or as super-repeats of Ig and Fn domains (Ferrara, Flaherty et al. 2005). Moreover, the arrangement of Fn and Ig domains and their actual sequences are highly conserved between twitchins and titin (Benian, Kiff et al. 1989; Labeit, Gautel et al. 1992). One major challenge in muscle biology is to understand how these modular domains function both at the individual and group scale, and how their mechanical properties vary to suit the type of muscle, or even

the location of a domain within a single sarcomere. Single molecule experiments have revealed the strength of Ig domains of titin and titin-like proteins to vary between 50-300pN, depending on the position of the domain within the sarcomere (Li, Linke et al. 2002; Bullard, Garcia et al. 2006). We studied the mechanical properties of five tandem Ig domains from *C. elegans* TTN-1 and twitchin. These 5 Ig domains are located in the C-terminal regions of the endogenous proteins and immediately follow the conserved kinase domain. We hypothesize that this region is functionally important because the architectural arrangement of the kinase domain plus 5 Igs is conserved in all of the giant titin-like proteins. Our data shows that these Ig domains, at least in the nematode, are slightly weaker (~93 pN for twitchin and ~85 pN for TTN-1) than the average Ig domain and those found in other muscle types such as cardiac titin (~200 pN) (Carrion-Vazquez, Oberhauser et al. 1999), insect kettin (~125-250 pN) and insect projectin (~109 pN) (Bullard, Garcia et al. 2006).

TTN-1 can be regarded as a twitchin/titin hybrid. At the sequence level, the TTN-1 kinase catalytic core is more similar to the kinase catalytic core of twitchin (54.4% identical) than it is to the kinase of human titin (39.2% identical) (Flaherty, Gernert et al. 2002). What makes TTN-1 "titin-like" is its enormous 2.2 MDa size and the presence of several regions consisting of tandem repeats that are likely to act as molecular springs (these are not found in twitchin, only in human titin). Using known crystal structures of the kinase domains of *C. elegans* twitchin (Figure 2.2A) (Kobe, Heierhorst et al. 1996), *Aplysia* twitchin (Kobe, Heierhorst et al. 1996), and human titin (Mayans, van der Ven et al. 1998), we built a molecular model of TTN-1 kinase (Figure 2.2B) to further compare the proteins and analyze the subtle characteristics that underlie the force activation

hypothesis. When visually comparing the TTN-1 model to the twitchin and titin crystal structures, TTN-1 seems most structurally related to twitchin (Figure 2.2A &B), however there does exist a few varying regions that seem to be upheld only between the TTN-1 model and the structure of human titin. The topology and the bond-structure that holds the regulatory domain in intimate contact with the catalytic core seem closely related for all four kinases. It seems likely that, if the force activation hypothesis is true, all of these enzymes could adhere to such an on/off mechanism.

In conclusion, we have provided evidence that in response to mechanical force, two giant titin-like protein kinase domains unfold in a step-wise manner. The first step is likely to be the movement of the autoinhibitory domain from the catalytic pocket, without complete unfolding of the domain. With the application of increased force to the protein, the kinase domain will rupture, in a biphasic manner. These data bolster the hypothesis that the autoinhibited giant kinases in muscle cells may be activated in response to the forces generated during each contraction/relaxation cycle. Further experiments are underway to test if applying small forces to these enzymes, forces sufficient only to remove the autoinhibitory region (Figure 2.6), would indeed result in the activation of these enzymes.

MATERIALS AND METHODS

MFS

The mechanical properties of single proteins were studied using a home-built single molecule force spectroscope as previously described (Rief, Gautel et al. 1997; Oberhauser, Marszalek et al. 1998; Carrion-Vazquez, Oberhauser et al. 1999;

Oberhauser, Hansma et al. 2001; Bullard, Linke et al. 2002; Oberhauser, Badilla-Fernandez et al. 2002; Phillips, Rosemary et al. 2005; Miller, Garcia et al. 2006). The spring constant of each individual cantilever (MLCT-AUHW: silicon nitride gold-coated cantilevers; Veeco Metrology Group, Santa Barbara, CA) was calculated using the equipartition theorem (Florin 1995). The cantilever spring constant varied between 10-50 pN/nm and rms force noise (1-kHz bandwidth) was ~10 pN. Unless noted, the pulling speed of the different force–extension curves was in the range of 0.4–0.6 nm/ms.

Single Protein Mechanics

In a typical experiment, a small aliquot of the purified proteins (~1-50 µl, 10-100 µg/ml) was allowed to adsorb to a clean glass coverslip (for ~10 min) and then rinsed with PBS pH 7.4. Proteins were picked up randomly by adsorption to the cantilever tip, which was pressed down onto the sample for 1-2 seconds at forces of several nanonewtons and then stretched for several hundred nm.

Analysis of Force extension curves

The elasticity of the stretched proteins were analyzed using the worm-like chain (WLC) model of polymer elasticity (Bustamante, Marko et al. 1994; Marko 1995):

$$\mathrm{F}(x) = \frac{kT}{p}\left[\frac{1}{4}\left(1-\frac{x}{L_c}\right)^{-2} - \frac{1}{4} + \frac{x}{L_c}\right]$$

where F is force, p is the persistence length, x is end-to-end length, and L_c is contour length of the stretched protein. The adjustable parameters are the persistence length, p, and the contour length, Lc.

Monte-Carlo simulations

The folding and unfolding of a domain was modeled as a two state Markovian process where the probability of unfolding was $P_u=N_f*ąDt$ where N_f is the number of folded domains anḍ Dt is the polling interval (Rief, Gautel et al. 1997; Oberhauser, Marszalek et al. 1998; Oberhauser, Hansma et al. 2001). The folding probability was $P_f=N_u*bḌt$ where N_u is the number of unfolded domains. The rate constants for unfolding, a, and refolding, b, are given by $a=a_o\exp(FDx_u/kT)$ and $b=b_o\exp(-FDx_f/kT)$ where F is the applied force and Dx_u and Dx_f are the unfolding and folding distances.

Homology Modeling of TTN1 Kinase and Regulatory Domains

Molecular modeling of *C. elegans* TTN-1 was completed using Modeler (version 7v7; University of California San Francisco; http://salilab.org/modeler/) for model construction and SYBYL (version 7.0; Tripos, Inc., http://www.tripos.com) for analysis and refinement. The TTN-1 model was built and minimized based on the existing structural information available from *C. elegans* twitchin (PDB entry 1KOA) (Kobe, Heierhorst et al. 1996). Initial sequence alignments were obtained from ClustalW (using TTN-1, human titin, *C. elegans*, *Aplysia*, and *Mytilus* twitchin, chicken smooth muscle MLCK, and *Drosophila* myosin MLCK, http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html), NCBI BlastP (http://www.ncbi.nlm.nih.gov/), CPHmodel (using TTN-1 and *C. elegans* twitchin, www.cbs.dtu.dk), SwissModel (using TTN-1 and *C. elegans* twitchin, http://swissmodel.expasy.org//SWISS-MODEL.html), CDDomain (http://www.ncbi.nlm.nih.gov/Structure/ cdd/wrpsb.cgi, hits = KOA, KOB, 1TKI), and

3Djigsaw (using TTN-1 and *C. elegans* twitchin, http://www.bmm.icnet.uk/~3djigsaw/). The alignment was optimized manually after each round of modeling using visual structural comparison of the previously published structural data of *C. elegans* twitchin (PDB entry 1KOA (Kobe, Heierhorst et al. 1996)), *Aplysia* twitchin (PDB entry 1KOB) (Kobe, Heierhorst et al. 1996), and human titin (PDB entry 1TKI) (Mayans, van der Ven et al. 1998). The proteins showed ~50% (for *C. elegans* and *Aplysia* twitchin) and ~40% (for titin) similarity against TTN-1 residues 15907–16373 (Figure 2.1B), the approximate boundaries of TTN-1 kinase and the flanking Ig in the full-length giant polypeptide.

We specifically studied the sequence alignment at the terminal beta sheets and the regulatory alpha helices, as these are the regions predicted to be most important in the force activation mechanism (Grater, Shen et al. 2005), and are also the regions with some of the highest variability (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002). The N-terminal beta sheets, beta C1-C3, are located at residues: 24-33 (betaC1), 46-53 (beta C2), and 48-56 (beta C3) in TKI; 53-62 (betaC1), 65-72 (beta C2), and 77-85 (beta C3) in KOB; 5942-5951 (betaC1), 5954-5961 (beta C2), and 5966-5974 (beta C3) in KOA; 28-37 (betaC1), 40-47 (beta C2), and 52-60 (beta C3) in TTN-1. The C-terminal beta sheets, beta C10-11 and beta R1 are located at residues: 174-182 (betaC10), 193-197 (betaC11), and 328-336 (betaR1) in TKI; 204-212 (betaC10), 224-228 (betaC11), and 362-370 (betaR1) in KOB; 6093-6100 (betaC10), 6113-6117 (betaC11), and 6253-6260 (betaR1) in KOA; 179-186 (betaC10), 199-203 (betaC11), and 334-341 (betaR1) in TTN-1. Alpha R1 and alpha R2 are located at residues: 292-306 (alphaR1) and312-318 (alphaR2) in TKI, 321-335 (alphaR1) and 344- 351(alphaR2) in KOB, 6211-6225 (alphaR1) and 6234-6241 (alphaR2) in KOA, and 292-307 (alphaR1) and 315-323 (alphaR2) in TTN-1.

All protein structure images were generated by PyMOL version 1.1beta1 (http://www.pymol.org).

Steered Molecular Dynamics Simulations

The giant kinase structures were analyzed by SMD (steered molecular dynamics) as implemented in NAMD (Lu, Isralewitz et al. 1998; Phillips, Rosemary et al. 2005). The CHARM22 force field was employed throughout. The structural coordinates for each kinase structure (1KOA and the homology model for TTN-1) were solvated in a 65Å x 65Å x 65Å box. Eighteen Na^+ ions were added, corresponding to a concentration of 0.1M. The system was then minimized with 1000 steps of conjugate gradient minimization from an initial temperature of 310K. This was followed by a 400ps MD simulation to equilibrate the entire system (protein, water, and ions). The backbone RMSD was evaluated at the completion of the equilibration step. The SMD protein-ion-water system contained ~30,000 atoms. Forces were applied by restraining a fixed termini point harmonically and moving the SMD atom with constant velocity (0.5A/ps) along a predetermined vector. Kinase domains were stretched at a constant speed of 0.5Å ps^{-1} until their extension exceeded 99% of the contour length. The trajectories were recorded every 2 fs and analyzed with VMD. Coulombic forces were restricted using the switching function from 10 Å to a cutoff at 12 Å. A spring constant (κ) of $10k_BT/ Å^2$ was used during each simulation. We ran three simulations of the extension of twitchin kinase and the homology model for TTN-1 kinase domains with similar results. To validate the accuracy of our *in silico* experiments we performed SMD simulations on other

protein domains, such as titin I27, ubiquitin and synaptotagmin C2A. Our SMD results are very similar to those published previously (data not shown).

Cloning and expression of TTN-1 and Twitchin constructs

The TTN-1 and twitchin constructs were amplified from the *C. elegans* random primed cDNA library RB2 (kindly provided by Robert Barstead, Oklahoma Medical Research Foundation, Oklahoma City, OK) using the primer pairs listed below. Each construct was first subcloned into the cloning vector bluescript pKS (+), sequenced and subcloned into the expression vector pET 28, fusing an in frame 6-his tag to the N-terminus.

TTN-1 Ig (38-42):

5’ GGTACGGATCCAGACTCACTATGGACGGAG,

3’GGACTGAATTCTTAGCAACAAGTCTTAGACAATCCCATATC;

TTN-1 FnKinIg:

5’ GGTACGAATTCGAGGACAAATATGCAATTGGTATTC,

3’GGATCAAGCTTTTAGCAACACTTCTCGATGACAGCTGGAG;

twc Ig (26-30):

5’ GGTAC GAGCTC GCCTTCTGGGATCGATCTGAAGC,

3’ GGACTTCTAGATTAGCAACAGACAAGGAGAAGAGC;

twc FnKinIg:

5’ GGTACGGATCCGACTCTGGAACTGTTAATGTC,

3’GGACTCTCGAGTTAGCAACATGGCTCGAATTTGAGTGGTTC

The following restriction sites were inserted and used for the subcloning procedure: TTN-1 Ig (38-42) 5' BamHI, 3' EcoRI; TTN-1 FnKinIg 5' EcoRI, 3'HindIII; twc Ig (26-30) 5'SacI, 3'XbaI; twc FnKinIg 5'BamHI, 3'XhoI.

The recombinant plasmids were transformed into E. coli BL21 (DE3) RIL (Stratagene). The cells were grown to mid-log phase in the presence of 25microg/ml kanamycin and 34 microg/ml chloramphenicol at 37°C. Protein expression was induced with 0.5 mM IPTG and continued overnight at 23°C. The cells were harvested and resuspended in 20mM TRIS, 500mM NaCl, and 5mM imidazole (pH7.9) with the addition of Roche complete EDTA-free protease inhibitor pellets. Lysis was completed by passage through a French press at 1000 psi. Detergents PEI (to 0.1%) and NP40 (to 0.01%) were added to the cell free extract. The soluble fraction was collected and loaded onto a pre-charged Novagen nickel column. The column was first washed with 25 column volumes of 20mM TRIS, 750mM NaCl, 5mM imidazole, 0.01% NP40 (pH7.9) followed by a second wash with 5 column volumes of 20mM TRIS, 750mM NaCl, 20mM imidazole, 0.01% NP40 (pH7.9). The proteins were eluted with 20mM TRIS, 500mM NaCl, and 1M imidazole. Protein purity of ~95% was confirmed using coomasie brilliant blue staining of 12% SDS PAGE.

Kinase Assays

The purified kinases were dialyzed against 20mM TRIS (pH8.0), 20mM NaCl, and 5mM betaME. The enzymes were added to reaction buffer (20mM TRIS pH7.4, 10mM magnesium acetate, 0.05% triton, 0.2mg/ml BSA) to a final concentration of 30 pg/microliter. The model substrate was a derivative of chicken smooth muscle regulatory

myosin light chain (kMLC 11-23) with the sequence KKRARAATSNVFS (Heierhorst, Tang et al. 1996) (synthesized by the Microchemical Facility, Emory University). The peptide substrate was added to the reaction mixture in excess (0.2mg/ml). The reactions were initiated by the addition of 400microM gamma^{32}P-ATP (0.25 microCi/microliter). Catalysis occurred for 10 minutes at 30°C. A portion of the reaction (25 microliters/40 microliters total volume) was removed and spotted onto Whatman P81 filters. The filters were washed in 75mM phosphoric acid. Once dry, the washed filters were placed into scintillation vials, the counts were measured, and the specific activity was calculated by the following equation:

$$\frac{(CPM_{sample} - CPM_{blank})}{((\text{Specific Activity } ^{32}\text{P-ATP})(\text{reaction time})([\text{enzyme}])(\text{reaction volume/spot volume}))}$$

Each reaction was done in triplicate and performed on freshly purified protein.

As expected from earlier results, specific activities were the highest for the reactions with the Fn-TTN-1Kin-Ig and significantly lower for the Fn-Twc kinase-Ig reactions (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002). The specific activity of the enzymatic labeling was normalized to reactions without the addition of kinase.

CHAPTER THREE

Identification and Characterization of the Substrate Interaction Between TTN-1 Kinase and UIG-1

INTRODUCTION

A number of extraordinarily large polypeptides exist in the striated muscles of probably all animals, the king of which is vertebrate titin. Ranging from ~800kDa- 3.7 MDa, and containing multiple domains, these proteins have been implicated in several functions including organization/assembly of the sarcomere, passive elasticity, and signaling (Granzier and Labeit 2004; Ferrara, Flaherty et al. 2005). One might find it surprising that between these proteins, functional variability exists, as the intracellular, giant polypeptides are seemingly similar, composed of multiple Ig and Fn domains and containing at least one protein kinase domain near the C-terminus. However, their specific localizations in the sarcomere, their unique regions, and the individual pattern of organization of their Ig and Fn domains likely provide individual roles for each of the giant polypeptides. Because these giant proteins are conserved, and were likely to have co-evolved with muscle mechanics, several model genetic organisms, including *C. elegans*, can be used profitably to study their functions.

Within *C. elegans* there are currently 3 known members of this giant, titin-like, protein family: UNC-89, twitchin (UNC-22), and TTN-1. UNC-89 exists in at least 6 different isoforms, all residing in the M-line (Benian, Tinley et al. 1996; Small, Gernert et al. 2004). The largest UNC-89 polypeptide, UNC-89-B, is ~900 kDa and contains 52 Ig domains and multiple regions involved in signal transduction, including two protein kinase domains, and single copies of SH3, DH, and PH domains. The DH domain specifically activates RhoA (Qadota et al., manuscript in revision), suggesting for the first time that the giant kinases are involved in downstream GTPase signaling. Twitchin

(~800kDa), named for its characteristic twitching mutant phenotype, has 30 Ig domains, 31 Fn domains, and a single kinase domain (Benian, Kiff et al. 1989; Benian, L'Hernault et al. 1993). Twitchin localizes to the outer portions of the muscle A-band (Moerman, Benian et al. 1988) and is thought to inhibit the rate of muscle relaxation in *Aplysia* and *Mytillus* muscle (Probst, Cropper et al. 1994; Siegman, Funabara et al. 1998). Located in sarcomeric I-bands, TTN-1 is significantly larger than twitchin and UNC-89, with a molecular weight of 2.2 MDa. In addition to 56 Ig domains and 11 Fn domains, TTN-1 has several regions predicted to be coiled-coil, and two regions consisting of tandem repeats (Flaherty, Gernert et al. 2002). The largest of these tandem repeat regions, called PEVT, is similar in amino acid composition and tandem repeat structure to the main elastic element of vertebrate titin called the PEVK region. Thus, this region of nematode TTN-1 is hypothesized to be elastic.

Similar to mammalian titin kinase, TTN-1 and twitchin have a single kinase catalytic domain and autoinhibitory region located near the C-terminal end of the giant molecules (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002). The kinase domains of both twitchin and TTN-1 show *in vitro* phosphotransferase activity when tested against a model peptide (Lei, Tang et al. 1994; Flaherty, Gernert et al. 2002). In contrast, UNC-89 has two kinase domains in the C-terminal region. Catalytic activity has thus far not been measurable for either of the kinase domains in UNC-89, but homology modeling suggests that one of the domains functions as a kinase, while the other is probably dormant (a psueudokinase) (Small, Gernert et al. 2004).

An important goal is to find substrates for the giant kinases. So far the only known substrates are for vertebrate titin: telethonin (Tcap) (Mayans, van der Ven et al.

1998) in developing muscle, and Nbr1 (Lange, Xiang et al. 2005) and MURF-1 (Centner, Yano et al. 2001) in mature muscle. The discovery of the signaling between titin kinase and Nbr1 provided pivotal insight into the pathology of a human myopathy, as mutations in titin kinase that abolish the binding result in a diseased state (Lange, Xiang et al. 2005). The substrates for the other giants are unknown, including vertebrate obscurin and any giant kinase from an invertebrate. This is despite the fact that protein kinase activity has been demonstrated for nematode and *Aplysia* twitchin (Lei, Tang et al., 1994; Heierhorst, Tang et al), *Drosophila* projectin (a twitchin homolog) (Ayme-Southgate, Vigoreaux et al., 1991) and TTN-1 (Flaherty, Gernert et al. 2002).

In an effort to identify binding partners and/or substrates for TTN-1 kinase a yeast-2-hybrid screen was used. The screen identified an interaction between the kinase domain of TTN-1 and the Cdc42 GEF, UIG-1. The protein-protein interaction was supported with *in vitro* binding and co-localization data. Further, using *in vitro* kinase assays, we were able to measure phosphotransferase activity by TTN-1 onto UIG-1, implying that there is *in vivo* signaling that occurs between TTN-1 and UIG-1. The interaction provides another link between a giant kinase and Rho GTPase signaling. Our results also demonstrate the first known substrate of an invertebrate giant protein kinase.

RESULTS

TTN-1 kinase interacts with the Cdc42 GEF UIG-1

Little is known about the functions of the giant protein kinases in *C. elegans*. In particular, TTN-1 kinase has no known protein-protein interactions and none of the giant a kinases have known substrates. In hopes of elucidating the molecular mechanisms of

TTN-1 kinase a yeast-2-hybrid screen was employed. The bait (Figure 3.1A) was segment of TTN-1 containing Ig-Fn-Kinase, located near the C-terminus of the 2.2 MDa polypeptide. As prey, both a comprehensive library of worm cDNAs and a collection ("bookshelf") of 35 proteins known to reside at either M-lines, dense bodies or both (Table 3.1) were utilized. Although the cDNA screening was not fruitful (data not shown), the bookshelf rendered a weak interaction between the Cdc42 GEF UIG-1 prey and the Ig-Fn-Kinase bait (Figure 3.1A). Further, the kinase domain alone is sufficient for the interaction; the Ig-Fn region is neither sufficient nor necessary (Figure 3.1A). In the original bookshelf screen, the full length UIG-1 was used as the prey. UIG-1 is a fairly large protein with a predicted molecular weight of 105 kDa. This encouraged us to determine the minimal regions required for interaction in the 2-hybrid system. The results shown in Figure 3.1B indicate that the C-terminal 280 residues of UIG-1 are critical for maintaining the interaction. Moreover, the N-terminal 187 amino acids are dispensable (Figure 3.1B). Interestingly, if only the N-terminal 36 amino acids are deleted, the interaction is the strongest relative to the alternative truncations and to the full-length protein (Figure 3.1B).

To confirm the yeast-2-hybrid results, a far western assay was employed (Figure 3.1C). 6his-TTN-1 IgFn3Kinase was incubated with a western blot containing MBP, MBP-UIG-1 N-terminal 2/3, or MBP UIG-1 C-terminal 2/3. Results show that TTN-1 kinase reacts with MBP UIG-1 C-terminal 2/3, but not with MBP. TTN-1 reacts much less strongly to MBP-UIG-1 N-terminal 2/3, a result consistent with the 2-hybrid results. Thus, the binding observed in the far western assay supports an interaction between TTN-1 kinase and UIG-1.

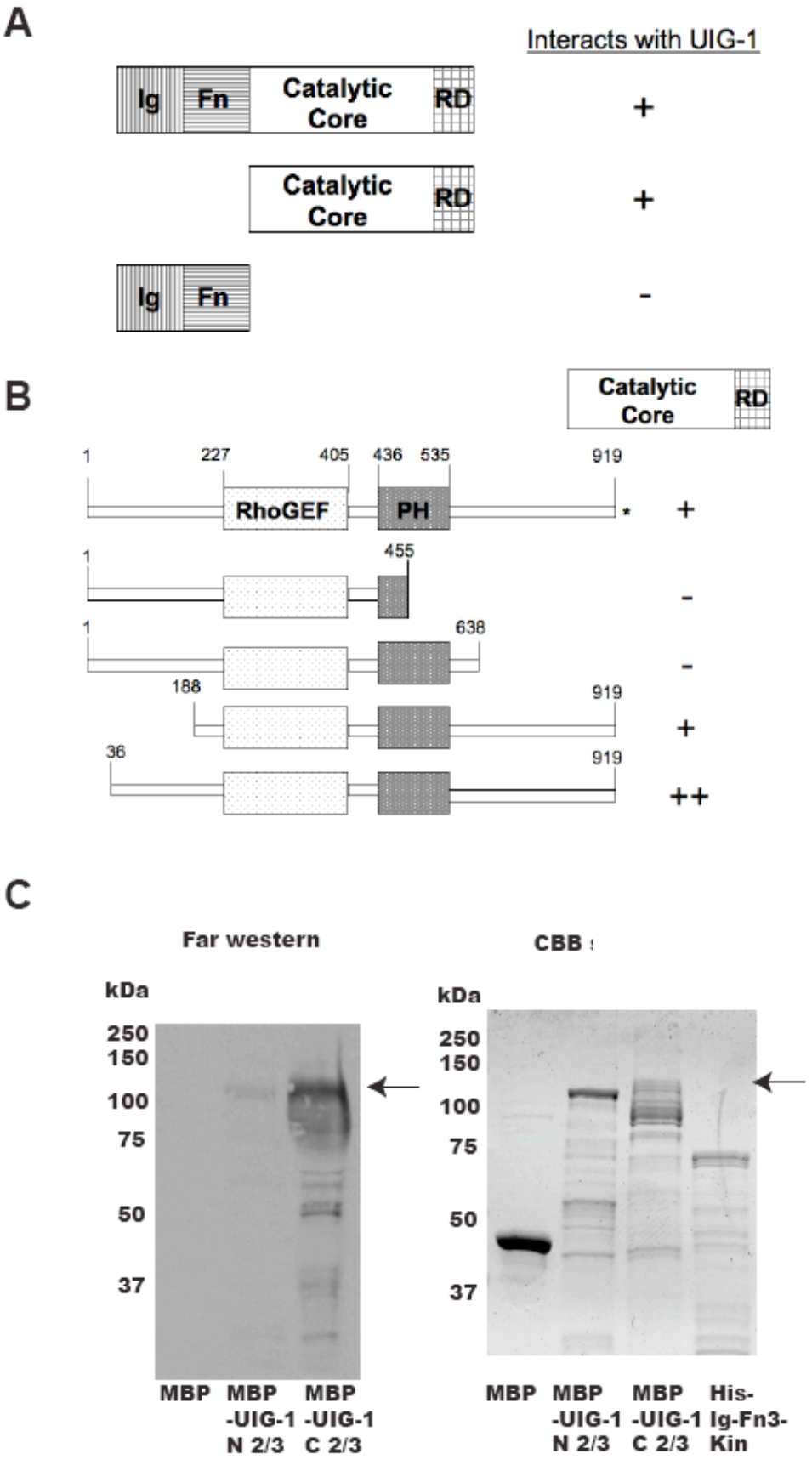

A
Interacts with UIG-1
Ig
Fn
Catalytic Core
RD
+
+
-
B
Catalytic Core
RD
1
227
405
436
535
919
RhoGEF
PH
*
+
1
455
-
1
638
-
188
919
+
36
919
++
C
Far western
CBB
kDa
250
150
100
75
50
37
MBP
MBP -UIG-1 N 2/3
MBP -UIG-1 C 2/3
His-Ig-Fn3-Kin

Figure 3.1- TTN-1 kinase interacts with the Cdc42 GEF UIG-1 A. Using yeast-2-hybrid screening of the M-line/dense body bookshelf, an interaction between TTN-1 IgFn3Kin (bait) and the full length UIG-1 (prey) was discovered. Deletion derivatives of the original bait showed that the minimal region for interaction with UIG-1 consists of kinase domain alone (catalytic core plus regulatory region). B. The 2-hybrid assay was used to determine which portions of UIG-1 are required for the interaction. Results show that the N-terminal 187 amino acids are dispensable. However, when only the N-terminal 36 amino acids are deleted, the interaction between the proteins is stronger in the 2-hybrid assay (measured by the ability of the yeast strain to grow on selective media; please see methods). C. A far western assay was employed to confirm the interaction between TTN-1 kinase and UIG-1. The N terminal 2/3 (aa 1 – 638) and C terminal 2/3 (aa 188 – 919) of UIG-1 were resolved on SDS PAGE and transferred to nitrocellulose membrane. The membrane was incubated with a solution containing 6-his-TTN-1 IgFn3Kin. Binding of the kinase to the blotted proteins was detected with an anti-6-his antibody. Results show that TTN-1 kinase interacts strongest with the C-terminal 2/3 of UIG-1, which is consistent with the 2-hybrid results. CBB staining shows the proteins used for the far western. Arrows indicate the positions of MBP-UIG-1 C-terminal 2/3 fragment.

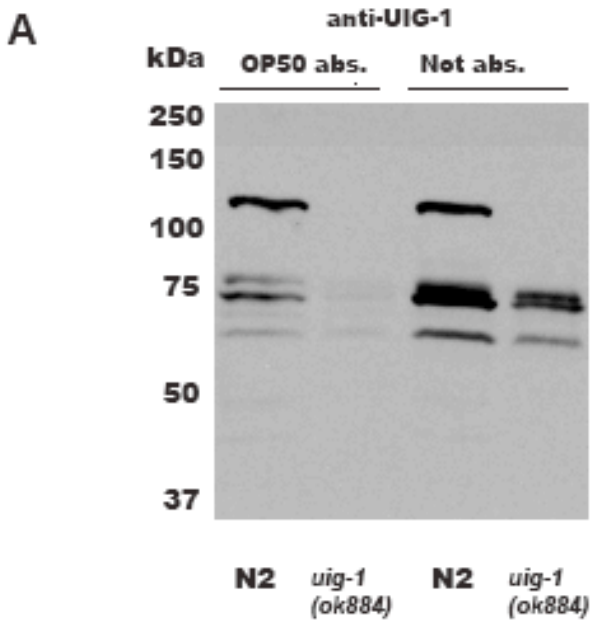
Plasmid name	Gene name	Gene ID	character
pGAD-u89 kin 15/14	*unc-89*	C09D1.1	Giant protein, pk2 region
pGAD-u89 KSP	*unc-89*	C09D1.1	Giant protein, KSP repeat
pGAD-u89 pk1 AB	*unc-89*	C09D1.1	Giant protein, pk1 region
pGAD-u89 pk1 CD	*unc-89*	C09D1.1	Giant protein, inter kinase
pGAD-UNC-98	*unc-98*	F08C6.7	Zinc finger
H11-10	*unc-96*	F13C5.6	no homology
pGAD-u96-62	*unc-96*	F13C5.6	no homology, C-terminus
pGAD-UNC-97	*unc-97*	F14D12.2	LIM domain, PINCH
i7-207	*unc-95*	Y105E8A.6	LIM domain
R1-5-2	*myo-3*	K12F2.1	MHC A C-term
pGAD-unc-54 C-term	*unc-54*	F11C3.3	MHC B C-term, not M-line
i6-102	*lim-8*	F28F5.3	LIM domain
i4-306	*lim-9*	F25H5.1	LIM domain
pGAD-UNC-112	*unc-112*	C47E8.7	ERM domain
pGAD-PAT-4	*pat-4*	C29F9.7	integrin linked kinase
pGAD-PAT-6	*pat-6*	T21D12.4	CH domain, actopaxin
pGAD-CeTalin		Y71G12B.11	Talin
pGAD-CeFAK	*kin-32*	C30F8.4	focal adhesion kinase
pGAD-Cepaxillin		C28H8.6	paxillin (LIM domain only)
pGAD-pat-3 cyto	*pat-3*	ZK1058.2	integrin beta, cytoplasmic
pGAD-u82-17	*unc-82*	B0496.3	kinase, kinase region
pGAD-u82-123456	*unc-82*	B0496.3	kinase, full length
pGAD-DEB-1	*deb-1*	ZC477.9	vinculin, full length
pGAD-DEB-1 head	*deb-1*	ZC477.9	vinculin, head region
pGAD-ATN-1 L	*atn-1*	W04D2.1	alpha-actinin, long form
pGAD-ATN-1 S	*atn-1*	W04D2.1	alpha-actinin, short form
pGAD-DIM-1	*dim-1*	C18A11.7	3 Ig domains
pGAD-unc-15	*unc-15*	F07A5.7	paramyosin
F6-1	*uig-1*	F32F2.1	Rho GEF
F9-2	*hum-6*	T10H10.1	Unconventional myosin
25B	*scpl-1*	B0379.4a	phosphatase, a form
135Aa	*scpl-1*	B0379.4b	phosphatase, b form

Table 3.1- Collection (bookshelf) of proteins used to screen for TTN-1 kinase binding partners. This bookshelf collection consists of known components of M-lines and/or dense bodies in *C. elegans* body wall muscle.

TTN-1 partially co-localizes with UIG-1

Previous studies of *C. elegans* body wall muscle have localized GFP-UIG-1 to the dense bodies (Hikita, Qadota et al. 2005) and TTN-1 to I-bands (Flaherty, Gernert et al. 2002). The published localizations suggested that the kinase region of TTN-1 and UIG-1 might overlap in their localizations. To investigate, antibody staining was employed. The rabbit serum, containing antibodies to UIG-1, was kindly provided by K. Kaibuchi (Nagoya University, Japan). After affinity-purification, the specificty of the antibodies was evaluated by performing a western blot on worm extracts. As shown in Figure 3.2, anti-UIG-1 antibodies react with a ~110kDa protein from wild type worms, but not from the *uig-1* deletion strain. Reactivity to additional lower molecular weight proteins is similar in wild type and the *uig-1* deletion strain, but was reduced upon pre-absorption of anti-UIG-1 with an acetone powder of OP50 *E. coli* (bacteria used to feed *C. elegans*). For localization of TTN-1, I used antibodies called EU143 previously generated to the C-terminal 353 residues of the ~18,500 residue TTN-1 polypeptide (Zastrow, Flaherty et al. 2006), a region that begins ~2000 amino acids C terminal of the kinase domain. These antibodies react on western blot with an ~2 MDa polypeptide from worm extracts (Ma et al., manuscript in preparation).

Indirect immunofluorescence microscopy (Figure 3.2B (top panel)) shows that UIG-1 localizes to dense bodies (marked by α-actinin), consistent with the localization of GFP-UIG-1 reported by Hikita et al (2005). The middle panel shows that UIG-1 and TTN-1 partially co-localize in regions surrounding dense bodies. These results support the hypothesis that there is physiological relevance to the *in vitro* data revealing an interaction between TTN-1 kinase and UIG-1. As further evidence for specificity of


A
anti-UIG-1
kDa
OP50 abs.
Not abs.
250
150
100
75
50
37
N2
uig-1 (ok884)
N2
uig-1 (ok884)

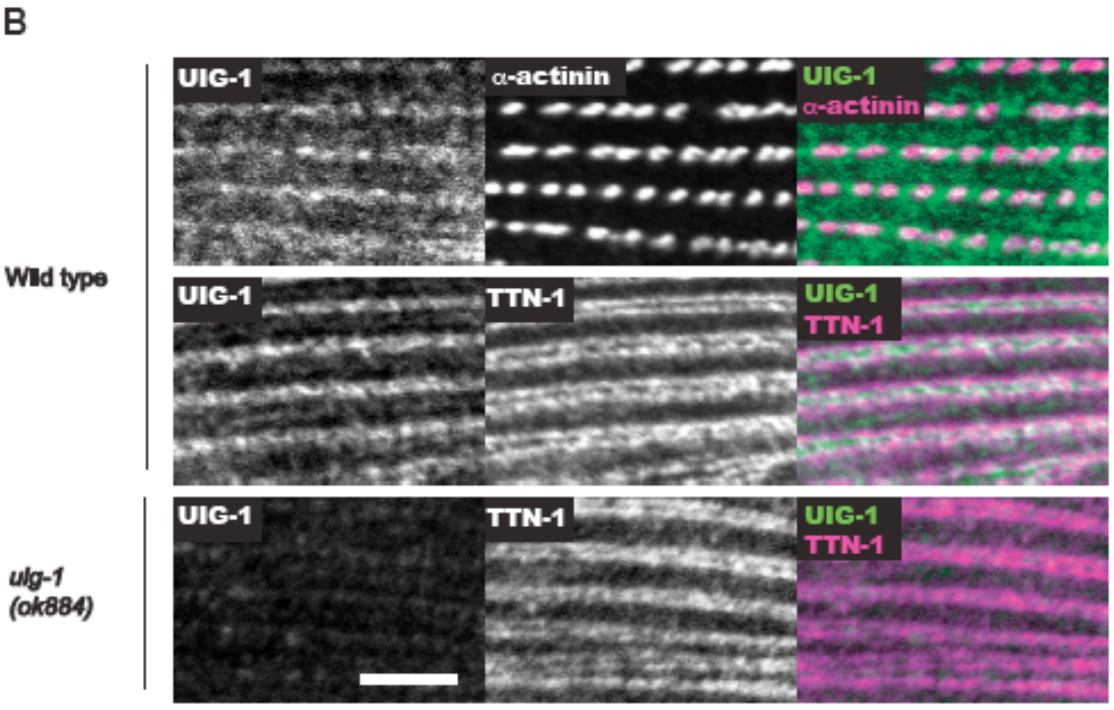

B
Wild type
ulg-1 (ok884)
UIG-1
α-actinin
UIG-1
α-actinin
UIG-1
TTN-1
UIG-1
TTN-1
UIG-1
TTN-1
UIG-1
TTN-1

Figure 3.2- UIG-1 and TTN-1 partially co-localize *in vivo* A. Verification of the specificity of the UIG-1 antibodies. Protein extracts from N2 (wildtype) and *uig-1(ok844)* (an intragenic deletion of *uig-1*) worms were separated by SDS-PAGE, transferred to a membrane, and reacted with affinity purified rabbit anti-UIG-1 antibodies. These antibodies react with a ~110 kDa protein from wild type but not from the *uig-1* deletion strain. Reactivity to additional lower molecular weight proteins is similar in wild type and the *uig-1* deletion strain, but was reduced upon pre-absorption of anti-UIG-1 with an acetone powder of OP50 *E. coli* (bacteria used to feed *C. elegans*). B. Immunoflourescent micrographs showing the localization of UIG-1 and TTN-1. The top panel shows that UIG-1 localizes to dense bodies (marked by α-actinin). The middle panel shows that UIG-1 and TTN-1 partially co-localize in regions surrounding dense bodies. The bottom panel shows the specificity of the anti-UIG-1 antibodies, in which no staining of UIG-1 was detected in the *uig-1* deletion strain. Bar represents 10 μm. Please note that areas of pink and green overlap (co-localization) appear white.

anti-UIG-1 antibodies, I found a greatly reduced fluorescent signal with anti-UIG-1 in the muscle of the *uig-1* deletion strain (Figure 3.2B, bottom panel).

UIG-1 is a substrate for TTN-1 kinase

Although there are no previously reported endogenous substrates of any of the giant protein kinases in *C. elegans*, the enzymatic activity of both TTN-1 and twitchin has been confirmed using model peptides (Lei, Tang et al. 1994; Heierhorst, Tang et al. 1996; Flaherty, Gernert et al. 2002; Greene, Garcia et al. 2008). This invited us to test the catalytic activity of TTN-1 against UIG-1. We first determined if TTN-1 could phosphorylate the nearly full-length UIG-1 fragment, F5-3 (residues 36-919) (Figure 3.3A), as the yeast-2-hybrid results implied a "strong" interaction between the proteins. Indeed, when purified, recombinant kinase and substrate were incubated with radioactively labeled ATP, transfer of the terminal (labeled) phosphate to F5-3 occurred (Figure 3.3B, + lanes). This transfer did not occur when TTN-1 kinase was omitted from the reaction conditions (Figure 3.3B, - lanes).

The large size and general *in vitro* insolubility of the F5-3 fragment prompted us to further break down UIG-1 to begin characterization of the substrate recognition sites. Full length UIG-1 was initially split into four parts (Figure 3.3A). The fragments were expressed and purified either as MBP fusions (UIG-1 A-C) or with an N-terminal 6-his tag (UIG1-D). When incubated with TTN-1 kinase plus reaction buffer, fragments A and C show ^{32}phosphate incorporation, while fragments B and D show no incorporation (Figure 3.3C). We then divided the A and C fragments in half, making A^1, A^{11} and C^1, C^{11}, respectively (Figure 3.3A). The assays revealed that residues targeted for

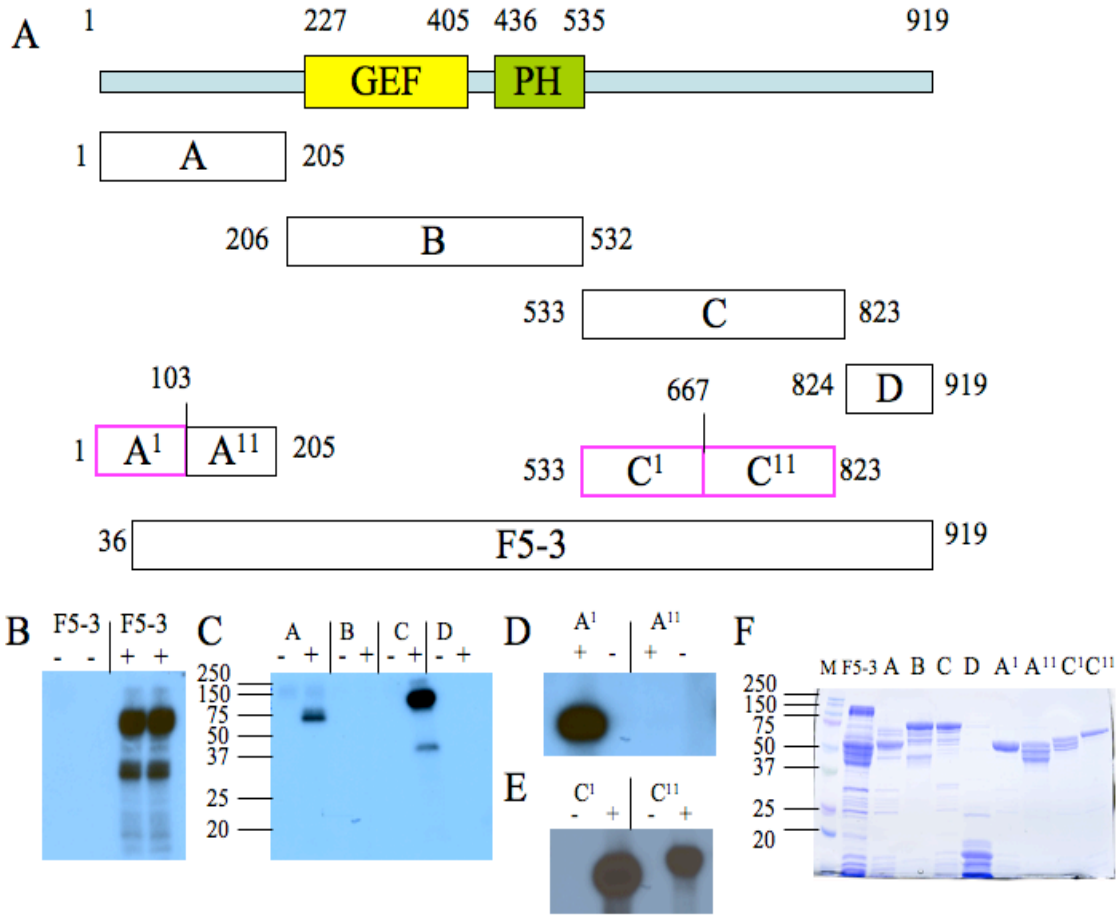

A
1
227
405
436
535
919
GEF
PH
1
A
205
206
B
532
533
C
823
824
D
919
103
667
1
A1
A11
205
533
C1
C11
823
36
F5-3
919
B
F5-3
F5-3
- - + +
C
A
B
C
D
- + - + - + - +
250
150
75
50
37
25
20
D
A1
A11
+ - + -
E
C1
C11
- + - +
F
M F5-3 A B C D A1 A11 C1C11
250
150
75
50
37
25
20

Figure 3.3- UIG-1 is a substrate for TTN-1 kinase A. Domain structure of UIG-1, and portions of UIG-1 expressed as MBP fusion proteins and tested as possible substrates using *in vitro* kinase assays with TTN-1 kinase. B. Phosphorylation of UIG-1 was first shown using the region F5-3 (amino acids 36-919). MBP-F5-3 was purified and incubated with TTN-1 FnKinIg reaction buffer, and radioactively labeled ATP. The reaction was resolved on SDS PAGE and exposed to film. F5-3 incubated with kinase (+ lanes) showed radioactive incorporation, while the reactions without kinase (- lanes) showed no incorporation. C. To begin to identify the phosphorylated regions, UIG-1 was broken down into 4 segments, A-D. The kinase assays showed that both the A and the C fragments are targets of phosphorylation, but the B and D fragments are not. D & E. To further characterize the enzyme/substrate relationship, the A and C fragments were each divided into two halves: A^{1}, A^{11}, C^{1}, and C^{11}. TTN-1 kinase phosphorylates A^{1}, but not A^{11} (Fig 3.3D). In contrast, the kinase phosphorylates both C^{1} and C^{11} (Fig 3.3E). These minimum substrate regions of UIG-1 are highlighted in pink in figure 3.3A. F. ~5μg of each UIG-1 segment tested as a substrate was resolved on SDS PAGE and visualized with CBB. The input gel shows that the fragments were fairly pure and of similar concentration.

phosphorylation lie in A^1 (amino acids 1-103), C^1 (amino acids 533-667), and C^{11} (amino acids 667-823) (Figure 3.3A, pink highlight).

Partial characterization of TTN-1 kinase

The most accurate interpretations of enzyme/substrate relationships are derived from studying their *in vitro* properties as functions of substrate concentrations, time, enzyme concentration, and product formation. Ultimately, the reactions kinetics are measured, defining the affinity of the enzyme for the substrate (K_M), the maximum velocity of the reaction (V_{max}), and the turnover of substrate to product per second (k_{cat}). In order to derive these important parameters, accurate, quantitative, and reproducible enzyme assays must be developed. We tested the ability of the enzyme to catalyze linear product formation and reproducible activity by monitoring the TTN-1 kinase catalyzed phosphotransfer of γ^{32}ATP onto the UIG-1 C^1 fragment as a function of time (Figure 3.4). The results support that the enzyme maintains stability on ice for at least 4 days and has consistent activity between reactions. As these results are highly encouraging, future work will continue to define the kinetics of the TTN-1 kinase catalyzed phosphotransferase reaction (please see chapter 3 discussion).

DISCUSSION

One ultimate goal of our lab is to understand the molecular mechanisms governing the assembly and maintenance of the sarcomere, the basic contractile unit of a muscle cell. We give particular focus to the muscle giant protein kinases that are found in muscles from throughout the animal kingdom. One of these giant kinases in *C. elegans* muscle

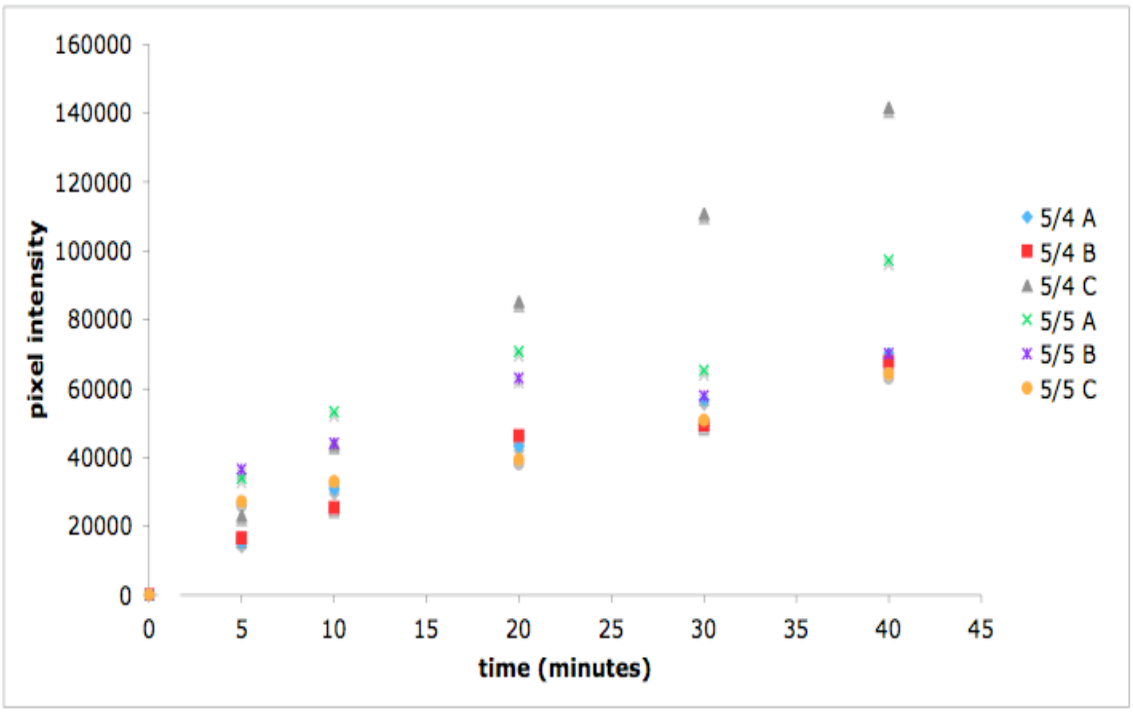


Figure 3.4- Time course of the enzymatic activity of TTN-1 using UIG-1 C[1] as a substrate. To test both the stability and the reproducibility of the enzyme assays, fresh TTN-1 kinase was purified and the activity against UIG-1 C[1] was measured as a function of time. The assays were performed on two separate days (labeled 5/4 and 5/5) in triplicate on each day. The results indicate that the enzyme maintains stability and has consistent activity between reactions.

is TTN-1, a twitchin/titin hybrid. We wished to identify proteins that interact with the kinase domain of TTN-1. After yeast-2-hybrid screening using TTN-1 IgFnKinase as bait, and a bookshelf (Table 3.1) consisting of M-line and dense body specific muscle proteins as prey an interaction between the Cdc42 GEF UIG-1 and the kinase domain of TTN-1 was identified (Figure 3.1A). This interaction was first appealing to us because there is precedence for the Rho-type signaling through the giant kinases. This is apparent in UNC-89 (*C. elegans*) and the mammalian homolog obscurin, each of which contain DH domains (Benian, Tinley et al. 1996; Russell, Raeker et al. 2002; Kontrogianni-Konstantopoulos, Jones et al. 2003; Small, Gernert et al. 2004), which at least in other proteins have been shown to have GEF activity for activating Rho family small GTPases. Indeed, our laboratory has preliminary data showing that the DH domain of UNC-89 specifically activates RhoA in nematode body wall muscle (H. Qadota et al., manuscript in revision).

Yeast-2-hybrid interactions can often yield false positives (i.e.-proteins that only interact under the constraints of the non-native setting). Experiments were conducted to acquire additional *in vitro* and *in vivo* evidence for the interaction. *In vitro*, a far western assay confirmed the regions of binding within UIG-1 that furnish the specificity for the interaction (Figure 3.1B). This assay provided the first evidence that the binding between TTN-1 kinase and UIG-1 is direct. Co-localization staining with polyclonal antibodies against UIG-1 and TTN-1 shows that in *C. elegans*, the proteins partially co-localize at the outer edges of the dense bodies (Figure 3.2B), supporting the hypothesis that physiologically there is functional merit to the interaction.

The idea that UIG-1 might act as a substrate for TTN-1 was inviting to us, and we therefore tested the ability of TTN-1 to phosphorylate UIG-1. Indeed, the kinase readily transferred terminal phosphates from ATP onto UIG-1 (Figure 3.3B), with specificity to different regions within UIG-1 (Figure 3.3C-E). We showed that the enzymatic reactions are linear over time and are also quantitatively reproducible (Figure 3.4). These results have identified the first substrate for a protein kinase domain of a muscle giant kinase in invertebrates. The data also provides the first evidence for the mechanism by which a muscle giant kinase signals to a Rho family small GTPase.

UIG-1 was originally discovered using a yeast-2-hybrid screen to identify proteins that interact with UNC-112 (Mig-2 in humans), a critical component of the focal adhesion-like dense bodies of nematode body wall muscle (Rogalski, Mullen et al. 2000; Hikita, Qadota et al. 2005). Decreasing UIG-1 expression using RNAi results in abnormal body wall muscle (Hikita, Qadota et al. 2005). Domain analysis of UIG-1 showed that the protein has a DH domain followed by a PH domain. Blast analysis indicates that UIG-1 has close homologs in other organisms. These include CDC24 (854 residues) from *S. cerevisiae*, CG30115 (1593 residues) from *Drosophila melanogaster*, and human KIAA1209 (1385 residues). Comparison of the DH/PH domains among these proteins shows that, as compared to nematode UIG-1, the yeast protein is 21% identical, the fly protein is 46% identical and the human protein is 33% identical (Figure 3.5). The ability of UIG-1 to catalyze GDP release (i.e.-to perform as a GEF) was tested *in vitro* (Hikita, Qadota et al. 2005). These assays confirmed that UIG-1 harbors Cdc42 specific GEF activity by promoting the dissociation of GDP and subsequent binding of GTP to

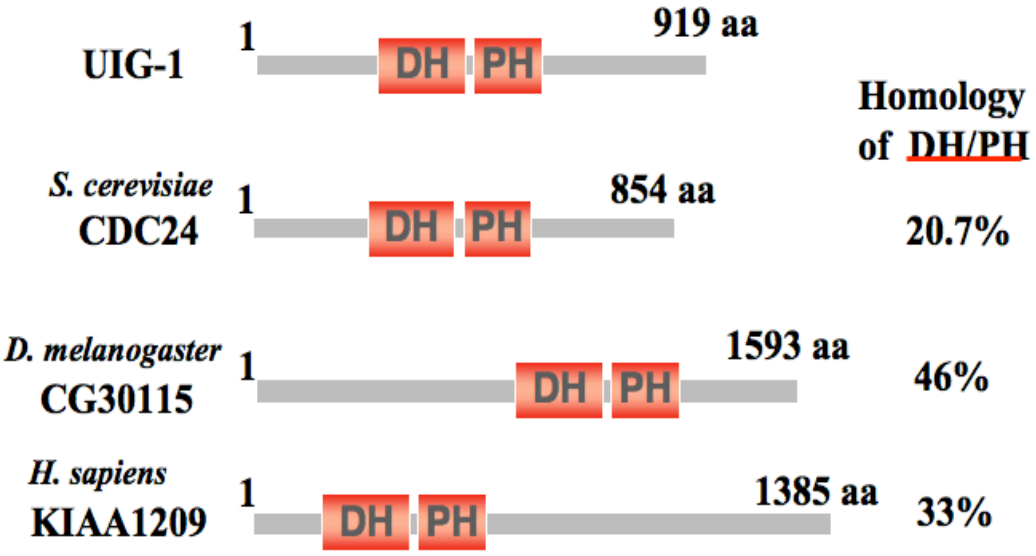


Figure 3.5- Sequence homology of the DH/PH region of UIG-1. Using homology searches and sequence alignments, the DH/PH region of UIG-1 was found to have similarity to other proposed GEFs from yeast, fly, and humans.

Cdc42. Further, the expression of ectopic UIG-1 in NIH3T3 cells induced filipodia formation (Hikita, Qadota et al. 2005), a known characteristic that results when constitutively active Cdc42 (Cdc42V12, a mutation that abolishes the GTP hydrolysis activity) is expressed in fibroblasts (Ridley and Hall 1992; Khosravi-Far, Chrzanowska-Wodnicka et al. 1994; Ridley and Hall 1994). An intragenic deletion of *uig-1* causes abnormal body wall structure that is similar, although not as severe as, Cdc42 depleted (by RNAi) body wall muscle, and is consistent with an important role for the GEF activity of UIG-1 in maintaining normal myofibrillar structure (Hikita, Qadota et al. 2005). It is thought that UNC-112 targets UIG-1 to its anatomical location at the dense bodies, however the mechanism by which UIG-1 is regulated is unknown. In addition to interacting with TTN-1, UIG-1 has been shown by our lab (and through a collaboration with Don Moerman's lab at University of British Columbia) to interact with talin (Ce Talin), vinculin (DEB-1), paxillin (PXL-1) and UNC-95.

Cdc42 is a low molecular weight GTP binding/hydrolyzing protein (GTPase) (Chen, Lim et al. 1993). When GTP is bound to the active site, the protein is considered biologically active; when GDP is bound, the protein is considered biologically dormant. Many of the closest homologs of Cdc42 (members of the Rho subfamily of the p21 ras superfamily) have critical roles in numerous biological processes, particularly cytoskeletal organization (Hall 1998). For Cdc42 specifically, these roles include regulation of the actin cytoskeleton (Hall 1998), cell adhesion (Fukata and Kaibuchi 2001), endocytic trafficking (Balklava, Pant et al. 2007), and cell polarization (Nobes and Hall 1999). In *C. elegans*, Cdc42 is expressed throughout the lifespan with peak expression occurring during embryogenesis (Chen, Lim et al. 1993). During

embryogenesis, Cdc42 has proven critical for cell polarity and spindle orientation (Nobes and Hall 1999; Kay and Hunter 2001) and therefore it is not surprising that RNAi against Cdc42 is embryonic lethal. Using L1 feeding to avoid the early developmental effects of Cdc42 knockdown, muscle abnormalities were assessed (Hikita, Qadota et al. 2005). The knockdown resulted in movement defects and disorganized, discontinuous actin filaments. The similar muscle phenotypes that occur when either Cdc42 or UIG-1 are diminished, and the substrate interactions between both UIG-1/Cdc42 and TTN-1 kinase/UIG-1 suggest that there might be an upstream regulatory role of the phosphorylation of UIG-1 by TTN-1 kinase.

The main function of a GEF is to regulate the on/off switch of a GTPase. Of much interest are the spatial and temporal mechanisms that the cell has designed to execute this precise control. Regulation via phosphorylation is a common cellular theme, given the reversibility of the transfer and the high intracellular ATP concentration (readily available substrate). Thus, many GEFs have adopted phosphorylation as a means to provide tight control over their structural conformations and localization, directly affecting the ability to bind their specific GTPase. Mammals (Yamauchi, Chan et al. 2005; Fujishiro, Tanimura et al. 2008), plants (Zhang and McCormick 2007), and flies (Forsthoefel, Liebl et al. 2005; Elwell, Ceesay et al. 2008) all have examples of targeted GEF phosphorylation effecting the activation of a GTPase, and these studies have been supported both *in vivo* (Forsthoefel, Liebl et al. 2005; Yamauchi, Chan et al. 2005; Zhang and McCormick 2007; Elwell, Ceesay et al. 2008; Fujishiro, Tanimura et al. 2008) and *in vitro* (Yamauchi, Chan et al. 2005). There is also precedence for phosphorylation of guanine nucleotide dissociation inhibitors (GDIs) to regulate the

binding of GDIs to GTPases (DerMardirossian, Rocklin et al. 2006; DerMardirossian and Bokoch 2006), further bolstering the hypothesis that UIG-1 phosphorylation has physiological implications in the behavior of the GEF.

Given the preliminary results, we are highly motivated to understand the physiological relevance of phosphorylation of UIG-1 by TTN-1. UIG-1 is known to have 3 characteristics: (1) it interacts with UNC-112; (2) it is localized to dense bodies, and (3) it has Cdc42 GEF activity. What is the effect of phosphorylation on each of these activities? We currently have 2 regions that are targeted for phosphorylation implying we can determine if the phosphorylation of separate regions is responsible for modifying separate activities.

We will first ask if phosphorylation affects the interaction of UIG-1 with UNC-112. Ideally, we will set-up 2-hybrid assays between TTN-1 and UIG-1 in which putative phosphorylated serines or threonines have been mutated to aspartic acid (phosphomimetic) and to alanine (nonphosphorylatable). According to Hikita et al. 2005, the minimal region of UIG-1 for interaction with UNC-112 is the N-terminus, just before the DH domain (our A segment, Figure 3.3A) (Hikita, Qadota et al. 2005). Thus, it is tempting to hypothesize that phosphorylation of A might be critical for the interaction of UIG-1 with UNC-112.

Another possibility is that phosphorylation of UIG-1 targets the protein to specific locations within the sarcomere. To determine if phosphorylation effects the localization of UIG-1 we will make transgenic worms carrying GFP fusions of UIG-1. In a *uig-1* deletion strain, we will inject constructs in which various sets of phosphorylatable residues have been converted to aspartic acid and/or alanine. Using fluorescence

microscopy we will determine whether these mutant UIG-1 proteins still localize to dense bodies.

Finally, we can examine if phosphorylation affects the GEF activity of UIG-1. For these experiments we will set-up *in vitro* GEF assays with Cdc42 and two portions of UIG-1, amino acids 1-532 (fragments A+ B) and amino acids 206-823 (fragments B + C). Previous studies used amino acids 1-537 of UIG-1 to show GEF activity (they could not express full length UIG-1, personal communication). However, if we used this fragment alone it would not allow us to investigate the effect of phosphorylation of the "C" region of UIG-1. We will therefore show that A+B and B + C have GEF activity towards Cdc42, but not Rho1. To evaluate the effect of phosphorylation of GEF activity we will first incubate UIG-1 (A + B or B + C) with TTN-1 (in kinase buffer plus ATP). After the kinase reaction is complete, we will incubate the reaction products with Cdc42 and look for a difference in the GEF activity relative to UIG-1 fragment incubated without kinase and reactions with kinase only (no UIG-1). If there is an effect of Cdc42 activity when UIG-1 is phosphorylated we will explore the following *in vivo* functional effects.

If phosphorylation decreases or inactivates the GEF activity we will make transgenic worms expressing variant UIG-1 under the heat shock promoter. UIG-1 will have the phosphorylated serines and/or threonines mutated to alanine. The variant UIG-1 would be incapable of being inactivated and would therefore be constitutively active. If an abnormal muscle phenotype is observed relative to overexpression of wild type UIG-1, then we will also examine if constitutively active (dominant active) Cdc42 has the same phenotype.

If phosphorylation increases or activates the GEF activity we will make transgenic worms expressing variant UIG-1 under the heat shock promoter. These variants will differ from the above because these will be phosphomimetic, changing the serines and/or threonines to aspartic acid. The resulting UIG-1 would be considered constitutively active. If this has a phenotype, then we will examine if constitutively active (dominant active) Cdc42 has the same phenotype.

In addition to understanding the physiological implications of the phosphorylation, we would also like to gain a better biochemical understanding of TTN-1 kinase. This will require experiments leading to the accurate determination of V_{max}, K_M and K_{cat} for TTN-1 against the substrate fragments. Preliminary data has supported the development of quantitative phosphotransferase assays in our laboratory (please see chapter 3 results). These assays will be further developed and the kinetic parameters of TTN-1 kinase for UIG-1 will be measured.

In conclusion, we have identified the first substrate interaction for a *C. elegans* giant protein kinase, and the first insight into a molecular mechanism by which a giant protein kinase is connected to Rho GTPase signaling. We have data to support the interaction and have located multiple regions within the full-length protein that are targeted for phosphorylation. We are actively trying to determine the physiological relevance and biochemical properties of the phosphorylation. These results have implication in understanding the intracellular signals involved in the maintenance of the sarcomere.

MATERIALS AND METHODS

Yeast-2-Hybrid Screen

To identify binding partners and/or substrates for TTN-1 kinase I screened a bookshelf collection of known M-line and/or dense body proteins (Miller, Qadota et al. 2006; Qadota, Mercer et al. 2007; Qadota, McGaha et al. 2008) by the 2-hybrid method. The Ig-Fn3-kinase segment of TTN-1 was used to screen the bookshelf both as bait and as prey. The bait plasmid construction had been described in Qadota et al (Qadota, McGaha et al. 2008). To prepare the corresponding prey plasmid, cDNA was amplified from the bait plasmid using synthesized primers

5' TTN-1 IgFn3Kinase GTACGGATCCAAGCTCACTATGGATC

3' TTN-1 Kinase GATCCTCGAGTTAGAGTCTTCTGAAAATGGCTCC

and cloned into the BamHI/XhoI sites of the pGAD-C1 vector (James, Halladay et al. 1996). Yeast strain PJ69-4A containing TTN-1 kinase bait was transformed separately with each of the bookshelf preys. Similarly, PJ69-4A was co-transformed with TTN-1 kinase prey and each of the bookshelf baits to obtain the opposite combination. In both cases, the transformants were streaked onto selective media lacking histidine or adenine. The yeast 2-hybrid assay was considered "strong" if growth was observed on both histidine (-) and adenine (-) plates; and was considered "weak" if growth was observed only on histidine (-). For a more detailed description of the yeast-2-hybrid method please see Mackinnon et al. (Mackinnon, Qadota et al. 2002).

Domain mapping of interacting proteins

To determine which portion of TTN-1 was minimally required for interaction with UIG-1, two deletion derivatives of TTN-1 Ig-Fn-kinase were prepared: kinase (catalytic core plus regulatory domain) and Ig-Fn3, using the following primers to amplify each fragment by PCR:

5' TTN-1 Kinase GGTACGGATCCAGAGTTGTCGAGGAGCCTAAG

5' TTN-1 IgFn3 GTACGGATCCAAGCTCACTATGGATC

3' TTN-1 Kinase GATCCTCGAGTTAGAGTCTTCTGAAAATGGCTCC

3' TTN-1 IgFn3 GGACTCTCGAGTTA TTGTTCTTTCACTTTAACTTC

These fragments were cloned into pGBDU-C1 using BamHI/XhoI or SalI and tested for interaction with UIG-1 full length prey. To determine which region of UIG-1 is required, we tested 4 deletion derivatives of UIG-1 originally described by Hikita et al. (2005) versus TTN-1 kinase (catalytic core plus regulatory domain).

Far Western Assay

Using recombinant, purified, TTN-1 kinase and UIG-1 the yeast-2-hybrid interaction was supported using a far western approach. TTN-1 Ig-Fn3-kinase was subcloned into pET28a+, fusing an in frame 6-histidine (6-his) tag to the N-terminus. The 6-his TTN-1 IgFn3-Kinase was purified as previously described for a similar 6-his TTN-1 kinase construct (Greene, Garcia et al. 2008). The N-terminal 2/3 (amino acids 1-638) or C-terminal 2/3 (amino acids 188-919) of UIG-1 was subcloned into the pMAL-KK1 expression vector, enabling the production of MBP fusion proteins. *E.coli* BL21 (DE3) RIL (Stratagene) harboring the MBP-UIG-1 expression vectors or the empty vector (for

MBP production) was grown for 2 hours at 37° C and induced with 1 mM IPTG for 4 hours at 25° C. MBP fusion proteins were purified using amylose resin as described (Mercer, Miller et al. 2006). The purity of the proteins was confirmed using SDS PAGE analysis and the protein concentration was estimated using the Bradford method. 5 mg of MBP, MBP-UIG-1-N-term 2/3, MBP- UIG-1-C-term 2/3 was resolved on a 10% SDS PAGE and transferred onto a nitrocellulose membrane. The membrane was first blocked overnight with 5% milk TBS-T before incubation with 5 mg / ml of 6-his-TTN-1-IgFn3Kinase in 5% milk TBS-T in a bag with a total volume of 4 ml for 1 hour at room temperature. After washing, the membrane was incubated with rabbit anti-6-his antibody (SC803 from Santa Cruz; 1:200 dilution) and washed again. Finally, the membrane was incubated with anti-rabbit-HRP (Amersham; 1:10,000 dilution), washed, and the binding was visualized using ECL (Pierce).

Worm Strains and Culture

Nematodes were grown at 20° C on NGM plates with *Escherichia coli* strain OP50 as food source (Brenner 1974). Bristol N2 was the wild-type strain and mutants included an intragenic deletion of *uig-1*, designated *uig-1(ok844)*, and obtained from the *C. elegans* Gene Knockout Consortium.

Western Blot and Immunofluorescence Microscopy

Procedures for preparing worm protein lysates and western blots were described previously (Mercer, Miller et al. 2006). Wild type (N2) and *uig-1(ok844*) were used to make lysates. The UIG-1 rabbit antibodies (T. Hikita and K. Kaibuchi, unpublished data)

were kindly provided by K. Kaibuchi (Nagoya University, Japan). These antibodies were affinity purified to an MBP fusion of residues 1-638 of UIG-1 using procedures described in Mercer et al. (Mercer, Flaherty et al. 2003). Affinity-purified anti-UIG-1 was used at 1:200 dilution. These antibodies were pre-absorbed with OP50 bacteria by mixing the purified antibodies with an acetone powder of OP50 bacteria at 25 mg/ml concentration, incubating for 1 hr at room temperature, and centrifugation for 5 minutes twice. Wild type adult worms were immunostained after fixation by the method described by Nonet et al. (Nonet, Grundahl et al. 1993). Anti-UIG-1 was used at 1:100 dilution, and anti-TTN-1 (EU143;(Zastrow, Flaherty et al. 2006); Ma et al., unpublished data) was used at 1:100 dilution. Anti-UIG-1 antibody was preincubated with *uig-1(ok844)* acetone powder to reduce background staining. Secondary antibodies for anti-UIG-1 and anti-TTN-1 are anti-rabbit antibody Alexa 488 conjugated and anti-rat antibody Alexa 594 conjugated, respectively. Confocal microscopy was carried out as described in Qadota et al. (Qadota, Mercer et al. 2007).

Kinase assays

The kinase assays were performed using recombinant 6his-TTN-1 FnKinIg (for purification details see (Greene, Garcia et al. 2008)) and MBP fusions of UIG-1 (except for one construct, fragment D, which was 6his tagged). To prepare the initial substrate, MBP-F5-3 (UIG-1 residues 36-919), the XhoI fragment of the F5-3 prey plasmid was cloned into pMAL-KK2. The A-D, A1, A11, C1, and C11 portions of UIG-1 were amplified using the bookshelf prey vector as a template and the primer pairs listed below. Each fragment, with the exception of D, was subcloned directly into the pMAL-KK1

expression vector fusing an inframe MBP to the N-terminus. UIG-1 D was subcloned into pET28a+ by inserting BamHI and XhoI restriction sites, allowing the fusion of an in frame 6-his tag to the N-terminus of the fragment.

UIG-1 A 5’ GCAGGCTCGAGATGATATCAGAACGGATACAACGAT,

3’ GCAGGTCTAGATTAAGATAAGGAAAATGTGTTTGATATAG;

UIG-1 B 5’ GCAGGCTCGAGTCAAACACATTTTCCTTATCTTCCC,

3’ GCAGGTCTAGATTATCGTTTCAATTCTTGCATCCAAGTTCTTTTC;

UIG-1 C 5’ GCAGGCTCGAGATGATGTTAGATCATTATTCCGTAGAG,

3’ GCAGGTCTAGATTATTGGTCTACTTTTGAATATCGTCTTTCAATG;

UIG-1 D 5’ GCAGGGGATCCCTGAAAAAGCGATCAAGAAAATATCAAG,

3’ GCAGGCTCGAGCTAAGCATTCTCGTATTGTCGTATTATAG;

UIG-1 A1 5’ GCAGGCTCGAGATGATATCAGAACGGATACAACGAT,

3’ GCAGGTCTAGATTACAAAGGTTCAATTAATTTTTTGCTG;

UIG-1 A11 5’ GCAGGCTCGAGCCAATGATCATCACTCCTC,

3’ GCAGGTCTAGATTAAGATAAGGAAAATGTGTTTGATATAG;

UIG-1 C1 5’ GCAGGCTCGAGATGATGTTAGATCATTATTCCGTAGAG,

3’ GCAGGTCTAGATTAAGGTCGGCAGTAGTTAGGATCAGTTG;

UIG-1 C11 5’ GCAGGCTCGAGCCAGTATGTCGAACGCAATCAG

3’ GCAGGTCTAGATTATTGGTCTACTTTTGAATATCGTCTTTCAATG;

Unless otherwise noted, 5’ XhoI and 3’ XbaI restriction sites were inserted and used for the subcloning procedures.

The MBP fusion proteins were expressed and purified as described above. To purify 6-his D, growth and induction paralleled the MBP fusions, however, a nickel

column was attached to an FPLC for purification, as described previously (used same protocol as for importin α) (Conti, Uy et al. 1998). The protein purity was confirmed using SDS PAGE analysis and the protein concentration was estimated using the Bradford method. 5-20 μg of UIG-1 substrate was added to reaction buffer (20mM TRIS pH7.4, 10mM magnesium acetate, 0.05% triton, 1mM DTT) containing 0.4mM ATP and 0.5μCi/μl γ-32ATP. The reactions were initiated by the addition of 50pg/μl kinase and incubated at 30°C for 30 minutes (unless otherwise noted). Upon completion, SDS buffer was added to stop the reactions, which were then heated at 95°C for 5 minutes. The samples were resolved on 10% SDS PAGE and fixed overnight in 40% methanol, 10% acetic acid. To visualize the ^{32}P incorporation, the gels were exposed to film at -80^{o} C using an intensifying screen for 15-30 minutes. To quantify the ^{32}P incorporation the gels were dried and exposed to a general purpose phosphor screen (GE healthcare). The excitation stored on the screen was detected with a Storm Phosphorimager (GE healthcare). The pixel intensity was quantified using ImageQuant software (GE healthcare).

CHAPTER FOUR

Conclusions and Future Directions

Enzymes are the primary catalysts of life. Within any organism, or any cell, thousands of enzymes are working. Enzymes perform in unison to facilitate functional complexity. In a sense, enzymes are somewhat similar to humans working in a factory. In a factory setting, all of the people are working to make the factory run smoothly; they all seemingly have the same job, but underneath such broad stereotyping, each of those workers is an individual, with unique qualities not found in their colleague. Enzymes, like people, all have their own unique biophysical properties that define them. Thus, to understand how biology works, or how life functions, the study of enzymes is critical. This thesis focused on three important aspects of enzymology. First, we examined the structure/function relationship of autoinhibition of giant protein kinases (chapter 2); second, we studied the interactions between a giant protein kinase and its substrate (chapter 3); finally, we evaluated how enzymes evolve when they are placed under selective pressure (appendix A). Together, these studies have enhanced not only the researchers knowledge and appreciation for biochemistry, but have also added to the scientific literature.

Enzyme regulation

Many enzymes require regulation. They must have a means of turning on and off for proper cellular homeostasis. In prokaryotes, this regulation is usually critical to avoid unnecessary energy expenditure (White 1999). In eukaryotes, especially higher animals, this regulation is more important for disease prevention and overall tissue health. A main focus of our lab is to understand the factors that govern sarcomere maintenance and assembly; how healthy muscle tissue is built and sustained. One common theme among

all types of muscle is the presence of giant muscle proteins. Titin, a hallmark giant muscle protein, is an intrasarcomeric filament that maintains the precise structural arrangement of thin and thick filaments (Labeit, Gautel et al. 1992). Titin also acts like a molecular spring, allowing for passive muscle stiffness (Labeit and Kolmerer 1995). In myogenic cells, if titin is absent, sarcomeres fail to assemble (Person, Kostin et al. 2000; van der Ven, Bartsch et al. 2000). Near the C-terminus of titin exists a single protein kinase domain (Labeit and Kolmerer 1995). When titin kinase is not active, the enzyme exists in an autoinhibited conformation. A downstream regulatory tail makes intimate contact with the kinase active site, blocking substrate binding, and therefore, blocking catalysis (Mayans, van der Ven et al. 1998). The signals that govern the removal of the regulatory tail (and hence, turn the enzyme on) remain a mystery. However, what is known is that when titin kinase is active, it sends signals to the nucleus that result in gene expression, including expression of genes involved in myofibril growth and maintenance (Lange, Xiang et al. 2005).

One hypothesis is that the forces generated from muscle activity impose strain on titin and titin-like kinases that allows for the removal of the regulatory tail (Grater, Shen et al. 2005; Lange, Xiang et al. 2005). This hypothesis has been supported nicely *in silico* using molecular simulation dynamics (Grater, Shen et al. 2005). This thesis employed single molecule techniques using MFS to study the unfolding of the giant, titin-like, protein kinases from *C. elegans,* TTN-1 and twitchin (Greene, Garcia et al. 2008). The data reveals that, indeed, the kinase domains can withstand forces greater than would be predicted for proteins that are not exposed to mechanical strain (such as barnase (Best, Li et al. 2001) and apo-calmodulin (Hertadi and Ikai 2002)). Further, the kinase

domains unwind in a step-wise fashion, suggesting that there is a hierarchy of mechanical resistance that is intrinsic to the enzyme. We believe that the results of these experiments bolster the force activation hypothesis, and motivate us to measure substrate binding is effected when low forces (~10 pN) are applied to the kinases.

The next step in testing the force activation hypothesis of giant protein kinases is to measure, in real time, substrate turnover increase as a result of force application. This will require an instrument that can apply force with more sensitivity and selectivity than the MFS. Magnetic tweezers utilize a magnetic field gradient to manipulate micron-sized paramagnetic objects. They are an inviting tool for this experiment, as (1) the instrument can be used to apply 10^{-3}-10^{2} pN (vs. 10-10^{2} pN for MFS) (Neuman and Nagy 2008) and (2) the macromolecule of interest must be tethered to the magnetic bead (the tether acts as a lever to apply force), thus providing specificity for pulling geometry and single molecule attachment. We have designed an experiment employing magnetic tweezers, which can be readily completed in collaboration with the labs of Laura Finzi (Emory Physics), David Dunlap (Emory Cell Biology), and Andres Oberhauser (UTMB Neuroscience and Cell Biology) (Figure 4.1). The project first requires a protein engineer to make a construct suitable for "pulling", which requires specific attachment to both the slide, and the DNA handle/lever that links the protein to the bead. Starting from the base of the construct, a 6-his tag attaches the N terminus of the protein to a nickel coated slide. The kinase and flanking domains follows, and at the C terminus of the protein is a small, monomeric DNA binding protein that has high affinity and specificity for its recognition sequence (we chose pPeng4 from (Montclare and Schepartz 2003)). The DNA attaches to the bead via digoxin/antidigoxin linkage. Modulated force can now

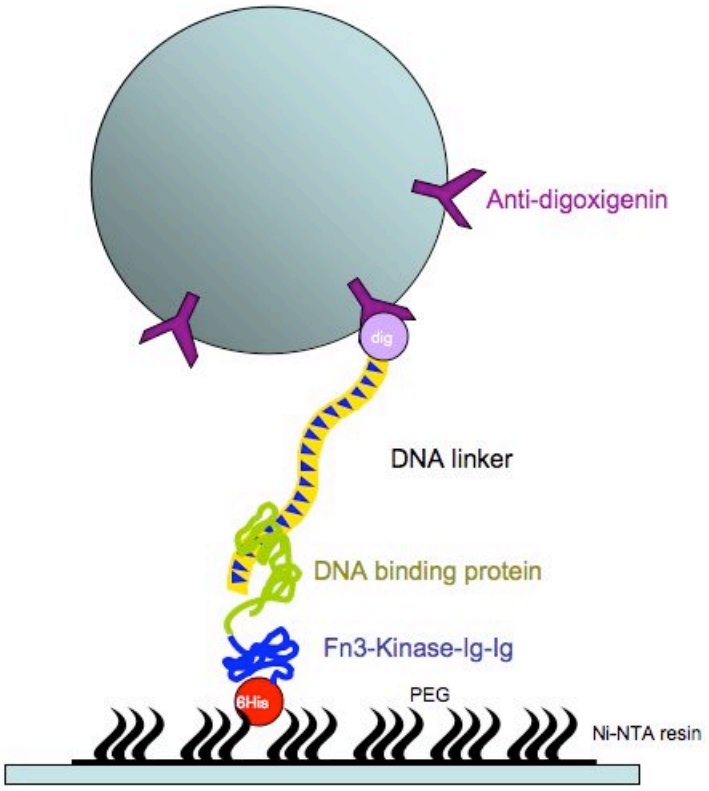


Figure 4.1- Experimental design: Measuring substrate binding of giant protein kinases when force is applied. Using molecular tweezers we will apply force to TTN-1 and twitchin kinase domains and measure the effect on binding to one of the substrates, ATP. The enzymes will be attached to the slide through an N terminal his tag, and attached to the bead through a DNA linkage of a DNA binding protein to its recognition sequence. The DNA linker will attach to the magnetic bead through the digoxin/antidigoxin interaction.

be applied to the kinase. To detect substrate turnover, we will use fluorescently labeled ATP and a total internal reflection fluorescence microscope. If successful, these results will be the first to experimentally test the ability of an enzyme to behave as a force sensor. They will also greatly enhance our understanding of the functions of giant protein kinases in muscle.

Enzyme substrate specificity

Perhaps equally important to understanding enzymes is discerning their substrates. For example, gaining insight into what individual molecules a particular kinase phosphorylates allows for a broader understanding of its cellular roles. Previous to this thesis, there were no known substrates for any of the giant protein kinases in *C. elegans*. In fact, there are very few known substrates for any of the giant muscle protein kinases (Heierhorst, Probst et al. 1995; Mayans, van der Ven et al. 1998; Lange, Xiang et al. 2005). Yeast-2-hybrid screening identified an interaction between TTN-1 kinase and the Cdc42 GEF, UIG-1. After confirming this interaction with an *in vitro* binding assay, we demonstrated the ability of TTN-1 to phosphorylate UIG-1 *in vitro*. Thus, potentially identifying the first protein substrate for any of the giant protein kinases in *C. elegans*.

Among other processes, Cdc42 is known for its roles in regulating the actin cytoskeleton (Hall 1998). Interestingly, an intragenic deletion of *uig-1*, which has no detectable UIG-1 protein by western blot, displays disorganized myofibrils. Given the ability of UIG-1 to activate Cdc42 (Hikita, Qadota et al. 2005), and the ability of Cdc42 to regulate actin (and hence, thin filaments) (Hall 1998) it is quite tempting to hypothesize a link between phosphorylation of UIG-1 by TTN-1 and thin filament

maintenance or assembly. *In vitro*, this hypothesis can begin to be tested by measuring changes in the GEF activity of UIG-1 for Cdc42 after pre-incubation of UIG-1 with TTN-1. We will further test the implications of the UIG-1/TTN-1 substrate interaction by identifying the residues of UIG-1 targeted for phosphorylation. Preliminary mass spectroscopy data on recombinant UIG-1 fragments reveals that there are multiple phosphorylation sites on UIG-1 after incubation with TTN-1 and ATP (data not shown). Moreover, all of these sites reside outside the known domains of UIG-1, namely the DH and PH domains. Once the exact sites are identified, we will employ *in vivo* methods to determine the physiological consequences of UIG-1 phosphorylation. Ultimately, we hope that these experiments provide insight into the functions of giant protein kinases within muscle.

Enzyme evolution

Without evolution, life would be unable to adapt to suit the ever-changing demands of the environment. Understanding how evolution proceeds on a macro level is quite mind-boggling – with so many variables between organisms, the task seems impossible. Studying evolution on a micro level simplifies the complexity, while still allowing for a better understanding of how life evolves. One way to examine evolution on the micro-scale is through enzymes: observing the changes that a single enzyme undergoes when a selective pressure is applied. Of course, the selective pressure must be one that the specific enzyme is capable of battling. For the purpose of this thesis, we created a metabolic dead end that resulted in *E. coli* death (Figure 4.2) (Parikh, Greene et al. 2006). The cells expressed phosphoribulokinaseA (prkA), which converts ribuolse 5

phosphate (a naturally occurring sugar in *E. coli*) into ribulose 1,5 bisphosphate (a toxic, foreign sugar to *E. coli*). To rescue the cells, we co-expressed an enzyme that could metabolize the dead-end product, as long as the enzyme's second substrate, CO_2, was readily available. The enzyme responsible for the sugar breakdown and CO_2 assimilation is ribulose 1,5 bisphosphate carboxylase/oxygenase, better known as Rubisco.

Our hope was to elicit decreases in the K_M of Rubisco for CO2 by decreasing the amount of CO_2 available to the cells. We, in fact, identified hypermorph mutations that conferred *E. coli* growth with decreased concentrations of CO_2. However, despite our desire for kinetic changes, the primary reason for the phenotype was an increased amount of Rubisco expression (Greene, Whitney et al. 2007). The next question became: what is the underlying cause of the increased protein concentration? RNA stability? Thermostability? Chaperone interactions? In fact, we discovered that Rubisco solubility increased by decreasing its natural propensity to mis-fold *in vivo*. In addition to the *in vivo* solubility increase, there were modest enhancements in the catalytic prowess of the variant enzymes. These results demonstrated the utility of using *E. coli*, a non-photosynthetic microbe, to study the primary enzyme responsible for photosynthesis, Rubsico. Rubisco, is responsible for most of the world's biomass but is a slow, non-specific catalyst. Our ultimate goal is to identify the chemical and biological constraints that limit Rubisco's evolutionary potential in nature, and find ways to overcome them.

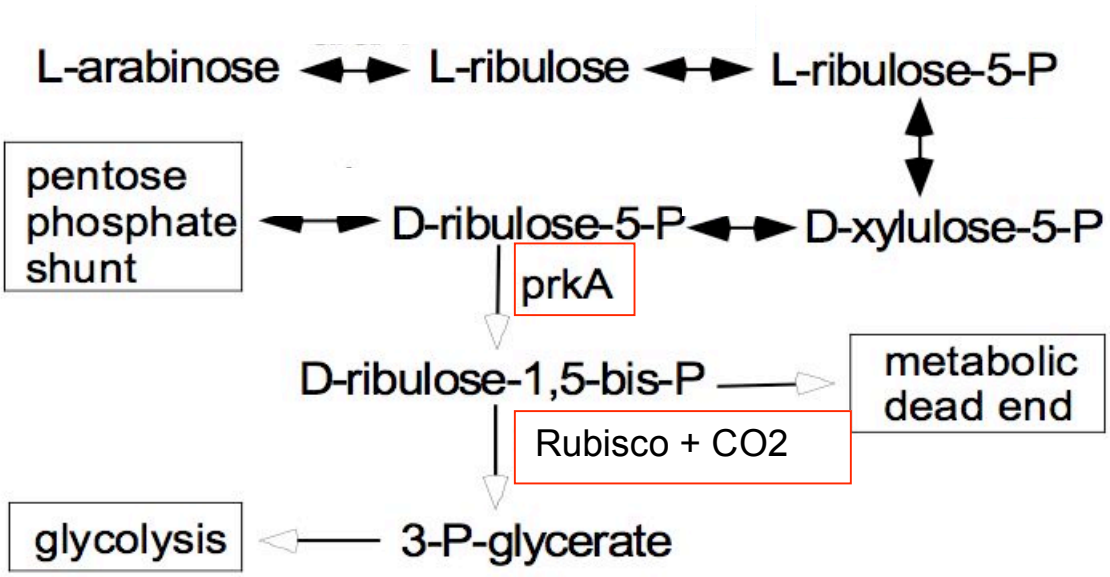


Figure 4.2- Engineering a metabolic dead end in *E. coli.* Cells expressing prkA build up ribulose 1,5 bisphosphate and cease to grow. Upon co-expression of Rubisco, and the addition of exogenous CO_2, the sugar is metabolized and vitality is restored.

REFERENCES

Ainavarapu, S. R., L. Li, et al. (2005). "Ligand binding modulates the mechanical stability of dihydrofolate reductase." Biophys J **89**(5): 3337-44.

Albright, C. F., B. W. Giddings, et al. (1993). "Characterization of a guanine nucleotide dissociation stimulator for a ras-related GTPase." Embo J **12**(1): 339-47.

Altmann, S. M. and P. F. Lenne (2002). "Forced unfolding of single proteins." Methods Cell Biol **68**: 311-35.

Balklava, Z., S. Pant, et al. (2007). "Genome-wide analysis identifies a general requirement for polarity proteins in endocytic traffic." Nat Cell Biol **9**(9): 1066-73.

Barral, J. M., C. C. Bauer, et al. (1998). "Unc-45 mutations in Caenorhabditis elegans implicate a CRO1/She4p-like domain in myosin assembly." J Cell Biol **143**(5): 1215-25.

Benian, G. M., J. E. Kiff, et al. (1989). "Sequence of an unusually large protein implicated in regulation of myosin activity in C. elegans." Nature **342**(6245): 45-50.

Benian, G. M., S. W. L'Hernault, et al. (1993). "Additional sequence complexity in the muscle gene, unc-22, and its encoded protein, twitchin, of Caenorhabditis elegans." Genetics **134**(4): 1097-104.

Benian, G. M., X. Tang, et al. (1996). "Twitchin and related giant Ig superfamily members of C. elegans and other invertebrates." Adv Biophys **33**: 183-98.

Benian, G. M., T. L. Tinley, et al. (1996). "The Caenorhabditis elegans gene unc-89, required fpr muscle M-line assembly, encodes a giant modular protein composed of Ig and signal transduction domains." J Cell Biol **132**(5): 835-48.

Bernards, A. and J. Settleman (2004). "GAP control: regulating the regulators of small GTPases." Trends Cell Biol **14**(7): 377-85.

Bernards, A. and J. Settleman (2005). "GAPs in growth factor signalling." Growth Factors **23**(2): 143-9.

Bernards, A. and J. Settleman (2007). "GEFs in growth factor signaling." Growth Factors **25**(5): 355-61.

Berwick, D. C. and J. M. Tavare (2004). "Identifying protein kinase substrates: hunting for the organ-grinder's monkeys." Trends Biochem Sci **29**(5): 227-32.

Best, R. B., S. B. Fowler, et al. (2002). "A simple method for probing the mechanical unfolding pathway of proteins in detail." Proc Natl Acad Sci U S A **99**(19): 12143-8.

Best, R. B., B. Li, et al. (2001). "Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation." Biophys J **81**(4): 2344-56.

Birkenfeld, J., P. Nalbant, et al. (2007). "GEF-H1 modulates localized RhoA activation during cytokinesis under the control of mitotic kinases." Dev Cell **12**(5): 699-712.

Bos, J. L., H. Rehmann, et al. (2007). "GEFs and GAPs: critical elements in the control of small G proteins." Cell **129**(5): 865-77.

Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.

Bullard, B., T. Garcia, et al. (2006). "The molecular elasticity of the insect flight muscle proteins projectin and kettin." Proc Natl Acad Sci U S A **103**(12): 4451-6.

Bullard, B., W. A. Linke, et al. (2002). "Varieties of elastic protein in invertebrate muscles." J Muscle Res Cell Motil **23**(5-6): 435-47.

Burr, A. H. and C. Gans (1998). "Mechanical significance of obliquely striated architecture in nematode muscle." Biol Bull **194**(1): 1-6.

Bustamante, C., J. F. Marko, et al. (1994). "Entropic elasticity of lambda-phage DNA." Science **265**(5178): 1599-600.

Carrion-Vazquez, M., H. Li, et al. (2003). "The mechanical stability of ubiquitin is linkage dependent." Nat Struct Biol **10**(9): 738-43.

Carrion-Vazquez, M., A. F. Oberhauser, et al. (2000). "Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering." Prog Biophys Mol Biol **74**(1-2): 63-91.

Carrion-Vazquez, M., A. F. Oberhauser, et al. (1999). "Mechanical and chemical unfolding of a single protein: a comparison." Proc Natl Acad Sci U S A **96**(7): 3694-9.

Centner, T., J. Yano, et al. (2001). "Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain." J Mol Biol **306**(4): 717-26.

Chahdi, A. and A. Sorokin (2008). "Protein kinase A-dependent phosphorylation modulates beta1Pix guanine nucleotide exchange factor activity through 14-3-3beta binding." Mol Cell Biol **28**(5): 1679-87.

Chen, W., H. H. Lim, et al. (1993). "The CDC42 homologue from Caenorhabditis elegans. Complementation of yeast mutation." J Biol Chem **268**(18): 13280-5.

Conti, E., M. Uy, et al. (1998). "Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha." Cell **94**(2): 193-204.

Costantin, L. L. (1975). "Contractile activation in skeletal muscle." Prog Biophys Mol Biol **29**(2): 197-224.

Costantin, L. L. (1975). "Electrical properties of the transverse tubular system." Fed Proc **34**(5): 1390-4.

Cox, S., E. Radzio-Andzelm, et al. (1994). "Domain movements in protein kinases." Curr Opin Struct Biol **4**(6): 893-901.

Crespo, P., K. E. Schuebel, et al. (1997). "Phosphotyrosine-dependent activation of Rac-1 GDP/GTP exchange by the vav proto-oncogene product." Nature **385**(6612): 169-72.

Davies, B. A., J. D. Topp, et al. (2003). "Vps9p CUE domain ubiquitin binding is required for efficient endocytic protein traffic." J Biol Chem **278**(22): 19826-33.

Davies, R. E. (1963). "A Molecular Theory of Muscle Contraction: Calcium-Dependent Contractions with Hydrogen Bond Formation Plus Atp-Dependent Extensions of Part of the Myosin-Actin Cross-Bridges." Nature **199**: 1068-74.

DerMardirossian, C., G. Rocklin, et al. (2006). "Phosphorylation of RhoGDI by Src regulates Rho GTPase binding and cytosol-membrane cycling." Mol Biol Cell **17**(11): 4760-8.

DerMardirossian, C. M. and G. M. Bokoch (2006). "Phosphorylation of RhoGDI by p21-activated kinase 1." Methods Enzymol **406**: 80-90.

Elphick, L. M., S. E. Lee, et al. (2007). "Using chemical genetics and ATP analogues to dissect protein kinase function." ACS Chem Biol **2**(5): 299-314.

Elwell, C. A., A. Ceesay, et al. (2008). "RNA interference screen identifies Abl kinase and PDGFR signaling in Chlamydia trachomatis entry." PLoS Pathog **4**(3): e1000021.

Ferrara, T. M., D. B. Flaherty, et al. (2005). "Titin/connectin-related proteins in C. elegans: a review and new findings." J Muscle Res Cell Motil **26**(6-8): 435-47.

Fersht, A. R. and V. Daggett (2002). "Protein folding and unfolding at atomic resolution." Cell **108**(4): 573-82.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-6.

Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-11.

Flaherty, D. B., K. M. Gernert, et al. (2002). "Titins in C.elegans with unusual features: coiled-coil domains, novel regulation of kinase activity and two new possible elastic regions." J Mol Biol **323**(3): 533-49.

Florin, E.-L., Rief, M., Lehmann, H. , Ludwig, M., Dornmair, C., Moy V.T., and Gaub, H.E. (1995). "Sensing specific molecular interactions with the atomic force microscope." Biosensors and Bioelectronics **10**(9-10): 895-901.

Forsthoefel, D. J., E. C. Liebl, et al. (2005). "The Abelson tyrosine kinase, the Trio GEF and Enabled interact with the Netrin receptor Frazzled in Drosophila." Development **132**(8): 1983-94.

Freiburg, A., K. Trombitas, et al. (2000). "Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity." Circ Res **86**(11): 1114-21.

Fujishiro, S. H., S. Tanimura, et al. (2008). "ERK1/2 phosphorylate GEF-H1 to enhance its guanine nucleotide exchange activity toward RhoA." Biochem Biophys Res Commun **368**(1): 162-7.

Fukata, M. and K. Kaibuchi (2001). "Rho-family GTPases in cadherin-mediated cell-cell adhesion." Nat Rev Mol Cell Biol **2**(12): 887-97.

Fukuda, N., Y. Wu, et al. (2005). "Phosphorylation of titin modulates passive stiffness of cardiac muscle in a titin isoform-dependent manner." J Gen Physiol **125**(3): 257-71.

Furst, D. O., M. Osborn, et al. (1988). "The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy: a map of ten nonrepetitive epitopes starting at the Z line extends close to the M line." J Cell Biol **106**(5): 1563-72.

Garcia-Mata, R. and K. Burridge (2007). "Catching a GEF by its tail." Trends Cell Biol **17**(1): 36-43.

Goldberg, J., A. C. Nairn, et al. (1996). "Structural basis for the autoinhibition of calcium/calmodulin-dependent protein kinase I." Cell **84**(6): 875-87.

Gossett, L. A., R. M. Hecht, et al. (1982). "Muscle differentiation in normal and cleavage-arrested mutant embryos of Caenorhabditis elegans." Cell **30**(1): 193-204.

Gotthardt, M., R. E. Hammer, et al. (2003). "Conditional expression of mutant M-line titins results in cardiomyopathy with altered sarcomere structure." J Biol Chem **278**(8): 6059-65.

Granzier, H., M. Helmes, et al. (2000). "Mechanical properties of titin isoforms." Adv Exp Med Biol **481**: 283-300; discussion 300-4.

Granzier, H., D. Labeit, et al. (2003). "Adaptations in titin's spring elements in normal and cardiomyopathic hearts." Adv Exp Med Biol **538**: 517-30; discussion 530-1.

Granzier, H. and S. Labeit (2007). "Structure-function relations of the giant elastic protein titin in striated and smooth muscle cells." Muscle Nerve **36**(6): 740-55.

Granzier, H. L. and S. Labeit (2004). "The giant protein titin: a major player in myocardial mechanics, signaling, and disease." Circ Res **94**(3): 284-95.

Grater, F., J. Shen, et al. (2005). "Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations." Biophys J **88**(2): 790-804.

Greaser, M. (2001). "Identification of new repeating motifs in titin." Proteins **43**(2): 145-9.

Greene, D., T. Garcia, et al. (2008). "Single-molecule force spectroscopy reveals a stepwise unfolding of C. elegans giant protein kinase domains." Biophys J.

Greene, D. N., S. M. Whitney, et al. (2007). "Artificially evolved Synechococcus PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency." Biochem J **404**(3): 517-24.

Gutierrez-Cruz, G., A. H. Van Heerden, et al. (2001). "Modular motif, structural folds and affinity profiles of the PEVK segment of human fetal skeletal muscle titin." J Biol Chem **276**(10): 7442-9.

Hall, A. (1998). "Rho GTPases and the actin cytoskeleton." Science **279**(5350): 509-14.

Han, J., B. Das, et al. (1997). "Lck regulates Vav activation of members of the Rho family of GTPases." Mol Cell Biol **17**(3): 1346-53.

Hanks, S. K. and T. Hunter (1995). "Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification." Faseb J **9**(8): 576-96.

Heierhorst, J., B. Kobe, et al. (1996). "Ca2+/S100 regulation of giant protein kinases." Nature **380**(6575): 636-9.

Heierhorst, J., R. J. Mann, et al. (1997). "Interaction of the recombinant S100A1 protein with twitchin kinase, and comparison with other Ca2+-binding proteins." Eur J Biochem **249**(1): 127-33.

Heierhorst, J., W. C. Probst, et al. (1995). "Phosphorylation of myosin regulatory light chains by the molluscan twitchin kinase." Eur J Biochem **233**(2): 426-31.

Heierhorst, J., X. Tang, et al. (1996). "Substrate specificity and inhibitor sensitivity of Ca2+/S100-dependent twitchin kinases." Eur J Biochem **242**(3): 454-9.

Hertadi, R. and A. Ikai (2002). "Unfolding mechanics of holo- and apocalmodulin studied by the atomic force microscope." Protein Sci **11**(6): 1532-8.

Heymann, B. and H. Grubmuller (2001). "Molecular dynamics force probe simulations of antibody/antigen unbinding: entropic control and nonadditivity of unbinding forces." Biophys J **81**(3): 1295-313.

Higashijima, T., K. M. Ferguson, et al. (1987). "The effect of GTP and Mg2+ on the GTPase activity and the fluorescent properties of Go." J Biol Chem **262**(2): 757-61.

Hikita, T., H. Qadota, et al. (2005). "Identification of a novel Cdc42 GEF that is localized to the PAT-3-mediated adhesive structure." Biochem Biophys Res Commun **335**(1): 139-45.

Hobert, O., D. G. Moerman, et al. (1999). "A conserved LIM protein that affects muscular adherens junction integrity and mechanosensory function in Caenorhabditis elegans." J Cell Biol **144**(1): 45-57.

Hodgkin, J. (1983). "Male Phenotypes and Mating Efficiency in CAENORHABDITIS ELEGANS." Genetics **103**(1): 43-64.

Hodgkin, J. (1983). "Two types of sex determination in a nematode." Nature **304**(5923): 267-8.

Hodgkin, J., R. H. Plasterk, et al. (1995). "The nematode Caenorhabditis elegans and its genome." Science **270**(5235): 410-4.

Hosono, R., Y. Mitsui, et al. (1982). "Life span of the wild and mutant nematode Caenorhabditis elegans. Effects of sex, sterilization, and temperature." Exp Gerontol **17**(2): 163-72.

Hresko, M. C., B. D. Williams, et al. (1994). "Assembly of body wall muscle and muscle cell attachment structures in Caenorhabditis elegans." J Cell Biol **124**(4): 491-506.

Hu, S. H., J. Y. Lei, et al. (1994). "Crystallization and preliminary X-ray analysis of the auto-inhibited twitchin kinase." J Mol Biol **236**(4): 1259-61.

Huang, Y. W., M. C. Surka, et al. (2006). "GTP binding and hydrolysis kinetics of human septin 2." Febs J **273**(14): 3248-60.

Huse, M. and J. Kuriyan (2002). "The conformational plasticity of protein kinases." Cell **109**(3): 275-82.

Huxley, H. E. (1957). "The double array of filaments in cross-striated muscle." J Biophys Biochem Cytol **3**(5): 631-48.

James, P., J. Halladay, et al. (1996). "Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast." Genetics **144**(4): 1425-36.

Jansen, G., E. Hazendonk, et al. (1997). "Reverse genetics by chemical mutagenesis in Caenorhabditis elegans." Nat Genet **17**(1): 119-21.

Jetten, M. S., M. E. Gubler, et al. (1994). "Structural and functional analysis of pyruvate kinase from Corynebacterium glutamicum." Appl Environ Microbiol **60**(7): 2501-7.

Jorgensen, E. M. and S. E. Mango (2002). "The art and design of genetic screens: caenorhabditis elegans." Nat Rev Genet **3**(5): 356-69.

Kamynina, E., K. Kauppinen, et al. (2007). "Regulation of proto-oncogenic dbl by chaperone-controlled, ubiquitin-mediated degradation." Mol Cell Biol **27**(5): 1809-22.

Kato, J., Y. Kaziro, et al. (2000). "Activation of the guanine nucleotide exchange factor Dbl following ACK1-dependent tyrosine phosphorylation." Biochem Biophys Res Commun **268**(1): 141-7.

Kay, A. J. and C. P. Hunter (2001). "CDC-42 regulates PAR protein localization and function to control cellular and embryonic polarity in C. elegans." Curr Biol **11**(7): 474-81.

Khosravi-Far, R., M. Chrzanowska-Wodnicka, et al. (1994). "Dbl and Vav mediate transformation via mitogen-activated protein kinase pathways that are distinct from those activated by oncogenic Ras." Mol Cell Biol **14**(10): 6848-57.

Knighton, D. R., R. B. Pearson, et al. (1992). "Structural basis of the intrasteric regulation of myosin light chain kinases." Science **258**(5079): 130-5.

Kobe, B., J. Heierhorst, et al. (1996). "Giant protein kinases: domain interactions and structural basis of autoregulation." Embo J **15**(24): 6810-21.

Kontrogianni-Konstantopoulos, A., E. M. Jones, et al. (2003). "Obscurin is a ligand for small ankyrin 1 in skeletal muscle." Mol Biol Cell **14**(3): 1138-48.

Labeit, S., M. Gautel, et al. (1992). "Towards a molecular understanding of titin." Embo J **11**(5): 1711-6.

Labeit, S. and B. Kolmerer (1995). "Titins: giant proteins in charge of muscle ultrastructure and elasticity." Science **270**(5234): 293-6.

Lange, S., E. Ehler, et al. (2006). "From A to Z and back? Multicompartment proteins in the sarcomere." Trends Cell Biol **16**(1): 11-8.

Lange, S., F. Xiang, et al. (2005). "The kinase domain of titin controls muscle gene expression and protein turnover." Science **308**(5728): 1599-603.

Leake, M. C., A. Grutzner, et al. (2006). "Mechanical properties of cardiac titin's N2B-region by single-molecule atomic force spectroscopy." J Struct Biol **155**(2): 263-72.

Lee, K. J., S. M. Mwongela, et al. (2008). "Determination of sphingosine kinase activity for cellular signaling studies." Anal Chem **80**(5): 1620-7.

Lei, J., X. Tang, et al. (1994). "Protein kinase domain of twitchin has protein kinase activity and an autoinhibitory region." J Biol Chem **269**(33): 21078-85.

Li, H. and J. M. Fernandez (2003). "Mechanical design of the first proximal Ig domain of human cardiac titin revealed by single molecule force spectroscopy." J Mol Biol **334**(1): 75-86.

Li, H., W. A. Linke, et al. (2002). "Reverse engineering of the giant muscle protein titin." Nature **418**(6901): 998-1002.

Li, H., A. F. Oberhauser, et al. (2001). "Multiple conformations of PEVK proteins detected by single-molecule techniques." Proc Natl Acad Sci U S A **98**(19): 10682-6.

Li, L., S. Wetzel, et al. (2006). "Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy." Biophys J **90**(4): L30-2.

Lim, C. C. and D. B. Sawyer (2005). "Modulation of cardiac function: titin springs into action." J Gen Physiol **125**(3): 249-52.

Linke, W. A. and A. Grutzner (2008). "Pulling single molecules of titin by AFM--recent advances and physiological implications." Pflugers Arch **456**(1): 101-15.

Linke, W. A., M. Ivemeyer, et al. (1998). "Nature of PEVK-titin elasticity in skeletal muscle." Proc Natl Acad Sci U S A **95**(14): 8052-7.

Linke, W. A., M. R. Stockmeier, et al. (1998). "Characterizing titin's I-band Ig domain region as an entropic spring." J Cell Sci **111 (Pt 11)**: 1567-74.

Liversage, A. D., D. Holmes, et al. (2001). "Titin and the sarcomere symmetry paradox." J Mol Biol **305**(3): 401-9.

Loomis, W. F., Jr., J. P. Wahrmann, et al. (1973). "Temperature-sensitive variants of an established myoblast line." Proc Natl Acad Sci U S A **70**(2): 425-9.

Lu, H., B. Isralewitz, et al. (1998). "Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation
." Biophys. J. **75**(2): 662-671.

Lu, H. and K. Schulten (1999). "Steered molecular dynamics simulations of force-induced protein domain unfolding." Proteins **35**(4): 453-63.

Ma, K., L. Kan, et al. (2001). "Polyproline II helix is a key structural motif of the elastic PEVK segment of titin." Biochemistry **40**(12): 3427-38.

Mackenzie, J. M., Jr. and H. F. Epstein (1980). "Paramyosin is necessary for determination of nematode thick filament length in vivo." Cell **22**(3): 747-55.

Mackinnon, A. C., H. Qadota, et al. (2002). "C. elegans PAT-4/ILK functions as an adaptor protein within integrin adhesion complexes." Curr Biol **12**(10): 787-97.

Manne, V., S. Yamazaki, et al. (1984). "Guanosine nucleotide binding by highly purified Ha-ras-encoded p21 protein produced in Escherichia coli." Proc Natl Acad Sci U S A **81**(22): 6953-7.

Marko, J. F. a. S., Eric D. (1995). "Stretching DNA." Macromolecule **28**(26): 8759-8770.

Maruyama, K. (1976). "Connectin, an elastic protein from myofibrils." J Biochem **80**(2): 405-7.

Maruyama, K., S. Kimura, et al. (1981). "Connectin, an elastic protein of muscle. Identification of "titin" with connectin." J Biochem **89**(3): 701-9.

Maruyama, K., S. Kimura, et al. (1984). "Molecular size and shape of beta-connectin, an elastic protein of striated muscle." J Biochem **95**(5): 1423-33.

Mayans, O., P. F. van der Ven, et al. (1998). "Structural basis for activation of the titin kinase domain during myofibrillogenesis." Nature **395**(6705): 863-9.

McLachlan, A. D. and J. Karn (1982). "Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle." Nature **299**(5880): 226-31.

Mercer, K. B., D. B. Flaherty, et al. (2003). "Caenorhabditis elegans UNC-98, a C2H2 Zn finger protein, is a novel partner of UNC-97/PINCH in muscle adhesion complexes." Mol Biol Cell **14**(6): 2492-507.

Mercer, K. B., R. K. Miller, et al. (2006). "Caenorhabditis elegans UNC-96 is a new component of M-lines that interacts with UNC-98 and paramyosin and is required in adult muscle for assembly and/or maintenance of thick filaments." Mol Biol Cell **17**(9): 3832-47.

Miller, E., T. Garcia, et al. (2006). "The mechanical properties of E. coli type 1 pili measured by atomic force microscopy techniques." Biophys J **91**(10): 3848-56.

Miller, R. K., H. Qadota, et al. (2006). "UNC-98 links an integrin-associated complex to thick filaments in Caenorhabditis elegans muscle." J Cell Biol **175**(6): 853-9.

Mirzoeva, S., S. Weigand, et al. (1999). "Analysis of the functional coupling between Calmodulin's calcium binding and peptide recognition properties." Biochemistry **38**(42): 14117-8.

Moerman, D. G., G. M. Benian, et al. (1988). "Identification and intracellular localization of the unc-22 gene product of Caenorhabditis elegans." Genes Dev **2**(1): 93-105.

Moerman, D. G. and B. D. Williams (2006). "Sarcomere assembly in C. elegans muscle." WormBook: 1-16.

Montclare, J. K. and A. Schepartz (2003). "Miniature homeodomains: high specificity without an N-terminal arm." J Am Chem Soc **125**(12): 3416-7.

Murphy, R. A. (1976). "Structural proteins in the myofilaments and regulation of contraction in vertebrate smooth muscle." Fed Proc **35**(6): 1302-6.

Nassar, N., J. Cancelas, et al. (2006). "Structure-function based design of small molecule inhibitors targeting Rho family GTPases." Curr Top Med Chem **6**(11): 1109-16.

Nave, R., D. O. Furst, et al. (1989). "Visualization of the polarity of isolated titin molecules: a single globular head on a long thin rod as the M band anchoring domain?" J Cell Biol **109**(5): 2177-87.

Neuman, K. C. and A. Nagy (2008). "Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy." Nat Methods **5**(6): 491-505.

Nobes, C. D. and A. Hall (1999). "Rho GTPases control polarity, protrusion, and adhesion during cell movement." J Cell Biol **144**(6): 1235-44.

Nonet, M. L., K. Grundahl, et al. (1993). "Synaptic function is impaired but not eliminated in C. elegans mutants lacking synaptotagmin." Cell **73**(7): 1291-305.

Oberhauser, A. F., C. Badilla-Fernandez, et al. (2002). "The mechanical hierarchies of fibronectin observed with single-molecule AFM." J Mol Biol **319**(2): 433-47.

Oberhauser, A. F. and M. Carrion-Vazquez (2008). "Mechanical biochemistry of proteins one molecule at a time." J Biol Chem **283**(11): 6617-21.

Oberhauser, A. F., P. K. Hansma, et al. (2001). "Stepwise unfolding of titin under force-clamp atomic force microscopy." Proc Natl Acad Sci U S A **98**(2): 468-72.

Oberhauser, A. F., P. E. Marszalek, et al. (1998). "The molecular elasticity of the extracellular matrix protein tenascin." Nature **393**(6681): 181-5.

Obermann, W. M., M. Gautel, et al. (1996). "The structure of the sarcomeric M band: localization of defined domains of myomesin, M-protein, and the 250-kD carboxy-terminal region of titin by immunoelectron microscopy." J Cell Biol **134**(6): 1441-53.

Ogata, T. and Y. Yamasaki (1990). "High-resolution scanning electron microscopic studies on the three-dimensional structure of the transverse-axial tubular system, sarcoplasmic reticulum and intercalated disc of the rat myocardium." Anat Rec **228**(3): 277-87.

Olson, M. F. (1996). "Guanine nucleotide exchange factors for the Rho GTPases: a role in human disease?" J Mol Med **74**(10): 563-71.

Parikh, M. R., D. N. Greene, et al. (2006). "Directed evolution of RuBisCO hypermorphs through genetic selection in engineered E.coli." Protein Eng Des Sel **19**(3): 113-9.

Person, V., S. Kostin, et al. (2000). "Antisense oligonucleotide experiments elucidate the essential role of titin in sarcomerogenesis in adult rat cardiomyocytes in long-term culture." J Cell Sci **113 Pt 21**: 3851-9.

Pham, N. and D. Rotin (2001). "Nedd4 regulates ubiquitination and stability of the guanine-nucleotide exchange factor CNrasGEF." J Biol Chem **276**(50): 46995-7003.

Phillips, J. C., B. Rosemary, et al. (2005). "Scalable molecular dynamics with NAMD." Journal of Computational Chemistry **26**(16): 1781-1802.

Probst, W. C., E. C. Cropper, et al. (1994). "cAMP-dependent phosphorylation of Aplysia twitchin may mediate modulation of muscle contractions by neuropeptide cotransmitters." Proc Natl Acad Sci U S A **91**(18): 8487-91.

Qadota, H., L. A. McGaha, et al. (2008). "A Novel Protein Phosphatase is a Binding Partner for the Protein Kinase Domains of UNC-89 (Obscurin) in C. elegans." Mol Biol Cell.

Qadota, H., K. B. Mercer, et al. (2007). "Two LIM domain proteins and UNC-96 link UNC-97/pinch to myosin thick filaments in Caenorhabditis elegans muscle." Mol Biol Cell **18**(11): 4317-26.

Ridley, A. J. and A. Hall (1992). "The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors." Cell **70**(3): 389-99.

Ridley, A. J. and A. Hall (1994). "Signal transduction pathways regulating Rho-mediated stress fibre formation: requirement for a tyrosine kinase." Embo J **13**(11): 2600-10.

Rief, M., M. Gautel, et al. (2000). "Unfolding forces of titin and fibronectin domains directly measured by AFM." Adv Exp Med Biol **481**: 129-36; discussion 137-41.

Rief, M., M. Gautel, et al. (1997). "Reversible unfolding of individual titin immunoglobulin domains by AFM." Science **276**(5315): 1109-12.

Rief, M., M. Gautel, et al. (1998). "The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy." Biophys J **75**(6): 3008-14.

Rogalski, T. M., G. P. Mullen, et al. (2000). "The UNC-112 gene in Caenorhabditis elegans encodes a novel component of cell-matrix adhesion structures required for integrin localization in the muscle cell membrane." J Cell Biol **150**(1): 253-64.

Russell, M. W., M. O. Raeker, et al. (2002). "Identification, tissue expression and chromosomal localization of human Obscurin-MLCK, a member of the titin and Dbl families of myosin light chain kinases." Gene **282**(1-2): 237-46.

Salojin, K. V., J. Zhang, et al. (1999). "TCR and CD28 are coupled via ZAP-70 to the activation of the Vav/Rac-1-/PAK-1/p38 MAPK signaling pathway." J Immunol **163**(2): 844-53.

Schmidt, A. and A. Hall (2002). "Guanine nucleotide exchange factors for Rho GTPases: turning on the switch." Genes Dev **16**(13): 1587-609.

Sebestyen, M. G., J. D. Fritz, et al. (1996). "Primary structure of the kinase domain region of rabbit skeletal and cardiac muscle titin." J Muscle Res Cell Motil **17**(3): 343-8.

Sharma, D., O. Perisic, et al. (2007). "Single-molecule force spectroscopy reveals a mechanically stable protein fold and the rational tuning of its mechanical stability." Proc Natl Acad Sci U S A **104**(22): 9278-83.

Siegman, M. J., D. Funabara, et al. (1998). "Phosphorylation of a twitchin-related protein controls catch and calcium sensitivity of force production in invertebrate smooth muscle." Proc Natl Acad Sci U S A **95**(9): 5383-8.

Small, T. M., K. M. Gernert, et al. (2004). "Three new isoforms of Caenorhabditis elegans UNC-89 containing MLCK-like protein kinase domains." J Mol Biol **342**(1): 91-108.

Sotomayor, M. and K. Schulten (2007). "Single-molecule experiments in vitro and in silico." Science **316**(5828): 1144-8.

Spiro, D. (1956). "The ultrastructure of striated muscle at various sarcomere lengths." J Biophys Biochem Cytol **2**(4, Suppl): 157-62.

Sprang, S. R. (1997). "G protein mechanisms: insights from structural analysis." Annu Rev Biochem **66**: 639-78.

Taylor, S. S. and E. Radzio-Andzelm (1994). "Three protein kinase structures define a common motif." Structure **2**(5): 345-55.

Trinick, J., P. Knight, et al. (1984). "Purification and properties of native titin." J Mol Biol **180**(2): 331-56.

Tskhovrebova, L. and J. Trinick (2003). "Titin: properties and family relationships." Nat Rev Mol Cell Biol **4**(9): 679-89.

Tskhovrebova, L., J. Trinick, et al. (1997). "Elasticity and unfolding of single molecules of the giant muscle protein titin." Nature **387**(6630): 308-12.

van der Ven, P. F., J. W. Bartsch, et al. (2000). "A functional knock-out of titin results in defective myofibril assembly." J Cell Sci **113 (Pt 8)**: 1405-14.

Vetter, I. R. and A. Wittinghofer (2001). "The guanine nucleotide-binding switch in three dimensions." Science **294**(5545): 1299-304.

von Castelmur, E., M. Marino, et al. (2008). "A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain." Proc Natl Acad Sci U S A **105**(4): 1186-91.

Wang, K., J. McClure, et al. (1979). "Titin: major myofibrillar components of striated muscle." Proc Natl Acad Sci U S A **76**(8): 3698-702.

Watanabe, K., P. Nair, et al. (2002). "Molecular mechanics of cardiac titin's PEVK and N2B spring elements." J Biol Chem **277**(13): 11549-58.

Waterston, R. H., H. F. Epstein, et al. (1974). "Paramyosin of Caenorhabditis elegans." J Mol Biol **90**(2): 285-90.

Wennerberg, K., K. L. Rossman, et al. (2005). "The Ras superfamily at a glance." J Cell Sci **118**(Pt 5): 843-6.

White, D. (1999). "The Physiology and Biochemistry of Prokaryotes." Oxford University Press(2nd edition).

Whiting, A., J. Wardale, et al. (1989). "Does titin regulate the length of muscle thick filaments?" J Mol Biol **205**(1): 263-8.

Williams, B. D. and R. H. Waterston (1994). "Genes critical for muscle development and function in Caenorhabditis elegans identified through lethal mutations." J Cell Biol **124**(4): 475-90.

Wilmann, M., M. Gautel, et al. (2000). "Activation of calcium/calmodulin regulated kinases." Cell Mol Biol (Noisy-le-grand) **46**(5): 883-94.

Witt, C. C., S. H. Witt, et al. (2008). "Cooperative control of striated muscle mass and metabolism by MuRF1 and MuRF2." Embo J **27**(2): 350-60.

Wood, W. B. (1988). "Determination of pattern and fate in early embryos of Caenorhabditis elegans." Dev Biol (N Y 1985) **5**: 57-78.

Yamaguchi, K., O. Ohara, et al. (2008). "Smurf1 directly targets hPEM-2, a GEF for Cdc42, via a novel combination of protein interaction modules in the ubiquitin-proteasome pathway." Biol Chem **389**(4): 405-13.

Yamauchi, J., J. R. Chan, et al. (2005). "The neurotrophin-3 receptor TrkC directly phosphorylates and activates the nucleotide exchange factor Dbs to enhance Schwann cell migration." Proc Natl Acad Sci U S A **102**(14): 5198-203.

Yang, X., E. J. Hubbard, et al. (1992). "A protein kinase substrate identified by the two-hybrid system." Science **257**(5070): 680-2.

Zastrow, M. S., D. B. Flaherty, et al. (2006). "Nuclear titin interacts with A- and B-type lamins in vitro and in vivo." J Cell Sci **119**(Pt 2): 239-49.

Zengel, J. M. and H. F. Epstein (1980). "Identification of genetic elements associated with muscle structure in the nematode Caenorhabditis elegans." Cell Motil **1**(1): 73-97.

Zhang, X., S. Zhang, et al. (2006). "Kinetic mechanism of AKT/PKB enzyme family." J Biol Chem **281**(20): 13949-56.

Zhang, Y. and S. McCormick (2007). "A distinct mechanism regulating a pollen-specific guanine nucleotide exchange factor for the small GTPase Rop in Arabidopsis thaliana." Proc Natl Acad Sci U S A **104**(47): 18830-5.

Appendix A

Artificially Evolved *Synechococcus* PCC6301 Rubisco Variants Exhibit Improvements in Folding and Catalytic Efficiency

(Most of the contents of this chapter were published in Greene et al. Biochemical Journal. 404(3): 517-24)

ABSTRACT

The photosynthetic CO_2-fixing enzyme, Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), is responsible for most of the world's biomass but is a slow, non-specific catalyst. We seek to identify the chemical and biological constraints that limit Rubisco's evolutionary potential in nature, and find ways to overcome them. Recently, the horizontal transfer of Calvin Cycle genes (*rbcL*, *rbcS* and *prkA*) from cyanobacteria (*Synechococcus* PCC6301) to gamma proteobacteria (*Escherichia coli*) was recapitulated in the laboratory. Three unique Rubisco variants containing single (M259T) and double (M259T/A8S, M259T/F342S) amino acid substitutions in the large (L) subunit were identified after random mutagenesis and selection in *E. coli*. Here we show that the Met-259-Thr mutation improved the yield of properly folded L subunit in *E. coli* 4 to 9-fold by decreasing its natural propensity to mis-fold *in vivo*. Moreover, the addition of osmolites to the growth media was able to further enhance productive folding of the Met-259-Thr L relative to wild-type L. The mutation does not increase steady-state levels of *rbc*L mRNA or L production, nor does it enhance productive interactions with *E. coli* molecular chaperones. The evolved enzymes showed improvement in their kinetic properties with the M259T variant showing a 12% increase in carboxylation turnover rate (k^c_{cat}), a 15% improvement in its K_M for CO_2 and no change in its K_M for ribulose-1,5-bisphosphate or its CO_2/O_2 selectivity. These results show that the directed evolution of the *Synechococcus* Rubisco in *E. coli* elicits improvement in folding and catalytic efficiency.

INTRODUCTION

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the nucleophilic carboxylation of ribulose-1,5-bisphosphate (ribulose-P_2). This conversion of inorganic CO_2 into carbohydrate is frequently the rate limiting step in the synthesis of most of the world's biomass (Parry, Andralojc et al. 2003). In spite of its biological importance, Rubisco is an inefficient catalyst, particularly at limiting CO_2 concentrations. Its turnover rate is less than one-thousandth that of many other plant enzymes (Morell, Paul et al. 1992). Moreover, the efficiency of Rubisco-catalyzed carbon assimilation is further compromised by the counter-productive fixation of O_2, which competes with CO_2 for addition to ribulose-P_2. The 2-phosphoglycerate product of the oxygenation reaction is recycled by photorespiration, an energy consuming process that releases one-quarter of the carbon incorporated into 2-phosphoglycerate (Wingler, Lea et al. 2000). It is not clear why such an important enzyme is so slow and non-specific. Our goals are to understand the physical, chemical and biological constraints to adaptive evolution, and to find ways to overcome them.

The Calvin Cycle is the major conduit for the fixation of carbon dioxide into the biosphere. It shares eight enzymes with the gluconeogenesis or pentose phosphate pathways. These enzymes are constitutively expressed in *E. coli* and most other organisms. The recruitment of four genes, those encoding phosphoribulokinase (PRK; *prkA*), sedoheptulose bisphosphatase (*sbpA* or *fbpA*) and the ribulose 1,5-bisphosphate carboxylase/oxygenase large (L; *rbcL*) and small (S; *rbcS*) subunits would thus be sufficient to initiate the Calvin cycle in most species (Shively, van Keulen et al. 1998). In all bacteria containing *rbcS* there is an associated *rbcL* physically linked in an operon

that can sometime also include *fbpA* and *prkA* genes (Gibson and Tabita 1996). In fact, these operons are sometimes encoded on mobile genetic elements (Shively, van Keulen et al. 1998). Since on its own the *rbcL* gene does not confer any selective advantage it seems likely that the entire cluster was transferred to ensure their stable acquisition. A phylogenetic analyses of *rbcL* sequences suggested that this gene was horizontally transferred at least four times during the evolution of proteobacteria, cyanobacteria and plastids (Delwiche and Palmer 1996).

Phylogenetic analyses can highlight patterns that reflect horizontal gene transfer or gene duplication and selective loss (Koonin 2003; Kurland, Canback et al. 2003). Others have identified natural mechanisms for the transfer DNA between genomes, but the adaptation of transferred genes to their new cellular micro-environments remains poorly understood (Thomas and Nielsen 2005). Gene sequence comparisons can sometimes identify instances of molecular adaptation, although most beneficial mutations are obscured by the accumulation of functionally neutral mutations (Kimura 1983). The "resurrection" and *in vitro* characterization of ancestral proteins can offer unambiguous evidence for molecular adaptation, but only if the investigator knows which biochemical tests to conduct (Thornton 2004). Directed evolution experiments aim to recapitulate hypothetical evolutionary events, thereby "resurrecting" ancestral proteins and the conditions in which they were selected.

Natural selection generally favors enzyme variants that evince speed (high turnover, k_{cat}), high substrate affinity (low Michaelis constant, K_M) and high substrate specificity (k_{cat}/K_M primary substrate divided by k_{cat}/K_M secondary substrate). Extant Rubisco homologues from different photosynthetic organisms show significant variation

in all three parameters (see summary in (Jordan 1981; Jordan and Ogren 1983; Tcherkez, Farquhar et al. 2006)). Their K_M for substrate CO_2 (K_c) varies from 3-340 mM and the corresponding carboxylation turn-over rates (k^c_{cat}) vary between 1-13 sec^{-1}. Rubisco's ability to differentiate CO_2 from O_2 is defined by its specificity factor ($S_{c/o}$), which equals its carboxylation efficiency (k^c_{cat}/K_c) divided by its oxygenation efficiency (k^o_{cat}/K_o), where k^o_{cat} is the O_2 saturated oxygenation rate and K_o is the K_M for O_2 (Jordan 1981).

We have studied the horizontal transfer and subsequent adaptation of Rubisco to a new cellular environment by emulating the process in the laboratory. Delwiche and Palmer observed inconsistencies between the phylogeny of Rubisco-containing organisms and the phylogenies of their *rbcL* genes; they hypothesized the horizontal transfer of the *rbcL* gene (and presumably the *rbcS* and *prkA* genes) from a cyanobacterium to a gamma proteobacterium (Delwiche and Palmer 1996). We biochemically tested this hypothesis by developing a high throughput genetic selection system for Rubisco function in the wild-type *E. coli* strain K-12 (a gamma proteobacterium). The *rbcL* from *Synechococcus* PCC6301 was randomly mutated by sequential error prone PCR reactions and hypermorphic variants were identified through selection in *E. coli* expressing *prkA*. Three unique Rubisco variants (M259T, M259T/A8S and M259T/F342S, Table 1 and Figure A.1) emerged after three rounds of random mutagenesis and genetic selection. Preliminary biochemical studies showed that *E. coli* cells expressing the evolvants exhibited 2 to 5-fold more carboxylase activity than those expressing the wild-type enzyme. Purified C-terminal-his_6-tagged versions of these mutant proteins showed similar increases in specific activity (Parikh, Greene et al. 2006). Here we extend these initial observations to underpin how mutations in the horizontally

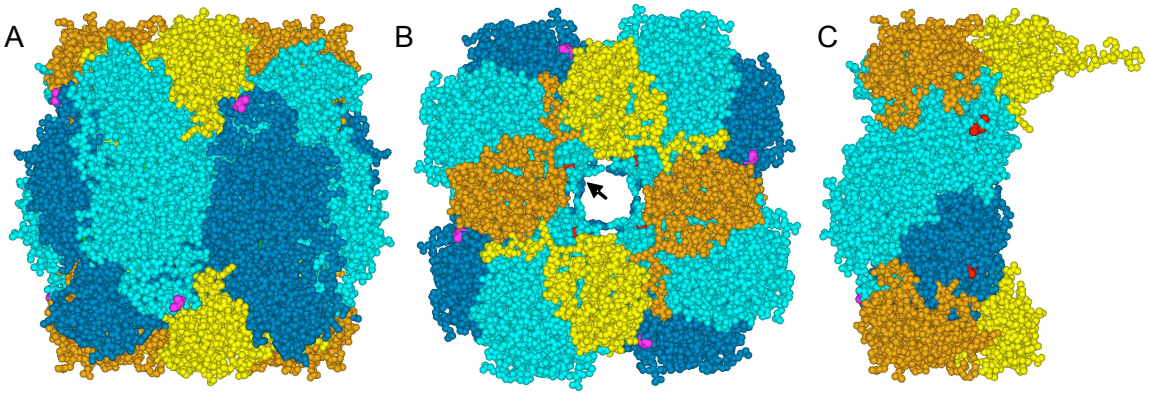


Figure A.1- The structure of L_8S_8 Rubisco is conserved. (A) Arrangement of the 8 large (L, light and dark blue) and 8 small (S, yellow and orange) subunits in the *Synechococcus* PCC6301 Rubisco hexadecamer viewed down the twofold and (B) fourfold axes (see (Newman and Gutteridge 1993) for details). (C) L_2S_4 unit of the Rubisco viewed from inside the central channel between the L subunits in the direction of the arrow in panel B. The position of the large subunit mutations Ala-8-Ser (pink) and Met-259-Thr (red) analyzed in this study are highlighted in each panel. Phe-342 is not a surface residue and is not visible.

Rubisco name	*rbcLS* mutant allele number [b]	Silent *rbcL* mutations	Amino acid modifications
(-)WT [a]	-	nil	nil
WT	-	c489t, c783t, t799c	nil
(-)M259T	-	nil	Met-259-Thr
M259T	2.29	c489t, c783t, t799c	Met-259-Thr
A8S/M259T	2.24	c783t, t799c	Ala-8-Ser, Met-259-Thr
M259T/F342S	3.54	t702c, c783t, t799c	Met-259-Thr, Phe-342-Ser

[a]Wild-type *rbcL-rbcS* operon, [b]clone number described in (Parikh, Greene et al. 2006)

Table A.1- Features of the *Synechococcus* PCC6301 Rubisco genes and proteins used in this study.

transfered *Synechococcus* PCC6301 *rbc*L increased the fitness of the host *E. coli* cell under the applied selective pressures.

RESULTS

Our objective is to elucidate the biochemical mechanism(s) of adaptation of *Synechococcus* PCC6301 Rubisco variants previously evolved in our *E. coli*-based selection system (see Table 1). Previous analyses indicated that compared with wild-type controls the evolved Rubisco variants were expressed 2 to 5-fold more efficiently in *E. coli* and the carboxylase activity of purified his_6-tagged enzyme preparations were ~5-fold greater (Parikh, Greene et al. 2006). Here we investigate fully the mechanisms that accounted for the expression differences and carefully characterize the steady-state kinetic parameters of the native (not tagged) versions of the wild-type and evolved Rubisco variants.

Catalytic properties

A comparison of the kinetic characteristics for the untagged wild-type *Synechococcus* PCC6301 Rubisco (WT) and the M259T, A8S/M259T and M259T/F342S variant enzymes were performed using recombinant enzyme preparations purified from *E. coli* by ammonium sulphate fractionation and ion exchange chromatography. The inclusion of affinity tag-fusions at either the N- or C-terminus of the Rubisco large (L) subunit was avoided since alterations to these regions can affect the kinetics properties of

Rubisco type[a]	$S_{c/o,}$ CO_2 / O_2 specificity[b]	k^c_{cat}, carboxylation turnover rate (s^{-1})	K_c, Michaelis constant for CO_2 (μM)	k^c_{cat}/ K_c Carboxylation efficiency (M^{-1} s^{-1})	K_{RuBP}, Michaelis constant for ribulose-P_2 (μM) (pure)
WT	43.9 ± 1.8	11.4 ± 0.1[b]	273 ± 10[b]	4.2 x 10^4	63.9 ± 4.5[b]
	43.0 ± 1.0[c]	11.6[c]	284[c]	4.1 x 10^4[c]	54 ± 3[d]
M259T	42.8 ± 3.9	12.8 ± 0.2[b]	237 ± 10[b]	5.4 x 10^4	67.0 ± 3.2[b]
A8S, M259T	40.5 ± 2.6	11.4 ± 0.1[e]	216 ± 9[e]	5.3 x 10^4	47.8 ± 4.6[e]
M259T, F342S	38.4 ± 0.7	10.0 ± 0.1[e]	207 ± 8[e]	4.8 x 10^4	33.7 ± 5.9[e]

Table A.2- Kinetic properties of purified wild-type and mutant *Synechococcus* PCC6301 Rubiscos identified by genetic selection.

[a]Refer to Table 1. [b]Average of measurements made in triplicate (±SD). [c]Kinetic parameters measured by (Morell, Paul et al. 1994) and [d]measured by (Whitney and Sharwood 2006). [e]Value from single measurement with error derived from Michaelis-Menten hyperbolic curve fit.

the enzyme (see (Kellogg and Juliano 1997) for summary). As shown in Table 2, the CO_2/O_2 specificity ($S_{c/o}$), the Michaelis constants of the Rubisco/substrate interactions (K_c, K_M for CO_2; K_{RuBP} , K_M for ribulose-P_2), and the substrate saturated carboxylation rate (k^c_{cat}) of the wild-type PCC6301 Rubisco were analogous to those reported previously. The replicate $S_{c/o}$ measurements were highly reproducible and revealed no change for the M259T enzyme and modest decreases (less than 15%) for the A8S/M259T and M259T/F342S Rubiscos. The M259T variant evinced a 12% increase in k^c_{cat}, while the A8S/M259T and M259T/F342S double mutants exhibited slightly diminished k^c_{cat} values that were wild-type-like and 12% less than wild-type, respectively. The K_c for the M259T, A8S/M259T and M259T/F342S Rubisco reactions were 13%, 21% and 24 % lower than the comparable wild-type value, respectively. Likewise, K_{RuBP} for the A8S/M259T and M259T/F342S enzymes were better than wild-type with respective reductions of 25% and 47% whilst the K_{RuBP} for M259T remained comparatively unchanged. With regard to their carboxylation efficiencies (k^c_{cat}/K_c,) the three evolved Rubiscos exhibited 15 to 28% improvements relative to the wild-type enzyme (Table 2).

Rubisco subunit synthesis and assembly in E. coli

To examine whether the mutations in the evolved Rubisco variants influenced the production of the Rubisco subunits in *E. coli* the total and soluble cellular protein were analysed by SDS PAGE. For each of the 6 Rubisco variants examined the L band was the predominant protein in the total cellular protein samples (Figure A.2A). To more accurately assess the relative amounts of L expressed, each sample was diluted 100-fold and re-analyzed by SDS PAGE (Figure A.2B). We observed little difference in the

amount of L synthesized, apart from a modest (< 2-fold) increase in the amount of total L in the *E. coli* expressing the A8S/M259T Rubisco. This indicated the nucleotide substitutions in *rbc*L coding for the non-silent mutations and the c783t and t799c silent mutations had little, if any, influence on the production of the variant L subunits in *E. coli*.

SDS PAGE analysis of comparable soluble cellular protein samples indicated >90% of the L-subunits produced was insoluble (Figure A.2C). Using immuno-detection the relative levels of the L and S subunits in the soluble *E. coli* protein samples were compared showing higher concentrations of soluble L and S in cells expressing the evolved Rubiscos than those producing the wild-type controls (Figure A.2D). These 2 to 4-fold differences in the amount of soluble L detected on the immunoblots correlated with the different levels of functional L_8S_8 Rubisco measured by CABP-binding (Figure A.2E) and that detected by non-denaturing PAGE (Figure A.2F) consistent with previous findings that most, if not all, of the soluble L detected in *E coli* extracts constitutes that assembled into functional L_8S_8 enzyme (Emlyn-Jones, Woodger et al. 2006; Whitney and Sharwood 2006). In contrast, the production of S in *E. coli* does not limit assembly of PCC6301 Rubisco since as much as 50% of the soluble S is not assembled in L_8S_8 complexes but can readily do so when supplied with L_8 cores (Andrews and Ballment 1984). Taken together, these data suggested the M259T mutation present in L of all three artificially evolved Rubisco variants improves the ability of the L to fold and/or assemble correctly with S into functional enzyme within *E. coli.* Notably the additional F342S mutation compromised the yield of functional Rubisco produced by approximately 2-fold, consistent with that shown previously (Parikh, Greene et al. 2006).

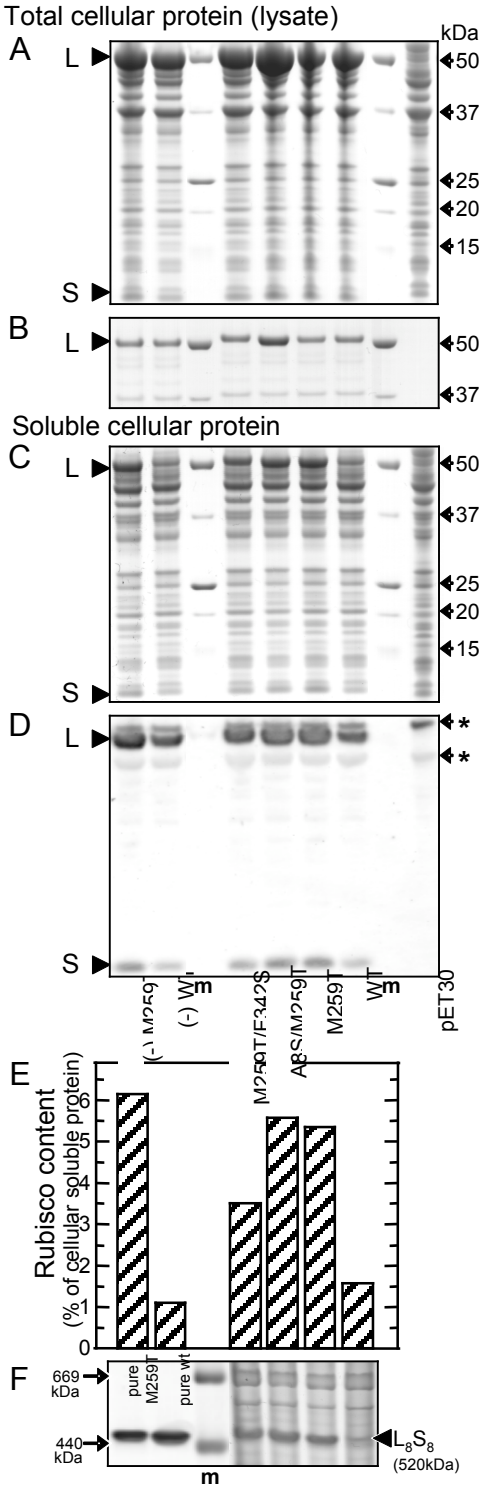

Total cellular protein (lysate)
kDa
A
L
S
50
37
25
20
15
B
L
50
37
Soluble cellular protein
C
L
S
50
37
25
20
15
D
L
S
*
*
(-) M259
(-) Wt
m
M259T/F342S
A8S/M259T
M259T
Wt
m
pET30
E
Rubisco content
(% of cellular soluble protein)
0
1
2
3
4
5
6
F
669
kDa
440
kDa
pure
M259T
pure wt
m
L8S8
(520kDa)

Figure A.2- The evolved Rubisco variants fold and assemble more efficiently in *E. coli*. Rubisco expression was induced with IPTG in BL21(DE3) *E. coli* cells transformed with *rbcLS*-pET30a+ plasmids coding for wild-type (WT) and the evolved *Synechococcus* PCC6301 Rubisco variants (see Table 1). To compare the amount, and partitioning, of Rubisco large (L, 52 kDa) and small (S, 13 kDa) subunits in the soluble and insoluble protein cellular fractions, comparable samples of total cellular protein (lysate) containing (A) 10 μg and (B) 0.1 μg of soluble cellular protein and (C) only the soluble cellular protein fraction (10 μg) were separated by SDS SPAGE and visualized with Coommassie staining. (D) Detection of the L and S expression level in the soluble cellular protein by immunobloting using antibodies to *Synechococcus* sp PCC7942 Rubisco. (E) The amount of functional Rubisco measured by [^{14}C]carboxyarabinitol-P_2 binding expressed as a percentage of the soluble cellular protein assuming a molecular mass of 65 kDa for each L active site (assuming one S for each L that binds the [^{14}C]carboxyarabinitol-P_2). (F) Coomassie staining of soluble cellular protein (30 μg) and purified L_8S_8 Rubisco separated by non-denaturing PAGE. m, molecular mass marker (sizes shown; 669 kDa, thyroglobulin: 440 kDa, ferritin); *, non-Rubisco *E. coli* proteins recognized by the antibody.

rbcL mRNA levels

The steady-state levels of *rbcL* mRNA that accumulated in *E. coli* expressing wild-type and M259T Rubisco (whose *rbcL* contains the M259T/t776c mutation common to all the evolved alleles in addition to 3 silent nucleotide mutations, Table 1) were highly comparable. Using Reverse Transcriptase/Real Time PCR, the amount of wild-type and M259T *rbcL* mRNA both increased ~10-fold upon induction with IPTG and accumulated to equivalent levels (data not shown) consistent with our findings that comparable amounts of both wild-type and M259T L were produced in the *E. coli* (Figure A.2B).

E. coli chaperone-assisted folding

As selection of the three Rubisco variants appeared to be based upon improved folding and assembly of the L in *E. coli* we examined whether this improvement resulted from an increased compatibility between the mutated L subunits and several key *E. coli* chaperone complexes. Indeed, 10-fold improvements in the amount of cyanobacterial Rubisco correctly folded and assembled in *E. coli* can be obtained by over-expressing the GroES/GroEL chaperonins that are essential for productive protein folding (Goloubinoff, Gatenby et al. 1989; Larimer and Soper 1993). Likewise, the 70kDa chaperone DnaK that works in concert with its co-chaperones DnaJ (a chaperone activating protein) and GrpE (a nucleotide exchange factor) is also requisite for productive folding of bacterial Rubiscos in *E. coli* (Checa and Viale 1997). BL21(DE3) cells expressing the wild-type and M259T Rubiscos were therefore co-transformed with plasmids directing the expression of GroES-GroEL and DnaK/J/GrpE, in addition to the 50-kDa trigger factor (TF) chaperone that, similarly to DnaK, co-translationally bind to nacent peptide chains emerging from polyribosomes to protect them from misfolding

and aggregation (Nishihara, Kanemori et al. 2000; Kaiser, Chang et al. 2006). Analogous to Figure A.2, PAGE and CABP binding analyses were performed on the total and soluble cellular *E. coli* proteins to confirm chaperone expression and measure the differences in Rubisco content and functional assembly (Figure A.3). A comparison of the relative banding intensities in the total and soluble protein samples analysed by SDS PAGE (Figure A.3A and B respectively) showed the chaperones were highly expressed and soluble while the majority of the abundant L produced was prone to aggregation and was predominantly insoluble.

Only the GroEL and its co-chaperonin GroES stimulated assembly of the M259T and wild-type PCC6301 Rubisco in *E. coli*. Differences in the amount of soluble L produced in the BL21 (DE3) cells after 5 hours induction at 23ºC (compared with 16 hours in Figure A.2) were confirmed by immunoblot analysis (Figure A.3C) and correlated with the amount of L_8S_8 enzyme measured by [^{14}C]-carboxyarabinitol-P_2 binding (Figure A.3D), and non-denaturing PAGE (Figure A.3E and F). Without additional GroES-GroEL the amount of functional M259T Rubisco produced was ~9-fold that of the wild-type. GroES-GroEL over-expression enhanced the yield of functional wild-type Rubisco by ~9-fold whereas the excess chaperonins evinced only a 40% improvement in the amount of assembled M259T enzyme recovered. In contrast, over-expression of the DnaJ/DnaK/GrpE (KJE) and TF molecular chaperones reduced the amount of soluble wild-type and M259T L produced (and accordingly the amount of L_8S_8 enzyme) by 20% and 40-60%, respectively, indicating incompatibilities between chaperones and

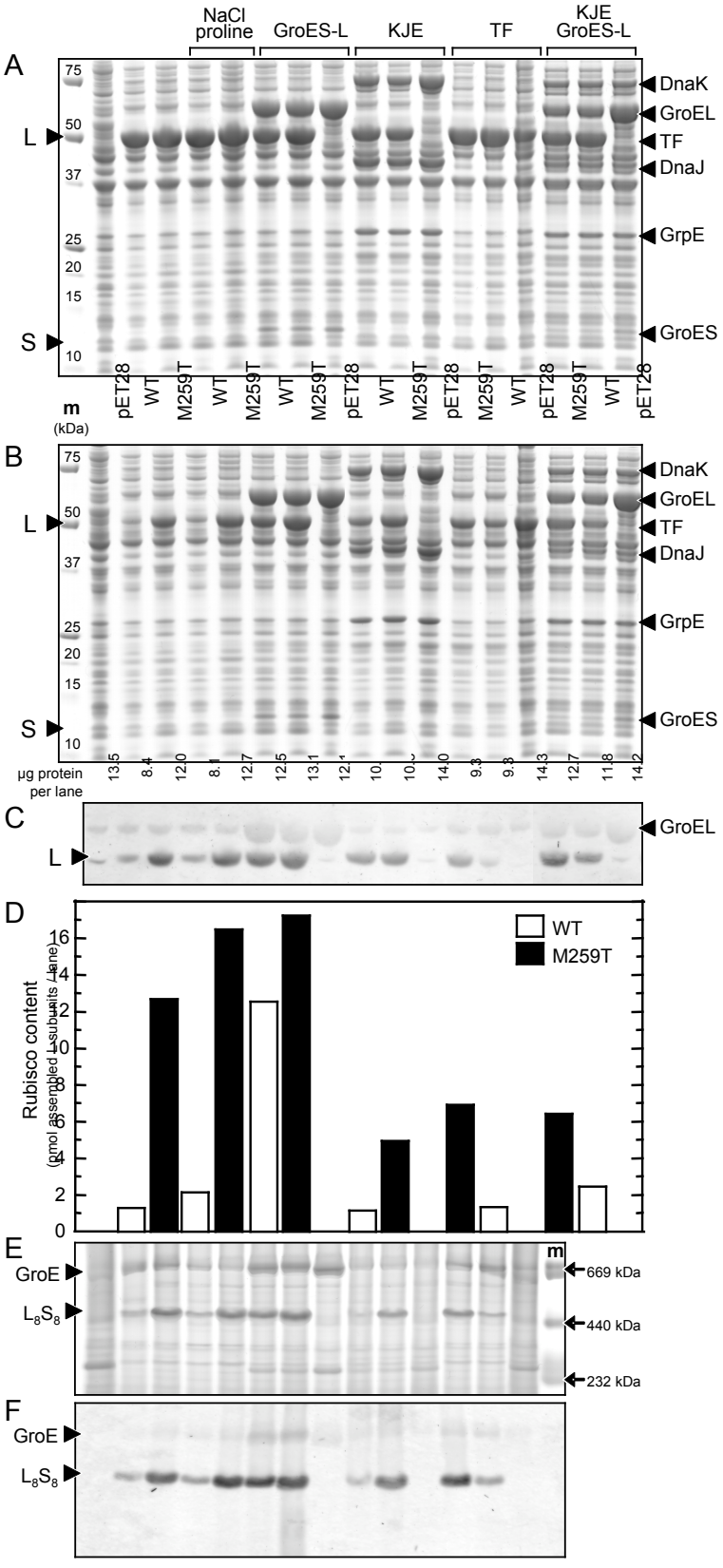

NaCl proline
GroES-L
KJE
TF
KJE GroES-L
A
B
C
D
E
F
L
S
m
(kDa)
75
50
37
25
20
15
10
DnaK
GroEL
TF
DnaJ
GrpE
GroES
pET28
WT
M259T
µg protein per lane
Rubisco content
(µmol assembled L-subunits / Lane)
WT
M259T
GroE
L_8S_8
669 kDa
440 kDa
232 kDa

Figure A.3- The influence of osmolites and *E. coli* chaperones on the expression and assembly of wild-type (WT) and the M259T *Synechococcus* PCC6301 Rubisco variants in BL21(DE3) *E. coli.* (A) Total and (B) soluble cellular protein from 1.5 x 10^6 cells separated by SDS PAGE and visualized with Coommassie staining. The relative amount of soluble protein loaded per lane is shown. (C) Detection of L in the soluble cellular protein by immunobloting using an antibody to tobacco Rubisco that also recognizes the GroEL chaperonin and an unknown 50kDa *E. coli* protein. (D) The amount of L assembled into functional Rubisco measured by [^{14}C]carboxyarabinitol-P_2 binding. (E) Soluble protein (twice that loaded in panel B) separated by non-denaturing PAGE and the position of correctly assembled L_8S_8 Rubisco and the GroES-GroEL chaperonin complex (GroE) identified by Coomassie staining and (F) immuno-detection with the tobacco Rubisco antibody. m, molecular mass markers (sizes shown). GroES-L, GroES (10.4 kDa) and GroEL (57.3 kDa) chaperonins; KJE, DnaK (70 kDa)-DnaJ (41 kDa)-GrpE (22 kDa) chaperones; TF, trigger factor (50 kDa).

the L subunits that hampered their productive folding (Figure A.3C to F). In cells over-expressing KJE modest improvements of 30% and 70% in the amount of functional M259T and wild-type Rubisco produced, respectively, were obtained by simultaneously over-expressing GroES-GroEL (Figure A.3), however no stimulation in functional Rubisco production by GroES-GroEL was evident in cells producing excess TF (data not shown). Clearly the improved solubility and functional assembly of the M259T mutated L is not due to an increased compatibility with the GroES-GroEL, KJE or TF chaperone complexes with over-expression of the latter chaperone systems actually perturbing productive folding of both the wild-type and mutated *Synechococcus* L.

Osmolite assisted folding

Addition of the osmolite, proline, to the growth media was used to examine the extent to which it can suppress self-aggregation of the L subunits within *E. coli*. The addition of proline to *E. coli* growth media produces an osmotic stress response that can decrease the propensity of proteins to aggregate *in vivo* (Ignatova and Gierasch 2006). Indeed, in cells grown in LB supplemented with 20 mM proline the solubility of both aggregation-prone L subunits for wild-type and M259T Rubiscos were both enhanced 35% (Figure A.3). In terms of net improvement in the amount of soluble L produced by the proline treatment, the amount of additional soluble M259T L exceeded that of the extra wild-type soluble L produced by ~5-fold indicating aggregation of the mutant L is more readily suppressed by proline.

In vitro solubility

Intrinsically soluble proteins will tolerate higher concentrations of ammonium sulfate before precipitating from solution (Jenkins 1998). The influence of ammonium sulfate on the solubility of purified wild-type and M259T Rubiscos were therefore compared. At low ammonium sulfate concentrations (< 15% (wt/vol)) the solubility of both Rubiscos was highly comparable (Figure A.4). At higher concentrations of ammonium sulfate the wild-type enzyme appeared more soluble indicating that the observed higher level of functional M259T Rubisco in *E. coli* are unlikely due to an intrinsic improvement in the solubility of the L_8S_8 complex.

In vitro thermostability

The structural integrity of the wild-type and M259T Rubiscos were compared by CD spectroscopy. Both enzymes showed matching CD spectra, indicating that their secondary structures were identical (Figure A.5A). Thermally induced unfolding of the enzymes was also monitored by CD at 222 nm (Figure A.5B). At this wavelength the melting temperature (T_M) for M259T Rubisco (T_M = 63.5 ºC) was slightly lower than the wild-type enzyme (T_M = 65 ºC) suggesting that the M259T mutation may slightly compromise the structural integrity of the L_8S_8 complex.

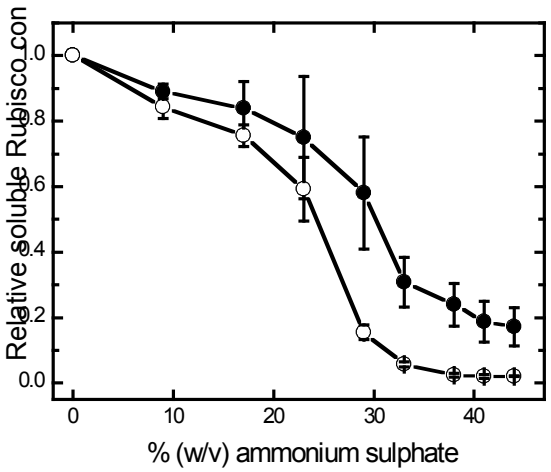


Figure A.4- The M259T Rubisco is less soluble than the wild-type protein *in vitro*. Equimolar concentrations of purified his_6-tagged WT (●) and M259T (○) Rubisco were incubated with varying concentrations of ammonium sulfate. Precipitated Rubisco was removed by centrifugation and the Rubisco content in the supernatant determined using a dye binding protein assay. The data is the average (±SD) of three separate experiments using different starting concentrations (0.4 to 1.7 μM) of Rubisco.

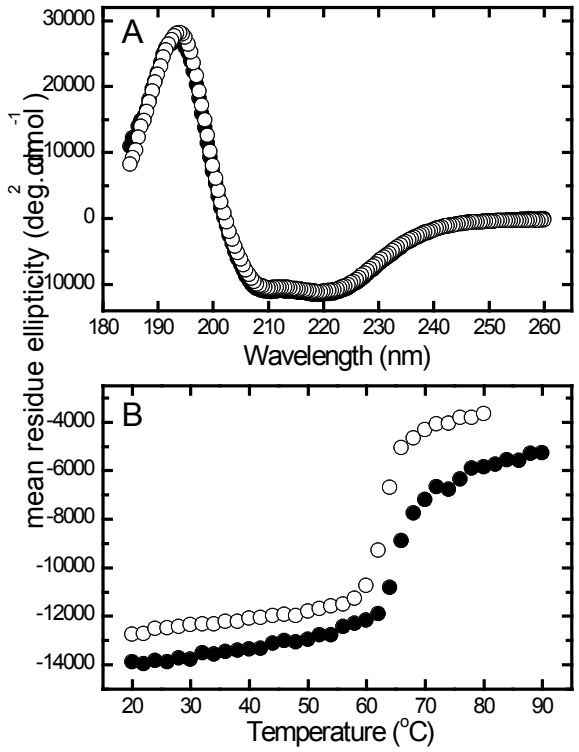


Figure A.5- Unfolding of the M259T and wild-type Rubisco's. (A) CD spectra of equimolar amounts (0.31 μM) of his_6-tagged WT (●) and M259T (○) Rubisco are indistinguishable in the far-UV region indicating equivalent structural folding. (B) Thermally induced unfolding transitions of the Rubiscos monitored by CD at 222nm.

DISCUSSION

The process of horizontal gene transfer (HGT) has been proposed as the best mechanism to explain the inconsistencies between the phylogeny of Rubisco-containing organisms and the phylogenies of their *rbcL* genes (Delwiche and Palmer 1996). In prokaryotes the HGT process generally correlates with the transfer of operational genes, for example those coding for traits such as antibiotic resistance, nitrogen fixation and photosynthesis. By recapitulating such a transfer event by transplanting cyanobacterial *rbcL*, *rbcS* and *prkA* genes into *E. coli* we have derived a means to direct the evolution of the Rubisco genes and study the adaptation of evolved variants to their new cellular environment (Parikh, Greene et al. 2006). By itself, uptake of a functionally expressed *prk*A gene by *E. coli* is detrimental to cell viability since the bacterium can not metabolize its catalytic by-product, ribulose-P_2, resulting in its accumulation to growth-arresting levels (Hudson, Morell et al. 1992). Removal of ribulose-P_2 therefore necessitated the co-transfer of the cyanobacterial Rubisco genes, providing a convenient means by which to screen for improvements in Rubiscos ability to increase the fitness of the host *E. coli* cell thereby enabling us to assess how Rubisco can adapt to the selective pressures of a new environment (Parikh, Greene et al. 2006).

The biochemical basis of ribulose-P_2 toxicity in *E. col*i remains unknown. The structural similarity of the pentose to fructose-P_2 suggests that it inhibits enzymes in glycolysis and gluconeogenesis or other enzymes regulated by fructose-P_2 (Hudson, Morell et al. 1992). Alternatively, the diversion of the carbon flux from the pentose phosphate shunt might be sufficient to cause a pyridoxal phosphate deficiency (Zhao and Winkler 1994). Irrespective of the toxicity mechanism, *E. coli* viability was made

dependent on the production of Rubisco variants that have adapted to the selection conditions. Here we demonstrated by careful characterization of the catalytic and biophysical properties of the evolved Rubiscos that the adaptive solution to overcome ribulose-P_2 toxicity evinced by the M259T amino acid substitution was to increase the pool of correctly folded Rubisco molecules by 4 to 9-fold (depending upon the duration of induction). This substitution is, to our knowledge, the first reported mutation that confers an improvement to productive Rubisco assembly within *E. coli*.

As we show here, the observed increase in Rubisco activity was not due to improvements in the steady state levels of the Rubisco mRNA (Figure A.4) or the amount of translated L subunits (Figures A.3 and A.5). Neither was the *in vitro* thermal stability or the solubility of the M259T Rubisco sufficiently altered to account for the difference. The modest improvements in K_{RuBP} and k^c_{cat} were also unlikely to account for selection of the variant enzymes under our stringent conditions. The selection of the three artificially evolved Rubisco variants was thus primarily based upon the M259T mutation reducing the propensity of the L subunits to form kinetically trapped, misfolded intermediates. Possibly the reduced aggregation of the L subunits may result from the mutation enhancing co-translational domain folding thus promoting the pool of L able to reach a conformation that enables productive assembly with the primarily soluble S subunits into L_8S_8 enzyme. A reduced propensity for unproductive aggregation of the mutant L subunits was further supported by the pronounced improvement in the amount of soluble mutated L subunits produced in cells treated with the osmolite proline relative to wild-type. These results suggest that the M259T mutation might enhance the folding of *Synechococcus* PCC6301Rubisco in other hosts.

Improvements in functional assembly of cyanobacterial Rubisco were elicited by over-expression of GroES-GroEL, but not the *E. coli* Trigger Factor (TF) or DnaK/DnaJ/GrpE (KJE) chaperones. The propensity for the L subunits to mis-fold and aggregate in *E. coli* indicates the bacterial chaperone machinery does not support the kinetically most efficient folding route available, although the M259T mutation appears to partially alleviate this kinetic impediment. Normally the overlapping functions of the TF and the KJE chaperones are to bind to nacent protein chains emerging from the polyribosomes and maintain them in a folding-competent state so as to enable productive post-translational folding of the entire protein (Kaiser, Chang et al. 2006). Curiously, although DnaK is essential for *Synechococcus* L_8S_8 production in *E. coli* (Checa and Viale 1997) we found over-expression of KJE and TF reduced productive folding of the wild-type and M259T L subunits by 20 to 50%, respectively. Clearly an incompatibility exists between the L subunits and the co-translational actions of the TF and KJE chaperone machinery such that increasing their expression promoted the formation of kinetically trapped, misfolded L intermediates. Even over-expression of the GroES/GroEL chaperonins had little, if any, influence on alleviating this incompatibility indicating the folding of the L subunits was irretrievably perturbed before interaction with GroEL.

As observed previously however, at normal physiological levels of TF and KJE the over-expression of GroES-GroEL substantially improved the pool of L subunits post-translationally folded correctly into L_8S_8 enzyme (Goloubinoff, Gatenby et al. 1989; Larimer and Soper 1993). These results indicate that incompatibilities with both the co-translational and post-translational chaperone folding machinery in *E. coli* hinder L

subunit folding and assembly. We suggest that improvements in L_8S_8 assembly might be further improved by the horizontal transfer of chaperone genes from *Synechococcus* PCC6301 into the new host bacterium or by directed evolution of the *E. coli* chaperones towards optimizing L_8S_8 assembly. A limitation to both strategies is the possibility that the folding and assembly of native *E. coli* proteins may be unfavorably impaired, as previously found when increasing GFP expression in *E. coli* through directed evolution of GroES-GroEL (Wang, Herman et al. 2002).

In addition to establishing that the primary phenotype of the selected Rubisco variants was encumbered by their unique propensity for improved productive folding, we have also established the M259T mutation uniquely improves the kinetic prowess of the enzyme. Previous amino acids substitutions to Rubiscos, introduced through site-directed or random mutagenesis, have sometimes conferred improvement in one or more catalytic properties, but always to the detriment of another (Harpel and Hartman 1992; Kellogg and Juliano 1997; Spreitzer and Salvucci 2002; Parry, Andralojc et al. 2003; Smith and Tabita 2003; Spreitzer, Peddi et al. 2005). Such catalytic tradeoffs are apparently commonplace for Rubisco in both natural and artificial selection. However, the M259T Rubisco characterized here is novel in that the improvements in k^c_{cat} and K_M for CO_2 (K_c) occurred with little, or no, detriment to its specificity ($S_{c/o}$) or its K_M for ribulose-P_2 (K_{RuBP}) (Table 2). Notably, the 12% improvement in k^c_{cat} for the M259T Rubisco contrasts with the 5-fold difference in specific activity previously measured for purified 6xHis-tagged wild-type and M259T Rubisco (Parikh, Greene et al. 2006). Previous mutagenic studies have highlighted that subtle modifications to the length and amino acid composition of the *Synechococcus* Rubisco L C-terminus significantly perturbs the

kinetics of the enzyme (Gutteridge 1986; Genhai Zhu 1998; Whitney and Sharwood 2006). The extent to which the C-terminally appended GHHHHHH sequence affected the kinetic properties of the wild-type and M259T Rubiscos relative to each other and the native (non-tagged) enzymes was not the focus of this study and was not examined.

Curiously Met-259 does not interact with residues from other subunits in the L_8S_8 complex and is positioned close to the surface within the solvent channel that traverses the hexadecamer along its fourfold axis (Figure A.1C). At this position Met-259 does not closely interact with any active site residues or with the α-carbon of the catalytic Lys-201 residue (located ~14.5 Ǻ away from the closest atom, the β-carbon, of Met259). Other mutagenic studies on form I Rubiscos, particularly those targeting modifications to residues at the L and S subunit interface, have also shown that perturbations to residues distant from the active site have significant influence on the kinetic properties of enzyme (Spreitzer and Salvucci 2002; Spreitzer, Peddi et al. 2005). Either the kinetic changes are induced by subtle changes to active site geometry, and/or changes in the conformational dynamics of the L_8S_8 holoenzyme (either in mobile structural elements such as loop 6 (Cleland, Andrews et al. 1998; Parry, Andralojc et al. 2003; Tcherkez, Farquhar et al. 2006) or in larger scale conformational changes (Schlitter and Wildner 2000)). Curiously the same Met-259-Thr mutation has been selected previously using a *trans*-complementation screen of an XL1-Red generated *Synechococcus* PCC6301 Rubisco mutant library using a Rubisco deficient strain of the bacterium *Rhodobacter capsulatus* (Smith and Tabita 2003). Notably the expression level, and kinetics, of the selected M259T variant in *R. capsulatus* were not quantified and therefore it remains to be

assessed whether, as in *E. coli*, the M259T Rubisco supported growth of the mutated bacterium due to its improved tendency toward productive assembly.

Modest changes were observed in the kinetic properties of the two other Rubisco variants selected in *E. coli*, both of which shared the Met-259-Thr mutation. The additional Ala-8-Ser mutation in the A8S/M259T Rubisco and the Phe-342-Ser in the M259T/F342S enzyme resulted in further improvements in their K_c and K_{RuBP}, with slight reductions in $S_{c/o}$ and k^c_{cat} compared with the parental M259T enzyme (Table 2). Like the Met-259-Thr mutation, it is not readily evident from the positioning of both these amino acids in the *Synechococcus* Rubisco crystal structure how they impart the changes in kinetics (Figure A.1). Ala-8 is within the highly divergent N-terminal region of the L subunit exposed to the external solvent that, in the crystal structure, resides at the interface between an L and S subunit (Figure 1A, B). Despite the high variability at the N-termini between different L subunits, this region has a significant influence on the catalytic prowess of both plant and bacterial Rubiscos (see (Kellogg and Juliano 1997) for summary).

Phe-342 is more highly conserved presumably due to its close location downstream of the highly conserved loop 6 region (Gly-326 to Gly-334 in PCC6301, which corresponds to residues 329 to 337 in the spinach L subunit) that plays an integral part during catalysis (Parry, Andralojc et al. 2003). When compared with other cyanobacteria L sequences (see below) only the aromatic phenylalanine and tyrosine amino acids are found at the equivalent position to codon 342. Again, two other Phe-324 mutants were identified in the aforementioned *R.capsulatus*-based selection of *Synechococcus* Rubisco; the kinetics of the F324V, but not the F324I, enzyme were

characterized (Smith and Tabita 2003). Comparable to that measured here for the M259T/F342S Rubisco, the K_{RuBP} for the F342V enzyme was reduced (improved) by ~50%, k^c_{cat} ~30% lower and the $S_{c/o}$ marginally compromised. However, in contrast to the 24% decrease in K_c for the M259T/F342S Rubisco the K_c was ~50% higher for the F342V enzyme. Since the individual A8S and F342S substitutions did not emerge in the selection we did not examine their effects upon the kinetic properties of *Synechococcus* PCC6301 Rubisco.

Alignment of other cyanobacterial L subunits shows that many natural homologues possess threonine at codon 259. The catalytic improvement in the *Synechococcus* PCC6301 Rubisco imparted by the Met-259-Thr substitution raises the question: "Why wasn't Thr 259 fixed in nature?" Comparison of the 21 different cyanobacterial genomes available from the JGI and NCBI databases (Figure A.6) found four species contained threonine, five glutamine and the remainder methionine at the position equivalent to codon 259 (data not shown). The kinetic improvements associated with the Met-259-Thr mutation might not be sufficient to warrant selection of the M259T Rubisco naturally. Alternatively, the mutation, whilst kinetically better, may be unfavorable in other ways such as reducing the compatibility of the L subunit with cyanobacterial chaperone complexes or other interacting proteins such as those involved in carboxysome formation. Clearly, future studies will have to evaluate the extent to which Rubiscos evolved in *E. coli* can be acclimated back into their native hosts. Implementing such a substitution will necessitate the production of a *Synechococcus* PCC6301 Rubisco mutant, similar to *Synechocystis* PCC6803 Syn6803Δrbc mutant

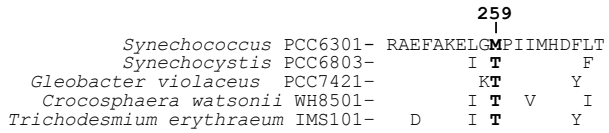


Figure A.6- Cyanobacterial L subunit comparison. Amino acids residues 250 to 266 in the *Synechococcus* PCC6301 Rubisco large subunit (PCC6301, GenBank #NC_006576, complement of nucleotides 139920-141338) compared with cyanobacterial Rubisco sequences containing threonine at the equivalent Met-259 codon (in bold). Only residues differing from the PCC6301 sequence are shown. PCC6803, (#NC_000911, nucleotides 2478414-9826); PCC7421, (#NC_005125, nucleotides 2307046-8470); WH8501, (JGI#400856410, nucleotides 12457-13875); IMS101, (JGI#403238160, complement of 249393-250823).

(Amichay, Levitz et al. 1993), which is amenable to the introduction of mutated *rbcLS* genes.

Using transition biased whole gene mutagenesis, we have shown that the directed evolution of *Synechococcus* PCC6301 Rubisco in *E. coli* can lead to improvements in its biophysical and kinetic parameters. As indicated previously (Parikh, Greene et al. 2006), selection for further kinetic improvements should be possible using Rubisco libraries constructed by more sophisticated mutagenic techniques. Modification of the selection process to favor the selection of mutants with improvements in other kinetic traits, such as its specificity for CO_2 over O_2 may also be possible. Importantly, the *E. coli*–based Rubisco selection may allow us to explore alternative sequence space not accessible to photosynthetic organisms whose survival is dictated by contingent physiological constraints.

MATERIALS AND METHODS

Materials

The *rbcLS*-pET30a+ expression vector and most of the molecular biology reagents and chemicals were previously described (Parikh, Greene et al. 2006). The sephadex G50 fine and HPX-87H resins, and the radiolabeled $NaH^{14}CO_3$ were from Pharmacia (GE Healthcare). The 0.2 micron nylon filters were from Alltech. All other chemicals were from Sigma Chemicals (St. Louis, MO).

Site-directed mutagenesis

The construction of the *rbcLS*-pET30a+ expression vectors containing the wild-type *Synechococcus* PCC6301 Rubisco *rbcL-rbcS* operon and those encoding the single (M259T) and double (M259T/A8S, M259T/F342S) amino acid mutations (with and without C-terminal his_6 tags fused to the large subunit) were described previously (Parikh, Greene et al. 2006) (summarized in Table 1). PCR-mediated site directed mutagenesis was applied to change the Thr-259 codon of Rubisco 2.29 (encoding M259T) back to methionine (to create an otherwise wild-type *rbcLS* with only silent mutations), and to introduce the M259T (t776c) mutation in the wild-type *rbcLS* gene (to create the (-) M259T Rubisco without silent mutations).

Expression of wild-type and artificially evolved Rubiscos

E. coli BL21(DE3) cells were transformed with the *rbcLS*-pET30 plasmids encoding the wild-type and mutant Rubisco variants. The transformed cells were separately grown at 23°C to mid-logarithmic phase in LB-kan and Rubisco expression induced for 16 hours with 0.5 mM IPTG. The cells were harvested by centrifugation (5 min at 5000 *g*), resuspended in extraction buffer (100 mM EPPS-NaOH pH 8.0, 1 mM EDTA, 2 mM DTT, 0.05% (wt/vol) *E. coli* protease inhibitor cocktail (Sigma)) and lysed by passage through a French pressure cell (140 MPa). An aliquot of lysate (total cellular protein) was taken for SDS PAGE analysis and the remainder centrifuged (10 min at 38,000 *g*) at 4ºC to remove the insoluble material. Aliquots of the supernatant were either assayed for protein content, Rubisco content (see below) processed for SDS PAGE and non-denaturing PAGE analysis as described previously (Whitney and Sharwood 2006).

Measuring Rubisco content by [2-^{14}C]carboxyarabinitol-P_2 binding - The concentration of Rubisco active sites was determined by the stoichiometric binding of the inhibitor [2-^{14}C]carboxyarabinitol-P_2. Ribulose-P_2 was synthesized and purified as described (Kane, Wilkin et al. 1998) and used to make *carboxy*-^{14}C-labeled carboxypentitol-P_2 (97,000 cpm / nmol) according to Pierce *et al*. (Pierce, Tolbert et al. 1980). Carboxypentitol-P_2 is an isometric mixture of carboxyribitol-P_2 and carboxyarabinitol-P_2. The latter is a transition-state analogue that preferentially binds to activated Rubisco in an almost irreversible manner (Schloss 1988). Samples containing Rubisco were activated with 25 mM $NaHCO_3$ and 20 mM $MgCl_2$ at 25°C for 30 minutes before incubating with 27 micromolar [2-^{14}C]carboxypentitol-P_2 for 15 to 45 min. The Rubisco-[2-^{14}C]carboxyarabinitol-P_2 complex was separated from unbound [2-^{14}C]carboxypentitol-P_2 by size exclusion chromatography and Rubisco content calculated as described previously (Ruuska, Andrews et al. 1998).

Purification of native (untagged) Rubiscos

The Rubisco variants were expressed in one liter cultures of *E. coli rbcLS*-pET30/BL21(DE3) cells as described above. After 16 hours of induction with 0.5 mM IPTG, the cells were harvested by centrifugation (5 min at 6000 *g*), resuspended in 45 mL ice cold extraction buffer, and lysed with a french pressure cell. The cellular debris was removed by centrifugation at 35,000 *g* for 15 min at 4 °C. The Rubiscos were purified by heating the extracts at 50 °C, precipitation with ammonium sulfate, followed by ion exchange chromatography using a Mono-Q (10/10) column (AP Biotech) as previously described (Morell, Paul et al. 1994). Collected fractions (5 mL) were assayed for

substrate-saturated ribulose-P_2 carboxylase activity using $NaH^{14}CO_3$ (2 C_i/mol, AP Biotech) (Andrews 1988). Appropriate fractions were pooled and a saturated ammonium sulfate solution (pH 7) slowly added to 60% (wt/vol) final concentration. The Rubisco precipitate was collected by centrifugation (22,000 *g*, 15 minutes at 4° C), dissolved in 3 mL storage buffer (15 mM EPPS-NaOH pH 8.0, 1 mM EDTA, 50 mM NaCl), and dialyzed using a Slide-a-lyzer cassette (Pierce Chemical Co.) against 1 L storage buffer overnight at 4°C. The dialysis was repeated against fresh storage buffer for 2 to 4 hours and then against storage buffer containing 20% glycerol (vol/vol) for 4 hours. The purified Rubisco was frozen in liquid nitrogen and stored at -70ºC. SDS and non-denaturing PAGE analysis showed that the Rubisco preparations were >95% pure.

Kinetic assays

The purified Rubisco preparations were used to measure the CO_2/O_2 specificity ($S_{c/o}$) at pH 8.3 as described (Kane, Viil et al. 1994). $S_{c/o}$ was measured in reactions equilibrated with an atmosphere of O_2 accurately mixed with 0.1% (vol/vol) CO_2 using vostoff pumps. The Michaelis constants for substrates CO_2 (K_c) and ribulose-P_2 (K_{RuBP}) were measured in $^{14}CO_2$-fixation assays at 25°C, pH 8 according to (Andrews 1988). The purified enzyme was pre-incubated at 25°C for 30 min in buffer containing 20 mM $MgCl_2$ and 25 mM $NaHCO_3$, and K_c measurements were performed in nitrogen sparged septum capped scintillation vials. The reactions were initiated by adding 10 microliters of purified enzyme to 1 mL of N_2 equilibrated assay buffer (100mM EPPS-NaOH, 20mM $MgCl_2$, 0.8mM ribulose-P_2, 0.1mg.ml^{-1} carbonic anhydrase) containing varying concentrations of $NaH^{14}CO_3$. K_{RuBP} measurements were performed in air equilibrated

assay buffer containing different concentrations of ribulose-P_2 (0 to 2 mM). The $^{14}CO_2$-assays were stopped after 1.5 or 2 min with 0.5 volumes of 25% (v/v) formic acid and dried at 80°C. The residue was dissolved in 0.5 ml water before adding 1 mL of scintillant (UltimaGold, Packard Biosciences) for scintillation counting. The Michaelis constants were determined by fitting the data to the Michaelis-Menten equation. From the K_c measurements, the catalytic turnover rate (k^c_{cat}) was calculated by dividing the extrapolated maximal carboxylase activity by the concentration of Rubisco active sites in the assay that was measured by [2-^{14}C]carboxyarabinitol-P_2 binding (see above).

PAGE and immunoblot analyses of protein expression in wild-type and artificially evolved Rubiscos

SDS PAGE, non-denaturing PAGE and immuno-blot analyses were performed according to Whitney and Sharwood (2007). Immunoblot analyses used polyclonal antisera raised in rabbits to purified *Synechococcus* PCC6301 Rubisco or tobacco Rubisco.

Reverse Transcriptase/Real Time PCR - The DH5Δlac(DE3) *E. coli* strain (Matsumura and Ellington 2001) was transformed with the *rbcLS*-pET30a+ plasmids coding for wild-type and M259T Rubisco and the transformed cells were propagated to mid-logarithmic stage in Luria-Bertani medium supplemented with 50 µg/mL kanamycin (LB-kan). Rubisco expression was induced with 0.5 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) overnight at 23°C. Total RNA was extracted using the RNeasy kit (Qiagen) and its concentration determined by absorbance measurements at 260 nm using a Shimadzu UV-1601 spectrophotometer. The RNA (30 or 0.3 ng) and varying amounts of the

rbcLS-pET30a+ plasmids were used to synthesize cDNA using the iScript one-step real time PCR kit (Bio-Rad) using the *rbcL*-specific primer 5'-TGGTGACCGACCCTTCTTC-3' (reverse). Real time PCR reactions were performed in duplicate using the reverse and forward (5'-GCTGACCGACATGGATCGGTACAA-3') *rbcL*-specific primers according to manufacturer instructions. Melt curve analysis confirmed the stability, accuracy, and specificity of the primer pair. The PCR amplification reactions were monitored with an iCycler real time PCR machine (Bio-Rad) and the amount of *rbcL* mRNA in each sample quantified using the accompanying software.

Analysis of WT and M259T Rubisco expression with the addition of osmolites and chaperones

E. coli BL21(DE3) cells were transformed with pET30 or the *rbcLS*-pET30 plasmids encoding wild-type and M259T Rubisco. The cells were grown in LB-kan in the presence or absence of 0.3 M NaCl and 20 mM proline at 37°C to an OD_{600} of 0.8 before inducing Rubisco expression with 0.5 mM IPTG at 23°C for 5 hours. As well, the cells were co-transformed with plasmids pGro7 (encodes GroEL/ES), pKJE7 (encodes DnaJ/K/GrpE), or pTF16 (encodes trigger factor) (chaperone plasmid set described in (Nishihara, Kanemori et al. 1998; Nishihara, Kanemori et al. 2000)) and similarly grown in LB-kan containing chloramphenicol (30 µg/ml). Protein expression was similarly induced at 23°C for 5 hours with 0.5 mM IPTG (Rubisco), 0.05% (w/v) L-arabinose (pGro7, pKJE7, pTf16, pG-KJE8) or/and 5 $\mu g.ml^{-1}$ tetracycline (pG-KJE8, pG-Tf2). Measurement of Rubisco content by SDS PAGE, non-denaturing PAGE and [^{14}C]carboxyaranitol-P_2

binding was standardized relative to cell density by measuring OD_{600} immediately prior to harvesting 15 ml of the cells by centrifugation (5 min at 4000 *g*). For BL21(DE3) cells transformed with pET30 or *rbcLS*-pET30 an OD_{600} of 1.0 was equivalent to 2.7×10^8 cells/ml.

In vitro solubility of purified Rubiscos

Wild-type and M259T his_6-tagged Rubiscos were expressed in *E. coli* and purified by immobilized metal affinity chromatography (IMAC) as previously described (Parikh, Greene et al. 2006). Enzyme concentration was measured by the Bradford dye-binding method using BSA as a standard. Equal concentrations of each enzyme were dispensed into separated tubes on ice and slowly mixed with different amounts of a saturated ammonium sulfate solution (pH 7.0) to final concentrations between 0 and 36% (wt/vol) ammonium sulfate. The samples were left on ice for 30 min before centrifuging (16,000 *g* for 30 min at 4°C), and the supernatants were assayed for protein content.

Circular dichroism spectra analyses

Far UV spectra and thermal denaturation were measured using a Jasco-810 spectropolarimeter. Far UV spectra were measured at 20°C between 185 and 260 nm in 0.5 nm increments using a 0.01 cm path length cell, a 2 nm bandwidth, a 4 second response time and a scanning speed of 20 nm/min. Thermal denaturation (unfolding) curves were measured at 2°C increments at a rate of 1 °C/min between 20-80 °C in a 0.1 cm path length cell at 222 nm.

REFERENCES

Amichay, D., R. Levitz, et al. (1993). "Construction of a Synechocystis PCC6803 mutant suitable for the study of variant hexadecameric ribulose bisphosphate carboxylase/oxygenase enzymes." Plant Molecular Biology **23**: 465-476.

Andrews, T. J. (1988). "Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits." J Biol Chem **263**(25): 12213-9.

Andrews, T. J. and B. Ballment (1984). "A rapid, sensitive method for quantitating subunits in purified ribulose bisphosphate carboxylase preparations." Plant Physiology **75**: 508-510.

Checa, S. K. and A. M. Viale (1997). "The 70-kda heat-shock protein dnaK chaperone system is required for the productive folding of ribulose-bisphosphate carboxylase subunits in Escherichia coli." European Journal of Biochemistry **248**(3): 848-855.

Cleland, W. W., T. J. Andrews, et al. (1998). "Mechanism of Rubisco: The Carbamate as General Base." Chem Rev **98**(2): 549-562.

Delwiche, C. F. and J. D. Palmer (1996). "Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids." Mol Biol Evol **13**(6): 873-82.

Emlyn-Jones, D., F. J. Woodger, et al. (2006). "RbcX Can Function as a Rubisco Chaperonin, But is Non-Essential in Synechococcus PCC7942." Plant Cell Physiol. **47**(12): 1630-1640.

Genhai Zhu, R. G. J., Hans J. Bohnert, Günter F. Wildner and Jürgen Schlitter (1998). "Dependence of catalysis and CO2/O2 specificity of Rubisco on the carboxy-terminus of the large subunit at different temperatures." Photosynth Res **57**(1): 71-79.

Gibson, J. L. and F. R. Tabita (1996). "The molecular regulation of the reductive pentose phosphate pathway in Proteobacteria and Cyanobacteria." Arch Microbiol **166**(3): 141-50.

Goloubinoff, P., A. A. Gatenby, et al. (1989). "GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli." **337**: 44-47.

Gutteridge, S. (1986). "Mutagenesis and expression of cloned Rubisco (ribulose-bisphosphate carboxylase)." Biochem Soc Trans **14**(1): 28-9.

Harpel, M. R. and F. C. Hartman (1992). "Enhanced CO2/O2 specificity of a site-directed mutant of ribulose-bisphosphate carboxylase/oxygenase." J Biol Chem **267**(10): 6475-8.

Hudson, G. S., M. K. Morell, et al. (1992). "Synthesis of spinach phosphoribulokinase and ribulose 1, 5-bisphosphate in Escherichia coli." Australian Journal of Plant Physiology **19**: 213-221.

Ignatova, Z. and L. M. Gierasch (2006). "Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant." Proc Natl Acad Sci U S A **103**(36): 13357-61.

Jenkins, W. T. (1998). "Three solutions of the protein solubility problem." Protein Sci **7**(2): 376-382.

Jordan, D. B. and W. L. Ogren (1983). "Species variation in kinetic properties of ribulose 1,5-bisphosphate carboxylase/oxygenase." Arch Biochem Biophys **227**(2): 425-33.

Jordan, D. B. a. O., W.L. (1981). "Species variation in the specificity of ribulose bisphosphate carboxylase/oxygenase." Nature **291**: 513-515.

Kaiser, C. M., H.-C. Chang, et al. (2006). "Real-time observation of trigger factor function on translating ribosomes." **444**(7118): 455-460.

Kane, H. J., J. Viil, et al. (1994). "An improved method for measuring the CO2/O2 specificity of ribulosebisphosphate carboxylase-oxygenase." Australian Journal of Plant Physiology **21**: 449-461.

Kane, H. J., J. M. Wilkin, et al. (1998). "Potent inhibition of ribulose-bisphosphate carboxylase by an oxidized impurity in ribulose-1,5-bisphosphate." Plant Physiology **117**(3): 1059-1069.

Kellogg, E. A. and N. D. Juliano (1997). "The structure and function of RuBisCO and their implications for systematic studies." American Journal of Botany **84**(3): 413-428.

Kimura, M. (1983). "Rare variant alleles in the light of the neutral theory." Mol Biol Evol **1**(1): 84-93.

Koonin, E. V. (2003). "Horizontal gene transfer: the path to maturity." Mol Microbiol **50**(3): 725-7.

Kurland, C. G., B. Canback, et al. (2003). "Horizontal gene transfer: a critical view." Proc Natl Acad Sci U S A **100**(17): 9658-62.

Larimer, F. W. and T. S. Soper (1993). "Overproduction of Anabaena 7120 ribulose-bisphosphate carboxylase/oxygenase in Escherichia coli." Gene **126**(1): 85-92.

Matsumura, I. and A. D. Ellington (2001). "In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates." J Mol Biol **305**(2): 331-9.

Morell, M. K., K. Paul, et al. (1992). "Rubisco: Maladapted or misunderstood?" Australian Journal of Botany **40**: 431-441.

Morell, M. K., K. Paul, et al. (1994). "Mutations of an active site threonyl residue promote á elimination and other side reactions of the enediol intermediate of the ribulosebisphosphate carboxylase reaction." Journal of Biological Chemistry **269**: 8091-8098.

Newman, J. and S. Gutteridge (1993). "The X-ray structure of Synechococcus ribulose-bisphosphate carboxylase/oxygenase-activated quaternary complex at 2.2-A resolution." J Biol Chem **268**(34): 25876-86.

Nishihara, K., M. Kanemori, et al. (1998). "Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli." Appl Environ Microbiol **64**(5): 1694-9.

Nishihara, K., M. Kanemori, et al. (2000). "Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli." Appl Environ Microbiol **66**(3): 884-9.

Parikh, M. R., D. N. Greene, et al. (2006). "Directed evolution of RuBisCO hypermorphs through genetic selection in engineered E.coli." Protein Eng Des Sel.

Parry, M. A., P. J. Andralojc, et al. (2003). "Manipulation of Rubisco: the amount, activity, function and regulation." J Exp Bot **54**(386): 1321-33.

Pierce, J., N. E. Tolbert, et al. (1980). "Interaction of ribulosebisphosphate carboxylase/oxygenase with transition-state analogues." Biochemistry **19**(5): 934-42.

Ruuska, S., T. J. Andrews, et al. (1998). "The interplay between limiting processes in C-3 photosynthesis studied by rapid-response gas exchange using transgenic tobacco impaired in photosynthesis." Australian Journal of Plant Physiology **25**(8): 859-870.

Schlitter, J. and G. F. Wildner (2000). "The kinetics of conformation change as determinant of Rubisco's specificity." Photosynth Res **65**(1): 7-13.

Schloss, J. V. (1988). "Comparative affinities of the epimeric reaction-intermediate analogs 2- and 4-carboxy-D-arabinitol 1,5-bisphosphate for spinach ribulose 1,5-bisphosphate carboxylase." J Biol Chem **263**(9): 4145-50.

Shively, J. M., G. van Keulen, et al. (1998). "Something from almost nothing: carbon dioxide fixation in chemoautotrophs." Annu Rev Microbiol **52**: 191-230.

Smith, S. A. and F. R. Tabita (2003). "Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase." J Mol Biol **331**(3): 557-69.

Spreitzer, R. J., S. R. Peddi, et al. (2005). "Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco

10.1073/pnas.0508042102." PNAS **102**(47): 17225-17230.

Spreitzer, R. J. and M. E. Salvucci (2002). "Rubisco: structure, regulatory interactions, and possibilities for a better enzyme." Annu Rev Plant Biol **53**: 449-75.

Tcherkez, G. G. B., G. D. Farquhar, et al. (2006). "Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized

10.1073/pnas.0600605103." PNAS: 0600605103.

Thomas, C. M. and K. M. Nielsen (2005). "Mechanisms of, and barriers to, horizontal gene transfer between bacteria." Nat Rev Microbiol **3**(9): 711-21.

Thornton, J. W. (2004). "Resurrecting ancient genes: experimental analysis of extinct molecules." Nat Rev Genet **5**(5): 366-75.

Wang, J. D., C. Herman, et al. (2002). "Directed evolution of substrate-optimized GroEL/S chaperonins." **111**(7): 1027-1039.

Whitney, S. M. and R. E. Sharwood (2006). "Linked rubisco subunits can assemble into functional oligomers without impeding catalytic performance 10.1074/jbc.M610479200." J. Biol. Chem.: M610479200.

Wingler, A., P. J. Lea, et al. (2000). "Photorespiration: metabolic pathways and their role in stress protection." Philos Trans R Soc Lond B Biol Sci **355**(1402): 1517-29.

Zhao, G. and M. E. Winkler (1994). "An Escherichia coli K-12 tktA tktB mutant deficient in transketolase activity requires pyridoxine (vitamin B6) as well as the aromatic amino acids and vitamins for growth." J Bacteriol **176**(19): 6134-8.