Structural Investigation of the Molecular Mechanisms underlying Titin Elasticity and Signaling

Inauguraldissertation

zur Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Eleonore von Castelmur

aus Basel, BS

Basel, 2010

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

PD Dr. M. Hennig,
Dr. O. Mayans,
Prof. T. Schirmer,

Basel, den 13.10.2009

Prof. Dr. Eberhard Parlow
Dekan

Abstract

Titin is a giant protein that spans >1μm from the Z-disc to the M-line, forming an intrasarcomeric filament system in vertebrate striated muscle, which is not only essential for the assembly of the sarcomere, but also critical for myofibril signaling and metabolism. Furthermore, it provides the sarcomere with resting tension, elasticity and restoring forces upon stretch, ensuring the correct positioning of the actin-myosin motors during muscle function. Titin is composed of ~300 immunoglobulin (Ig) and fibronectin-III (FnIII) domains, arranged in linear tandems. They are interspersed by an auto-inhibited Ser kinase (TK) close to its C-terminus as well as several unique sequences, most prominently a differentially spliced stretch rich in PEVK residues which localizes to the I-band part of titin where its elastic properties reside. There, the PEVK segment is flanked by a long Ig tandem, which together act as serial molecular springs that determine titin elastic response.

The focus of this work lay in the elucidation of the molecular mechanisms governing titin I-band elasticity and the recruitment of the M-line signalosome around TK involved in the control of myofibril turnover and the trophic state of muscle. To that effect, we have elucidated the crystal structure of a six-Ig fragment representative of the elastic Ig-tandem at 3.3Å resolution. The model reveals the molecular principles of Ig-arraying at the skeletal I-band of titin as mediated by conserved Ig-Ig transition motifs. Regular domain arrangements within this fragment point at the existence of a high-order in the fine structure of the filament, which is confirmed by EM data on a 19-mer poly-Ig segment. Our findings indicate a long-range, supra-order in the skeletal I-band of titin, where assembly of Ig domains into dynamical super-motifs is essential for the elastic function of the filament. We propose a novel model of spring mechanism for poly-Ig elasticity in titin based on a “carpenter ruler” model of skeletal I-band architecture. Furthermore, we have focused on the recruitment of the ubiquitin ligase MURF1 to the M-line signalosome through its specific interaction with titin domains A168-A170. MuRF1 contains several oligomerization motifs in succession, which indicates a possible need for tight regulation. We have therefore analyzed their influence on the oligomeric state of the protein. Our SEC-MALS data showed that the α-helical region of MuRF1 is dimeric in isolation, while in combination with the preceding B-Box domain, itself a dimerization motif, higher-order assembly is induced, which might be of physiological importance.

We could also show that higher-order assembly of MuRF1 did not disrupt binding to A168-A170 in pull-down assays. Further biophysical or structural characterization of the complex of A168-A170 with MuRF1 constructs was hindered by the severely compromised solubility of the complex. Finally, we have successfully solved the crystal structure of the FnIII-Kin-Ig region of twitchin, which corresponds to titin A170-TK-M1. The N-terminal linker wraps around the kinase domain and positions the preceding FnIII domain in such a way that it blocks the autoregulatory tail in its inhibitory positon. Thus, from the structure we could conclude that stretch-activation of Twc kinase seems unlikely and instead propose phosphorylation of Y 104 as a possible activation mechanism.

Our findings illustrate how the structural and functional diversity in titin's modular architecture has evolved not only on the basis of individual domains. Rather, functionality often involves adaptation of several neighboring domains or even whole Ig tandems/super-repeats. This is reflected in variations in mechanical and dynamic properties observed in different parts of the chain and highlights the necessity of working with representative multi-domain fragments to gain a comprehensive understanding of the titin chain.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

βME	β-mercaptoethanol
ABF	Adaptive Biasing Force
AFM	Atomic Force Microscopy
AK	Adenylate Kinase
asu	asymmetric unit
ATP	Adenosine Triphosphate
CAD	Coronary Artery Disease
CAM-K	Ca2+/Calmodulin-dependent Kinase
CARP	Cardiac Ankyrin Repeat Protein
CC	Coiled Coil
CD	Circular Dichroism
CHES	2-(*N*-Cyclohexylamino)ethane sulfonic acid
DARP	Diabetes Ankyrin Repeat Protein
DCM	Dilated Cardiomyopathy
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
FJC	Freely Jointed Chain
FnIII	Fibronectin type III
HCM	Hypertrophic Cardiomyopathy
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMERF	Hereditary myopathy with early respiratory failure
Ig	Immunoglobulin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kin	Kinase
LB	Luria Bertani broth
M-CK	Muscular isoform of Creatine Kinase
MARP	Muscle Ankyrin Repeat Protein
MD	Molecular Dynamics
MES	2-(*N*-morpholino)ethanesulfonic acid

MFC	MuRF Family Conserved
MIR	Multiple Isomorphous Replacement
MLC	Myosin Light Chain
MLCK	Myosin Light Chain Kinase
MMC	Methyl mercury Chloride
MOPS	3-(*N*-morpholino)propanesulfonic acid
MPD	2-Methyl-2,4-pentanediol
MR	Molecular Replacement
MuRF	Muscle Ring Finger
M_W	Molecular Weight
NMR	Nuclear Magnetic Resonance
OD_X	Optical Density at x nm
OT	Optical Tweezers
PCR	Polymerase Chain Reaction
PDH	Pyruvate Dehydrogenase
PEG	Polyethylene glycol
PEVK	Titin segment rich in proline (P), glutamate (E), valine (V), lysine (K)
PFK	Phosphofructokinase
PKA	Protein Kinase A
PKG	Protein Kinase G
PMF	Potential of Mean Force
R-factor	Crystallographic residual for working set of reflections
R-free	Crystallographic residual for test set of reflections
RBCC	Ring B-Box Coiled Coil
RDC	Residual Dipolar Couplings
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC-MALS	Size Exclusion Chromatography Multi-Angle Light Scattering
SMD	Steered Molecular Dynamics
TEV	Tobacco Etch Virus
TK	Titin Kinase
TMD	Tibialis Muscular Dystrophy

Tn-I	Troponin I
TnT	Troponin T
TNFα	Tumor Necrosis Factor α
TRIM	Tripartite Motif
Tris	Tris-(hydroxymethyl)-aminomethane
Twc	Twitchin
UBA	ubiquitin-associated domain
V_M	Matthews coefficient
WLC	Worm-Like Chain
Y2H	Yeast Two Hybrid

Acknowledgements

I am very grateful to my supervisor Dr. Olga Mayans for giving me the opportunity to work in her group and for taking us on an enlightening journey to Liverpool. With her guidance and support I was able to pursue an interesting and often challenging project and develop my research interests and skills in a constructive environment. In this vein, I would also like to thank Dr. Dietmar and Prof. Siegfried Labeit for a long-standing and fruitful collaboration on the titin and MuRF project as well as Dr. Guy Benian for our more recently established collaboration on twitchin.

I would also like to thank PD Dr. Michael Hennig for agreeing to coreferee this thesis and to Prof. Martin Spiess for taking on the role as chair. My thanks go also to Prof. Tilman Schirmer for taking over as faculty representative in Basel.

A big thank you to the current and former group members, both in Basel and Liverpool: Dr Marco Marino for teaching me biochemistry, Zohre Ucurum-Fotiadis for help with cloning and Dr. Arnaud Basle for all his troubleshooting in and out of the lab, especially during our time in Liverpool. Paul Wassmann and Christophe Wirth from the Schirmer lab were a big help in the SEC-MALS experiments. Pilar, Micha, Rainer, Sonja, Darko, Cian in Basel and Arnaud, Barbara, Marc and Sonia in Liverpool have provided and interesting and stimulating environment to work in as well as good times outside of work. A special thank you also to Dr. Laurent Kreplak for his involvement in the project and many helpful discussions.

People from other labs that have proved helpful and been good company in and out of the lab include: Marija, Marko & little Jana, Esther, Yvonne, David, Sara, Wanda, Julia, Herve, Rod, Jochen and Sophia in Basel. In Liverpool Alessandro, Kat, Martyna, Yann, the three Pauls, Nicolas, Gaz, Laurence, Chris, Santy, Mark and Pryank as well as others I might have forgotten here.

A big thank you also to my friends Alexandra and especially also Hannah, for keeping me grounded in real life and our good times without Science in sight.

Last, but definitely not least, I would like to express my deep gratitude to my family, for their continued support, encouragement and belief in me. To you I dedicate this thesis.

1 Introduction

1.1 The Sarcomere

The sarcomere, basic functional unit of skeletal and cardiac muscle, is one of the most complex cellular bio-machineries. Here, hundreds of proteins form an intricate web of dynamic interactions necessary to differentiate and maintain its ultrastructure and function (for a review see Clark et al., 2002). The sarcomere generates force via a contractile mechanism, with motion on the macroscopic scale being attained through the concerted action of the hundreds of sarcomeres forming each myofibril (see Figure 1.1 for the organization of muscle). In order to fulfill this task efficiently, sarcomeres within neighboring fibrils have to be correctly aligned parallel to each other. This near-crystalline arrangement of sarcomeres gives rise to a repetitive pattern of striations when observing cardiac or skeletal muscle under the light microscope (hence their collective term “striated muscle”).

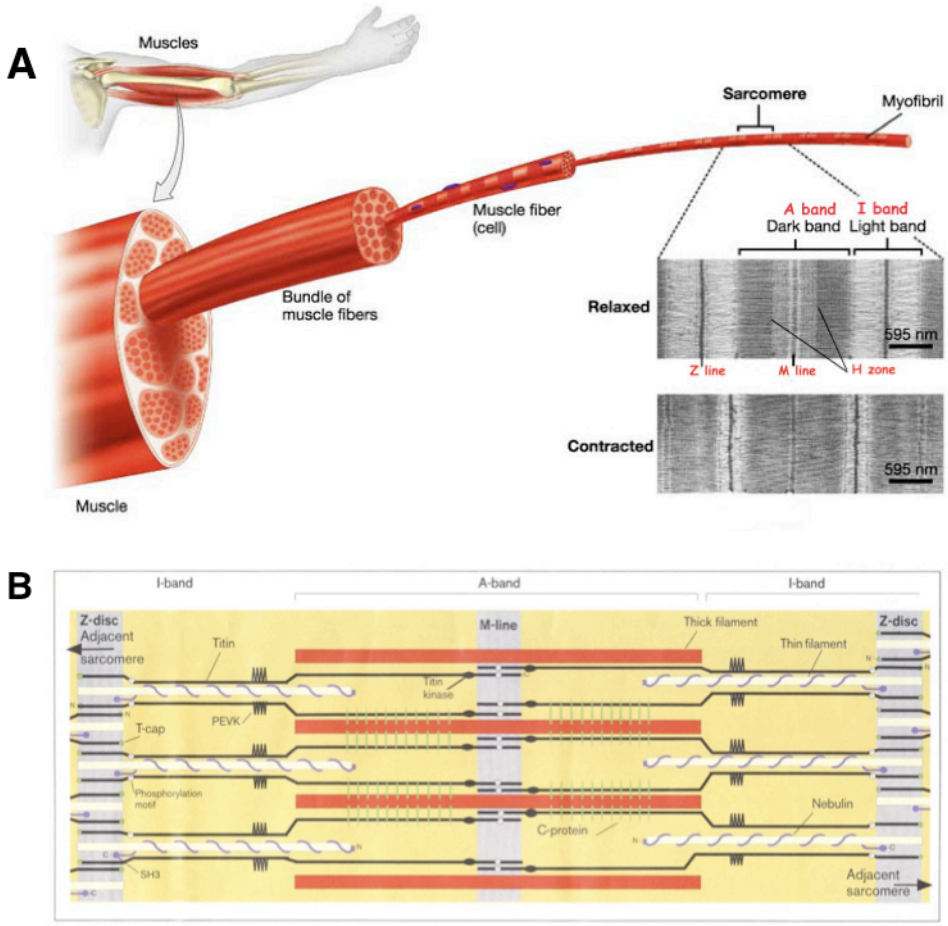


Figure 1.1 Overview of skeletal muscle organization
A) Skeletal muscle constitution from its macroscopic anatomy to the microscopic level. The inserted electron micrographs show the sarcomere in its relaxed and contracted states. As can be seen, the sarcomere shortens through the increased interdigitation of thick and thin filaments. The macroscopic shortening of skeletal muscle is achieved by the simultaneous and accumulative shortening of all its constituent sarcomeres. (Figure adapted from http://fig.cox.miami.edu/~cmallery/150/neuro/muscle.htm).
B) Principal components of the sarcomere. Four distinct filament systems (actin, myosin, titin and nebulin) maintain sarcomere integrity and function. The central A-bands contain myosin thick filaments, anchored in the M-line, and the I-bands contain actin thin filaments that extend into the A-band from their anchoring in the Z-disc. Titin spans half of a sarcomere, from the Z-disc to the M-line, where the C-terminal regions of titin molecules from opposite sarcomeres overlap. Nebulin closely interacts with thin filaments and is thought to act as a ruler for thin filament assembly. The C-terminal ends of nebulin molecules insert into the Z-disc; their amino-terminal ends are located at, or in close proximity to, the pointed ends of the thin filaments. (Figure adapted from Gregorio et al., 1999).

The sarcomere is delimited on either side by the Z-disc (“Zwischen”), seen in electron micrographs as a dense dark line bisecting a lighter region, termed the I-band (“Isotropic” in polarized light). Thus, each half of the I-band falls into different sarcomeres. In the middle of the sarcomere lies the darker A-band (“Anisotropic” in polarized light). Within the A-band the M-line (“Mitte”) demarcates both the midpoint of the sarcomere and the middle of the A-band and is surrounded by a region of lesser density termed the H-zone (“Hell”). Based on fine-structure in electron

micrographs the A-band can also be further divided into three zones, the C-zone encompasses the region of 11 striations, 14.3nm apart, where C-protein (also termed MBP-C for myosin binding protein C) binds to myosin (see Figure 1.1B), and is surrounded by the D-zone (distal) and P-zone (proximal) as seen from the M-line midpoint of the thick filaments (Sjöström and Squire, 1977).

The main components of the sarcomere are organized into three filament systems: the thin filaments, the thick filaments and titin. The thin filaments are anchored in the Z-disc, through their interaction with **α-actinin**, and extend through the I-band into the A-band where they interdigitate with the thick filaments with a symmetry of 6 thin filaments per thick filament. **Actin** is the main component of the thin filaments, together with the regulatory proteins **tropomyosin** and **troponin**. The filamentous protein **nebulin** has been postulated to act as a ruler for thin filament assembly. The bi-polar thick filaments on the other hand are made up of around 300 **myosin** molecules each. The myosin heads reverse their polarity at the M-line in a region called the bare zone, allowing for interaction with actin filaments from the other half of the sarcomere. In electron micrographs M lines show a fine structure of up to 5 parallel, transverse lines, referred to as M-bridges (Sjöström and Squire, 1977) that help to cross-link the thick filaments and keep them in register both laterally and transversally (see below, section 1.2.1.1). It is in the zone of interdigitation of thin and thick filaments, where the concerted actin-myosin interaction – regulated by troponin and tropomyosin and powered by ATP consumption – slides the thin filaments past the thick filaments, shortening the sarcomere and thereby generating muscle contraction. Intricate protein networks at the Z-disc and M-line anchor the thin and thick filaments, respectively, maintain them in register during contraction and link them to the sarcolemma, the membrane surrounding the muscle cell, in case of the Z-disk.

The third filament system is composed of a head-to-head and tail-to-tail arrangement of the giant protein titin, where single molecules extend from the Z-disc to the M-line, spanning over 1μm in relaxed sarcomeres. Titin, its architecture and various essential functions in muscle are the subject of this dissertation.

1.2 Multitasking titan

1.2.1 Titin architecture

Titin is a very large (M_W > 3MDa), highly modular protein composed of ~300 Ig and Fn-III domains linked in tandem (Bang et al., 2001; Labeit and Kolmerer, 1995), which gives the molecule its beaded appearance in electron micrographs (Figure 1.2) (Labeit et al., 1990; Trinick et al., 1984; Tskhovrebova and Trinick, 2002). Additionally, the protein contains a single catalytic domain, a serine/threonine kinase (abbreviated TK for titin kinase), the unique PEVK segment – thus named for the prevalence of proline (P), glutamate (E), valine (V) and lysine (K) – as well as some other unique sequences and non-repetitive insertions (see Figure 1.2). Structurally, the protein can be roughly divided into N-terminal and C-terminal halves. The N-terminal half contains the N-terminus anchored in the Z-disc lattice through its interaction with **telethonin** and a host of other Z-disc proteins as well as the I-band region that is mostly free from lattice interactions. The C-terminal half localizes to the A-band region where titin interacts with the thick filaments, while the C-terminus itself is firmly embedded in the M-line lattice.

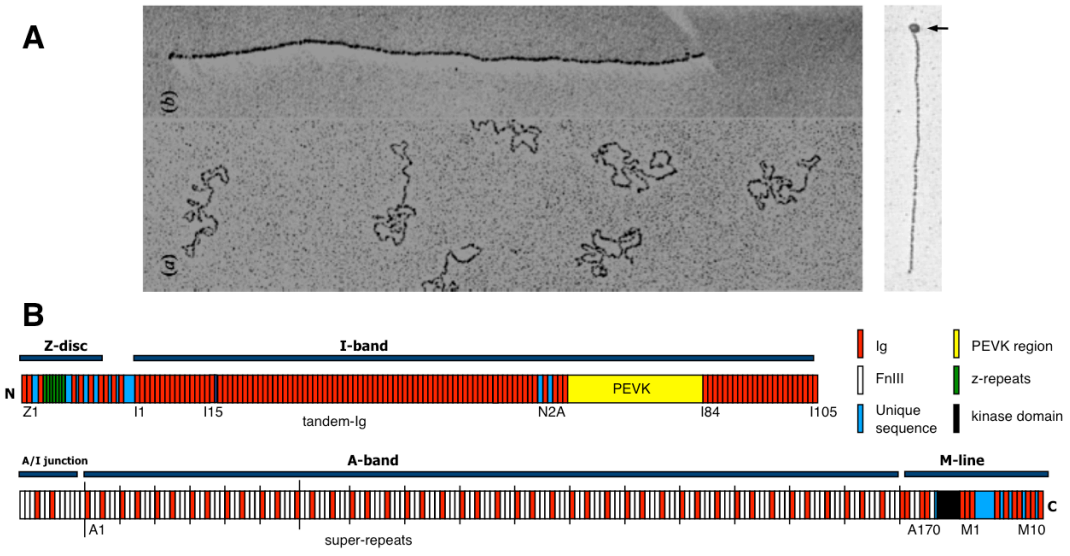


Figure 1.2: Titin architecture
A) Electron micrographs of full length titin molecules in stretched and coiled conformation visualized using electron microscopy (taken from Tskhovrebova and Trinick, 2002) Arrows point to a globular head that can often be seen and localizes to the C-terminus of the filament. B) Schematic representation of the modular organization of titin (skeletal N2A isoform).

The general architecture is modulated by extensive differential splicing that, depending on muscle type and developmental stage, gives rise to a large variety of titin isoforms ranging from ~3.0 to ~3.7 MDa in size (Bang et al., 2001; Freiburg et al., 2000; Labeit and Kolmerer, 1995) as well as a much shorter **novex-III** isoform (~700kDa) that are coexpressed within the same sarcomere. The latter contains an alternative C-terminus that encompasses a binding site for **obscurin** in the unique sequence located N-terminal to the novex-III domain I21 (Bang et al., 2001).

1.2.1.1 Sarcomeric regions of titin:

Z-disc - The two most N-terminal Ig domains of titin, Z1Z2, extend through the Z-disc into the border of the neighboring sarcomere, where they interact with the small sarcomeric protein telethonin (Gregorio et al., 1998). This interaction helps anchor titin to the Z-disc while telethonin further acts as an adaptor protein linking signaling and structural proteins to titin. Adjacent to this site lie the differentially spliced Z-repeats. These constitute binding sites for α-actinin, which cross-links titin with actin, further integrating titin in the Z-disc lattice. Bordering the I-band lie the two Ig domains Z9-Z10 that serve as primary binding site for obscurin (Young et al., 2001).

I-band – This is the elastic region of titin underlying the role of the molecule in muscle passive mechanics. The I-band region consists of constitutively expressed proximal (I1-I15) and distal (I84-I105) Ig tandems that flank a differentially expressed central tandem (I28-I79) and the unique PEVK segment, which is also differentially expressed (Figure 1.3). Spliced variants of the PEVK region range in size from 186 residues in cardiac isoforms to 2174 in soleus muscle (Bang et al., 2001), with their length directly correlated to the length of the spliced central Ig-tandem (Labeit and Kolmerer, 1995). In addition, there are two unique sequence insertions, the cardiac specific N2B (I24-I26) and the N2A (I80-I83) region, expressed both in skeletal and cardiac muscle that act as protein interaction hotspots.

The several titin variants predominate in different muscle types and different physiological stages. Skeletal muscles commonly express long titin variants (termed N2A forms), while heart muscle usually contains shorter filaments (N2B or N2BA titin types, where N2B is very short and stiff and N2BA is longer and more compliant). These variations in the contour length of titin I-band directly affect the passive elasticity and working range of the sarcomere depending on muscle type.

One particular variation is that of the novex-III titin isoform, where the I-band region contains a few sparse Ig domains (I18-I23), not in tandem and interspersed by unique sequences (Figure 1.3). The novex-III variant is ultra-short, with its C-terminus extending only ~100nm into the I-band (Bang et al., 2001). This isoform is present both in cardiac and skeletal muscle, but much less abundant than full-length conventional titin. Its coexpression together with the conventional, longer titin isoforms might be necessary to accommodate the different symmetries in the sarcomere (the hexagonal lattice of the thick filaments in the A-band *vs.* a square lattice arrangement of the thin filaments in the Z-disc, (see (Knupp et al., 2002) for potential scheme to reconcile this symmetry paradox).

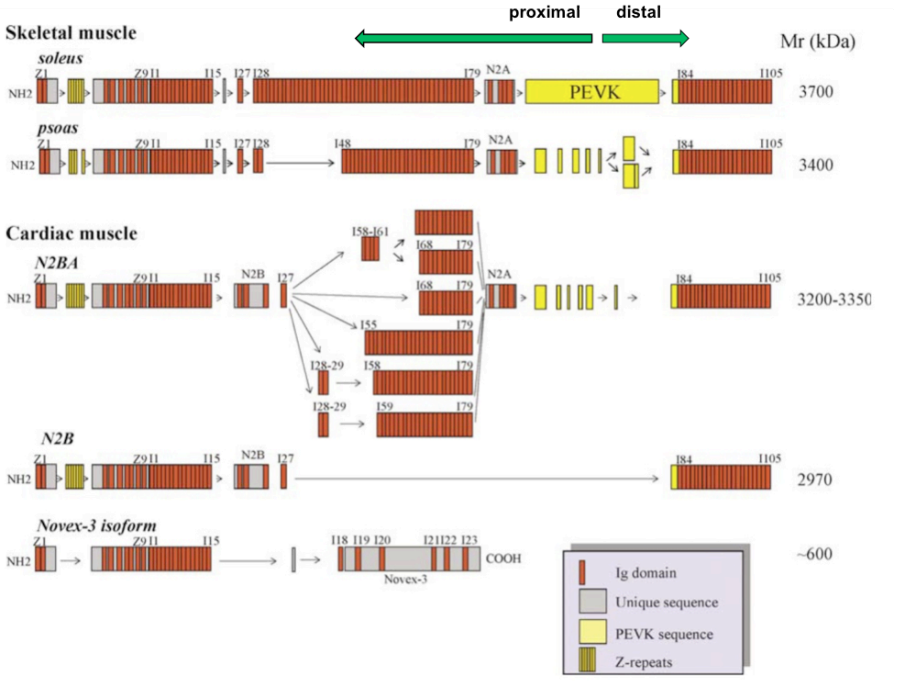


Figure 1.3: Spliceoform composition of I-band titin
The central part of titin I-band undergoes extensive differential splicing of its Ig-tandems and PEVK sequences. This results in multiple filament variants with different stiffness. This central region is flanked by constitutively expressed Ig-tandems. (Adapted from Granzier and Labeit, 2002).

A-band - The largest part of titin, namely around 2.1 MDa, is localized in the A-band region. This region shows a very high degree of conservation both across muscle types and different species. It is composed of FnIII (65%) and Ig (35%) domains arranged into patterns of super-repeats (Muhle-Goll et al., 2001). There are two types of super-repeats, composed of seven (Ig-$(FnIII)_2$-Ig-$(FnIII)_3$) and eleven domains (Ig-$(FnIII)_2$-Ig-$(FnIII)_3$-Ig-$(FnIII)_3$) respectively (Figure 1.4). At the edge of the

A-band lie six copies of the short super-repeat contiguous with eleven copies of the larger one located towards the C-terminus of the molecule. The short super-repeats coincide with the D-zone of the A-band whereas the longer ones span its C-zone. It has been suggested that the shorter 7-domain repeat is the evolutionary ancestor of the longer 11-domain repeat that formed by duplication of the last 4 domains (Kenny et al., 1999). Furthermore, domains at corresponding positions in the 7- and 11-domain super-repeats share higher sequence similarity than other domains of the same type within the same super-repeat (Kenny et al., 1999; Labeit et al., 1992). This indicates that the regular domain arrangement in the A-band arose from duplication on the level of whole super-repeats. The periodicity of the 11-domain super-repeat correlates with the 42.9nm thick filament helical repeat as well as with the pattern of the 11 C-zone stripes (see Figure 1.1B) (Craig and Offer, 1976; Labeit et al., 1992).

It has been shown that titin binds both to myosin (Labeit et al., 1992; Muhle-Goll et al., 2001) and to MBP-C (Freiburg and Gautel, 1996; Houmeida et al., 1995; Whiting et al., 1989). Electron microscopy of myosin thick filaments isolated from mice gave rise to a three-dimensional reconstruction with a resolution of 4nm (Zoghbi et al., 2008). In this reconstruction, 11 extra densities, not explained by the myosin heads, with an axial separation of 4nm are seen in every 42.9nm repeat. They appear to form a longitudinal strand running along the axis of the thick filament, consistent with titin's 11-domain super-repeat. This would place titin directly on the surface of the thick filament, running approximately parallel to the filament axis. Furthermore, comparison of wt and MyBP-C knockout filaments enable localization of three MyBP-C domains within the thick filament reconstruction, adjacent to both myosin heads and the titin-assigned density.

Taken together, this data strongly supports the proposed role of titin as a molecular ruler for thick filament assembly.

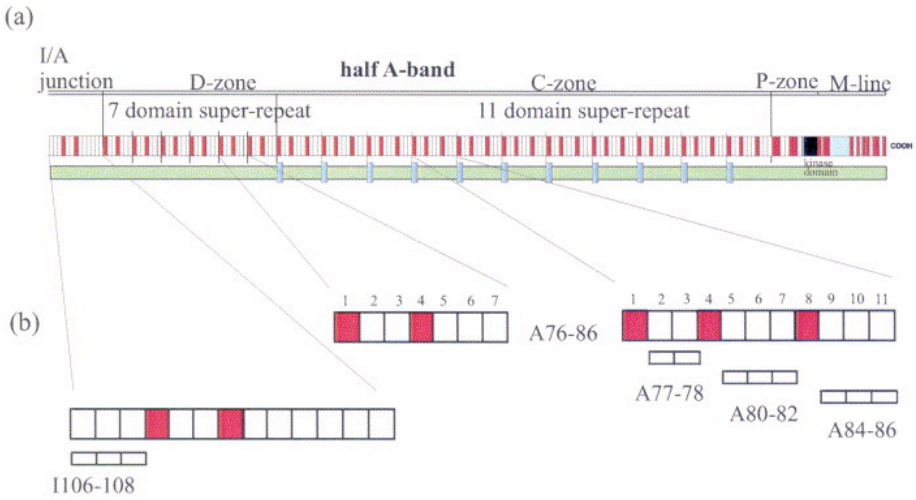


Figure 1.4: **Modular organization of the A-band of titin**
Both the arrangement (a) and the composition (b) of the super-repeats found in this region are shown. Red boxes represent Ig-domains, FnIII domains are shown in white. Underneath, the myosin filament is sketched in green with bound C-protein in blue. (Taken from Muhle-Goll et al., 2001).

M-line - In the boundary between the A-band and the M-line, i.e. in the P-zone, the regular pattern of super-repeats breaks down into a unique Ig-Ig-Fn-Fn-Ig-Ig-Fn motif that leads onto the C-terminal 250kDa of the protein. The titin kinase lies at the border of this region and is followed by 10 Ig domains that are interspersed with 7 unique insertions (is1-is7) of varying size and predicted secondary structure (Kolmerer et al., 1996; Labeit and Kolmerer, 1995) (see Figure 1.2 and 1.4).

Titin is an integral part of the M-line matrix where it interacts with other filamentous M-line proteins, namely myomesin (190kD protein) and M-protein (165kD protein) that are part of the globular head domain[1] seen in isolated titin molecules (Nave et al., 1989; Vinkemeier et al., 1993). Both these proteins are involved in the formation of the M-bridges and bind both to the central bare zone of myosin filaments (Bähler et al., 1985) and to the C-terminus of titin (Nave et al., 1989). Thus they form a transversal network of filamentous proteins involved in transversal and longitudinal organization of the sarcomeric lattice in the M-line similar to alpha-actinin in the Z-disc (for a review see Agarkova and Perriard, 2005).

[1] Oriented molecules of purified titin extracted from myofibrils uniformly visualize in electron microscopy as elongated thin rods with a globular head domain (see Figure 1.2). Immuno-electron microscopy revealed that the head domain localizes to the M-line in sarcomeres and is composed of M-line proteins (Nave et al., 1989).

1.2.1.2 Invertebrate titin homologs

Invertebrate sarcomeres show a much larger variation in size and structure than in vertebrates, as they have adapted to widely varying functions of muscles often not found in vertebrates. While no giant molecule that fulfills all the roles of vertebrate titin has been found, several titin-related proteins of varying size have been identified in phyla ranging from arthropods to mollusks and nematodes (for reviews see Bullard et al., 2002; Ferrara et al., 2005; Tskhovrebova and Trinick, 2003). These proteins, also termed mini-titins, share the filamentous, multi-domain composition of titin and can be roughly divided into two classes – those that resemble the titin I-band in their composition and those that reproduce the conserved, repetitive domain composition of the titin A-band region (see Figure 1.5). Their modular architecture makes the proteins very adaptable through step-wise modification. Thus, the proteins can be tailored to the specific requirements of the species/phylum through sequence variations and insertions, obviating the need for new genes or gene duplications and gene evolutions to gain new functions. To that effect, large stretches of these mini-titins exhibit conserved domain arrangements while other regions have diverged to accommodate special requirements.

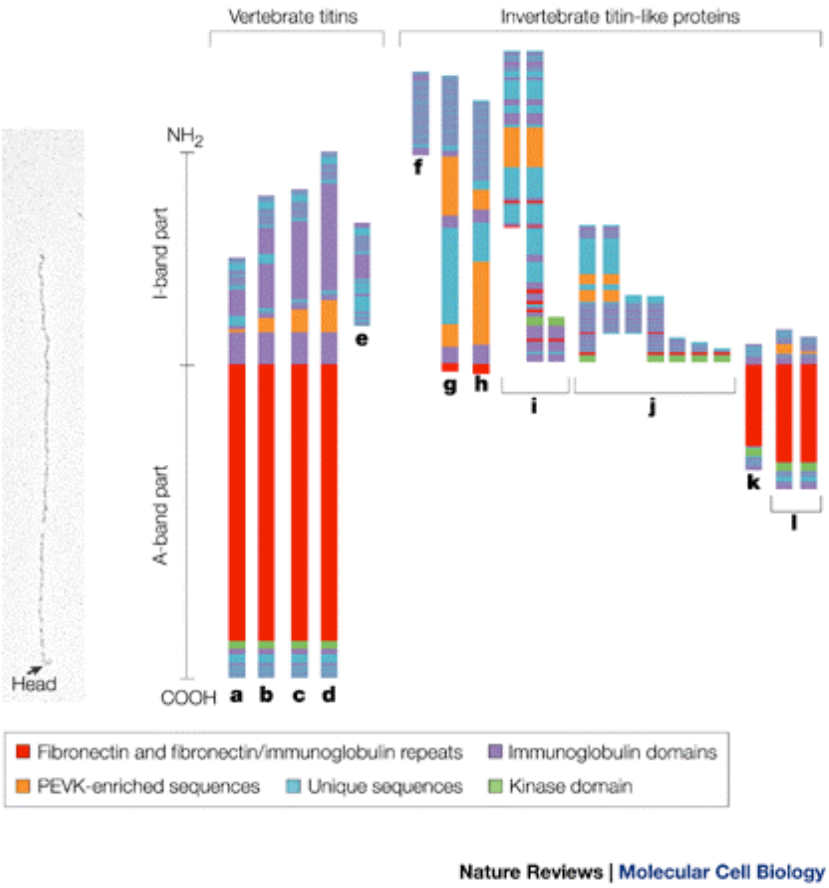


Figure 1.5: Overview of the titin family
The titin protein family comprises a range of filamentous intra-sarcomeric muscle proteins that exhibit morphological, sequence and functional similarities. In vertebrates, titin stretches all the way from Z-disc to M-line (a-d), with the exception of Novex-III (e). In invertebrates a number of related proteins, termed mini-titins, can be grouped according to their sarcomeric location, spanning either the I-band region or located in the A-band, associated to the thick filaments
Kettins (f) are 0.5-0.7 MDa in size and are most likely expressed as a truncated isoform of larger proteins by alternative splicing, e.g. from D-titin (2.0MDa, g), found in *D. melanogaster*, or I-connectin (~2.0MDa, h) found in crayfish. The larger isoforms span the I-band and form elastic connections between he Z-disk and the thick filaments in giant invertebrate sarcomeres. The smaller kettin isoforms might fulfill the same function in sarcomeres with short I-bands, e.g. in indirect flight muscle.
The stretchins (j, found in Drosophila) and *C. elegans* titins (i) are expressed in many isoforms from single genes. They mainly contain Ig and unique sequences, but both groups also contain a kinase domain close to their C-terminus, that belong to the same family as titin and smooth muscle myosin light chain kinase.
Finally, the twitchins (k) and projectins (l) resemble the A-band region of titin, as they are around 0.8-1.0MDa and are composed of FnIII and Ig repeats and a C-terminal kinase domain. Twitchin is mainly localized in the A-band region of the sarcomere, while projectin is found either in the I-band or the A-band depending on muscle type. (taken from Tskhovrebova and Trinick, 2003)

1.2.2 Titin function

Titin fulfils many essential functions in muscle both during myofibrillogenesis and in mature tissue. However, it is best known for being the main source of passive tension in the sarcomere (Granzier and Irving, 1995; Horowits et al., 1986). Such passive tension restores the sarcomere to its resting or slack length upon over-stretch or compression, ensures a regular distribution of force and controls the position of the actin-myosin motors, centering the thick filaments in the sarcomere. Titin also plays a role in the assembly of sarcomeres during myofibrillogenesis where it is thought to act as a molecular ruler (for reviews see Granzier and Labeit, 2004; Miller et al., 2004; Tskhovrebova and Trinick, 2003). Furthermore, the titin filament system acts as a scaffold for the recruitment and compartmentalization of other sarcomeric components, many of which are involved in signaling pathways or metabolic enzymes. Because of its inherent elasticity and prominent location in the sarcomere titin is thought to be a biomechanical stress sensor, involved in the triggering of signaling pathways that regulate the adaptive remodeling of muscle to mechanical load.

1.2.2.1 Titin elasticity:

At physiological forces, the titin filament is the primary molecular determinant of elasticity in the skeletal and cardiac muscles from vertebrates. These store mechanical energy and stretch over twice their resting length without disrupting their structural integrity. In this context, titin acts as a bi-directional spring that stretches and recoils, ensuring that the sarcomere is restored to its resting length after contraction or extension[2], promoting a regular distribution of force and supporting the correct positioning of actin-myosin motors during muscle function.

The spring components of titin are located in its I-band fraction, which forms an elastic connection between the ends of the thick myosin filaments and the Z-disc (Granzier et al., 1996; Labeit and Kolmerer, 1995; Linke et al., 1996; Linke et al., 1999; Linke et al., 1998). The I-band segment behaves like two (isoform N2A) or

[2] Muscles work in antagonistic pairs, e.g. the biceps and triceps in the upper arm that are responsible for bending and stretching the arm respectively. This is due to the fact that sarcomeres as underlying working units can only actively contract through the action of the acto-myosin motors. Thus, through the antagonistic arrangement, while one muscle (actively) contracts the antagonistic muscle, and therefore its constituent sarcomeres, gets stretched while in a passive state, i.e. without any energy consumption. Hence the name passive stretch.

three (isoforms N2B, N2BA) sequentially linked entropic springs that extend at different forces (Figure 1.6). In this region, titin consists of two main springs, the PEVK segment and the flanking poly-Ig tandems. Both segments straighten upon myofibril stretch causing the development of passive entropic tension in the sarcomere. Poly-Ig arrays extend at low force while PEVK-repeats unravel at higher load, with the combined action of both springs defining the mechanical stiffness of the sarcomere (Granzier and Labeit, 2004; Trombitás et al., 1998; Trombitás et al., 2003 among many others). In cardiac muscle, the N2B element acts as a third, non-modular spring, which extends last, if needed, and only at high sarcomere stretch (Linke et al., 1999). The N2B spring might compensate for the intrinsically higher passive stiffness of the shorter cardiac isoforms.

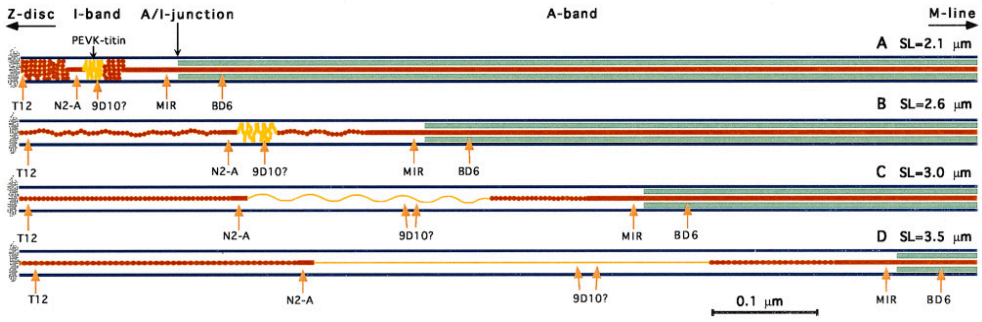


Figure 1.6: Model of titin extension during sarcomere stretch.
Shown is a major portion of the half-sarcomere including the whole I-band region at four different stages of stretch. A) slack, B) minor stretch, C) moderate stretch and D) extreme stretch, with sarcomere lengths (SL) representative for the situation in psoas muscle indicated at the right side. Color code: dark blue, actin; green, myosin; yellow, PEVK, red, non-PEVK titin, with the I-band Ig domains that are relevant to stretch represented as circles. Arrows indicate known binding sites for titin antibodies used in labeling studies of stretched muscle. (Figure taken from Linke et al., 1996)

The importance of stretch-recoil and stiffness control in muscle function is further demonstrated by the finely tuned composition of both poly-Ig and PEVK segments in titin that, through alternative splicing, undergo constant and extensive adaptation to the mechanical load in the different physiological and pathological states of muscle (Granzier and Labeit, 2002) (Figure 1.3). The turnover and adaptive remodeling of titin isoforms is particularly crucial in heart, where it impacts diastolic filling by modulating myocardial stiffness. This causative relation is proven by patients suffering from Coronary Artery Disease (CAD) or dilated cardiomyopathy (DCM) who exhibit altered titin isoform expression patterns in their cardiac tissue.

The passive elasticity of titin can also be modified short-term in a dynamic fashion. Phosphorylation of the N2B unique sequence ($N2B_{us}$) insertion by protein

kinase A (PKA) or G (PKG) at Ser469 (of the $N2B_{us}$) in response to β-adrenergic stimulation was shown to lower passive tension (Fukuda et al., 2005; Krüger et al., 2009; Krüger and Linke, 2006; Yamasaki et al., 2002). Binding of Ca^{2+} to Glu-rich repeats found in the PEVK segment on the other hand increased passive tension in experiments on skinned muscle preparations (Labeit et al., 2003).

1.2.2.2 Titin recruitment of sarcomeric proteins

Scaffold proteins play an important role in the crowded environment of the cell in localizing and facilitating the interactions between other molecules (reviewed in Vondriska et al., 2004). Binding to scaffold proteins can provide a means for selective spatial orientation as well as temporal coordination (e.g. sequential binding) of individual components involved in complex formation. Scaffolding can also influence the specificity and kinetics of interactions by enforcing localized higher concentrations and by preventing interactions with components of other pathways through compartmentalization. Alternatively, binding can also help to sequester proteins and keep them inactive. Both the binding to the scaffold and activity of the individual molecules can be influenced e.g. by conformational changes, post-translational modifications as well as competitive or non-competitive interactions with other components.

In this regard its modular architecture and central position in the sarcomere predispose the titin filament as an ideal candidate to act as a scaffolding protein. Indeed, along its length titin is known to bind numerous sarcomeric proteins in addition to all the interactions with structural components that help build and maintain the sarcomeric lattice. Mapping of the binding sites shows that they are clustered in three regions to form protein interaction nodes or compartments (see Figure 1.7) (for reviews see Granzier and Labeit, 2004; Krüger and Linke, 2009; Lange et al., 2006; Miller et al., 2004).

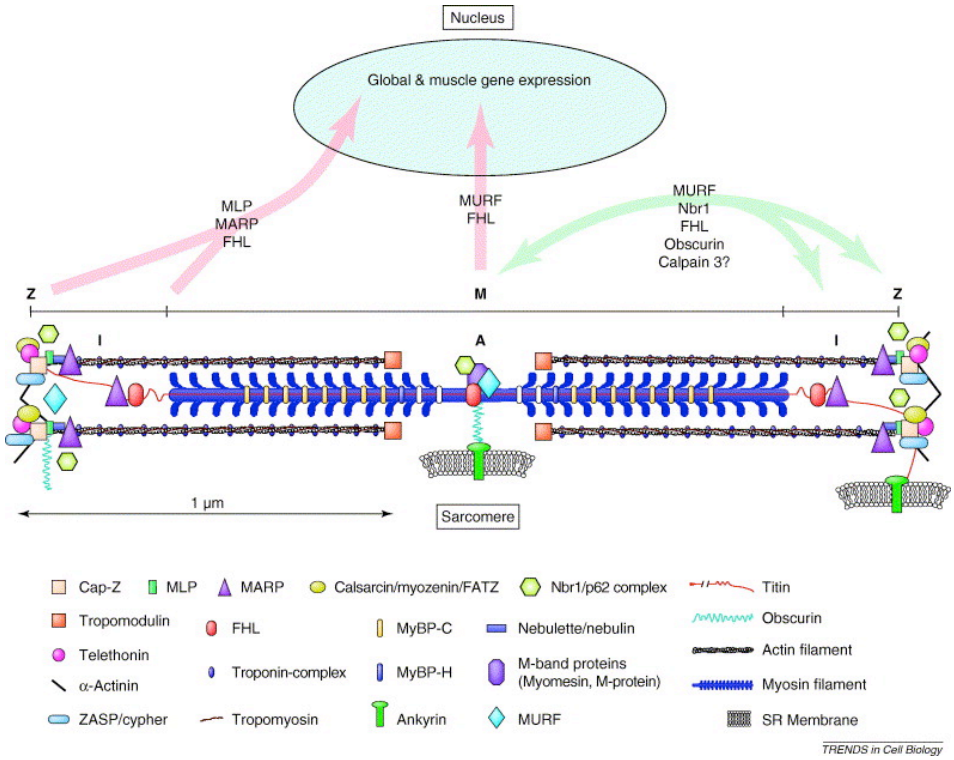


Figure 1.7: Scheme of titin protein interactions in the sarcomere
Simplified scheme of titin within the sarcomere. Several sarcomeric proteins show multiple binding sites along titin, indicated with green arrows, as well as the potential to relocate to the nucleus, indicated with pink arrow. (Taken from Lange et al., 2006)

In the Z-disc the titin binding partner telethonin interacts with a range of proteins including **muscle LIM protein** (MLP), that shuttles to the nucleus in response to stretch and is thought to play a role as a positive regulator in myogenic differentiation and hypertrophy (Arber et al., 1994; Boateng et al., 2007).

In the I-band both the N2B and N2A segment accommodate several protein interaction motifs. The N2B segment was shown to bind **DRAL/FHL-2**, which has been implicated to act as an adaptor protein involved in cardiac stress management and signaling (Lange et al., 2002). DRAL/FHL-2 interacts with the metabolic enzymes creatine kinase (M-CK), adenylate kinase (AK) and phosphofructokinase (PFK) (Lange et al., 2002). DRAL/FHL-2 was also found in the M-line where it binds to the unique sequence insertion is2 (Lange et al., 2002). Further, N2B also interacts with **αB-crystallin**, which is known to function as a chaperonin in assisting protein folding (Bullard et al., 2004).

In the N2A segment domains I82/I83 constitute a binding site for calpain protease **P94**, also known as calpain-3, which is involved in protein degradation (Ono et al., 2004). Binding of P94 to titin was shown to fine-tune and regulate the activity of

P94 (Sorimachi et al., 1995; Sorimachi et al., 2000), prevent P94 autolysis and to protect titin from proteolysis (Hayashi et al., 2008). The unique insertion between I80 and I81 contains a Tyr rich motif that interacts with a conserved motif found in the three homologous muscle ankyrin repeat proteins (MARPs), **cardiac ankyrin repeat protein** (CARP), **ankrd2** (also Arpp) and **diabetes ankyrin repeat protein** (DARP) (Miller et al., 2003). All three proteins show cytokine-like induction in response to cardiac injury and muscle denervation (CARP), skeletal muscle stretching (ankrd2) or after starvation (DARP), indicating that they might play a role in muscle stress response. **Myopalladin**, which interacts with the SH3 domain of nebulin and the EF-hands of α-actinin in the Z-disc, also localizes to the complex formed at the N2A node, through its interaction with CARP. Interestingly, all proteins of the N2A complex can also be found in the nucleus (Lange et al., 2006; Miller et al., 2003).

In the M-line titin domains A168-A170 adjacent to TK represent a binding site for **MuRF-1**, a member of the muscle ring finger protein family that acts as an ubiquitin E3 ligase and plays a role in muscle protein turnover and myofibril degradation (Bodine et al., 2001; Kedar et al., 2004). **MuRF-2**, a second member of the family, together with **nbr-1** and **p62**, two zinc-finger proteins, was shown to be part of a signaling complex (signalosome) that assembles around the TK and is involved in the control of gene expression and protein turnover in mature muscle (Figure 1.8) (Lange et al., 2005).

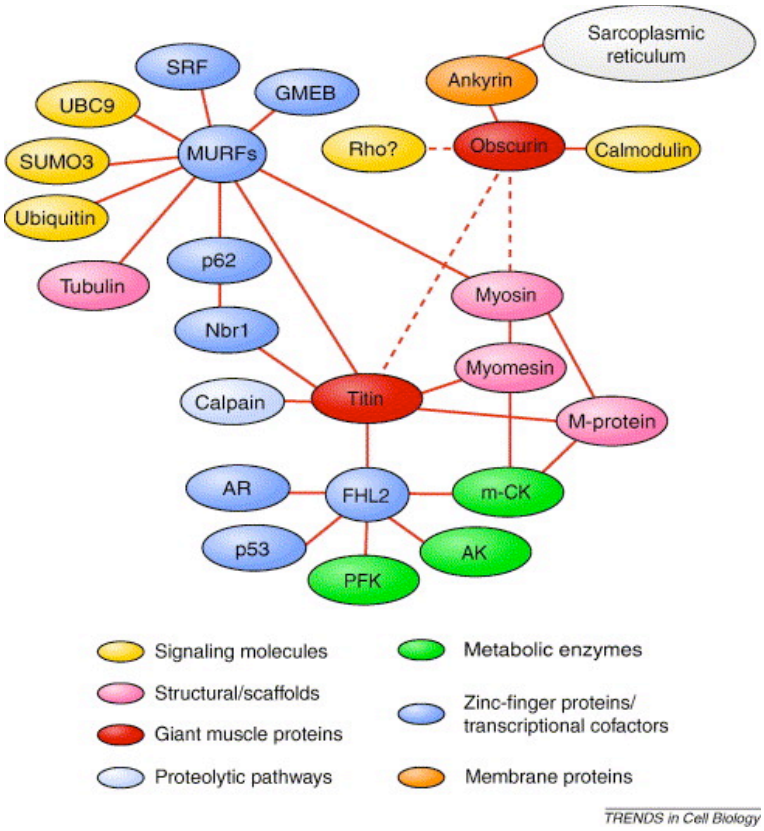


Figure 1.8: M-line interaction network centered around titin
The scheme shows the complex network of protein interactions, where many components also exhibit multiple cellular localizations. Besides the network of protein interactions that make up the M-line lattice that helps provide structural stability, many of the interactions involve signaling molecules or transcriptional cofactors and thus provide feedback to transcriptional pathways. Putative or indirect links are shown as dashed lines. (Taken from Lange et al., 2006)

Thus, in summary, titin binds to a range of metabolic enzymes, potentially helping to localize them in areas of high energy demand, and also facilitates interactions of a host of proteins, many of which are known signaling domains acting in muscle pathways. These proteins are believed to sense the level of mechanically induced conformational changes in the filament and, thus, are though to be involved in stretch sensing and tissue turnover by regulating gene expression (Krüger and Linke, 2009; Lange et al., 2006; Miller et al., 2004)

1.2.3 Titin kinase and the M-line signalosome

In developing myocytes, after activation by an as yet unidentified, developmentally regulated kinase, TK in turn phosphorylates telethonin, which in mature muscle is located >1μm away in the Z-disc and only transiently colocalizes with the kinase (Mayans et al., 1998). This has led to the conclusion that TK plays a role in myofibrillogenesis, which has been confirmed by the fact that expression of a constitutively active kinase construct leads to the breakdown of the normal cytoskeletal architecture (Mayans et al., 1998). Its role in mature muscle on the other hand is still unknown. Therefore, in an effort to identify potential targets or activators of TK, yeast two-hybrid (Y2H) screening was performed using TK as bait (Centner et al., 2001). Since the arrangement of domains flanking the kinase is conserved in other giant protein kinases like myosin light chain kinase (MLCK), and invertebrate titin homologs projectin and twitchin (Bullard et al., 2002) (see also Figure 1.5), the bait construct was designed to include domains A168-A169-A170-TK-M1-M2-M3 (Centner et al., 2001). This identified a new RING-finger protein, termed MuRF1 for Muscle RING-Finger protein, that binds N-terminally of TK to the domains A168-A169 via its central α-helical region. Two more family members were found by further Y2H screening using MuRF1 as bait, and were termed MuRF2 and MuRF3 (Centner et al., 2001). MuRF1 was shown to be an ubiquitin E3 ligase (Bodine et al., 2001) that is upregulated during muscle atrophy and targets a range of myofibrillar as well as metabolic enzymes (see section 4.1.2 and Table 4.1). Hence, MuRF1 is thought to be involved in ubiquitin-depended degradation of muscle protein (Bodine et al., 2001)

Expression of GFP fusion constructs of either MuRF1 or its binding site on titin, A168-A170, led to the disruption of sarcomeric integrity (Mcelhinny et al., 2002), showing that this interaction and its proper regulation are important for the maintenance for an ordered M-line structure (Mcelhinny et al., 2002). Furthermore, using a range of deletion constructs, it was shown that the central region of MuRF1 was responsible for M-line targeting, while the RING domain targets the protein to the nucleus. Further, *in vitro* interaction studies showed that MuRF1 interacts with ubiquitin-conjugating enzyme 9 (Ubc9) and isopeptidase T-3 (ISOT-3), enzymes involved in SUMO modification, as well as glucocorticoid modulatory element binding protein-1 (GMEB-1), a nuclear protein implicated in transcriptional regulation (Mcelhinny et al., 2002). This suggests that MuRF1 plays an important role in maintaining M-line structure, as its binding to A168-A170, adjacent to TK, suggests that it might be involved in kinase-based signaling pathways. At the same time its

nuclear localization and interaction with GMEB-1 suggest a nuclear function, potentially in the regulation of gene expression (Mcelhinny et al., 2002). Taken together with its role as ubiquitin E3 ligase, this suggests that MuRF1 is involved in the control of muscle protein turnover and degradation, possibly under the control of TK.

Recently it has been proposed that TK might be activated in a stretch-dependent fashion through structural rearrangements, and thus acts as a stress sensor (Gräter et al., 2005). Yeast two-hybrid (Y2H) screening of a TK construct designed to mimic this mechanically activated kinase identified the zinc-finger protein **nbr1** as a protein that interacts specifically with this potentially semi-opened state of TK but not the fully inhibited TK (Lange et al., 2005).

In a subsequent Y2H screen nbr1 was found to interact with **p62**, a related multi-compartment Zn-finger protein that together with nbr1 acts as a prominent scaffolding unit of large signaling complexes (Pawson and Scott, 1997). Both nbr1 and p62 were *in vitro* substrates of TK, with substrate sites in their N-terminal region. They could also be localized at the M-line in close proximity to TK muscle cells (Lange et al., 2005). P62 functions as a multivalent signaling platform, which interacts with many more kinase signaling pathways in addition to the one of TK. In a search to identify muscle-specific response factors MuRF2 was found to specifically interact with the ubiquitin-association domain (UBA) of p62 (Lange et al., 2005). Under atrophy or mechanical arrest, MuRF2 was shown to translocate to the nucleus, where it can interact with **serum response factor (SRF)**, and suppress SRF dependent muscle gene expression (Lange et al., 2005) (see Figure 1.9). SRF controls expression of immediate early genes, *e.g.* c-fos, which are responsible for the response of muscle to hypertrophic stimuli, like mechanical stress (Fuller et al., 1998; Sadoshima et al., 1992).

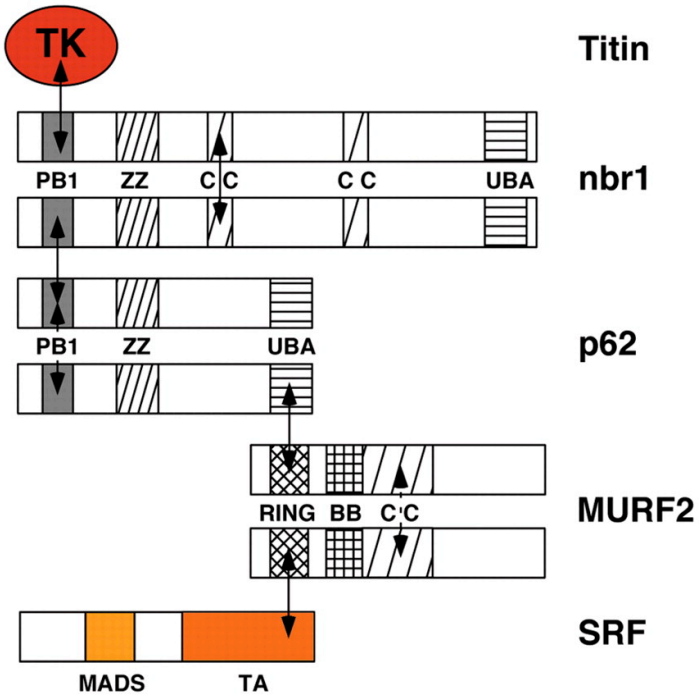


Figure 1.9: The M-line signalosome
Y2H screening using TK as bait identified nbr1 as a scaffold that in turn recruits p62 and MuRF2 to the signalosome. Interactions were further mapped biochemically and interacting domains are indicated by arrows. (Taken from Lange et al., 2005)

Support for the involvement of the signalosome assembling around TK in the control of muscle turnover came from analysis of the underlying mutation causing hereditary myopathy with early respiratory failure (HMERF), an autosomal dominant muscle disease. Patients suffering from this disease were found to carry a heterozygous point mutation CGG → TGG, leading to the exchange of an absolutely conserved Arg for a Trp in the regulatory tail of TK at position 279 (R279W) (Lange et al., 2005). The mutation maps to the binding site of nbr1 and indeed, the interaction with TK was much reduced and both nbr1 and MuRF2 were abnormally localized in patient biopsies.

Furthermore, in a mouse model conditional knockout of the M-line exons Mex1 and Mex2, coding for the kinase as well as the surrounding domains encompassing the binding sites for MuRF1/2/3, DRAL/FHL2 and myomesin, is embryonic lethal (Gotthardt et al., 2003) when activated at early stages of development. If the conditional knockout was activated at later stages of development, severe progressive myopathy led to the disassembly of existing sarcomeres and to an early death at five weeks of age (Gotthardt et al., 2003). In a study using a complete knockout of these M-line exons, it could be shown that the titin M-line region was dispensable for early sarcomeric assembly, but necessary for complete integration of

titin into the M-line lattice and the formation of a continuous filament system (Weinert et al., 2006). Since not only TK but also other protein binding sites are affected by this knockout, the effects seen in these experiments can therefore not be attributed to abolished kinase signaling alone.

In conclusion, the upstream activators, cellular substrates and role for TK in mature muscle are still not elucidated. The identification of the signalosome of nbr-1, p62 and MuRF-2 assembling around TK (Lange et al., 2005) and the binding of MuRF-1 to adjacent domains A168-A170 (Centner et al., 2001), however, point to its involvement in the regulation of protein expression and degradation in striated muscle (Centner et al., 2001; Lange et al., 2005; Mcelhinny et al., 2002).

2 Aim

The focus of this work is to understand the molecular mechanisms by which titin mediates muscle elasticity and concomitantly acts as a biomechanical stress sensor, involved in the adaptive remodeling of muscle in response to sarcomere activity. To date, no structural data is available on the spring elements of the I-band beyond single Ig domains. We therefore set out to investigate the structural organization and dynamics of representative fragments of the Ig tandem, I65-I70 and I101-I106. Successful structure elucidation should allow for the first time the formulation of a structure-based model of titin elasticity as well as probing titin chain dynamics by MD simulations. In the M-line, meanwhile, TK and its adjacent domains A168-A170 have been implicated in the recruitment of MuRF proteins to form a signalosome involved in the regulation of the trophic state of muscle. The structural and biophysical characterization of this complex will allow analyzing the molecular basis of the interaction and aid studies on the blockage of complex formation *in vivo* by pharmacological means, which could be a potential treatment in patients suffering from muscle atrophy. Finally, elucidation of the crystal structure of the FnIII-Kin-Ig region of the C. elegans titin homolog twitchin should for the first time allow the visualization of the kinase domain within the context of the filament and allow the exploration of potential activation mechanisms of this auto-inhibited kinase. Taken together this range of representative structures should shed light on how conformational changes in the titin chain are responsible for passive elasticity as well as the stretch-dependent recruitment of signaling complexes to the filament.

3 Titin elasticity

The crystal structure of titin I-band domains I65-I70, which is presented in this chapter, allows us for the first time to formulate a structure-based model of titin I-band elasticity, which we term the carpenter ruler model and which will be critically discussed. MD simulations probed the mechanical properties of I65-I70 under weak stretching forces in order to explore the secondary and tertiary structure elasticity of this construct. Furthermore, a study that explored the applicability of ultra-low *ab initio* methods for the phasing of filamentous protein structures based on I65-I70 is presented. Finally, we present the characterization and crystallization of a fragment of the distal, constitutive Ig tandem.

3.1 Introduction

3.1.1 Identification of spring elements in titin I-band

In the titin I-band the Ig tandems and PEVK segment behave like two independent springs, linked in series, that extend at different forces (Gautel and Goulding, 1996; Linke et al., 1996). In slack sarcomeres the I-band fraction is not straight but rather contracted, as shown by stretching experiments on antibody-labeled myofibrils (Granzier et al., 1996; Trombitás et al., 1998; Trombitás et al., 1995). These studies in combination with experiments on single titin molecules as well as recombinant fragments using AFM (recently reviewed in Linke and Grützner, 2008) and Optical Tweezers (OT) (Di Cola et al., 2005; Kellermayer et al., 1998, 2001; Leake et al., 2004 and many more) showed that upon application of directed, external force first the Ig region of the molecule is straightened, then, at higher forces, the PEVK domain unravels (Figure 3.1 and Figure 3.2). In cardiac muscle, the N2B element acts as a third, non-modular spring, which extends last, if needed, and only at high sarcomere stretch (Linke et al., 1999). The N2B spring might compensate for the intrinsically higher passive stiffness of the shorter cardiac isoforms (Granzier and Irving, 1995; Trombitás et al., 2000). Finally, in single molecule experiments, high forces and/ or high speeds of extension lead to sequential unfolding of the β-sandwich domains of the Ig, giving rise to a characteristic saw-tooth pattern in their force-extension curves. Ig-domains begin to unravel at 20-300pN, with the exact value depending on stretch speed, (Kellermayer et al., 1997; Rief et al., 1997; Tskhovrebova et al., 1997) and only refold once the applied force is lowered to a few piconewtons (~2.5pN; Kellermayer et al., 1997). This hysteresis shows unfolding-refolding to be a very inefficient and improbable process in muscle, especially the heart, which undergoes continuous stretch and contraction cycles, and cannot afford such an enormous energy loss.

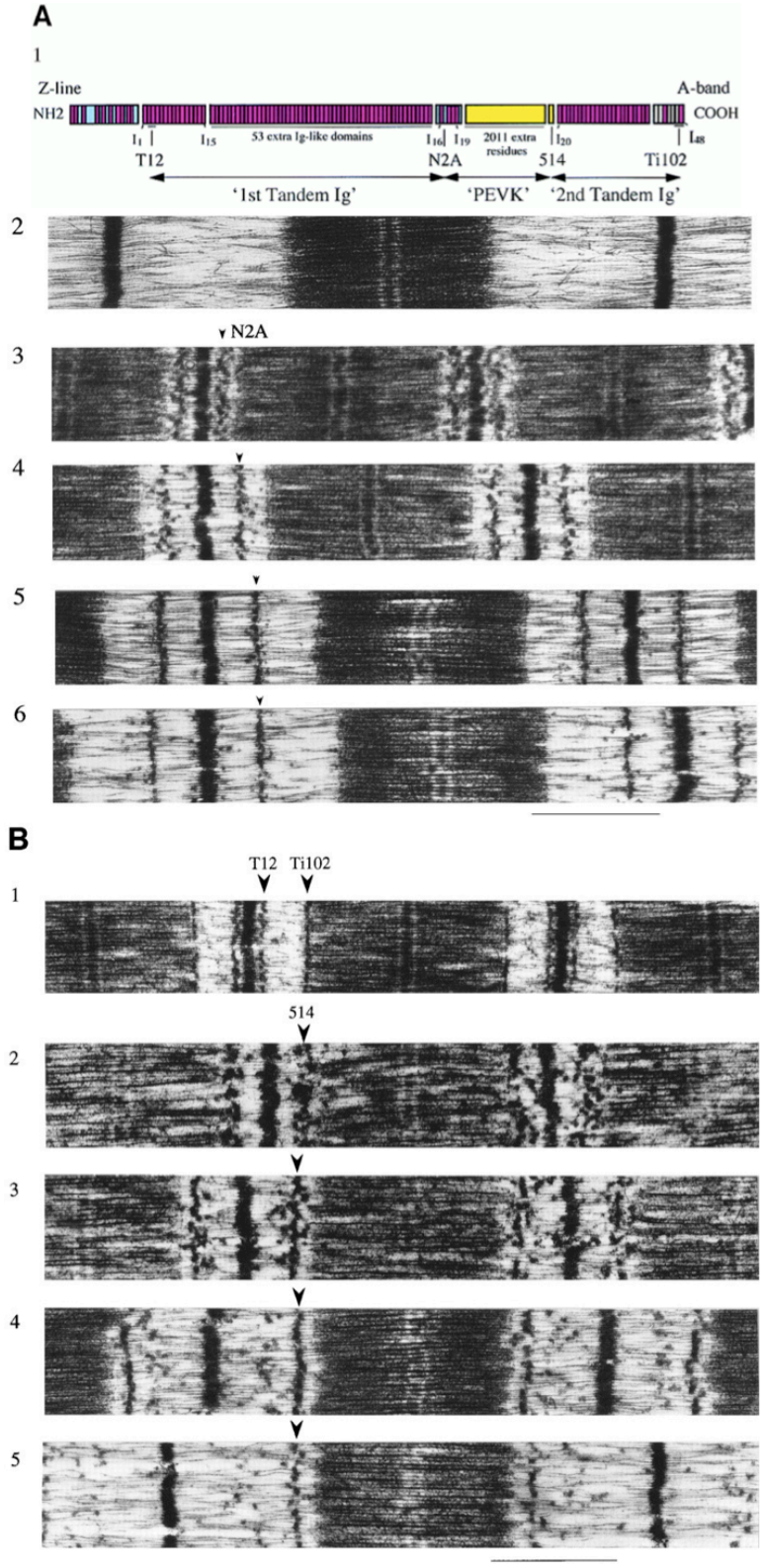


Figure 3.1: Electron micrographs of sarcomeres stretched and labeled with antibodies Labeling used anti-titin antibodies T12, N2A, 514, and Ti102 (binding sites indicated in A1). (A1) schematic of I-band segment of human soleus muscle titin (Labeit and Kolmerer, 1995). Color scheme: Red - Ig; White - FnIII; Blue - unique sequence; yellow - PEVK. Domain numbering according to that of cardiac titin with extra domains and amino acid residues of human soleus titin indicated. (A2) Control. (A3–A6) Labeled with N2A. (B1) Labeled with T12 and Ti102. (B2–B5) Labeled with 514.). Bar, 1.0 mm. (taken from Trombitás et al., 1998)

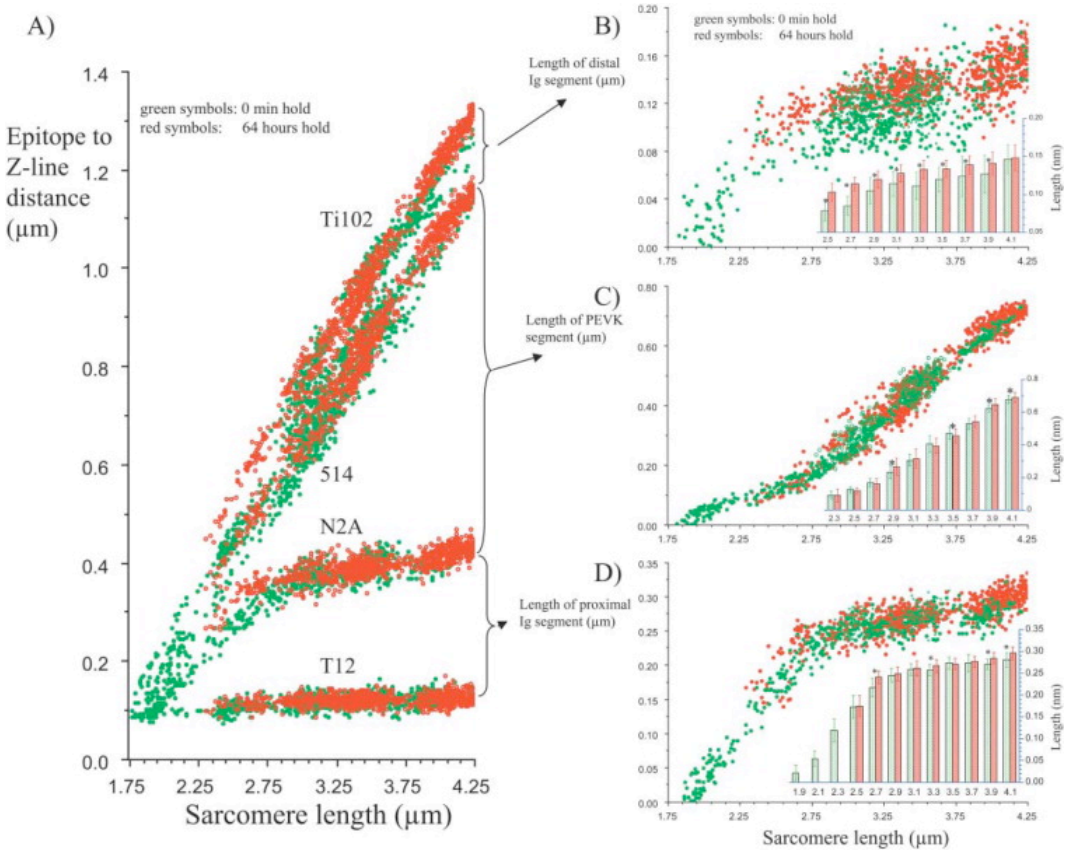


Figure 3.2: Sarcomere length dependence of epitope to Z-line distance
(A) Mid-epitope to mid-Z-line distance of T12, N2A, 514, and Ti102 epitopes in fibers stretched and held for 0 min (green symbols) or 64 h (red symbols). (See Figure 3.1 for antibody labeling of sarcomeres). Length of proximal and distal tandem-Ig segment: distance from T12 to N2A and from 514 to Ti102, respectively. Length of PEVK: distance from N2A to 514 epitopes. (B–D) Lengths of tandem-Ig and PEVK segments versus sarcomere length (SL) of fibers stretched and held for 0 min or 64 h. Insets show mean and SD of results binned in 0.1 mm SL bins. Asterisks denote statistically significant differences ($P < 0.05$). (Taken from Trombitás et al., 2003).

The availability of the first titin single domain structures allowed for Molecular Dynamics (MD) studies to model the unfolding of titin Ig domains under applied external force. This helped to understand the behavior and appearance of force-extension curves, identifying a potential energy barrier, protecting the Ig-domains from unraveling. Multiple H-bonds connecting the A-B and A'-G strands together, i.e. glueing the start and end of the domains, have to be ruptured simultaneously before the core amino acids become exposed (see Figure 3.3B).

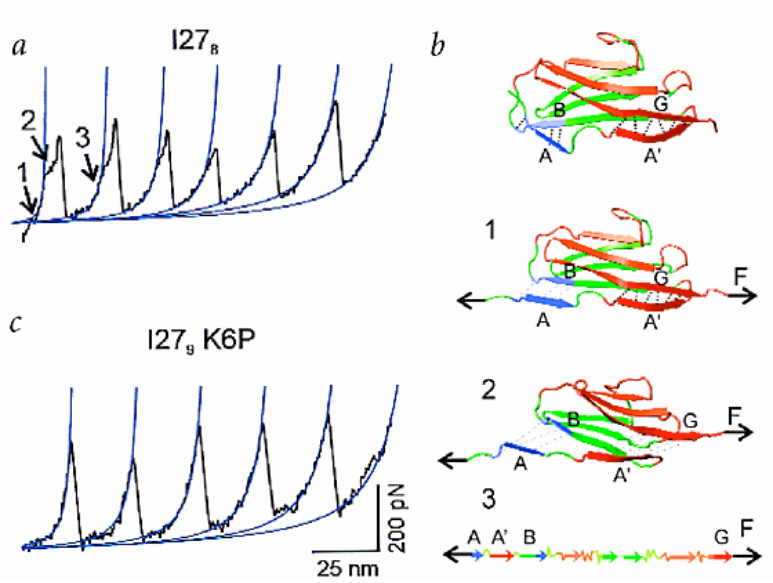


Figure 3.3: Force extension curves of engineered titin poly-Ig fragments
Single molecule force spectroscopy by AFM on a poly-protein consisting of several copies of a single domain of titin (I91, formerly I27). The protein is adsorbed to a gold-coated surface and picked up by the cantilever tip. As the gold substrate is withdrawn the protein is stretched. As the molecule resists extension the cantilever is deflected. At pulling forces higher than the mechanical stability of the protein, tension is relieved by the sequential unraveling of domains, which leads to a temporary release of tension. In modular proteins, this chronological unfolding of one domain after the other leads to a force extension curve with a characteristic saw-tooth pattern (A, C). This pattern can be analyzed quantitatively using the worm-like chain (WLC) and freely jointed chain (FJC) models developed for polymer elasticity (see section 3.4.1). The amplitude of force at which a domain unfolds reflects its mechanical stability and the spacing between the peaks is an indicator of the contour length gain as the domains unfold. B) shows steered molecular dynamics simulations of the unfolding of domain I91 (formerly I27). From the top the molecule is shown without any force applied and increasing force until the lowest diagram shows the completely unfolded domain. (Taken from Fisher et al., 2000)

While these titin unfolding studies provided insight into the mechanical stability of titin's Ig domains and proved essential for the development of the AFM/ OT techniques it has been shown that titin is able to accommodate all physiological working lengths of the sarcomere without the need to unfold a single domain (Bang et al., 2001), as the longest sarcomere length studied so far, 4.2μm, can be accommodated by the native structure of the Ig domains and their linkers (Improta et al., 1996; Trombitás et al., 1998). Furthermore, the physiological forces acting on titin *in vivo* have been estimated to be in the range of 0-10pN/molecule (Watanabe et al., 2002b), which is below the forces needed for titin unfolding in single molecule experiments (see above).

3.1.2 Characterization of I-band spring components

3.1.2.1 Titin I-band Ig domains

The first single domain structures of titin (see Table 3.1), M5 in the M-line (Pfuhl and Pastore, 1995), I91 (originally I27; Improta et al., 1996) and I1 (Mayans et al., 2001) showed that titin Ig domains belong to the intermediate or I-set of Ig that was first identified in telokin (Harpaz and Chothia, 1994). The I-set combines features from the constant (C-set) and variable (V-set) set (the classical Ig types identified in antibodies, as defined by (Williams, 1987; Williams and Barclay, 1988)). The I-set can be regarded as a truncated version of the V-set missing the C'' strand – the two antiparallel β-sheets are formed by β-strands ABED (purple) and A'GFCC' (cyan) respectively, as seen in Figure 3.4. The first strand, A, splits into two halves separated by a classic β-bulge, where the first half-strand belongs to one sheet and the second switches to the second sheet, aligning itself in a parallel fashion to the extended G-strand. This switching as well as the number and size of the β-strands of these Ig-domains is characteristic for the I-set and any insertions or deletions of amino acids are clustered in the loop regions.

Figure 3.4: Crystal Structure of domain I1 of titin
Ribbon representation where strands have been labeled according to the convention by (Harpaz and Chothia, 1994). The strands colored in purple form one sheet and the strands in cyan the second. (Figure taken from Mayans et al., 2001)

Table 3.1: Atomic structures of titin and twitchin components

Domain	Type	Location	Source	PDB code	Method	Reference
Titin:						
M5	Ig	M-line	human	1NCT/1NCU/1TNM/1TNN	NMR	(Pfuhl et al., 1995)
I91 (I27)	Ig	Constitutive, distal I-band	human	1TIT/1TIU/1WAA	NMR/X-ray	(Improta et al., 1996)
A71	Fn3	A-band (11 domain superrepeat)	human	1BPV	NMR	(Muhle-Goll et al., 1998)
TK	Kinase	M-line	human	1TKI	X-ray	(Mayans et al., 1998)
I1	Ig	Constitutive, proximal I-band	human	1G1C	X-ray	(Mayans et al., 2001)
Z7:a-actinin	Z-repeat	Z-disc	human	1H8B	NMR	(Atkinson et al., 2001)
M1	Ig	M-line	human	2BK8	X-ray	(Müller et al., unpublished)
Z1Z2	Ig-Ig	Z-disc	human	2A38	X-ray	(Marino et al., 2006)
Z1Z2:telethonin	Ig-Ig	Z-disc	human	1YA5/2F8V	X-ray	(Zou et al., 2006)
A168-A170	Ig-Ig-Fn3	A-band	human	2NZI	X-ray	(Mrosek et al., 2007)
A168-A169	Ig-Ig	A-band	human	2J8H/2J8O/2ILL	X-ray	((Müller et al., 2007)
Twitchin:						
I18	Ig	A-band	C. elegans	1WIT/1WIU	NMR	(Fong et al., 1996)
I26	Kinase - Ig	A-band	C. elegans	1KOA	X-ray.	(Kobe et al., 1996)

For a discussion of the first multidomain structures of titin, which were only reported after initiation of this study, see section 3.4.3. They have been included here for completeness sake. (Table adapted from Marino et al., 2005)

Indeed, Ig from different parts of titin vary slightly in their structure and show distinct molecular properties and stabilities (Fraternali and Pastore, 1999; Pfuhl and Pastore, 1995; Politou et al., 1996; Politou et al., 1994; Watanabe et al., 2002a). This has been speculated to reflect the varied functions of the different regions within titin.

A study combining the structure of I91 with homology modeling of I92, NMR data on I91-I92 dynamics and SAXS data on both I91-I92 and I91-I94 came to the conclusion that titin consists of independently folding and only weakly interacting Ig domains that preferentially adopt extended conformations and only have restricted motion around the linkers in solution (Improta et al., 1998), which was later confirmed by a study on related constructs from the I91-I96 region (Scott et al., 2002). This led to the (subsequent) modeling of titin domains arranged like beads on a string with arbitrary orientations but limited conformational flexibility around the linker (Improta et al., 1998; Improta et al., 1996).

Careful sequence alignments and analysis together with the first atomic resolution structures revealed that in the titin I-band Ig-domains display high levels of conservation (Improta et al., 1998; Pfuhl and Pastore, 1995; Witt et al., 1998) and can be grouped into super-repeats, derived from evolutionary duplication of gene stretches (Gautel, 1996; Kenny et al., 1999). Furthermore, two distinct subclasses, termed “N-conserved” and “N-variable”, can be distinguished based on features clustered around domain interfaces and linkers (Marino et al., 2005; Witt et al., 1998). The classification is based on the following criteria:

i) the presence of a **PPxf** motif at the domain N-terminus in strand A (where P is proline, x any residue and f a hydrophobic group – usually a phenylalanine. Upper and lower case reflects the rate of conservation of these groups)

ii) the presence of a lengthened FG β-hairpin housing a conserved **NxxG** motif at its turn (where N is asparagine, x any residue and G glycine)

iii) conservation in length and sequence of the BC turn, regarding a proline-rich motif, usually PPh or PxP (where h is any hydrophobic residue)

iv) the conservation of a negatively charged group, often a glutamate, at the domain C-terminus after strand G

Representation of the pattern of conservation is depicted in Figure 3.5.

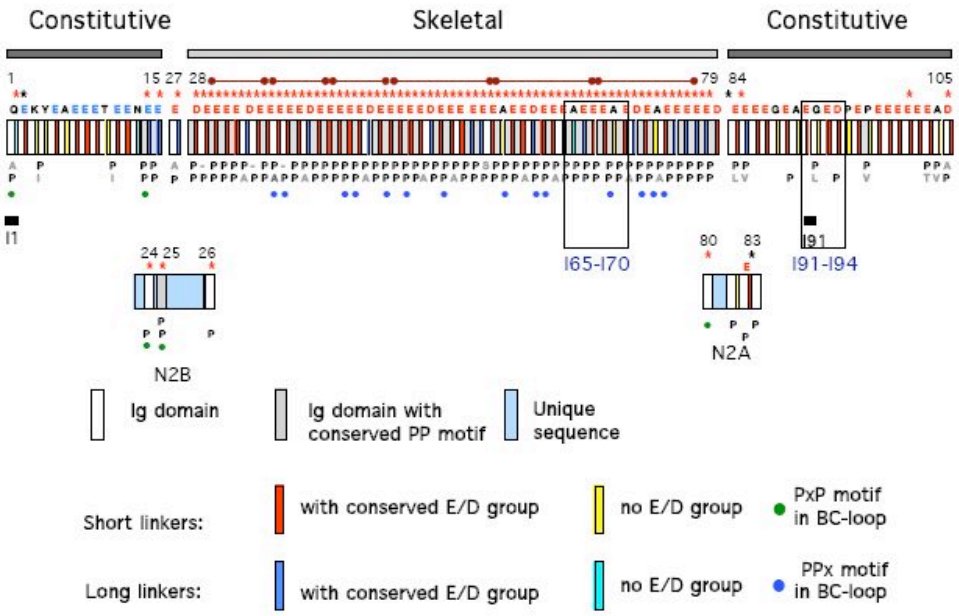


Figure 3.5: Conservation in the I-band region of titin
Ig domains and linker segments are represented as filled boxes (PEVK region excluded for simplicity). The presence of an extended FG β-hairpin is given as stars, with red indicating a conserved NxxG motif and black any other sequence. The conserved E/D group is displayed on top of the corresponding linker boxes, where blue refers to the C-terminus of strand G and red indicates one additional position (*i*+1). Linkers are classified as short if containing less than three residues and long otherwise. In the differentially expressed tandem a two-residue linker corresponds to a zero residue inter-domain separation. The conservation of a PP motif in strand A is displayed, as well as proline-rich sequences at the BC loop. Black bars mark modules for which atomic structures have been elucidated. Segments for which domain arrangement has been analyzed by SAXS are boxed. Super-repeats of 6-10 Ig domains within the central tandem are marked with a capped-bar. (Taken from Marino et al., 2005)

The “N-conserved” group includes all Ig of the differentially expressed tandem Ig, as well as the flanking domains within constitutive regions (I1, I14-I15, I84-I85, I105) and I25 from the N2B element. This group is characterized by the presence of a long FG β-hairpin, the PP motif at the N-terminus within strand A as well as a longer BC-loop, often containing either PPh or PxP motif, which all cluster to the N-terminus of the domain (Figure 3.6).

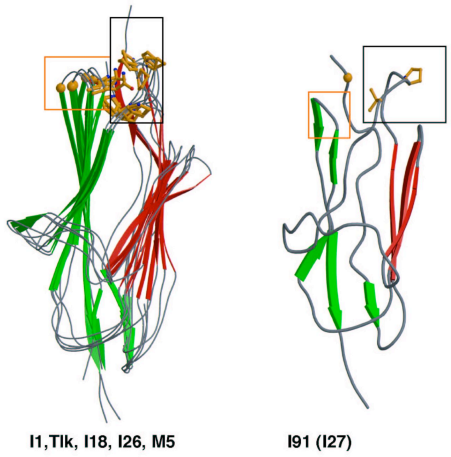


Figure 3.6: Ribbon representation of Ig grouped according to their N-terminal loop cluster
The BC and FG loop are boxed in black and orange A) shows the N-terminally extended group while B) shows a representative of the second group. The structures correspond to 1GIC (I1; Mayans et al., 2001), 1NCT (M5; Pfuhl and Pastore, 1995), 1TIT (I91; Improta et al., 1996), 1WIT (I18; Fong et al., 1996), 1KOA (I26; Kobe et al., 1996) and 1FHG (Telokin; Holden et al., 1992). (Taken from Marino et al., 2005)

The second group of I-band domains, which comprises the large majority of constitutive domains, shows more variation in its features. However, it can be defined by shortened BC and FG elements as well as by fewer prolines involved in interactions at the N-terminal loop cluster. This group has therefore been named “N-variable” (Figure 3.6).

Inter-domain linker sequences also exhibit a regular arrangement of (recurrent) linker features, wherein groups of domains are connected by short linkers (up to three residues) that contain conserved glutamate residues and are interspersed by longer linkers housing aspartate residues (Figure 3.5). The repeat structure of the pattern agrees with the previously reported existence of 6-10 Ig super-repeats in this tandem (Gautel, 1996; Kenny et al., 1999) and might be speculated to result in a local higher-order structural arrangement of the repeats in the tandem, which would have consequences for the mechanical dynamics of the array.

In order to investigate potential functional implications for the high conservation of the glutamate residues in linker sequences, Marino and colleagues performed SAXS measurements on a triple mutated I65-I70$^{3E/A}$ construct (Marino et al., 2005). They found the bending rigidity not significantly altered, which was later confirmed by the x-ray structure of I67-I69^{E93A}, which was part of this study (von Castelmur et al., 2008). Rotary shadowing EM on both the native and mutated constructs gave similar results for their end-to-end extension, <R>= 24.6 +/- 3.3nm and 22.7 +/- 3.4nm respectively. In the calculations, the 6Ig fragments were treated as beads on a string and modeled using a six-module freely jointed chain (FJC; the segment length of 5nm is large compared to the contour length of ~29nm) with a restricted cone of motion. This gave a resulting opening angle of 143°, which agrees with the results of the study on I91-I94 and I91-I96 mentioned above.

3.1.2.2 The PEVK spring element

The second spring element, the unique PEVK segment, contains two main sequence motifs, namely the PPAK motif and polyE repeat (Greaser, 2001; Gutierrez-Cruz et al., 2001). The PPAK motifs are 28-33 residues long and named after the first four amino acids of the consensus sequence (**PPAK**VPEVPK KPVPE[E/K]KVPV PVPKKPEA; Greaser, 2001). They are suggested to be the individual structural subunits that are arranged in super-repeats of 2-12 motifs, interspersed with acidic glutamate-rich polyE repeats of variable length (Greaser, 2001).

In early immuno EM and OT experiments the PEVK segment was assumed to most probably form a quasi-unfolded, random-coil chain (Trombitás et al., 1998) and modeled as a permanently unstructured polypeptide applying the WLC model. Subsequent CD and NMR experiments on different PEVK constructs revealed the presence of secondary structure elements. Namely this region contains polyproline type II helices and β-turns — speculated to potentially be linked in series to form S-shapes or even β-turn helices — in addition to unordered coil. The three forms appear to interconvert without long-range cooperativity *in vitro* without the need for proline trans/cis isomerization (Gutierrez-Cruz et al., 2001; Ma et al., 2001; Ma and Wang, 2003). Still, the PEVK clearly has no tertiary structure and in AFM and OT experiments displays a smooth force extension and relaxation curve lacking any obvious unfolding peaks. Both ionic strength and Ca^{2+} concentration have been

shown to influence the elastic properties, i.e. persistence length, but not conformation (assessed by circular dichroism) of human fetal TP1 PEVK in AFM experiments (Forbes et al., 2005). Increasing concentrations of trifluoroethanol on the other hand resulted in clear changes of secondary structure but not persistence length, leading Forbes and colleagues to postulate the PEVK region to be a flexible polyampholyte behaving like a gel-like enthalpic spring, regulated in its elasticity through varying sequence-specific charge interactions.

Furthermore, the PEVK has been reported to interact with F-actin, retarding thin filament sliding (Yamasaki et al., 2001), and interaction forces have been measured in OT experiments to be around 6-8pN before rupturing. These proline-rich sequences also contain a multitude of both canonical class I and class II as well as novel SH3 binding motifs, consisting of one or more overlapping copies of each of the canonical motifs. These SH3 binding motifs appear to be involved in the integration of stress signaling pathways by binding e.g. to the nebulin SH3 domain (Ma et al., 2006; Ma and Wang, 2002). So, overall, the PEVK segment can be seen as a mostly unstructured, medium-compliant spring element with interspersed protein binding sites, making it a potential stress/ stretch sensor in signaling pathways (Ma et al., 2006; Nagy et al., 2005).

3.1.2.3 The cardiac N2B segment

For completeness sake, the N2B segment, which is the cardiac specific third spring element, will also be mentioned here. Its elastic properties lie in its unique sequence insertion of 572 residues that behave as a non-modular spring, extending up to ~60nm in single myofibril experiments (Linke et al., 1999).

Thus, the titin I-band contains three spring elements of vastly different characteristics, linked in series in order to provide elasticity for the whole working range of muscle. Since domain unfolding is unlikely to drive titin elasticity at physiological mechanical loads, elucidating the structure and dynamics of the titin chain (at a structural level higher than the Ig module) is crucial to unravel its mechanistic principles. Here we set out to solve the crystal structure of I65-I70 as a representative 6 Ig fragment of the skeletal Ig tandem. The crystal structure of I65-I70 should allow the formulation of the first structure-based model of titin mechanics.

3.2 Crystal structure of titin I-band domains I65-I70

Personal contribution: Experimental design and execution (protein expression, purification, crystallization, heavy atom derivatization, data collection and processing, elucidation and analysis of I65-I70 structure, design and characterization of I67-I69 mutants, including structure solution of I67-I69^{E93A}, modeling of the skeletal I-band tandem)

A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain

Eleonore von Castelmur*, Marco Marino*, Dmitri I. Svergun[†‡], Laurent Kreplak[§], Zöhre Ucurum-Fotiadis*, Petr V. Konarev[†‡], Alexandre Urzhumtsev[¶‖]**, Dietmar Labeit[††], Siegfried Labeit[††], and Olga Mayans*[‡‡]

*Division of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland; [†]European Molecular Biology Laboratory, Hamburg Outstation, c/o Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, D-22603 Hamburg, Germany; [‡]Institute of Crystallography, Russian Academy of Sciences, Leninsky Prospect 59, Moscow 117333, Russia; [§]M. E. Müller Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland; [¶]University-Nancy, 54506 Vandoeuvre-les-Nancy, France; [‖]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique-Institut National de la Santé et de la Recherche Médicale-Université Louis Pasteur, 67404 Illkirch, France; **Institut de Biologie Moléculaire et Cellulaire, Université Louis Pasteur, 67084 Strasbourg, France; and [††]Institut für Anästhesiologie und Operative Intensivmedizin, Universitätsklinikum Mannheim, 68167 Mannheim, Germany

Edited by John Kuriyan, University of California, Berkeley, CA, and approved November 21, 2007 (received for review July 31, 2007)

Myofibril elasticity, critical to muscle function, is dictated by the intrasarcomeric filament titin, which acts as a molecular spring. To date, the molecular events underlying the mechanics of the folded titin chain remain largely unknown. We have elucidated the crystal structure of the 6-Ig fragment I65–I70 from the elastic I-band fraction of titin and validated its conformation in solution using small angle x-ray scattering. The long-range properties of the chain have been visualized by electron microscopy on a 19-Ig fragment and modeled for the full skeletal tandem. Results show that conserved Ig–Ig transition motifs generate high-order in the structure of the filament, where conformationally stiff segments interspersed with pliant hinges form a regular pattern of dynamic super-motifs leading to segmental flexibility in the chain. Pliant hinges support molecular shape rearrangements that dominate chain behavior at moderate stretch, whereas stiffer segments predictably oppose high stretch forces upon full chain extension. There, librational entropy can be expected to act as an energy barrier to prevent Ig unfolding while, instead, triggering the unraveling of flanking springs formed by proline, glutamate, valine, and lysine (PEVK) sequences. We propose a mechanistic model based on freely jointed rigid segments that rationalizes the response to stretch of titin Ig-tandems according to molecular features.

electron microscopy | poly-Ig tandem structure | small angle x-ray scattering | titin elasticity | x-ray crystallography

The striated muscle of vertebrate is characterized by a striking elasticity that allows it to store mechanical energy and stretch over twice its resting length without disrupting its structural integrity. At physiological amounts of stretch, most of the elastic response of the myofibril is generated by the intrasarcomeric titin filament (≈3.2 MDa, >1-μm length). This protein functions as a bidirectional spring that stretches and recoils during muscle function to return the myofibril to its resting length (1). The spring components of titin are located in its I-band fraction, which forms an elastic connection between the ends of the thick filaments and the Z-disk. Titin contains two main elastic components, a proline-rich PEVK (proline, glutamate, valine, and lysine) segment of up to 2,200 residues length and a poly-Ig array formed by up to 95 modules (2). Both segments straighten upon myofibril stretch developing a passive entropic tension in the sarcomere. Poly-Ig arrays extend at low force whereas PEVK-repeats unravel at higher load, with the combined action of both springs defining the mechanical stiffness of the sarcomere (1, 3). The importance of stretch–recoil control in muscle function is emphasized by the finely tuned composition of both poly-Ig and PEVK segments in titin, which through splicing undergo an extensive adaptation to the different physiological and pathological states of muscle.

The molecular basis of titin chain elasticity is currently unknown. The response to stretch of its I-band components, particularly poly-Ig arrays, has been extensively studied using nanotools and molecular simulations (refs. 1 and 4 and references within). These methodologies finger-print unfolding phenotypes determined by the secondary structure elements of the Ig fold (i.e., analyze events at the module level) but do not report on the behavior of the chain. Because domain unfolding is unlikely to drive titin elasticity at physiological loads (3, 5), elucidating the structure and dynamics of the titin chain (at a level higher than the Ig module) is crucial to establish its mechanistic principles. To date, atomic models of domain tandems from titin have been scarce and limited to the N-terminal Ig-doublet Z1Z2 (6, 7) part of the Z-disk, and A168–A170 (8) (subfragment in 9) located at the M-line near the C terminus of the filament. These tandems act as docking sites for sarcomeric binding partners and do not relate to I-band elasticity. In the current study, we report atomic models of poly-Ig tandems from the spring region of titin that can aid to elucidate the mechanistic principles of sarcomere elasticity.

Results

Atomic Structures of Poly-Ig Tandems from I-Band Titin. We have elucidated the crystal structure of a six-Ig fragment, I65–I70, from the skeletal I-band of soleus titin at 3.3-Å resolution [Fig. 1*A* and supporting information (SI) Movie 1]. X-ray data statistics and model parameters are given in Table 1. Its modules belong to the "N-conserved" I-type of Ig folds (10), share a pairwise sequence identity of ≈35% and a high structural similarity [average rmsd 1.2 ± 0.1 Å for main chain atoms, calculated by using SPDBV (11)]. They are serially connected by linkers of diverse sequence composition and 0- to 3-residue length (Table 2). I65–I70 adopts a semiextended arrangement (≈21-nm end-to-end distance corresponding to ≈28-nm contour length) in L-shaped conformation, where the four C-terminal Ig lie almost perfectly coaxial, forming a straight section, and the N-terminal I65–I66 domains are bent

Author contributions: E.v.C., M.M., D.I.S., D.L., S.L., and O.M. designed research; E.v.C., M.M., D.I.S., L.K., Z.U.-F., P.V.K., A.U., D.L., S.L., and O.M. performed research; D.L. and S.L. contributed new reagents/analytic tools; E.v.C., M.M., D.I.S., L.K., and O.M. analyzed data; and E.v.C., M.M., D.I.S., L.K., S.L., and O.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Model coordinates and experimental x-ray amplitudes for I65–I70, I67–I69, and I67–I69^{E93A} have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3B43, 2RIK, and 2RJM, respectively).

[‡‡]To whom correspondence should be addressed. E-mail: Olga.Mayans@liv.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0707163105/DC1.

© 2008 by The National Academy of Sciences of the USA

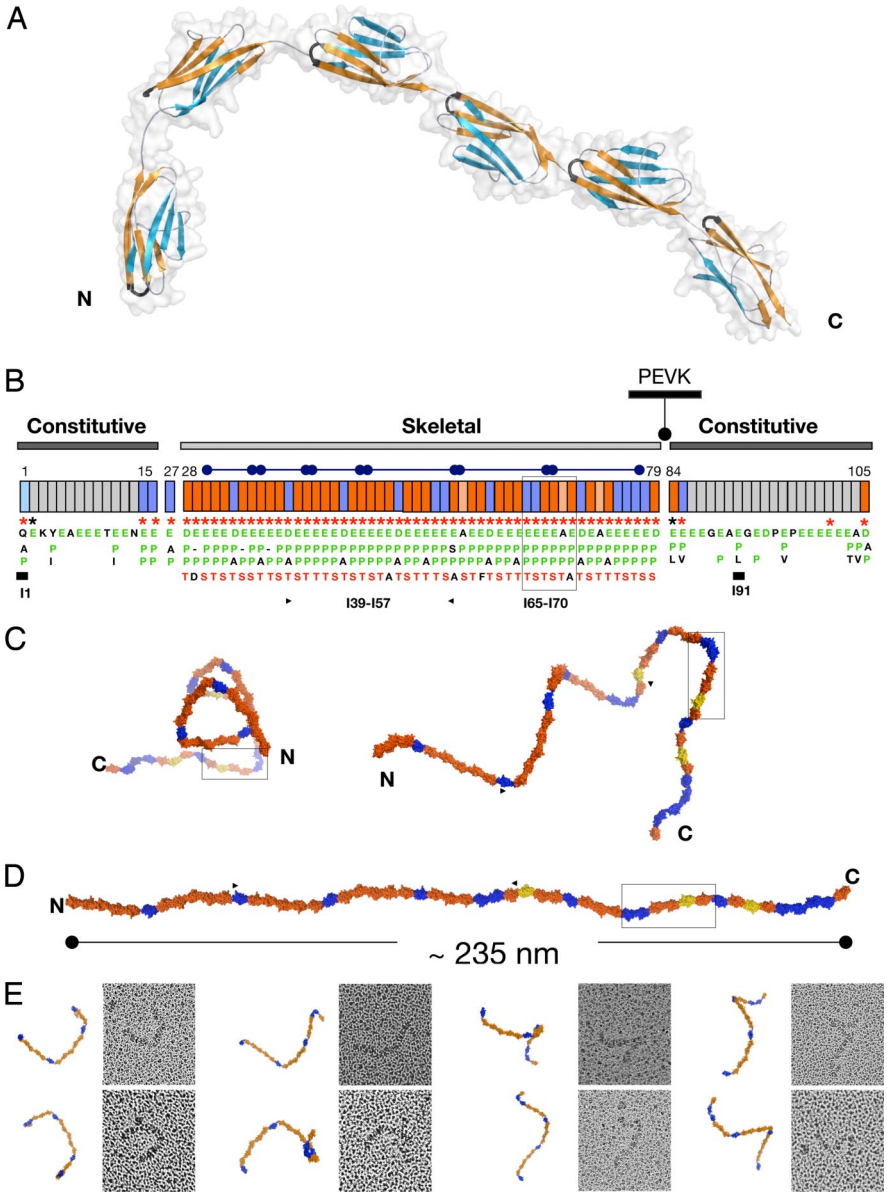


Fig. 1. Structural order in the poly-Ig from I-band titin. (*A*) Crystal structure of I65–I70. β-sheets are color coded to emphasize domain torsions. The FG β-hairpin, which claps against the Ig–Ig transition motif EPP, is colored black. (*B*) Modular composition of the I-band of soleus titin from rabbit (N2A and N2B elements are omitted). Ig domains are represented as boxes, where orange indicates Ig tightly connected and blue represents Ig containing a C-terminal three-residue linker. Annotations refer to conserved features at the Ig–Ig interfaces, where (*i*) an FG β-hairpin containing an NxxG sequence is marked by red asterisks, (*ii*) interdomain EPP motifs in green are listed vertically under each domain (Ig exhibiting a natural E-to-A mutation in this motif are colored salmon), and (*iii*) the conserved S/T residue in the BC loop is shown in red. These features are characteristic of the skeletal but not of the constitutive Ig tandems. Super-repeats of 6 or 10 Ig are indicated by capped bars. Domains with previously known structure are marked with a thick bar. (*C*) Frontal (*Left*) and lateral (*Right*) views of a predicted model of the complete skeletal Ig-tandem in one of its putative slack conformations in solution as calculated from linker arrangements in I65–I70. (*D*) Model in extended conformation (*C* and *D* are color coded as in *B*). (*E*) EM images (70 × 70 nm^2) of glycerol-sprayed/rotary shadowed I39–I57 accompanied by its corresponding model. The model (fragment indicated by arrows in *C*) has been oriented to match the micrographs, but no other manipulation has been applied (the 3D conformation of the I39–I57 model can be visualized in SI Movie 2).

away from the molecular axis, resulting in a frontal curvature of 114° (angle defined by the centers-of-mass of I65-I66-I67). Whereas domains in the linear fraction are joined tightly, the N-terminal Ig causing chain bending house longer three-residue linkers (Table 2). The long-range conformational definition of I65–I70 is unexpected because its contour length (≈28 nm) is about three times the persistence length estimated for this region of titin (≈9 nm) (12).

The Ig domains in I65–I70 are finely ordered, showing regular intermodular spacings, hinge openings, and torsional angles (Table 2). The torsional pattern is particularly striking. The C-terminal modules I67–I70, joined by tight linkers, display relative torsions of ≈180° that result in an up-and-down domain alternation along the chain and give a straight appearance to this segment. In contrast, domains I65 and I66, connected by longer linkers to their successor Ig, consistently exhibit an ≈90° torsion that deviates the filament path from linearity. This regularity in modular parameters indicates the existence of an architectural order in the fine structure of the titin filament.

Conservation of Intermodular Transition Motifs. The I-band of skeletal-muscle titin (X90569) contains 52 specific Ig connected by either a tight linker or a three-residue hydrophilic sequence (Fig. 1*B*). The model of I65–I70 includes representatives of both types of domain interfaces, revealing the molecular principles of Ig arraying in the complete I-band tandem. The structure of I65–I70, however, was affected by partial crystallographic disorder in its I68–I69 subsection; thus, to analyze this region reliably, we elucidated independently the crystal structure of the three-Ig subfragment I67–I69 at 1.6-Å resolution (Table 1). Even though crystallization media and lattice parameters were unrelated, the conformation of I67–I69 agreed excellently with that of the equivalent fraction of I65–I70 (Table 2), confirming the well defined architecture of the tandem. In this array, tight modular transitions are mediated by a highly conserved EPP sequence motif N-terminal to each Ig domain (Figs. 1*B* and 2*A*). This motif packs against the groove formed by β-hairpin FG and the BC loop, which clap onto it via conserved N and S/T residues, respectively (Figs. 1*A* and 2*D*). In these tight connections, the regular torsional alternation of domains appears due to a combination of favorable steric orientation and interactions of the last residue of Ig$_i$, often a lysine, with the moderately variable positions of the NxxG motif of β-hairpin FG in Ig$_{i+1}$. The latter

Table 1. Native x-ray data and model refinement statistics

	I65–I70		I67–I69	
Spacegroup	$P6_522$		C2	
Unit cell dimensions	$a = b = 141.43$ Å, $c = 166.01$ Å $\alpha = \beta = 90°$, $\gamma = 120°$		$a = 86.68$ Å, $b = 86.29$ Å, $c = 44.70$ Å $\alpha = 90°$, $\beta = 104.97°$, $\gamma = 90$	
Solvent content, %	67		52	
X-ray data				
X-ray source	ESRF ID-29–1		ESRF ID-29–1	
Detector	ADSC Quantum Q210		ADSC Quantum Q210	
Wavelength, Å	0.9792		0.9792	
Resolution, Å	17–3.3	(3.35–3.30)	19–1.6	(1.62–1.60)
Unique reflections	14,953	(639)	41,644	(1,312)
R_{sym} (I)	8.0	(47.8)	5.7	(39.5)
Multiplicity	6.84	(6.83)	4.57	(3.75)
Completeness, %	98.6	(97.3)	99.4	(87.9)
I/σ (I)	19.6	(4.3)	15.25	(3.53)
Refinement				
No. of reflections in working/test set	14,953/757		41,644/1,042	
No. of protein atoms/solvent molecules	4,412/0		2,214/479	
R factor/R_{free}, %	22.0/26.7		17.1/21.6	
rmsd bond length, Å/bond angle, °	0.006/0.94		0.006/1.51	

Values in parentheses correspond to the highest resolution shell.

commonly contributes a small hydrophobic residue as well as an acidic group that leads to a salt bridge formation across domains (Fig. 2*D*). Tight Ig connections effectively result in a zero-length linker that can be predicted to sterically oppose the sharp bending and kinking of the titin chain.

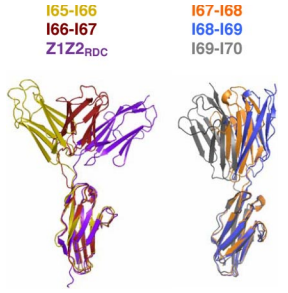
Longer linker sequences, as those in I65-I66-I67, can be regarded as structural variants of the tight connections, where a three-residue insertion of variable sequence has occurred at E◊PP. The interactions of the PP motif with the FG and BC loops at Ig_{i+1} are maintained, but the inserted residues as well as the conserved E group are now free from interactions (Fig. 2*C*). No direct Ig–Ig contacts are observed. Thus, long linkers could be expected to allow modular motions of large amplitude, probably acting as effective hinges in the chain. The Ig-doublet Z1Z2 from titin, which has been characterized recently through experiment (6) and simulation (13), contains a three-residue linker related to those of I65-I66-I67 (Table 2). At equilibrium in solution, the conformation of Z1Z2 is not random but it adopts a preferential, metastable semiextended state that resembles that of I65–I66 and I66–I67 in this study (Table 2). Nevertheless, its domain arrangement was shown to be adaptable, ranging from fully extended to compact "V"-shapes. It is conceivable that the conformational dynamics of long-linkers from I-band titin might approximate that of Z1Z2, thus introducing conformational freedom in the filament and permitting the pronounced bending of the chain at those points.

The conservation of EPP residues as Ig–Ig transition motifs is striking (Fig. 1*B*). To investigate the role of these residues in tandem conformation, the mutated variants I67–I69^{E93A} and I67–I69$^{P94A/P95I}$ were studied using x-ray crystallography and EM, respectively (*SI Text*). This analysis revealed that, whereas the conserved glutamate is dispensable and does not influence significantly interdomain arrangement, the proline doublet is central to tandem architecture. The fact that variant I67–I69$^{P94A/P95I}$ had its solubility severely compromised and resulted in fibril formation (SI Fig. 5) indicated that the N-terminal PP motif acts as a gatekeeper of the Ig fold. It probably ensures the structural integrity of the tandem through stacking interactions that prevent fraying of β-strand A at domain interfaces.

Solution Structure of I65–I70 and Its Subfragments I67–I69 and I66–I69. Given the potential flexibility of filamentous poly-Ig arrays and to ensure that crystalline conformations were not fortuitous, we investigated the structure of I65–I70 and its three- and four-Ig subfragments I67–I69 and I66–I69 in solution using small angle x-ray scattering (SAXS). Experimental scattering patterns are shown in Fig. 3, and calculated parameters are given in Table 3. A comparison of R_g and D_{max} values with those calculated from theoretical linear models indicated that I67–I69 and I66–I69 are nearly linear in solution, whereas I65–I70 must be somewhat bent. Further, SAXS models of I65–I70 that were independently constructed *ab initio* using rigid-body refinement of six individual Ig by simultaneous fitting to scattering data from I65–I70, I67–I69, and I66–I69, were in remarkably good agreement with the crystal structure of I65–I70 (Table 3 *Inset*). Calculations on I67–I69 and I66–I69 yielded nearly linear models. Accordingly, the scattering patterns computed from the crystal structures agreed well with the experimental data (Fig. 3 and Table 3). These results suggest that the crystalline conformation of I65–I70 (and its subsections) is a good representative of this tandem in solution.

Table 2. Domain arrangements in the crystal structures of I65–I70 and I67–I69

Tandem	Linker*	Opening angle†, °	Torsion‡, °	Distance§, Å
I65–I66$_{6IG}$¶	**IK**-E**R**K-L**PP**	136	+88	47.6
I66–I67$_{6IG}$¶	**LS**-E**H**E-V**PP**	133	+88	49.8
I67–I68 $_{6IG/3IG}$¶	**VQ**-**EPP**	173/155	−175/−160	44.6/44.26
I68–I69 $_{6IG/3IG}$¶	**VK**-**EPP**	155/165	+160/+158	42.55/42.76
I69–I70$_{6IG}$¶	**LK**-**APP**	136	−122	43.69
Z1Z2$_{RDC/SAXS}$‖	**VK**-AET-**APP**	136	+85	44.5



*Linker is defined as those residues whose main chain is free from interactions. Residues within Ig domains are in bold. Domain boundaries are indicated by a hyphen. The conserved Ig-Ig transition pattern of skeletal I-band domains is boxed. The rabbit (but not human) sequence of I69–I70 hosts a natural E-to-A mutation at this point.
†Angle definition as in ref. 6.
‡Angle definition as in ref. 6.
§Distance between the centers of mass of individual domains.
¶Subscripts 6IG and 3IG refer to the crystal structure of I65–I70 and I67–I69, respectively.
‖Domain arrangement of Z1Z2 in solution as calculated from SAXS data and NMR residual dipolar couplings (6). It should be noted that Ig-doublets containing long linkers exhibit similar opening and torsion angles but differ in their swing angle.

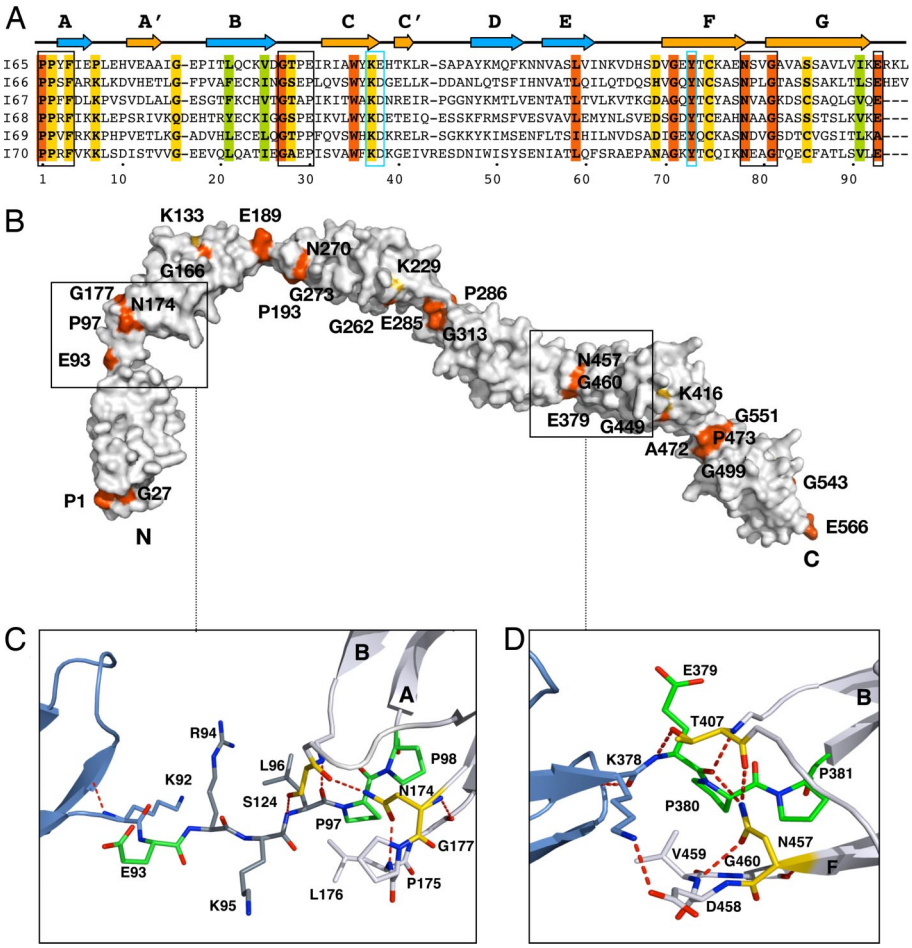


Fig. 2. Molecular features of I65–I70. (*A*) Structure-based sequence alignment where orange and yellow represent >90% and >70% sequence identity, respectively, across all Ig of the skeletal tandem. Green indicates conserved hydrophobic positions. The EPP motif, NxxG sequence in β-hairpin FG, and the BC loop are boxed in black. To ease comparison, the E group in tight linkers is given as the last residue of the preceding domain. A conserved set of residues (KD at the CC′ region and Y at β-strand F) responsible for the conformation of the CC′D loop characteristic of this Ig type is boxed in blue. (*B*) Molecular surface of I65–I70 colored according to sequence conservation as in *A*. (*C*) I65–I66 long linker interface. The three inserted residues are in dark gray. The conserved E is now an integral part of the linker, whereas L, the last of the inserted residues, has replaced it at the N terminus of the following Ig. (*D*) I68–I69 interface representative of tight connections. The transition motif EPP, an integral part of the N terminus of I69, is in green. The conserved N residue from β-hairpin FG and the T group from the BC loop are in yellow; their interactions are conserved in all Ig constituents of I65–I70. Hydrogen bonds are shown as dashed lines (experimental electron density for both linkers is shown in SI Fig 6).

In addition, we estimated the modular flexibility of these fragments by analyzing their apparent volume as a hydrated particle, V_p. For globular proteins, V_p in Å^3 is ≈1.5–2 times the molecular mass (MM) in Da, but this ratio will noticeably exceed 2 if proteins adopt multiple conformations in solution (14). V_p values for all poly-Ig fragments in this study were 1.6–1.9 times their MM (Table 3), pointing to a restricted conformational variability in solution. This conclusion is in excellent agreement with a previous study of I65–I70 that, based on a distribution of end-to-end molecular distances measured in a sample population using EM, concluded that this tandem is a relatively stiff formation (10). Furthermore, the study estimated the average opening angle between any two consecutive Ig to be ≈143°, in close agreement with the average value of 146° in the current crystallographic model (Table 2). Taken together, SAXS and EM data suggest that the overall modular flexibility of I65–I70 is limited and that its conformation in solution closely resembles that of the crystal structure.

Model of the I-Band Tandem of Skeletal Titin. The skeletal I-band of titin is arranged into domain super-repeats consisting of 6 or 10 Ig (15, 16) (Fig. 1*B*). The structure of I65–I70 (and its subfragment I67–I69) provide atomic data for representative domain and linker types of this section of titin and allow estimating the structural dynamics of the complete skeletal Ig-array. A constructed model suggests the presence of a supra-order in titin that results from a succession of modules joined through tight connections, interrupted at regular intervals by interspersed long linkers that alter the smooth "linearity" of the chain and lead to the formation of dynamic segmental units in the filament (Fig. 1*C*). Modules tightly connected are likely to be conformationally restrained, whereas Ig pairs joined by longer linkers can be expected to allow long-range intermodular motions and, thus, the bending of the chain. This conclusion is supported by EM images of a recombinant 19-Ig fragment, I39–I57, comprising three I-band super-repeats (Fig. 1*E*). Although all linkers might have a certain flexibility, images suggest that the shorter connections allow only moderate deformations whereas long linkers permit acute bending. This points to the existence of an architectural and dynamic supra-pattern in the filament caused by a regularity of molecular features. Speculatively, the inferred filament pattern might correspond to the observation of local pseudohelical substructure in EM (17) and atomic force microscopy (AFM) (18) studies of full-length titin molecules.

Segmental Flexibility as Mechanism of Titin Chain Elasticity. These findings suggest that the elasticity of the skeletal I-band tandem is mediated by a regular pattern of domain pairs with two different bending stiffnesses, namely pairs that resist bending and have a tendency to retain an extended conformation and pairs that permit complex interdomain motions of large amplitude as observed in the doublet Z1Z2 (6). The stretch and recoil of the titin chain could then be expected to first induce conformational changes in the long permissive linkers, likely to act as weak "flexor" points. The deformation of these few points in the filament would be sufficient to achieve a complete extension of the chain (Fig. 1*D*). This implies that the entropic properties of titin poly-Ig are not homogeneous but that a regular distribution of variably entropic points exists along the filament resulting in a segmental distribution. We term this concept of dynamic super-repeats, where relatively rigid segments are intercalated with pliant connectors, the "carpenter ruler" model of titin elasticity.

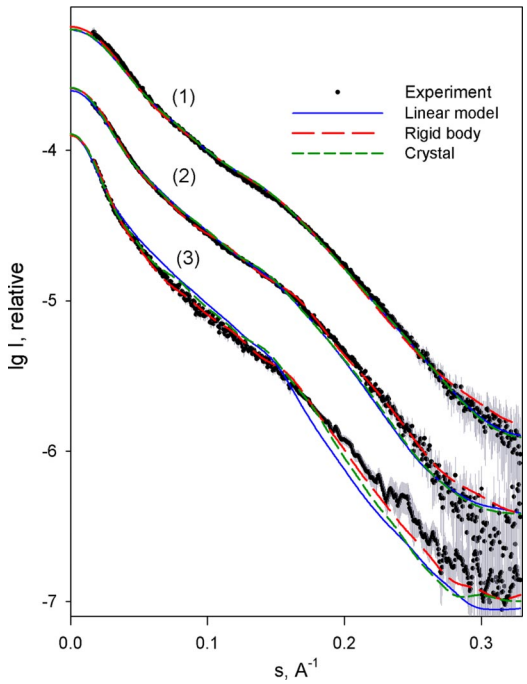


Fig. 3. Scattering patterns of I65–I70, I67–I69, and I66–I69. Experimental scattering is displayed as dots with error bars. Patterns computed from linear, rigid-body, and crystallographic models are blue, red, and green, respectively. For better visualization, patterns are displaced along the logarithmic intensity scale. Patterns 1, 2, and 3 correspond to I67–I69, I66–I69, and I65–I70, respectively.

Based on this concept, the elasticity of this poly-Ig tandem should be expressed in terms of a discrete formalism where the chain is represented by freely jointed rigid-segments (FJC) and not through continuous models such as the worm-like chain (WLC) form currently applied to this system (e.g., refs. 3 and 18–20) that assumes a smoothly flexible chain of random conformation and homogeneous composition. According to the segmental FJC model, the skeletal poly-Ig tandem of titin (comprising 52 Ig) can be described as containing 15 segments. On average, each segment consists of 3.5 Ig domains of 4.5-nm length each, resulting in an average segment length of ≈15.6 nm. For an FJC, the mean square end-to-end distance $\langle R^2\rangle_{FJC}$ is defined as $\langle R^2\rangle_{FJC} = n{\cdot}l^2$, where n is the number of bonds (i.e., segments) and l is the average bond length (21). For this tandem ($n = 15$, $l = 15.6$ nm), $\langle R^2\rangle_{FJC} = 3{,}650$ nm^2. This value allows now calculating the persistence length λ of a WLC with a contour length L of 52·4.5 = 234 nm. For a WLC in 3D equilibrium, $\langle R^2\rangle$, L and λ are related by $\langle R^2\rangle_{WLC} = 2\lambda L[1 - (\lambda/L)(1 - e^{-L/\lambda})]$ (21). The persistence length so estimated is 8 nm, which is in excellent agreement with experimental values for this section of titin (≈9 nm) (12). Thus, our model, which now incorporates primary molecular information, explains overall statistical parameters currently available for the titin chain. It also shows that a chain can have fairly long stiff segments and be highly flexible as long as large freedom of motion is possible at discrete points. This conclusion emphasizes the fact that great care must be taken when attempting to interpret statistical concepts of polymer dynamics in terms of structural features of a chain.

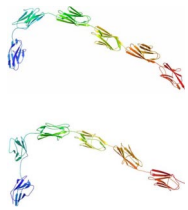
Mechanistically, it could be speculated that the dominant role of pliant linkers is to permit the molecular shape variations that take place in the filament at moderate stretch (this being the regime where the end-to-end chain distance is much smaller than the contour length of the folded tandem). Thus, they primarily contribute conformational entropy to the system. Compared with the commonly used WLC model, which overestimates molecular shape contribution by assuming a continuous chain bending where entropic force arises from a narrowing in the distribution of end-to-end distances upon extension, our model has fewer accessible conformations and favors a lower conformational entropic component. In contrast, the stiffer segments seem to permit only minor molecular shape rearrangements. However, their mechanical contribution might be essential in resisting high forces that develop at a late stretch stage when the chain is fully extended. Considering that these segments must undergo thermal fluctuations of a certain amplitude and frequency (libration), additional stretch upon full chain extension would lead to a damping of the amplitude of the libration [as proposed for elastin (22)]. Here, librational entropy would oppose the over-stretch of the poly-Ig chain, acting as an energy barrier that protects it from domain unfolding by triggering, instead, the unraveling of the more compliant PEVK sequences in neighboring springs. As a result, the interplay of poly-Ig and PEVK serial springs causes a nonlinear response to stretch that defines the mechanical properties of titin (3, 5).

Discussion

The molecular events underlying titin mechanics remain largely unknown. Whereas techniques such as AFM have amply reported on the unfolding of secondary structure elements in its Ig domains, these data do not enlighten us on the primary mechanism of titin elasticity *in vivo* that resides on the properties of its folded chain.

Table 3. Conformation of I65–I70 (and subfragments) in solution by SAXS analysis

Sample	I65–I70	I67–I69	I66–I69
$MM^{exp/calc}$, kDa	65 ± 5/63.8	34 ± 4/31.3	44 ± 5/41.6
R_g^{exp}, Å	64 ± 3	39 ± 1	49 ± 2
$R_g^{linear,RB,Xtal}$, Å	75.5/62.4/68.8	37.9/39.4/38.0	51.0/52.0/51.0
D_{max}^{exp}, Å	220 ± 10	130 ± 10	170 ± 10
$D_{max}^{linear,RB,Xtal}$, Å	260/210/218	135/130/138	175/160/170
V_p^{exp}, 10^3 Å^3	128 ± 8	58 ± 5	75 ± 7
$\chi^{linear,\ RB,\ Xtal}$	2.46/1.11/1.80	1.48/1.39/1.42	2.14/1.26/1.83



MM, R_g, D_{max}, and V_p denote molecular mass, radius of gyration, maximum dimension, and hydrated volume, respectively. Superscript "exp" refers to experimental values. The rest are calculated from models where "linear" indicates a poly-Ig model with a linear domain arrangement, "RB" corresponds to a model obtained by rigid body refinement against SAXS data, and "Xtal" refers to crystallographic models. I66–I69 and I67–I69 are fractions extracted from the crystal structure of I65–I70. χ is the discrepancy between the experimental pattern and that computed from the model. D_{max} values calculated from models include a 6-Å correction to account for molecular hydration. V_p, the apparent volume of a hydrated particle, reports on the modular flexibility of the samples. For rigid proteins, V_p is ≈1.5–2 times the molecular mass (MM) in Da, whereas this ratio notably exceeds 2 in samples with multiple conformational states in solution. (*Inset*) Crystal structure of I65–I70 (*Upper*) and SAXS-derived model obtained by rigid-body refinement (*Lower*).

Given that the structure of this filament is poorly understood, its mechanics have remained described in terms of a first-approximation statistical model of polymer entropy, the WLC model, which considers titin as a continuous chain of random conformation and homogeneous composition. Although concerns about this model were echoed at an early stage (23), no alternative mechanistic principle has been proposed so far. Based on atomic structures, SAXS, and EM data of poly-Ig components from the I-band of titin, we now propose a model of poly-Ig elasticity based on a discrete organization of the chain into finely structured super-motifs displaying a segmental dynamics. Structural data indicate that the entropic properties of these tandems are not homogeneous, but that variably entropic points exist along the filament. We identify the location of flexible points in the chain and model them within the context of the full skeletal I-band tandem. This mechanical model, here termed "carpenter-ruler," is based on freely jointed rigid segments of variable segment length. The model allows calculating the physical properties of this spring and offers great promise for the future study, rationalization, and modeling of the stretch-response phenotype of skeletal myofibrils.

Experimental Procedures

Cloning. Domains I65–I70 (amino acids 7946–8511), I67–I69 (amino acids 8137–8417), and I66–I69 (amino acids 8137–8511) from rabbit soleus titin, and I39–I57 (amino acids 5498–7287) from the human variant (X90569) were cloned independently into pET-M11 (EMBL collection), containing a His_6-tag and a TEV protease cleavage site before the target gene. The mutated variant I67–I69$^{P94A/P95I}$ was generated using the QuikChange method (Stratagene).

Protein Production. Expression was in *Escherichia coli* Rosetta (DE3) grown in LB medium supplemented with 34 μg/ml chloramphenicol and 25 μg/ml kanamycin up to an OD_{600} of 0.6 at 30°C. Induction used 1 mM isopropyl β-D-thiogalactoside (IPTG), and growth was continued at 20°C overnight. Bacterial cells were harvested by centrifugation, resuspended in 50 mM Tris·HCl (pH 7.5), 50 mM NaCl and sonicated in the presence of lysozyme, DNase, and a protease inhibitor mixture (Roche Diagnostics). The homogenate was applied to a HiTrap column (GE Healthcare) equilibrated in lysis buffer. Tag removal was by incubation with TEV protease overnight at 4°C in dialysis against 50 mM Tris·HCl (pH 7.5), 150 mM NaCl, and 1 mM DTT. Subtractive purification of the protease (His_6-tagged) was carried out on a further chelating step. The eluted protein was purified to homogeneity by ion exchange chromatography using a Resource Q column and gel filtration on a Superdex 200 HiLoad 16/60 column (both from GE Healthcare) in 50 mM Tris·HCl (pH 7.5). Samples were then concentrated and stored at 4°C. Tags were removed in samples used in crystallization. Samples used in SAXS and I39–I57 retained the tag and were not incubated with protease.

Crystal Structure Elucidation of I65–I70. Crystals grew as large hexagonal obelisks from 2 M NaCl, 100 mM sodium acetate (pH 4.9–5.1), and 3–5% (vol/vol) glycerol or NDSB-201 using a hanging drop setting at protein concentrations of 10–20 mg/ml. Cryoprotection for data collection used 20% (vol/vol) glycerol as supplement. Data processing used XDS/XSCALE (24) (statistics are listed in Table 1). Phasing by molecular replacement failed. Experimental phasing used MIRAS on CH_3HgCl, Ta_6Br_{12}, and Se-Met derivatized crystals (SI Table 5). Heavy atom sites were identified with CNS (25), and phases were calculated in SHARP (26). The positions of Se-Met and Hg scatterers (the latter predictably vicinal to Cys residues) were used as sequence markers for the recognition and positioning of individual Ig during model building. Manual model building was in O (27). Initial refinement used CNS (25) and applied bulk solvent correction, overall anisotropic B-factor, grouped isotropic B-factor, and energy minimization. At a later stage, once the high-resolution structure of I67–I69 became available, its individual domains were used to replace equivalent Ig in I65–I70 on a final series of refinement that used TLS and B-group refinement in PHENIX (28). Model parameters are given in Table 1.

Crystal Structure Elucidation of I67–I69. Crystals of I67–I69 grew as thin plates ($\approx$20 μm in thickness) from 0.2 M ammonium acetate, 0.1 M sodium acetate (pH 4.6), and 30% (wt/vol) PEG-4000 in hanging-drops. For data collection, crystals were shock-frozen at 100 K in solutions supplemented with 20% (vol/vol) glycerol. Data processing used XDS (24). Phasing was by molecular replacement in Phaser (29), using a model derived from telokin (PDB code 1FHG) as search model, accounting for 1/3 of the asymmetric unit and with 28% average sequence identity to the target domains. Phase improvement used ARP/wARP (30), and model building was in O (27). Model refinement used Refmac5 (31). Data and model statistics are given in Table 1.

Small Angle X-Ray Scattering. Synchrotron SAXS data from I65–I70, I67–I69, and I66–I69 were collected at X33 (DESY, Hamburg, Germany) using a linear gas detector and a MAR Image plate. All samples were measured at least at three concentrations (1.0–20.0 mg/ml) in the range of momentum transfer $0.01 < s < 0.35$ Å^{-1} ($s = 4\pi \sin\theta/\lambda$, where the wavelength λ is 1.5 Å and 2θ is the scattering angle). Radiation damage, monitored by repetitive 1- or 2-min exposures of the sample, was not detected. Data were processed using PRIMUS (32) and GNOM (33). Standard procedures were applied to the computation of the forward scattering $I(0)$, radius of gyration R_g, maximum dimension D_{max}, and the hydrated volume V_p. The molecular mass of the solutes was evaluated by comparison with reference solutions of BSA.

The scattering amplitudes of individual Ig domains and the intensities of I65–I70, I67–I69, and I66–I69 were calculated from atomic coordinates using CRYSOL (34). SASREF (35) was used to determine the conformation of I65–I70 in solution through rigid body refinement, which assembled individual Ig while maintaining chain connectivity and avoiding steric clashes. Starting from random arrangements, SASREF used simulated annealing to simultaneously fit the three experimental patterns by minimizing the overall discrepancy $\chi^2_{tot} = \chi^2_{I65–I70} + \chi^2_{I66–I69} + \chi^2_{I67–I69}$ (definition of χ^2 as in ref. 35).

Electron Microscopy. Aliquots (5 μl) of protein samples were adsorbed (1 min) onto glow-discharged carbon-coated copper grids and negatively stained with 2% (wt/vol) uranyl acetate. Alternatively, aliquots (20 μl) of protein samples at a concentration of 0.1 mg/ml were glycerol-sprayed and rotary shadowed according to standard procedures (36). Visualization used a Hitachi H-7000 transmission electron microscope (Hitachi, Tokyo, Japan).

ACKNOWLEDGMENTS. We thank the staff at ID29-1 (European Synchrotron Radiation Facility) and X06SA (Swiss Light Source) for excellent support during data collection and Pavel Afonine and Garib Murshudov for helpful discussions. Special thanks go to Prof. Henk Granzier for critical reading of this manuscript. Z.U.-F. was supported by SNF Grant 3100A0-112595. A.U. was supported by Pole Intelligence Logicielle CPER Lorraine. D.S. and P.K. were supported by the European Union Design Study SAXIER (011934). D.L and S.L. were supported by Deutsche Forschungsgemeinschaft Grants La668/9-1 and La1619/1-1.

1. Granzier HL, Labeit S (2004) *Circ Res* 94:284–295.
2. Bang ML, Centner T, Fornoff F, Geach AJ, Gotthardt M, McNabb M, Witt CC, Labeit D, Gregorio CC, Granzier H, Labeit S (2001) *Circ Res* 89:1065–1072.
3. Trombitás K, Greaser M, Labeit S, Jin JP, Kellermayer M, Helmes M, Granzier H (1998) *J Cell Biol* 140:853–859.
4. Kellermayer MS, Grama LJ (2002) *J Muscle Res Cell Motil* 23:499–511.
5. Linke WA, Ivemeyer M, Olivieri N, Kolmerer B, Rüegg JC, Labeit S (1996) *J Mol Biol* 261:62–71.
6. Marino M, Zou P, Svergun D, Garcia P, Edlich C, Simon B, Wilmanns M, Muhle-Goll C, Mayans O (2006) *Structure* 14:1437–1447.
7. Zou P, Pinotsis N, Lange S, Song YH, Popov A, Mavridis I, Mayans O, Gautel M, Wilmanns M (2006) *Nature* 439:229–233.
8. Mrosek M, Labeit D, Witt S, Heerklotz H, von Castelmur E, Labeit S, Mayans O (2007) *FASEB J* 7:1383–1392.
9. Müller S, Lange S, Gautel M, Wilmanns M (2007) *J Mol Biol* 371:469–480.
10. Marino M, Svergun DI, Kreplak L, Konarev PV, Maco B, Labeit D, Mayans O (2005) *J Muscle Res Cell Motil* 26:355–365.
11. Guex N, Peitsch MC (1997) *Electrophoresis* 18:2714–2723.
12. Di Cola E, Waigh T, Trinick J, Tskhovrebova L, Houmeida A, Pyckhout-Hintzen W, Dewhurst C (2005) *Biophys J* 88:4095–4106.
13. Lee EH, Hsin J, Mayans O, Schulten K (2007) *Biophys J* 93:1719–1735.
14. Heller WT (2005) *Acta Crystallogr D* 61:33–44.
15. Gautel M (1996) *Adv Biophys* 33:27–37.
16. Kenny PA, Liston EM, Higgins DG (1999) *Gene* 232:11–23.
17. Tskhovrebova L, Trinick J (2001) *J Mol Biol* 310:755–771.
18. Kellermayer MS, Bustamante C, Granzier HL (2003) *Biochim Biophys Acta* 1604:105–114.
19. Leake MC, Wilson D, Gautel M, Simmons RM (2004) *Biophys J* 87:1112–1135.
20. Rivetti C, Guthold M, Bustamante C (1996) *J Mol Biol* 264:919–932.
21. Flory PJ (1969) *Statistical Mechanics of Chain Molecules* (Interscience, New York).
22. Urry DW (1988) *J Protein Chem* 7:1–34.
23. Politou AS, Thomas DJ, Pastore A (1995) *Biophys J* 69:2601–2610.
24. Kabsch W (1993) *J Appl Crystallogr* 26:795–800.
25. Brunger AT *et al* (1998) *Acta Crystallogr D* 54:905–921.
26. de La Fortelle E, Bricogne G (1997) *Methods Enzymol* 276:472–494.
27. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) *Acta Crystallogr A* 47:110–119.
28. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) *Acta Crystallogr D* 58:1948–1954.
29. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) *Acta Crystallogr D* 61:458–464.
30. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS (2001) *Acta Crystallogr D* 57:1445–1450.
31. Murshudov GN, Vagin AA, Dodson EJ (1997) *Acta Crystallogr D* 53:240–255.
32. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) *J Appl Crystallogr* 36:1277–1282.
33. Svergun DI (1992) *J Appl Crystallogr* 25:495–503.
34. Svergun DI, Barberato C, Koch MHJ (1995) *J Appl Crystallogr* 28:768–773.
35. Petoukhov MV, Svergun DI (2005) *Biophys J* 89:1237–1250.
36. Fowler WE, Aebi U (1983) *J Ultrastruct Res* 83:319–334.

SUPPLEMENTARY MATERIAL

A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain

Eleonore von Castelmur[1], Marco Marino[1], Dmitri I. Svergun[2,3], Laurent Kreplak[3], Zöhre Ucurum-Fotiadis[1], Petr V. Konarev[2,3], Alexandre Urzhumtsev[4,5], Dietmar Labeit[6], Siegfried Labeit[6], Olga Mayans[1*]

[1]Division of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland; [2]European Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany & Institute of Crystallography; [3]Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia; [3]M.E. Müller Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland; [4]B.P. 239, Department of Physics, University-Nancy, 54506 Vandoeuvre-les-Nancy & IGBMC, 1 rue L.Fries, 67404 Illkirch; [5]IBMC, 15 rue R.Descartes, 67084 Strasbourg, France; [6]Institut für Anästhesiologie und Operative Intensivmedizin, Universitätsklinikum Mannheim, Mannheim 68167, Germany;

Section S1: Crystal structure elucidation of I67-I69^{E93A}

Methods

Sample production protocols were as for wild-type I67-I69 and crystals were obtained in related conditions (1.05M ammonium chloride, 0.1M sodium acetate pH 4.6, and 26% PEG-4000) at a protein concentration of 10 mg/ml. While the crystallization yield of the wild-type form was high, trials on I67-I69^{E93A} consistently resulted in heavy precipitation of the sample and only two small crystals (~70 μm longest axis) could be obtained. Structure elucidation was by difference Fourier using phases derived from the I67-I69 model followed by rigid-body optimization of the position of individual Ig. Further model building and refinement was as for the wild-type. Data statistics and model parameters are given in **Table S1**.

Table S1: X-ray data and model refinement statistics

	I67-I69^{E93A}	
Spacegroup	C2	
Unit cell dimensions	a=85.28 Å, b=86.48 Å, c=44.41 Å, β=104.94°	
Solvent content	51 %	
X-ray data		
X-ray source	SLS X06SA	
Detector	MAR-CCD 225	
Wavelength (Å)	1.008	
Resolution (Å)	18-2.0	(2.05-2.0)
Unique reflections	20955	(1372)
R_{sym} (I)	8.0	(47.2)
Multiplicity	4.74	(3.5)
Completeness (%)	99.2	(93.1)
I/σ (I)	13.8	(3.2)
Refinement:		
Number of reflections in working / test set	20955 (1071)	
Number of protein atoms / solvent molecules	2186/ 206	
R-factor / R-free (%)	17.8/ 22.8	
Rmsd bond length (Å) / bond angle (°)	0.016/1.60	
Ramachandran plot: favoured / disallowed (%)	91.9 / 0	

Results

The conservation of glutamate residues at Ig-Ig linker regions across the full I-band of titin is striking (**Fig 2A** in main text). The structures of I65-I70 and I67-I69 in this study reveal that the lateral carboxylic groups of these conserved residues are free from interactions. In order to examine whether they might play a role in interdomain conformation, the mutated variant I67-I69^{E93A} was investigated using X-ray crystallography. Its structure, elucidated at 2.0 Å resolution, revealed no significant differences when compared to that of the wild-type (**Fig S1**), indicating that interdomain glutamate residues are not a key determinant of tandem architecture. Thus, the physiological role of these exposed, highly conserved groups remains an enigma. Further studies will be required to establish whether they might play a mechanical role in force response in these poly-Ig arrays.

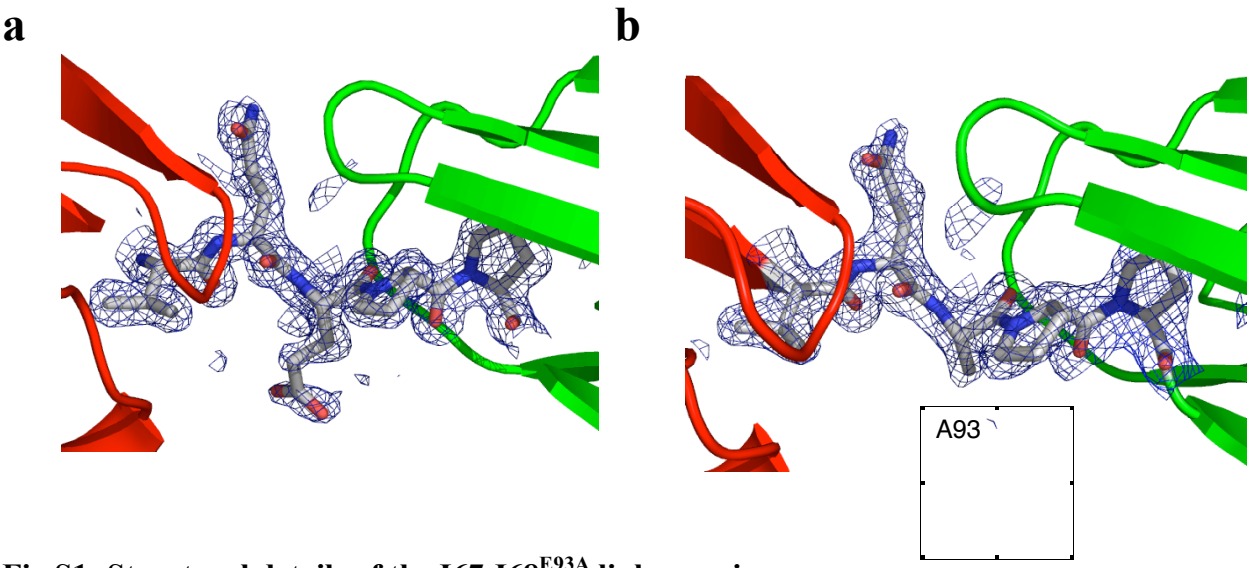


Fig S1: Structural details of the I67-I68^{E93A} linker region

a. Wild-type and **b.** E93A mutated variant. A $(2F_{obs}-F_{cal})\alpha_{calc}$ electron density map contoured at 1σ level is shown for linker groups.

SECTION S2: EM analysis of variant I67-I69$^{P94A/P95I}$

Methods

Electron microscopy

Overexpression and purification protocols were as for wild-type I67-I69. Aliquots (5 μl) of I67-I69$^{P94A/P95I}$ were adsorbed (1 minute) onto glow-discharged carbon-coated copper grids and negatively stained with 2% (w/v) uranyl acetate. Visualization used a Hitachi H-7000 transmission electron microscope (Hitachi, Ltd, Tokyo, Japan).

Results

The double variant I67-I69$^{P94A/P95I}$ with an altered N-terminal PP motif had its solubility severely compromised exhibiting irreversible aggregation and precipitation upon storage. EM analysis showed that the sample had formed extensive fibrils that co-existed with unordered aggregates (**Fig S2**). It should be emphasized that this severe phenotype arises from the alteration of a single PP motif in this 3Ig fragment, where the second domain interface in the array had not been mutated. This result suggests that the conserved PP motif is key to the structural integrity of the Ig domains in this tandem. A previous study (1) defended that the capability of tandem Ig of titin to avoid aggregation and misfolding during synthesis resides on a diversification of sequences across serial modules. However, our data show now that the N-terminal PP motif acts as a gatekeeper of the domain fold and ensures the structural integrity of the array.

The crystal structures of I65-I70 and I67-I69 show that both proline residues are in trans conformation. They are located at the N-terminus of each Ig domain, prior to β-strand A, and pack into the groove formed by the BC and FG loops via conserved interactions. Their structural roles appear to be varied, namely i) fixing of the relative orientation of an Ig module respect to the previous one in the chain; ii) limiting of the conformational space of the preceding residue through sterical hindrance derived from the properties of their imino group, possibly contributing to provide certain indirect linker stiffness; and iii) securing the packing of the initial β-strand A against the core of the fold, probably preventing its fraying and preserving the Ig framework. A related observation was made by Steward and co-

workers (2) when altering conserved N-terminal prolines in FnIII doublets by site-directed mutagenesis. These authors found that the replacement of those residues with alanines increased the tendency of the samples to aggregate. All data taken together, it could be concluded that proline residues at domain interfaces in multimodular proteins play an essential structural role.

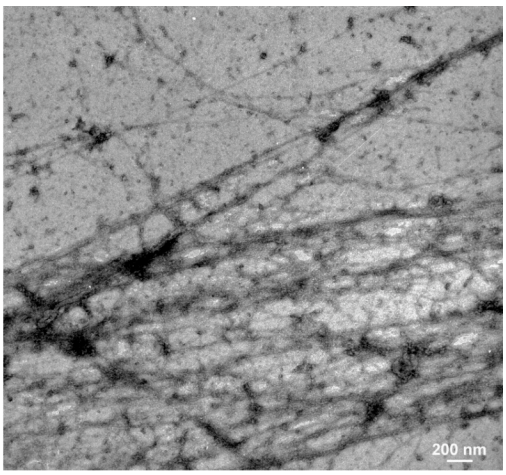


Fig S2: Electron microscopy image of I67-I69$^{P94A/P95I}$ obtained by negative staining

Fibrils are 8-10 nm wide in cross-section in non-bundled forms.

SECTION S3: Electron density map of long and short linker representatives

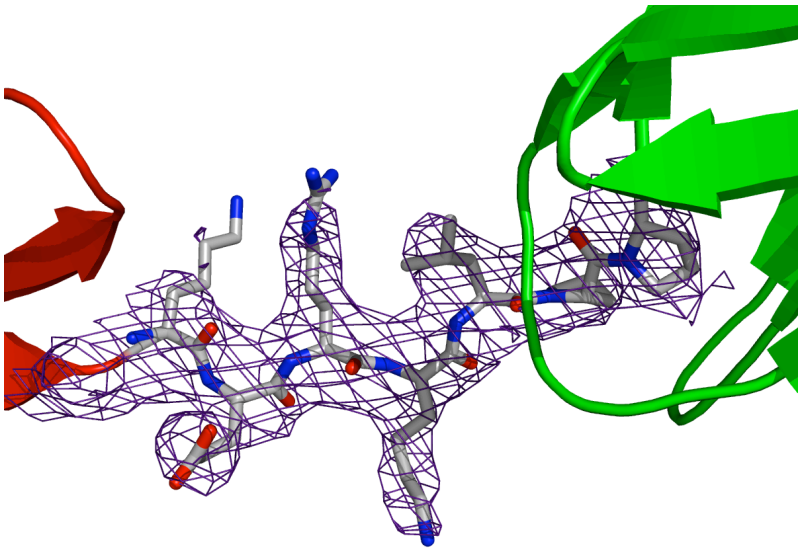


Fig S3a: I65-I66 long linker region

A $(2F_{obs}\text{-}F_{cal})\alpha_{calc}$ electron density map at 3.3 Å resolution contoured at 1σ level is shown for linker groups.

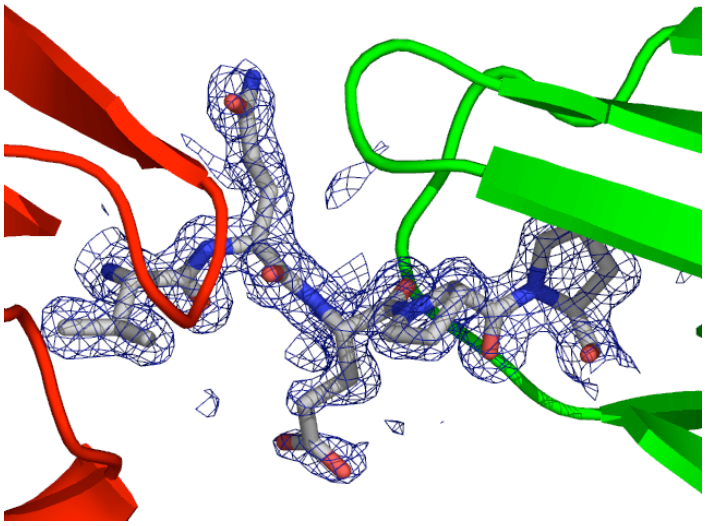


Fig S3b: I67-I68 short linker region

A $(2F_{obs}\text{-}F_{cal})\alpha_{calc}$ electron density map at 1.6 Å resolution contoured at 1σ level is shown for linker groups.

SECTION S4: phasing statistics for I65-I70

Table S4: X-ray data and phasing statistics for derivatized crystals of I65-I70

Spacegroup		$P6_522$	
Data type	Methyl-mercury chloride	Ta_6Br_{12}	Se-Met
Soaking (time / concentration)	2:10 hr / saturation	1 hr / saturation	-
X-ray source	ESRF ID-29-1	ESRF ID-23-1	SLS X06SA
Detector	ADSC Quantum Q210	MARCCD 225	MARCCD 165
Wavelength (Å)	0.979	1.059	0.9793
Unit cell (a, b, c; Å)	141.79, 141.70, 168.01	142.31, 142.31, 166.12	141.37, 141.27, 164.98
Resolution shells	15-4.2 (4.4-4.2)	18-5.5 (6.0-5.5)	17-4.3 (4.4-4.3)
Unique reflections	13265 (1741)	5906 (1381)	12690 (823)
R_{sym}(I) (%)	17.7 (38.5)	9.4 (35.5)	13.5 (32.8)
Multiplicity	9.06 (9.13)	5.86 (5.88)	9.74 (9.75)
Completeness (%)	96.2 (97.4)	96.4 (98.1)	97.9 (99.3)
I/σ(I)	11.31 (6.16)	14.68 (7.17)	16.93 (8.41)
Phasing statistics*:*			
Number of heavy atom sites	11	3	4
Phasing power (iso/ano)	1.50/1.33	0.42/1.14	0.37/0.11
R_{cullis} (iso/ano)	0.71/0.71	0.90/0.90	0.99/0.99

References

1. Wright CF, Teichmann SA, Clarke J, Dobson CM (2005) *Nature* **438**:878-81.
2. Steward A, Adhya S, Clarke J (2002) *J Mol Biol* **318**:935-40.

3.2.1 Extended Materials and Methods

Preliminary work for this study, *i.e.* the protein production, crystallization & derivatization, data collection, processing and initial phase calculation were part of my master thesis (Structural Characterization of I-Band Spring Elements from the Muscle Filament Titin, University of Basel, 2005). Here, a brief account is given for completeness.

3.2.1.1 Protein production and crystallization

Protein over-expression and purification was conducted in *E. coli* strain Rosetta (*DE3*) (Novagen) following published protocols (Marino et al., 2005; von Castelmur et al., 2008). Figure 3.7 shows a typical size exclusion chromatogram as well as the typical final purity of the protein sample used for crystallization. Pure protein (in 50mM Tris-HCl pH 7.5) was concentrated as required for crystallization (10-20mg/ml as measured by A_{280nm}) and stored at 4°C.

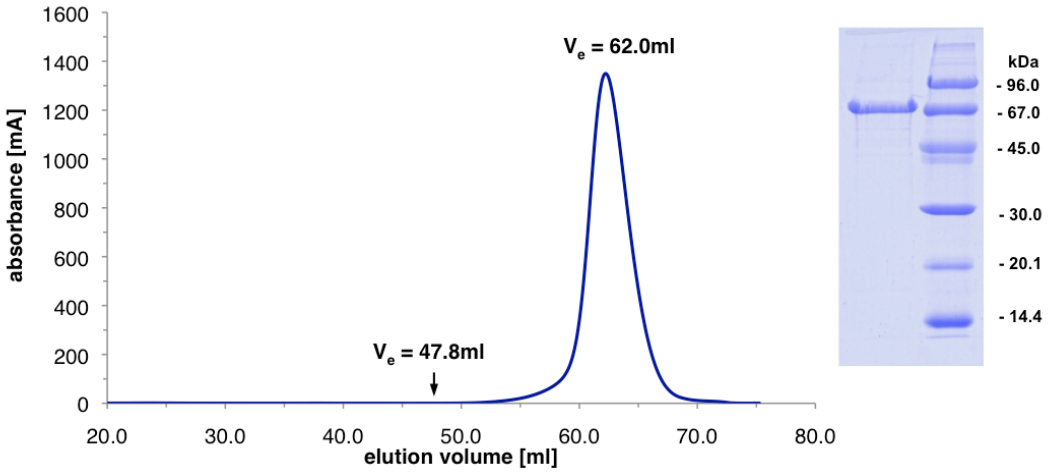


Figure 3.7: Size exclusion chromatogram of titin I65-I70
Size exclusion was performed in 50mM Tris pH 7.5 using a Superdex200 HiLoad 16l60 column (GE Healthcare). The signal shown corresponds to A_{280nm}. The elution volume is relatively high for a monomeric sample (Mr=63862 Da), but reflects the elongated shape of the molecule (estimated Mr for a globular protein would be >500kDa, calibration curve in appendix 6.2). A representative lane from SDS-PAGE illustrates the purity of the sample.

Initial crystallization screening yielded crystals under various precipitants, but nevertheless crystals always displayed the same hexagonal morphology and similar diffraction properties. The best crystals were obtained from 2M NaCl, 100mM sodium acetate pH 4.9-5.1 and 3-5% glycerol or NDSB-201 as additives. They were grown

from protein solutions of 10-20mg/ml using the hanging drop, vapor diffusion method and reached their final size of 250-300μm in their longest axis after 4-12 weeks growth (see Figure 3.8). Crystals were harvested and flash-frozen in liquid nitrogen, using mother liquor supplemented with 20% glycerol as cryoprotectant.

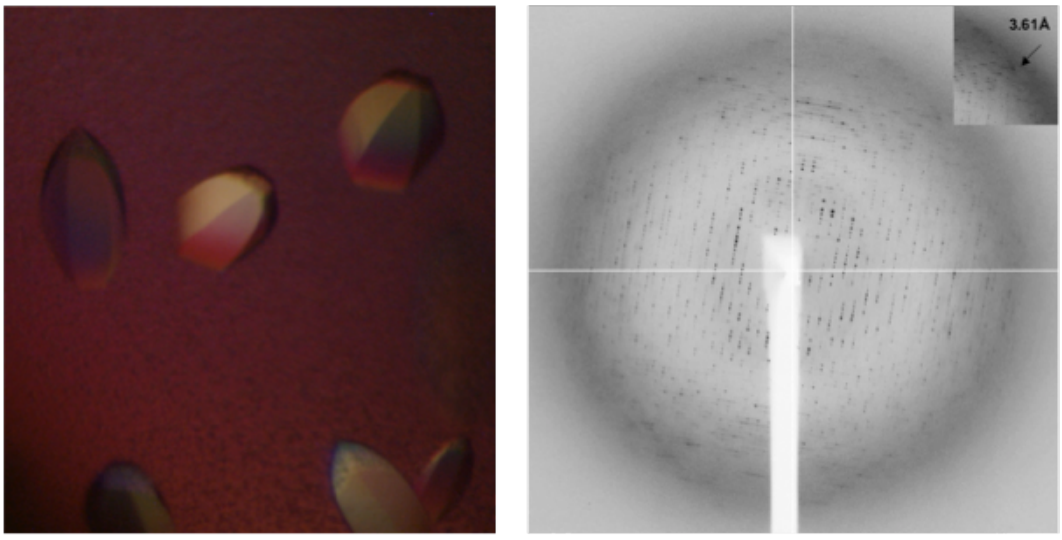


Figure 3.8: I65-I70 crystals and corresponding diffraction pattern.
Crystals were grown from 0.1M Na-acetate pH 5.1, 2M NaCl, 5% glycerol. The diffraction pattern illustrates the diffuse scattering affecting spot background and shape. The edge of the detector corresponds to a maximum resolution of 3Å

3.2.1.2 Data collection and processing

Native data was collected at the beamline ID-29 (ESRF, Grenoble) under cryoconditions and in several wedges of non-overlapping frames. Frequent recentering was necessary as the crystal suffered acutely from radiation damage, even when the beam was attenuated to 23% of its full flux.

Crystals of a Seleno-methionine I65-I70 derivative were small, showed only limited diffraction and suffered severely from radiation damage, preventing the collection of a complete three-wavelength MAD (Multiple Anomalous Dispersion) data set (data collection at beamline X06SA at SLS, Villingen, Switzerland). As a result, only the peak data (collected at 12.660keV) were of sufficient quality to be used in phasing.

The primary phasing power derived from two heavy atom data sets obtained from CH_3HgCl (MMC) and Ta_6Br_{12} derivatives, collected at ESRF beamlines ID29 and ID23 respectively. Data sets for these derivatives were collected at appropriate wavelengths to exploit the anomalous scattering of the heavy atom compounds for

phasing using the MIRAS method (Multiple Isomorphous Replacement with Anomalous Scattering). For data collection of MMC derivatives, a wavelength of 0.9792Å (12.66keV) was chosen, to make use of the strong anomalous signal of the Hg L-III edge (12.283keV) while the Tantalum derivative data was collected at a wavelength of 1.0597Å (11.7keV) in order to exploit the anomalous signal of the Ta L-I edge (11.68keV). Phasing used the three sets (Se-Met, MMC, Ta_6Br_{12}) in combination.

Data processing used the XDS software package (Kabsch, 1993) and the statistics for all data sets are reported in (von Castelmur et al., 2008). The crystal lattice was determined to be $P6_x22$ (x=1 / 5), based on systematic absences, and contained one molecule per asymmetric unit, corresponding to a solvent content of 67% (V_M = 3.76).

3.2.1.3 Phasing

Several attempts at phasing by molecular replacement with Phaser (Mccoy et al., 2005; Storoni et al., 2004) and AMoRe (Navaza, 2001) failed. No conclusive solution could be found using either program, even when using the structure of I67-I69, that became available in our laboratory during the course of this research (von Castelmur et al., 2008), sharing 100% identity and covering 50% of the molecule. This was due to a number of factors: *i)* the quality and limited resolution of the data; *ii)* the filamentous shape and *iii)* inherent flexibility of the molecule where small local deviations in conformation propagate along the chain and add up to appreciable long-range deviations. Thus, structure elucidation using experimental phasing was pursued instead.

For all derivative data sets, analysis of heavy atom binding used the CCP4 program suite (Collaborative Computational Project, 1994). Data from the MMC derivative was used for the initial identification of sites using Patterson searches implemented in CNS (Brünger et al., 1998). In total, 11 Hg sites were identified in this derivative in recurrent cycles. Phases were calculated and refined in SHARP (de La Fortelle and Bricogne, 1997) and the handedness of the screw axis determined to be $P6_522$. The resulting electron density map was further refined in SOLOMON (Abrahams and Leslie, 1996), through a solvent flipping procedure (assuming a solvent content of 60%, approximating one molecule per asymmetric unit).

Phases thus obtained from the MMC derivative were used to calculate both isomorphous and anomalous difference Fourier maps of the Ta_6Br_{12} derivative for identification of heavy atom sites. After refinement together with the previously identified sites in SHARP, the improved phases from this cycle were then used to identify sites in the Se-methionine peak data set by calculating an anomalous difference Fourier map with the new phases. The four identified Se sites were subsequently used as sequence markers to serve as guidance during model building and domain identification. Another cycle of SHARP was run, refining all sites simultaneously and applying solvent flipping assuming 60% solvent. The examination of the residual maps output by SHARP and difference Fourier maps calculated with these refined phases for all derivatives showed that all possible phase information had been extracted (Figure of merit after the last cycle of solvent flipping was 0.83).

3.2.1.4 Model building and refinement

Automated skeletonization, followed by manual skeleton editing, of the electron density map calculated using SHARP in combination with the molecular graphics software O (Jones et al., 1991) allowed the identification of individual domains (**Figure 3.9**). As a next step, individual immunoglobulin domains (telokin, PDB code 1TLK) with poly-alanine sequences were manually positioned in skeletonized maps and “pulled into position” by rigid body refinement in CNS. It proved challenging to establish the identity of the individual Ig and work out the connectivity of the chain, particularly in the areas around the six-fold, where the packing was tighter and symmetry related copies clustered closely together. For this, the position of Se-Met peaks and Hg binding sites to Cys served as guides due to the characteristic distribution of Met and Cys residues among the domains and later also proved useful to ensure the correct sequence registry during model building.

After assignment of the correct position within the model, individual domains were conjoined into a single model with chain breaks for the missing linkers.

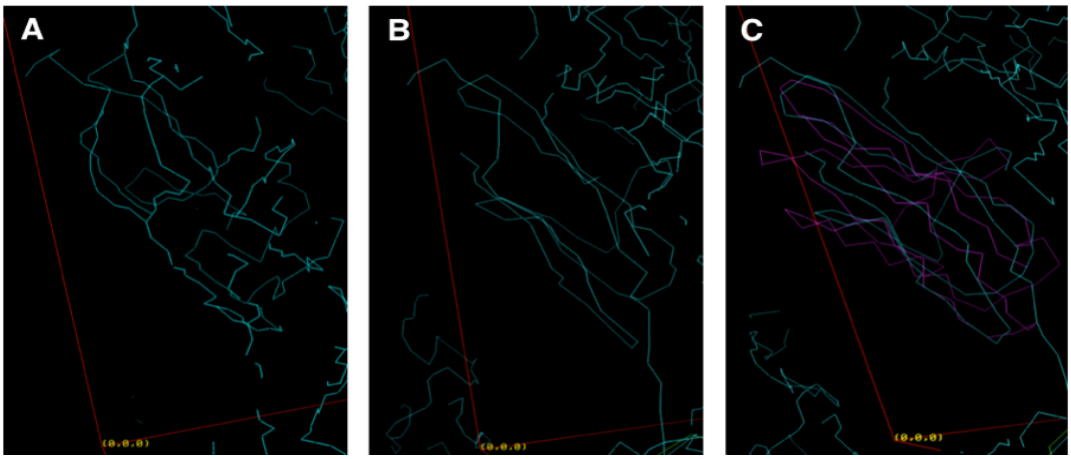


Figure 3.9: Bones from skeletonization of electron density maps before and after editing and superposition of Ig
Electron density maps obtained from SHARP after density modification were skeletonized in O. Bones are shown in cyan. A) Initial bones before manual editing allowed the recognition of domains. B) Bones were manually edited to remove any branching and to establish the connectivity of the protein backbone. C) Ig domain (telokin poly-Ala, magenta) manually superposed onto the edited bones.

Subsequently, poly-alanine Ig were replaced with domains from the high resolution structure of I67-I69 that were trimmed down to identity, i.e. identical and conserved residues were kept, while all the other residues were mutated to alanine. Side-chains were mutated to reflect the correct sequence and fit into the electron density and missing linker residues localized and built in iterative cycles of model building and refinement. Partial disorder of the molecule and the resulting poor electron density within domains I68-I69, affected building and refinement and led to bad stereochemistry within this region. Thus, the domains were replaced with the individual Ig from the high-resolution structure of I67-I69 and main chain restraints applied in refinement.

The model was refined in CNS against native data between 12-3.3Å and refinement included overall anisotropic B-factor scaling, bulk solvent correction, conjugate gradient minimization of atomic coordinates and grouped B-factor refinement with two B-factor values per residue (one for the main chain and one for the side chain). R-free was used as a tool for cross-validation of refinement (Brünger, 1992). For the refinement of initial models, a maximum-likelihood target function taking experimental phases into account was employed – such phases were excluded at later stages. Due to the limited resolution of the data, no solvent molecules were built into the model. The presence of correct hydrogen bonding pattern amongst β-sheets was used as a sanity check for the model in conjunction with Ramachandran statistics and rmsd values for bond lengths and bond angles. A

final round of refinement was performed in Phenix (Adams et al., 2002) using translation-libration-screw (TLS) refinement, treating each domain as an individual TLS group (see Figure 3.10).

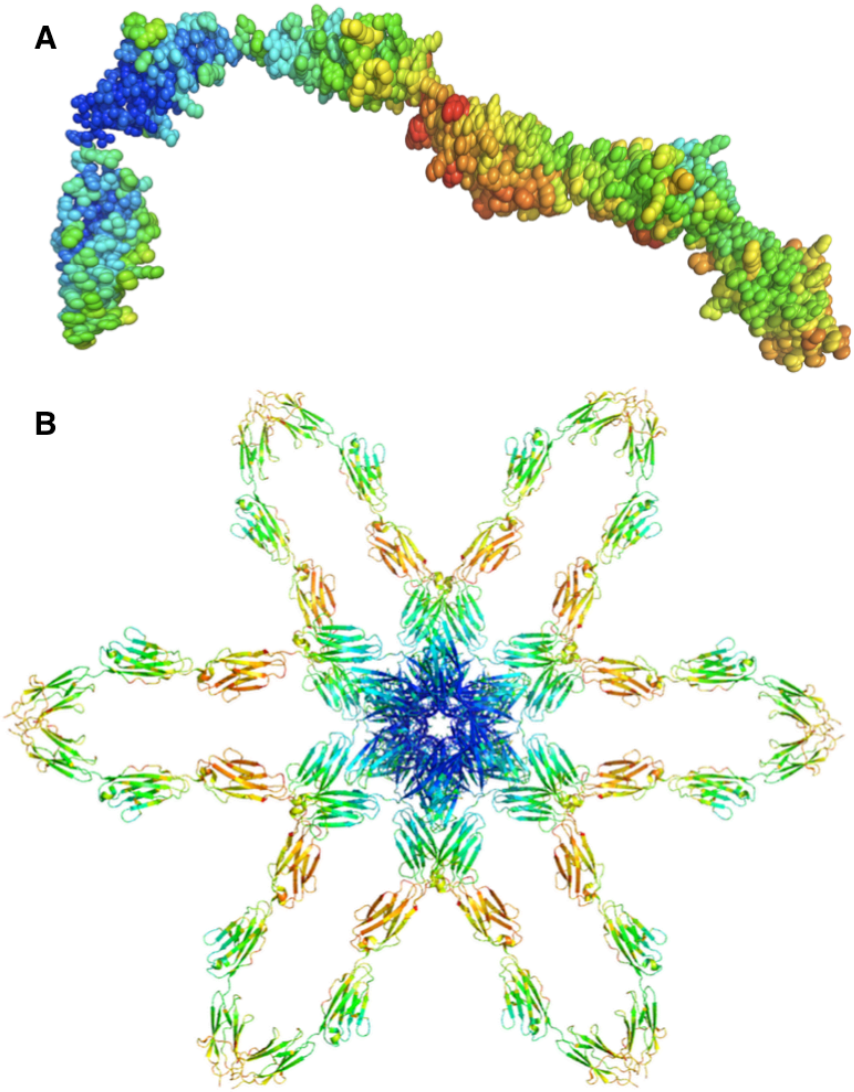


Figure 3.10: Structure and crystallographic lattice of I65-I70
A) Crystal structure of I65-I70 displayed as thermal ellipsoids, derived from the TLS refinement. Atomic displacement parameters are colored according to the magnitude of the displacements, ramped from blue (smallest) to red (largest).
B) Crystallographic lattice of I65-I70, illustrating the arrangement into a network of crisscrossing molecules that delineate vast solvent channels (67% solvent content). The 'heads' of the molecules wrap around the six-fold axis, stacking upon each other to form a continuous intermolecular β-sheet and leading to a column of densely packed molecules. The tails extend away from the axis and form loosely packed regions in the lattice that suffer from partial disorder. For clarity molecules are displayed as ribbon diagrams, again colored according to the magnitude of the temperature factors.

3.2.1.5 Modeling of the central skeletal I-band region

Modeling of the differentially spliced central I-band tandem made use of the fact that I65-I70 is representative of this region, which is entirely made up of Ig domains from the N-conserved type that are connected by conserved linker motifs and are arranged into super-repeats. Modeling and manipulation was performed in O, using fragments of I65-I70 and I67-I69. The structures were subdivided into shorter fragments to obtain representative models for domains joined either by long linkers, short linkers containing the conserved Glu or short linkers containing the Ala mutation. Fragments were kept as long as possible in order to minimize the number of steps necessary to assemble the model (e.g. by using fragment F rather than fragments E plus G, see Figure 3.11). The model was assembled from the fragments by least squares superposition of the first domain of the following fragment on the last domain of the preceding fragment. Figure 3.11 gives a schematic overview of the fragments used for I-band model assembly.

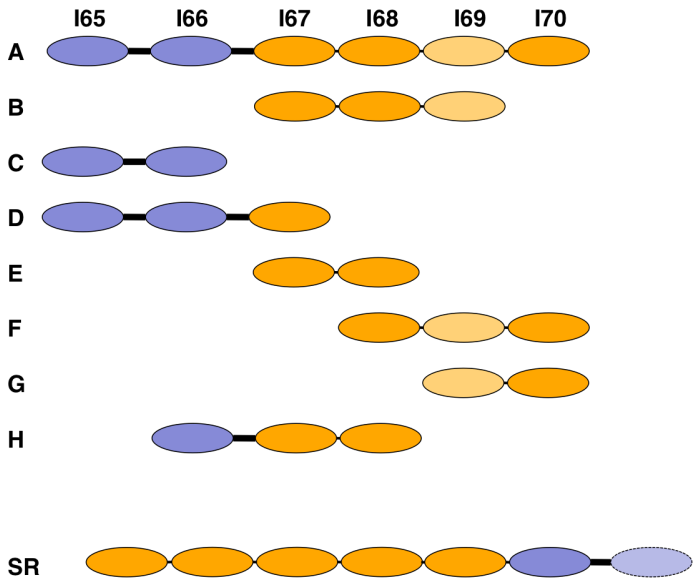


Figure 3.11: Overview of fragments used for the modeling of the skeletal I-band tandem
The crystal structures of I65-I70 (A) and I67-I69 (B) were subdivided into fragments to account for all necessary scenarios. Coloring was in accordance with Figure 3.12 where domains shaded in orange represent tightly connected Ig (lighter shade indicates the presence of E-to-A mutation in transition sequence) while blue domains contain a C-terminal three-residue linker. Fragment C is for the modeling of a single long linker, fragment D for two consecutive long linkers, fragment E for a single short linker, fragment F for the modeling of a normal tight connection followed by a E-to-A mutated linker, fragment G is for a single E-to-A mutated linker and fragment H allows the modeling of a long linker followed by a tight connection. Finally, SR represents the Super-Repeat fragment obtained by appending fragments C-C-D-B into a single fragment model. The last, shaded domain is not part of the super-repeat, but necessary for the super-position and addition of the following fragment to the growing chain.

In order to facilitate the modeling, a first step was to generate the 6-domain super-repeat (SR) at the beginning of the tandem and use that for assembly of the four super-repeat copies. After the regularity of the first four super-repeats break down, modeling relied on the smaller fragments extracted from I65-I70 and I67-I69. Thus, for the modeling of the full chain, fragments were assembled in the order SR-SR-SR-SR-B-E-D-F-B-H-B-A-E-H-G-E-D-D. The superposed fragments were merged into one structure model, representing the whole skeletal tandem spanning from I28 to I79 (see Figure 3.12a).

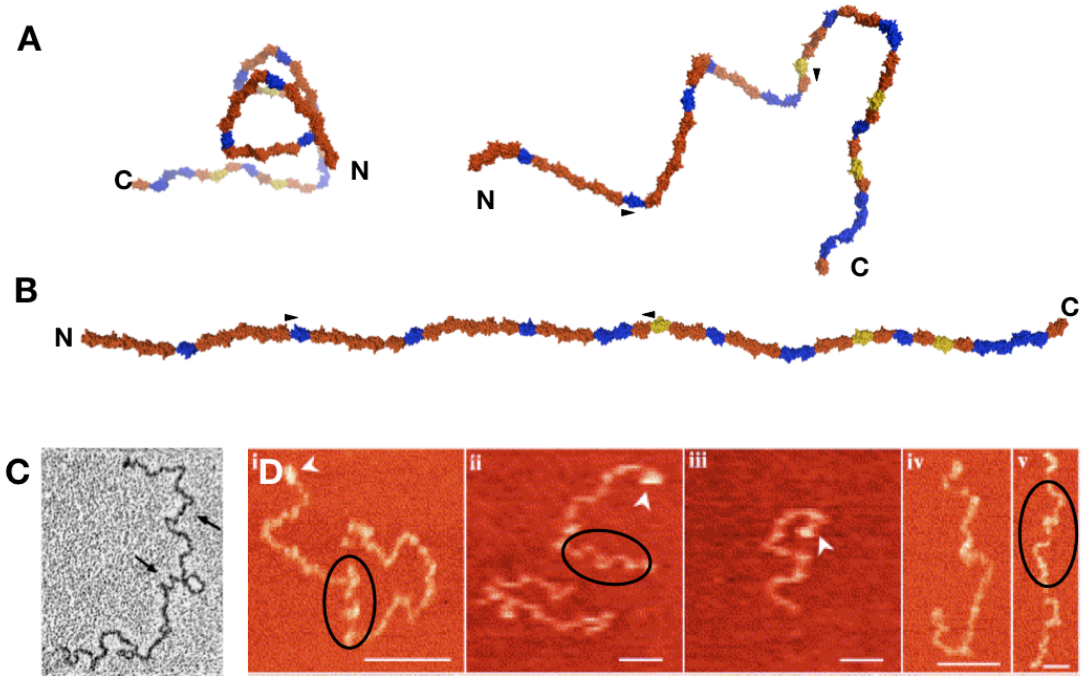


Figure 3.12: Structural model of the skeletal I-Band tandem
A) Frontal (left) and side (right) view of the resulting model for I28-I79, representing a possible slack conformation, extrapolated from the structure and conformation of I65-I70.
B) Model in extended conformation. Coloring is as for Figure 3.11, representing the conservation pattern of the linkers. Arrows indicate the position of I39-I57.
C) Local pseudo-helical substructure observed in EM of full-length titin molecules (Tskhovrebova and Trinick, 2001).
D) Local substructure observed in AFM of full-length titin, hinting at a possible preferred arrangement of super-repeats (Kellermayer et al., 2003). Scale bar, 0.1μm.
Speculatively, this local substructure observed in C and D could correspond to the distinctive pattern observed in our modeling of the I-band filament.

Modeling of the extended state was done manually in O, where straight sections connected by short linkers were rotated around a residue in each long linker acting as a pivot point to generate a “straight” molecule (Figure 3.12B).

For comparison with electron micrographs of recombinantly expressed I39-I57, the corresponding fragment was extracted from the model of I28-I79, as indicated in Figure 3.12. The model was re-oriented in space so that projections match the micrographs as well as possible, but no other manipulations were performed (see Figure 1E of article).

3.2.2 Ultralow-resolution *ab initio* phasing of I65-I70

Structure elucidation of I65-I70 was complicated by diffraction of poor quality and low resolution. MR did not work, even with a model covering 50% of the molecule. Thus, experimental phases were needed. The process turned out to be slow and problematic, since Se-Met labeled protein crystals were affected by pronounced radiation sensitivity and identification of suitable heavy atom derivatives was difficult and phasing protracted.

Ultra-low resolution *ab initio* methods have been applied successfully to the phasing of globular proteins and macromolecular complexes. However, filamentous proteins, characterized by marked anisometry and idiosyncratic crystal lattices, have not been addressed before using this approach. Since it is desirable to broaden the palette of methods applicable to the elucidation of these difficult structures, we have explored the applicability of *ab initio* techniques to I65-I70, with the hope of accelerating phasing and structure elucidation. Here we present an *a posteriori* analysis of this study and discuss the source of the difficulties encountered in this study and the extent to which viable results could be obtained.

We found that the essential difficulty in this study lay in our inability to predict the 'correctness' of the *ab initio* phase sets and, thus, to implement suitable scoring functions. Current phase evaluation protocols largely rely on the interpretability of the resulting maps in terms of individual molecular features and lattice connectivity – often by visual inspection. In the case of filamentous proteins, map interpretation at ultra low-resolutions can be unworkable. The reason for this is the small dimensions (particularly in cross-section) of individual molecules and their tendency to form tight intermolecular contacts where the assembly area is large in relation to the molecular volume. This distorts the resulting electron density images to the extent that lattice features strongly dominate any individual shapes, i.e. the peaks in the syntheses become displaced from the molecular centers toward intermolecular positions. This complicated enormously the scoring of resulting phase sets, troubled confidence estimations and led to the impossibility of defining correct phase selection criteria that would allow extending the phases to resolution limits suitable for recognition of individual molecules. Nevertheless, even under these circumstances, *ab initio* phasing was able to determine the region of the unit cell occupied by the ensemble of protein molecules – *i.e.* the protein mask. These *ab initio* phases improved significantly the quality of initial maps when combined with experimental phases.

In summary, our study suggests that filamentous molecules are challenging targets for low-resolution *ab initio* methodologies, partly because of the lack of voluminous domains, well resolved in space and of simple identification, and partly because of the peculiarities of their atypical crystal lattices. But even if the position of individual molecules in the unit cell might not be determined through this methodology, this can still lead to the successful determination of a lattice mask that can be exploited in density modification approaches, indicating that – albeit limitations - *ab initio* low-resolution phasing can be applied productively to this class of proteins

Personal contribution to the work: provision of data and models for analysis

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Ultralow-resolution *ab initio* phasing of filamentous proteins: crystals from a six-Ig fragment of titin as a case study

Alexandre Urzhumtsev,[a,b]* Eleonore von Castelmur[c] and Olga Mayans[c,d]

[a]IGBMC, CNRS-INSERM-ULP, 1 rue Laurent Fries, 67404 Illkirch, France, [b]Department of Physics, Faculty of Science and Technologies, Nancy-University, 54506 Vandoeuvre-les-Nancy, France, [c]Division of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland, and [d]School of Biological Sciences, Biosciences Building, University of Liverpool, Crown Street, Liverpool L69 7ZB, England

Correspondence e-mail: sacha@igbmc.u-strasbg.fr

Received 28 September 2007
Accepted 31 January 2008

Low-resolution diffraction data (resolution below 12 Å) from crystals of a filamentous six-Ig fragment of titin, I65–I70, were used in *ab initio* phasing with the aim of calculating its lattice packing and molecular envelope. Filamentous molecules, characterized by marked anisometry and idiosyncratic crystal lattices, have not been addressed before using this methodology. In this study, low-resolution phasing (19–122 Å) successfully identified the region of the unit cell occupied by the molecule. Phase extension to a higher resolution (12 Å) yielded regions of high density that corresponded either to the positions of individual Ig domains or to zones of dense intermolecular contacts, hindering the identification of individual domains and the interpretation of electron-density maps in terms of a molecular model. This problem resulted from the acutely uneven packing of the molecules in the crystal and it was further accentuated by the presence of partially disordered regions in the molecule. Addition of low-resolution reflections with phases computed *ab initio* to those obtained experimentally using MIRAS improved the initial electron-density maps of the atomic model, demonstrating the generic utility of low-resolution phases for the structure-elucidation process, even when individual molecules cannot be resolved in the lattice.

1. Introduction

Filamentous proteins are intrinsically flexible and have a poorly defined long-range order, which often renders them unsuitable for crystallographic analysis either by preventing crystal growth or by yielding crystalline formations of low diffraction quality. An examination of crystal lattices for cases in which successful structure elucidation has been achieved shows that the anisometry of these molecules results in two types of frequently observed packing arrangements: (i) tight lateral associations of molecules with only minimal interstitial bulk solvent or (ii) loose arrangements of crisscrossed molecules that delineate vast solvent channels (Fig. 1). In the latter, crystallographic contacts involve only a small fraction of the molecule, so that most of it remains free from interactions and suspended in solution. Owing to the inherent flexibility of filamentous proteins, the unrestrained fraction often exhibits crystalline disorder to a greater or lesser extent. In brief, these high-solvent lattices tend to be characterized by an alternation of well defined areas rich in crystal contacts, which are usually coincident with primary crystallographic axes, and loose 'bridging' regions of poor crystallographic definition.

© 2008 International Union of Crystallography
Printed in Singapore – all rights reserved

Table 1
Statistics for low-resolution X-ray data from I65–I70.

Space group $P6_522$, unit-cell parameters $a = b = 140.15$, $c = 164.32$ Å. The Matthews coefficient V_M of 3.8 Å^3 Da^{-1} corresponds to a solvent content of 67%.

	Whole data set	Highest resolution shell
Resolution zone (Å)	122.0–8.0	9.0–8.0
Unique reflections	1176	334
R_{merge}	7.5	8.3
Multiplicity†	22.8	14.7
Completeness (%)	100.0	100.0
$I/\sigma(I)$	31.9	19.6

† Data were collected in three sweeps that extended to different resolutions, thus the multiplicity for the whole set is noticeably higher than that for the highest resolution shell.

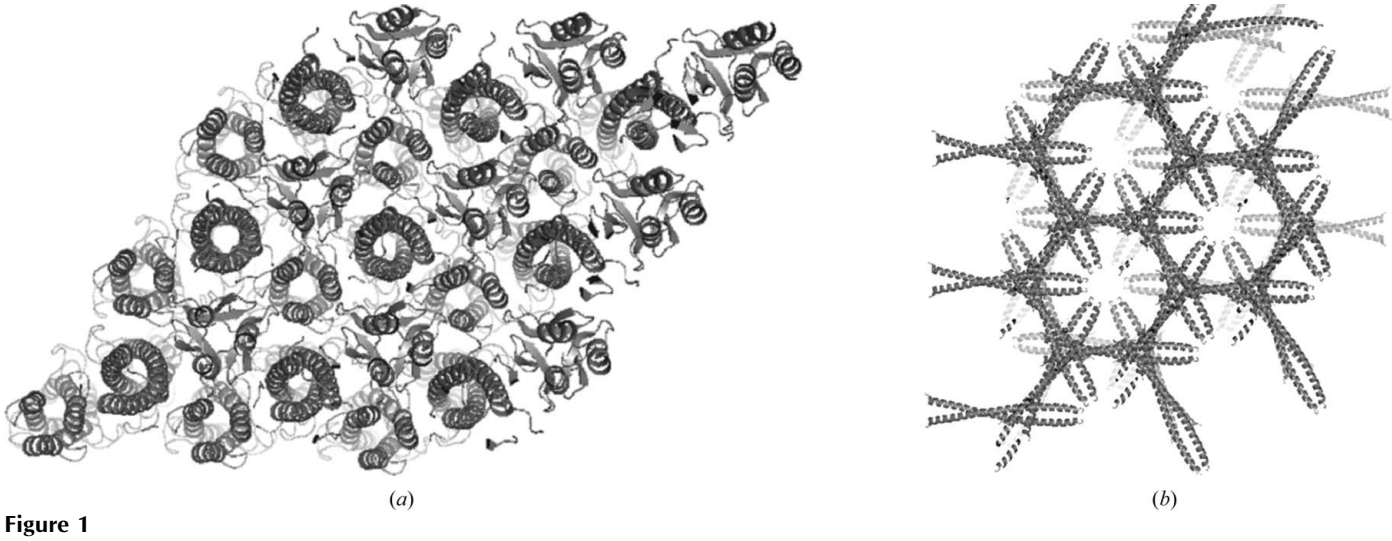
The phasing of filamentous proteins can also be affected by the specific characteristics of these molecules. For example, molecular-replacement approaches can be unsuccessful because of long-range structural deviations along the molecular axis. Thus, it is desirable to broaden the palette of methods that are applicable to the determination of these structures. To this effect, we have investigated the potential of *ab initio* ultralow-resolution phasing for the calculation of a molecular envelope that could serve as an initial source of phases in the elucidation process.

The current work is an exploratory *ab initio* study of a six-Ig fragment, I65–I70, from the muscle filament titin. This protein is composed of ∼300 serially connected Ig and FnIII domains and acts as a molecular spring in the sarcomere (reviewed by Granzier & Labeit, 2004). The crystal structure of I65–I70 has recently been elucidated at 3.3 Å resolution using MIRAS (von Castelmur *et al.*, 2008). Diffraction of poor quality and limited maximal resolution (ranging from 6.0 to 3.3 Å for individual crystals) prevented the steady progress of experimental phasing in this case. Thus, we decided to explore independently and in parallel the capabilities of ultralow-resolution *ab initio* phasing in aiding the process. Obviously, evaluation of the results could only be carried out *a posteriori* once the atomic model had been completed. Even though we anticipated that the composition of I65–I70 by Ig globular domains would make the molecule methodologically more tractable than the strongly anisometric motifs of minimal cross-sections, such as coiled coils, the final analysis showed that I65–I70 posed an extreme challenge to this technique. This text discusses the source of the difficulties encountered and the extent to which viable results could be obtained. Even though the *ab initio* calculations did not allow us to make conclusions on molecular features, they led us to identify with reasonable accuracy the fraction of the unit cell occupied by the protein fragment. This information, which is particularly suitable for exploitation in density-modification approaches, indicates that *ab initio* low-resolution phasing can be productively applied to this class of proteins, albeit with limitations.

2. Data and tools

2.1. Ultralow-resolution diffraction data and molecular characteristics of I65–I70

The cloning, expression and crystallization of domains I65–I70 (amino acids 7946–8511) from rabbit soleus titin (human titin at TrEMBL Q8WZ42) have been reported previously (Marino *et al.*, 2005; von Castelmur *et al.*, 2008). In brief, crystals belonged to space group $P6_522$, with unit-cell parameters $a = b = 140.15$, $c = 164.32$ Å and one molecule per asymmetric unit ($V_M = 3.8$ Å^3 Da^{-1}; 67% solvent content). Low-resolution diffraction intensities (122–8 Å resolution) were recorded on beamline X06SA at SLS (Villigen) independently from those finally used in structure elucidation,



(*a*) (*b*)

Figure 1
Crystallographic lattices characteristic of filamentous proteins. (*a*) Tightly packed lattice exemplified by a fragment of fibritin that includes a coiled-coil motif and a C-terminal foldon (PDB code 1aa0; Strelkov *et al.*, 1996). The space group is *P*3, with an estimated solvent content of 36% (V_M = 2.35 Å^3 Da^{-1}). (*b*) Loose crisscrossed crystal packing illustrated by a coiled-coil fraction from lamin A in space group $P6_522$ (PDB code 1x8y; Strelkov *et al.*, 2004). The estimated solvent content is 71% (V_M = 4.2 Å^3 Da^{-1}). V_M is the Matthews coefficient (Matthews, 1968).

which originated from a different crystal. Data were collected at a crystal-to-detector distance of 475 mm on a MAR CCD 165 detector at $\lambda = 0.9537$ Å. The images corresponded to non-overlapping 1° oscillation steps, amounting to a total of 100° rotation. Three passes were carried out with decreasing exposure times and using an increasingly attenuated beam to ensure optimal recording of strong reflections. Data were processed using the *XDS*/*XSCALE* suite (Kabsch, 1993). Table 1 lists the data-processing statistics.

I65–I70 is composed of six Ig domains that share an average pairwise sequence identity of ∼35%. Individual domains are small (composed of about 100 amino acids) and have an elliptical shape; their axial dimensions are approximately 45 × 30 × 20 Å. The interdomain linker sequences were predicted to consist of three, three, zero, zero and zero residues in length based on sequence data (Marino *et al.*, 2005). This suggested that the N-terminal half of the fragment has a high potential for flexibility and conformational heterogeneity, while the C-terminal part may be better defined structurally. Previous small-angle X-ray scattering (SAXS) data and an electron-microscopy analysis of the end-to-end molecular distance distribution of I65–I70 indicated that it is semi-rigid and predominantly adopts a semi-extended arrangement in solution, with an estimated length of about 23 nm (Marino *et al.*, 2005). However, it should be borne in mind that parameters corresponding to molecular averages and estimated in solution are not necessarily indicative of the overall molecular conformation of I65–I70 in a crystalline lattice.

2.2. Phasing methods

Diffraction data from native crystals were recorded to 3.3 Å resolution. Repeated attempts to elucidate the structure by molecular replacement or MAD on SeMet-labelled crystals were unsuccessful. Therefore, experimental phasing using heavy-atom derivatives was instead pursued. Because this became a demanding task with slow progress, we initiated a parallel search for the overall conformation and packing of I65–I70 in the crystal using ultralow-resolution data with *ab initio* phasing.

We applied two previously reported low-resolution *ab initio* phasing methods: the *FAM* (Lunin *et al.*, 1995, 1998) and image-connectivity approaches (Lunin *et al.*, 2000). A reminder of these methods is given in Appendix *A*. The study of I65–I70 was a challenge for both methodologies, but especially for the connectivity-based approach. Filamentous proteins can present unusual lattice arrangements (Fig. 1) in which locally dense packing makes it scarcely possible to recognize individual molecules in contact regions in low-resolution electron-density maps. As a consequence, it is difficult to predict a correct scoring rule for the selection of best phase variants based on the interpretability of connectivity in these maps. Thus, the *FAM* approach was chosen as being the most suitable for the initial stages of *ab initio* phasing in this project.

The programs *SF2CNS* (Urzhumtsev & Urzhumtseva, 2002) and *PyMOL* (DeLano, 2002) were used to calculate and display the Fourier maps in this study.

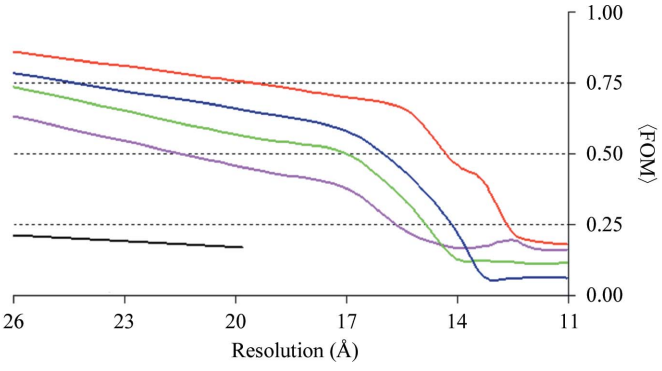


Figure 2
Figure of merit and effective resolution of *ab initio* phase sets. The mean figure of merit ⟨FOM⟩ is shown as a function of resolution. Phase sets were obtained by selecting maps at resolutions of 19 Å (set FAM19; black), 16 Å (set conn16; magenta), 15 Å (set conn15; green), 14 Å (set conn14; blue) and 12 Å (set conn12; red). The resolution at which ⟨FOM⟩ falls below 0.5 can be considered as the effective value.

3. Results of *ab initio* phasing

3.1. Initial phasing with the *FAM* method

The *FAM* phasing procedure depends on two key parameters: the composition of the models (*i.e.* the number of spheres in the asymmetric unit of the cell) and the number of reflections used. The result of the *FAM* method is not the position of the spheres, but a set of phases and their figures of merit. Thus, it should be emphasized that the positions of spheres for the best *FAM* models do not necessarily correspond to the centres of molecules or domains (Lunin *et al.*, 1995). The larger the number of spheres, the better the model shape can be approximated and the more closely the experimental structure factors may be fitted. Increasing the number of spheres (each characterized by the three coordinates of its geometrical centre) requires increasing the number of structure-factor magnitudes used for phasing in order to avoid over-parametrization. In the case of I65–I70, initial models were composed of six spheres, which were explored using all 90 reflections with resolution below 21 Å.

Two other important parameters in *FAM* phasing are the B value and the minimal distance $D_{\rm min}$ between the centers of the positioned spheres. The parameter B represents the size of the Gaussian isotropic peak of the electron density for each *FAM* sphere. To find suitable values for these parameters, we screened them systematically using several short series generations each consisting of 1000 models with given B and $D_{\rm min}$ values. Those series containing a large number of models whose calculated structure-factor magnitudes bore a relatively high correlation to the experimental values determined the optimal values of B and $D_{\rm min}$. This analysis estimated $B = 10\,000$ Å^2 and $D_{\rm min} = 25$ Å; the latter is consistent with the smallest dimension of the Ig domain (about 20–25 Å).

Selected B and $D_{\rm min}$ values were used to calculate electron-density maps for I65–I70 by three consecutive iterations of *FAM*. The phase set $\{\varphi_{\rm FAM21}\}$ selected from the first iteration

extending to 21 Å resolution was treated using clustering techniques (Lunin *et al.*, 1990). However, since the main cluster gave an image that was basically identical to that obtained by the averaging of all selected phase sets, clustering was not performed in subsequent iterations and simple averaging was instead applied. The correct sign of the density cannot be calculated directly by *FAM*; therefore, we applied a generalized maximum-likelihood technique proposed by Lunin *et al.* (1998) and Petrova *et al.* (2000) (further details are given in Appendix *A*). After a generic first cycle with models generated uniformly across the unit cell, two further iterations (data resolution 122–19 Å; resulting in phase set $\{\varphi_{\rm FAM19}\}$) were directed to improve the molecular envelope by generating new sphere models inside the envelope obtained from the previous iteration and defined by a volume of $0.8 \times V_{\rm cell}$ (see Appendix *A* for the choice of the density-cutoff level in defining molecular envelopes). The values of all other parameters in these cycles were kept as in the initial cycle. In subsequent iterations, the number of spheres was increased from 6 to 12, $D_{\rm min}$ was decreased from 25 to 15 Å, the *B* value was decreased slightly from 10 000 to 8000 Å^2 and the envelope volume was reduced. The formal resolution of the newly obtained phase set was increased stepwise up to 16 Å. However, the fine refinement of parameters and the more detailed 12-sphere *FAM* models did not result in a significant image improvement and thus *FAM* phasing was interrupted at the phase set $\{\varphi_{\rm FAM19}\}$ after the third iteration.

3.2. Phase extension with connectivity

The low-resolution image of I65–I70 computed with the primary *FAM* phase set $\{\varphi_{\rm FAM19}\}$ was not interpretable in terms of individual domains. To improve the image, we applied a connectivity-based phase-extension procedure in which data resolution was gradually increased to 16, 15, 14 and 12 Å (phase sets $\{\varphi_{\rm conn16}\}$, $\{\varphi_{\rm conn15}\}$, $\{\varphi_{\rm conn14}\}$ and $\{\varphi_{\rm conn12}\}$, respectively). Fig. 2 shows the variation of the mean figure of merit with resolution for the intermediate and final phase sets in this study. Since the loss of phase information at increasing resolution is reflected by a corresponding decrease in the associated mean figure of merit, the effective resolution of a given phase set is lower than its nominal resolution and can then be considered as that where the mean figure of merit falls below 0.5 (Lunin & Woolfson, 1993). In particular, it should be noted that the effective resolution of the final set $\{\varphi_{\rm conn12}\}$ is approximately 15 Å rather than the calculated 12 Å.

The initial expectation in this study was that a correct low-resolution Fourier synthesis would reveal local maxima near the centre of each individual Ig domain. For this reason, we formulated the topological requirement of identifying six independent globules of similar size (one globule per domain), each with 12 copies (the number of symmetry copies in the space group), as a rule for map selection during phase extension and improvement. Indeed, at a resolution of 19 Å the molecular envelope defined by the unit-cell volume $0.30 \times V_{\rm cell}$ in $\{\varphi_{\rm conn12}\}$ maps was more detailed than a similar envelope for $\{\varphi_{\rm FAM19}\}$ maps at the same resolution (Figs. 3*a* and 3*b*). At a resolution of 15 Å, $\{\varphi_{\rm conn12}\}$ maps were too noisy and when applying higher cutoff levels, for example $0.15 \times V_{\rm cell}$, the molecular envelopes became split into separate globules (Fig. 3*c*). All efforts to interpret these maps in terms of six linked Ig domains failed. In retrospect, the high density peaks corresponded to regions of tight crystallographic contacts and not to individual Ig domains. This complication primarily arose

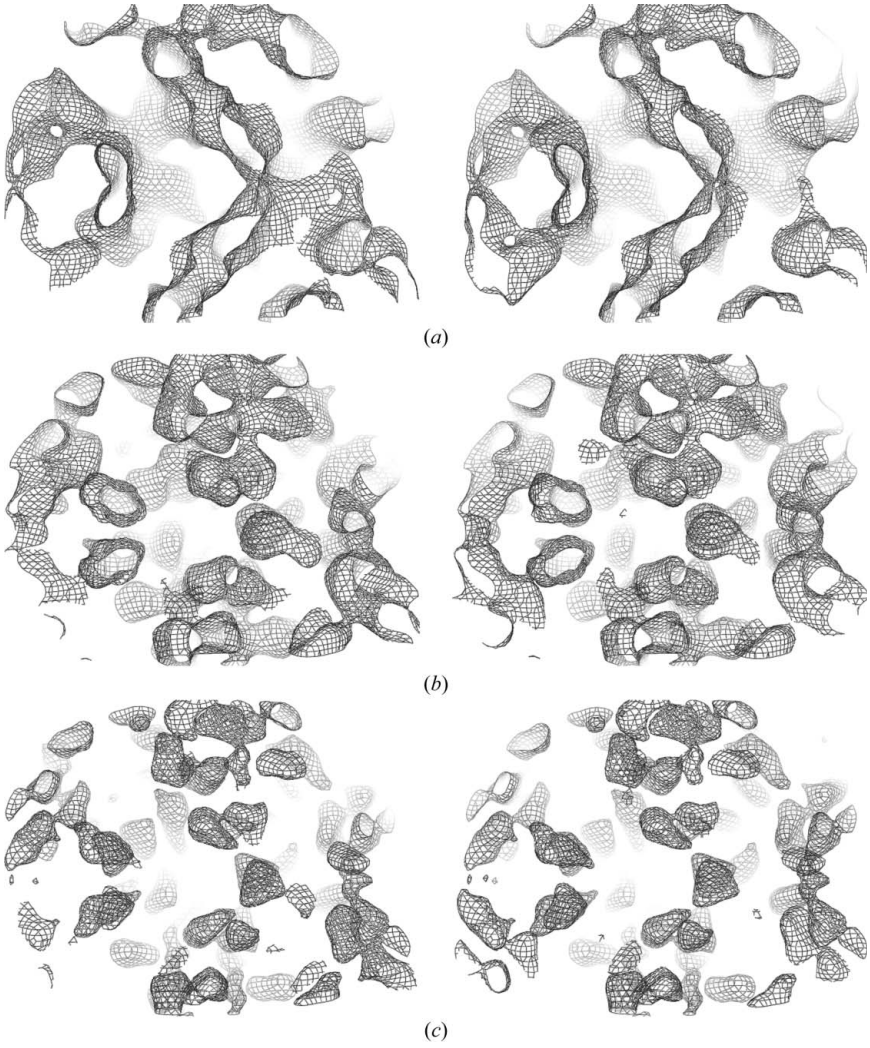


Figure 3
Fourier synthesis calculated with *ab initio* phase sets. Electron-density maps corresponding to (*a*) phase set $\{\varphi_{\rm FAM19}\}$, resolution 19 Å, $0.30 \times V_{\rm cell}$, contour level 0.6σ, (*b*) phase set $\{\varphi_{\rm conn12}\}$, resolution 19 Å, $0.30 \times V_{\rm cell}$, contour level 0.4σ, (*c*) phase set $\{\varphi_{\rm conn12}\}$, resolution 15 Å, $0.15 \times V_{\rm cell}$, contour level 0.8σ.

from the crystal lattice features of I65–I70, a phenomenon of low-resolution phasing that has been observed previously for globular proteins (see, for example, Lunin *et al.*, 1995).

4. *A posteriori* comparison with the atomic model

4.1. Atomic model and molecular packing

The crystal structure of I65–I70 has been elucidated at 3.3 Å resolution using MIRAS independently of this *ab initio* study (von Castelmur *et al.*, 2008). In agreement with SAXS and EM data (Marino *et al.*, 2005; von Castelmur *et al.*, 2008), I65–I70 adopts a semi-extended conformation with an end-to-end distance of ∼212 Å (280 Å contour length). The molecule is bent at its N-terminal 'head' composed of two Ig domains, while the C-terminal half has a straight appearance (Fig. 4*a*). In the crystal, each molecule is roughly contained in a plane normal to the *z* axis, in which the bent 'heads' wrap around the sixfold axis and the 'tails' extend away from it. The 'heads' stack laterally upon each other along the *z* axis, building a continuous intermolecular β-sheet that results in the formation of an infinite left-handed helix within the crystal (Fig. 4*b*). In this arrangement, the N-terminal 'heads' form a densely packed region with numerous intermolecular contacts. In contrast, the molecular 'tails' form a loosely packed region in the form of 'bridges' that connect the dense 'columns' centred at the sixfold axes. The unrestrained 'tails' of the molecule suffered from poor crystalline order (probably owing to chain dynamics), further increasing the contrast between the vast continuous zones of high density and the rest of the unit cell.

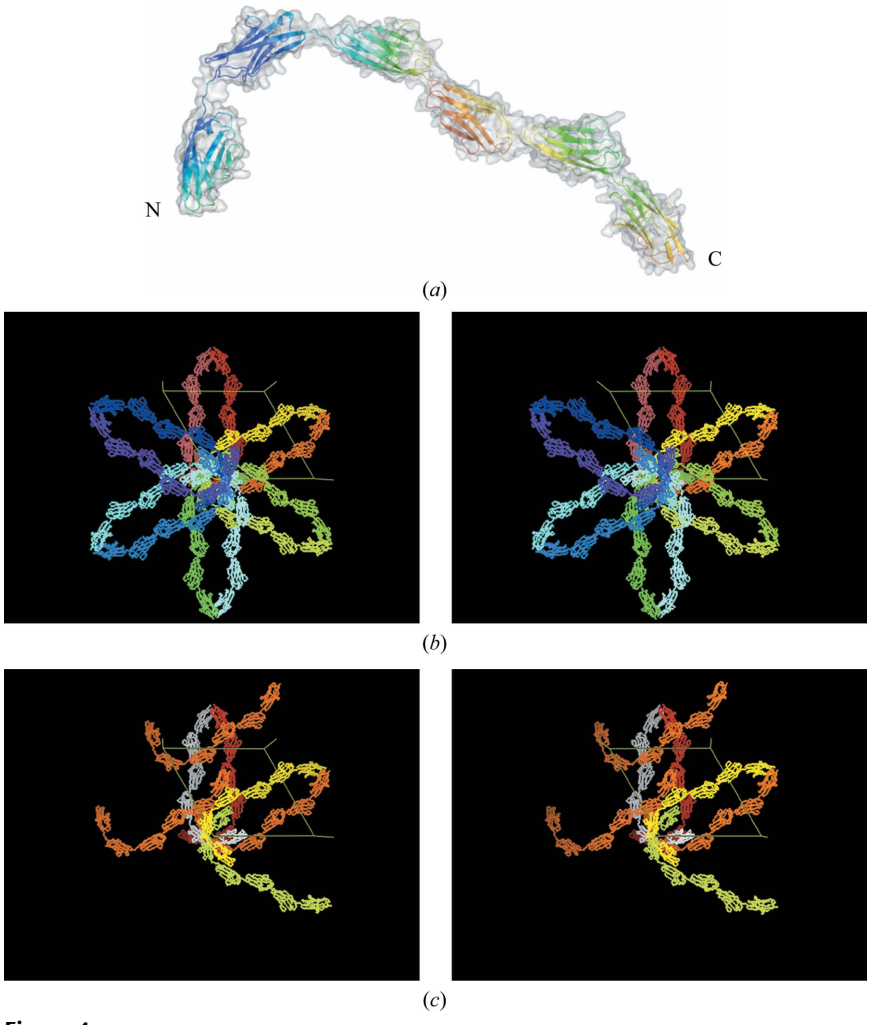


Figure 4
Crystallographic lattice of I65–I70 and molecular contacts. (*a*) Crystal structure of I65–I70 (PDB code 3b43, von Castelmur *et al.*, 2008); the molecular cartoon is coloured in a blue-to-red gradient according to model temperature factors, where blue represents the lowest values and thus indicates well ordered regions of the molecule; the molecular envelope calculated from the atomic model is shown in grey. (*b*) Stereoview of 12 molecules around the sixfold axis forming a left-handed helix; the colour, from red to blue, reflects the *z* position of the centre of the molecule. (*c*) Example of molecular contacts: an I65–I70 molecule (white) in contact with as many as three identical crystallographic copies (orange) that belong to three neighbouring unit cells (a stereoview); the 'head' of the molecule in white is inserted between the symmetry copies forming a left-handed molecular helix (in yellow and bottom-right in orange).

4.2. Analysis of low-resolution maps calculated with model phases

In order to analyze the origin of the difficulties in interpreting low-resolution maps, we calculated structure factors $F_{\rm mod}\exp(i\varphi_{\rm mod})$ from the atomic model and applied a bulk-solvent correction using the flat-mask model (Jiang & Brünger, 1994) and bulk-solvent values of $k_{\rm bulk} = 0.32$ e Å^{-3} and $B_{\rm bulk} = 70$ Å^2 (Fokine & Urzhumtsev, 2002; Afonine *et al.*, 2005). The phases $\varphi_{\rm cryst}$ of the resulting structure factors $F_{\rm cryst}\exp(i\varphi_{\rm cryst})$ were a better approximation to the unknown phases $\varphi_{\rm obs}$ associated with $F_{\rm obs}$ than the phases $\varphi_{\rm mod}$ merely calculated from the model. Three Fourier syntheses were then calculated at a resolution of 19 Å with coefficients $F_{\rm mod}\exp(i\varphi_{\rm mod})$, $F_{\rm obs}\exp(i\varphi_{\rm mod})$ and $F_{\rm obs}\exp(i\varphi_{\rm cryst})$. In the first synthesis, calculated using $F_{\rm mod}\exp(i\varphi_{\rm mod})$, the envelope with a unit-cell volume $0.30 \times V_{\rm cell}$, roughly corresponding to the protein fraction suggested by the value of the Matthews coefficient $V_{\rm M}$, included the atomic model reasonably well (only about 20% of the model was missed). At high cutoff levels, for example corresponding to $0.02 \times V_{\rm cell}$, the peaks of the synthesis generally corresponded to the centres of individual Ig domains. However, syntheses calculated from experimental $F_{\rm obs}$ and model phases ($\varphi_{\rm mod}$ or $\varphi_{\rm cryst}$) showed a deterioration of the image. The envelopes of a volume near $0.3 \times V_{\rm cell}$ missed more atoms than the model synthesis, practically 40% for the synthesis with $F_{\rm obs}\exp(i\varphi_{\rm cryst})$. At higher cutoff levels, the density at the centres of individual Ig positions progres-

sively split or disappeared. In addition, spurious peaks appeared at the interface between Ig domains belonging to the same or different molecules.

A more detailed inspection of density peaks was performed in order to examine whether they presented the expected connectivity features, *e.g.* one peak per domain. This would mean the presence of six peaks of similar volume (individual Ig domains), each with 12 copies (reflecting the space-group symmetry). However, the Fourier syntheses calculated with the 'best possible' set of coefficients $F_{\rm obs}\exp(i\varphi_{\rm cryst})$ at a resolution of 19, 15 and 12 Å and explored using different cutoff levels did not reveal the desired features. In other words, even the best possible phases $\varphi_{\rm cryst}$ did not lead to molecular envelopes with optimal coverage of the model and with readily interpretable characteristics in terms of individual domains. Thus, we concluded that an interpretation of low-resolution maps in terms of a molecular structure of I65–I70 was unfeasible, even when using very accurate phases and at a resolution higher than that actually applicable to *ab initio* phasing in this study.

4.3. Comparison of results

The primary Fourier synthesis $\rho_{\rm FAM19}(\mathbf{r})$ obtained by *ab initio* phasing was that calculated with the coefficients $m_{\rm FAM19}F_{\rm obs}\exp(i\varphi_{\rm FAM19})$ at 19 Å resolution using experimental structure-factor magnitudes weighted by the corresponding figures of merit. In order to perform a comparison with the model, an optimal relative choice of the unit-cell origin had to be made as well as a decision on whether density $\rho_{\rm FAM19}(\mathbf{r})$ was to be flipped or not. Visual inspection as well as numerical analysis of two allowed origin shifts, namely (0, 0, 0) and $(0, 0, \frac{1}{2})$, in both $\rho_{\rm FAM19}(\mathbf{r})$ and $-\rho_{\rm FAM19}(\mathbf{r})$ unambiguously indicated that a best superposition required shifting the origin of the *ab initio* calculated image onto the point $(0, 0, \frac{1}{2})$ but that density flipping was not required, confirming the likelihood-based choice of the sign of the density.

A superposition of the atomic model with $\rho_{\rm FAM19}(\mathbf{r})$ shifted by $(0, 0, \frac{1}{2})$ revealed that a molecular envelope of volume 0.30 $\times$ $V_{\rm cell}$ showed reasonably continuous molecular packing (Fig. 5*a*) and covered half of the model. The enlargement of the envelope volume to 0.6 $\times$ $V_{\rm cell}$ reduced the missed part of the model to 20%. The imperfection in the model coverage is not surprising since maps calculated using the model phases $\varphi_{\rm cryst}$ also failed to fully enclose the molecule (see §4.2). A synthesis calculated using final refined phases, $m_{\rm conn12}F_{\rm obs}\exp(i\varphi_{\rm conn12})$, at 19 Å showed that at the same cutoff level the molecular envelope became split into globules (Fig. 5*b*). However, these did not entirely correspond to individual Ig domains and this new envelope did not improve model coverage. At the higher resolution of 15 Å and using a cutoff volume of 0.15 $\times$ $V_{\rm cell}$ the envelope did not systematically correlate with individual domains either (Fig. 5*c*). In summary, *ab initio* phasing in this study succeeded in correctly identifying the volume of the unit cell occupied by the protein fraction but could not resolve individual molecules within the lattice or individual domain components within the molecule.

5. Low-resolution-based image improvement

During the current study, it became evident that the crystallographic case of I65–I70 was extremely unfavourable for low-resolution *ab initio* phasing. It was not possible to identify individual molecular envelopes, as was the case, for example, in Lunin *et al.* (2001), Fokine, Morales *et al.* (2003) and Müller *et al.* (2006), where the molecules or their core parts were more compact and had less pronounced intermolecular contacts. However, even the identification of the fraction of the unit-cell volume occupied by the ensemble of molecules (*i.e.* the protein mask) can be helpful in structure elucidation. For example, it can aid MIR-based experimental phasing, which traditionally ignores structure factors at a resolution below approximately 15–30 Å. For I65–I70, experimental phases became available in the resolution range 4.2–18.0 Å (von Castelmur *et al.*, 2008). However, the mean figure of merit of these phases was low for resolutions lower than 17 Å. We substituted the poorly defined MIRAS phases for *ab initio* phases in the resolution range 17–18 Å and completed the diffraction data down to 122 Å resolution using the low-resolution *ab initio* set including the calculated phases. Such composite map calculation, mixing phase information, can be performed at early stages of structure investigation. In the case of I65–I70, a Fourier synthesis with coefficients

$$\begin{cases} m_{\rm conn12}F_{\rm obs}\exp(i\varphi_{\rm conn12}), & 122.0\ \text{Å} > d > 17.0\ \text{Å} \\ m_{\rm MIRAS}F_{\rm obs}\exp(i\varphi_{\rm MIRAS}), & 17.0\ \text{Å} > d > 4.2\ \text{Å} \end{cases}$$

substantially improved the starting images derived from the experimental phases alone, exhibiting an enhanced solvent–protein delimitation (Fig. 6). This confirmed that the *ab initio* low-resolution phases contained correct and useful structural information, even when this was scarcely recognizable by visual inspection. Thus, it could be inferred that phase sets obtained through the combination of low-resolution *ab initio* approaches and other phasing methods might facilitate significantly the structure-elucidation process by aiding density-modification protocols (Podjarny *et al.*, 1981; Urzhumtsev, 1991). This would particularly overcome bottlenecks in those cases in which only phases of poor quality were obtained by conventional means.

6. Discussion

The six-Ig fragment I65–I70 from titin was extremely unfavourable for low-resolution *ab initio* studies. Given that low-resolution *ab initio* phasing has been applied successfully to a number of globular proteins, we saw the need to (i) understand the parameters that had hindered the application of the technique in this case and (ii) evaluate to what extent these parameters were generic for filamentous proteins and likely to affect the overall applicability of the technique to this class of proteins. In I65–I70, the irregularity of the crystal packing, the multi-domain composition, the large number of symmetry copies in the unit cell and the numerous intermolecular contacts of each copy worked together to compromise the success of the technique. Nonetheless, the essential difficulty

in this study lay in our inability to predict the 'correctness' of the *ab initio* phase sets and thus to implement suitable scoring functions. Current phase-evaluation protocols largely rely on the interpretability of the resulting maps in terms of individual molecular features and lattice connectivity, often by visual inspection. In the case of filamentous proteins, such map interpretation at ultralow resolutions can be unworkable. The reason for this is the small dimensions (particularly in cross-section) of individual molecules (or domains) and their tendency to form tight intermolecular contacts where the assembly area is large in relation to the molecular volume. This distorts the resulting images to the extent that lattice features strongly dominate any individual shapes. In the case of I65–I70, the interpretability of the molecular envelope in terms of structural features was precluded by the inability to identify individual Ig domains. The grounds for this appeared to be the dense molecular packing of multiple domains near the sixfold axis, which caused individual molecular envelopes not to be resolved and the peaks in the syntheses to become displaced from the domain centres toward interdomain positions. This complicated enormously the scoring of resulting phase sets, affected confidence estimations and led to the impossibility of defining correct phase-selection criteria that would allow extension of the phases to resolution limits suitable for the recognition of individual domains. An *a posteriori* analysis of Fourier syntheses calculated from model phases showed that the difficulty of map interpretation was high even when maps had been computed using phases derived from the model. Nevertheless, even under these circumstances, *ab initio* phasing was able to predict the phases for the lowest resolution reflections with reasonable accuracy, leading to the determination of the region of the unit cell occupied by the ensemble of I65–I70 molecules, *i.e.* the protein mask, and resulting in a significant improvement of experimental maps.

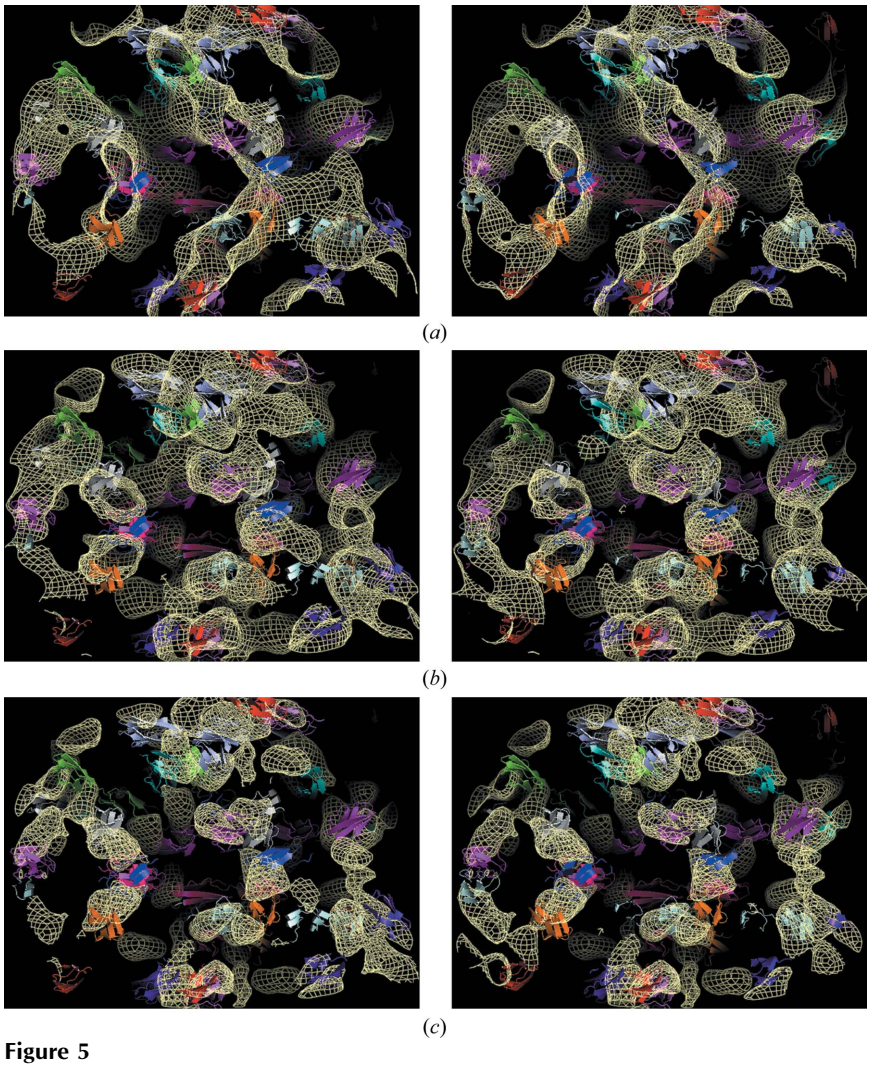


Figure 5
Stereoview of the superposition of the atomic model with the *ab initio* phased Fourier syntheses. The same view is shown as in Fig. 3. (*a*) Phase set $\{\varphi_{\rm FAM19}\}$, resolution 19 Å, $0.30 \times V_{\rm cell}$, contour level 0.6σ, (*b*) phase set $\{\varphi_{\rm conn12}\}$, resolution 19 Å, $0.30 \times V_{\rm cell}$, contour level 0.4σ, (*c*) phase set $\{\varphi_{\rm conn12}\}$, resolution 15 Å, $0.15 \times V_{\rm cell}$, contour level 0.8σ.

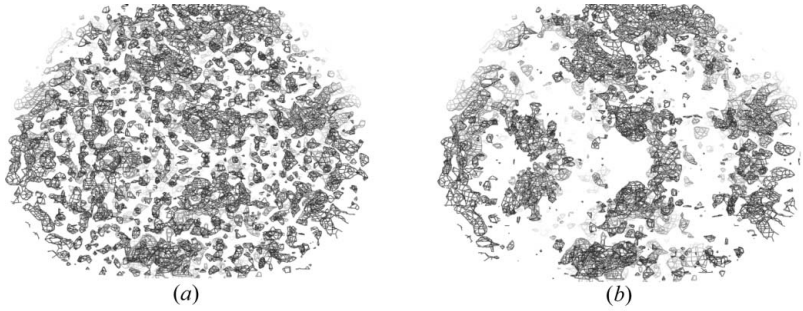


Figure 6
MIRAS-phased syntheses at 4.2 Å resolution; the volume of the selected region was $0.15 \times V_{\rm cell}$. The same view is shown as in Figs. 3 and 5. (*a*) Experimental phases only, no phase information below 18 Å, contour level 0.8σ. (*b*) Experimental phases at 4.2–17.0 Å and phases from the *ab initio* set $\{\varphi_{\rm conn12}\}$ at 17.0–122 Å, contour level 0.9σ.

In summary, this study suggests that filamentous molecules are likely to be challenging targets for low-resolution *ab initio* methodologies, partly because of the lack of voluminous domains that are well resolved in space and of simple identification and partly because of the peculiarities of their atypical crystal lattices (Fig. 1*b*). This highlights the need to implement scoring parameters that incorporate lattice considerations into these calculations. However, even if the position of individual molecules within the unit cell may not be determined using this methodology, it can still lead to

the successful determination of a lattice mask that can aid significantly in the structure-elucidation process of these molecules by improving the initial map quality.

APPENDIX *A*
Main features of low-resolution *ab initio* phasing

A1. *FAM* and connectivity-based *ab initio* phasing methods

The *FAM* protocol (Lunin *et al.*, 1995, 1998) generates multiple random models each composed of a small number of large spherical scatterers ('few-dummy-atoms models'). For each model, structure factors are calculated and their magnitudes are compared with the experimental values. Typically, several hundred sets exhibiting high correlation are selected. The corresponding phase values are then either averaged over all the sets or processed using clustering techniques (Lunin *et al.*, 1990). This procedure gives one or a few sets of phase values and corresponding figures of merit for each reflection. These are then associated with the experimental structure-factor magnitudes to compute a Fourier synthesis.

The protocol of 'image-connectivity' applies phase selection based on the connectivity features of high-density regions in Fourier syntheses. These syntheses are calculated using experimental structure-factor magnitudes and randomly generated phase sets (Lunin *et al.*, 2000). Here again one or a few possible phase sets are obtained per reflection by averaging selected phases. The two methods, which are complementary to a certain extent, have been shown to be capable of phasing several hundred of the lowest resolution structure factors and have been applied efficiently in combination to the *ab initio* phasing of large macromolecular complexes, such as, for example, LDL (Lunin *et al.*, 2001).

A2. Tools for *a posteriori* map scoring and analysis

The calculation of a conventional correlation coefficient is unsuitable for the comparative assessment of low-resolution electron-density maps. This is because the main goal of low-resolution phasing is to reproduce the molecular envelope or shape of the target protein and a parameter that merely quantitates the difference in low and high density values is irrelevant to molecular-shape definition. Thus, to evaluate the results of this study, we used the similarity of the envelopes as a function of their volume. For a given Fourier synthesis, an envelope is defined as a set of grid points whose value is superior to a given cutoff level. Such a level is usually quoted in different units according to the map resolution. Studies at subatomic resolution use the absolute scale of e Å^{-3}, while resolutions common in conventional macromolecular crystallography (1–4 Å) traditionally define cutoff levels in σ values. Neither of these units is convenient for low-resolution studies, where the sharpness of the maps varies strongly from one synthesis to another, especially when varying the resolution or weighting function. As a consequence, the same σ level may produce envelopes with significantly different volume that are difficult to compare directly. To make the comparison adequate, we chose cutoff levels such that a given volume fraction of the unit cell $V_{\rm cell}$ was selected consistently across tests.

A3. General features of low-resolution images

The experience derived from conventional crystallographic studies at resolutions near 1–4 Å cannot be applied to the evaluation of low-resolution Fourier syntheses in a straightforward manner. The previous section illustrates the necessity of switching from σ units to fractional volumes, which indirectly reflects the fact that low-resolution maps contain a much weaker signal. It should be considered that low-resolution envelopes cover the molecular chains as well as the space between them, making the volume of such envelopes larger than the sum of strictly atomic volumes. While Fourier maps at usual resolutions are often displayed at the level of 1–2σ, this would produce unreasonably small envelopes in low-resolution phasing. Another important feature of low-resolution envelopes is that they cannot represent sharp molecular features well and thus cannot cover all macromolecular atoms (unless the fractional volume approaches 1) even when the synthesis is calculated with exact structure-factor values.

There is a common belief that low-resolution Fourier syntheses represent molecular envelopes when the cutoff level is relatively low and the centres of the molecules when this level is high. However, density peaks are often shifted from molecular centres toward regions of close intermolecular contacts. Moreover, owing to the relatively small number of reflections used in the calculation, a small change in phases can significantly modify the image and essentially the position of the peaks. It should also be borne in mind that bulk-solvent correction decreases the contrast between two neighbouring peaks and can result in their merging. Together with the difficulty of establishing reliable and unbiased assessment parameters, the shifting of features complicates the evaluation of results enormously.

Finally, it should be noted that an increase in the resolution of the Fourier maps is rarely able to resolve the arising conflicts. On the contrary, an increase in resolution from 20–25 to 10–12 Å often makes maps even less suitable for visual inspection since they stop showing molecular-envelope features but are not yet capable of revealing secondary-structure elements.

A4. Phase-ambiguity problems and their solution

There are several possible transformations of an electron density $\rho(\mathbf{r})$ that do not change the structure-factor magnitudes but that do affect the calculated phases and images. During structure elucidation, choices of corresponding parameters have to be made *a priori* and arbitrarily, although this does not preclude the phasing process. One of these transformations refers to the shift $\mathbf{t}$ of the origin permitted for a given space group (Lunin & Lunina, 1996), which modifies electron density according to $\rho(\mathbf{r}) \rightarrow \rho(\mathbf{r} - \mathbf{t})$. This does not change the image itself, but should be taken into account during map alignment in comparative processes. In space

groups $P6_522/P6_122$ the only permitted shift is $\mathbf{t} = \frac{1}{2}\mathbf{c}$, where $\mathbf{c}$ is the corresponding base vector of the unit cell.

The enantiomer density transformation $\rho(\mathbf{r}) \rightarrow \rho(-\mathbf{r})$ conserves the connectivity features of the Fourier maps and the magnitudes of structure factors and substitutes the phases φ by $-\varphi$. In the current study, since $P6_522$ and $P6_122$ are enantiomeric space groups, this means that phasing cannot identify the true space group. It is not possible to resolve this ambiguity at low resolution and high-resolution images are required to reveal the 'hand' through the features of the secondary-structure elements. Thus, for this *ab initio* study we arbitrarily chose space group $P6_522$. This was shown at a later stage to be the correct symmetry. Although this facilitated *a posteriori* comparison with the model, it did not influence the phasing process itself.

Finally, a density transformation $\rho(\mathbf{r}) \rightarrow -\rho(\mathbf{r})$ keeps the magnitudes, changes the phases by π and changes the envelopes for a given cutoff level. Again, in contrast to Fourier maps at conventional resolutions, the overall features of the flipped map at ultralow resolution are very similar to those of the direct map. This might make it difficult to visually identify the correct sign of the image. In order to assess which of the two molecular envelopes is more probable, $\rho(\mathbf{r})$ or $-\rho(\mathbf{r})$, a generalized maximum-likelihood technique can be used (Lunin *et al.*, 1998; Petrova *et al.*, 2000). If two (or several) possible molecular envelopes are known, this technique randomly generates multiple models consisting of the same number of dummy atoms inside each of the envelopes. These models are composed of a relatively large number of identical scatterers of a size (determined by the *B* value) much smaller than that of spheres used for *FAM* phasing. For each model, the correlation of its structure-factor magnitudes with the experimental data is calculated. The envelope for which the number of generated models with high correlation is largest is considered to be the most likely to be correct. Since phasing at low resolution requires a complete set of data at a given resolution, such a control set now includes reflections of a resolution slightly higher than that used for phasing. Thus, the correlation is calculated for structure factors excluded from previous calculations.

Our gratitude goes to Marco Marino for providing crystal samples for this study and to Natalia Lunina and Vladimir Y. Lunin for access to *ab initio* phasing programs. Special thanks are also given to Clemens Schulze-Briese (SLS) for his assistance during data collection. We acknowledge support by SNF 3100A0-112595.

References

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.
Castelmur, E. von, Marino, M., Svergun, D. I., Kreplak, L., Labeit, D., Ucurum-Fotiadis, Z., Konarev, P. V., Urzhumtsev, A., Labeit, S. & Mayans, O. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 1186–1191.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.
Fokine, A., Morales, R., Contreras-Martel, C., Carpentier, P., Renault, F., Rochu, D. & Chabriere, E. (2003). *Acta Cryst.* D**59**, 2083–2087.
Fokine, A. & Urzhumtsev, A. (2002). *Acta Cryst.* D**58**, 1387–1392.
Granzier, H. L. & Labeit, S. (2004). *Circ. Res.* **94**, 284–295.
Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Lunin, V. Yu. & Lunina, N. L. (1996). *Acta Cryst.* A**52**, 365–368.
Lunin, V. Y., Lunina, N. L., Petrova, T. E., Urzhumtsev, A. G. & Podjarny, A. D. (1998). *Acta Cryst.* D**54**, 726–734.
Lunin, V. Y., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A. G. & Podjarny, A. D. (1995). *Acta Cryst.* D**51**, 896–903.
Lunin, V. Y., Lunina, N. L., Ritter, S., Frey, I., Berg, A., Diederichs, K., Podjarny, A. D., Urzhumtsev, A. & Baumstark, M. W. (2001). *Acta Cryst.* D**57**, 108–121.
Lunin, V. Y., Lunina, N. L. & Urzhumtsev, A. G. (2000). *Acta Cryst.* A**56**, 375–382.
Lunin, V. Yu., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* A**46**, 540–544.
Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
Marino, M., Svergun, D. I., Kreplak, L., Konarev, P. V., Maco, B., Labeit, D. & Mayans, O. (2005). *J. Muscle Res. Cell Motil.* **26**, 355–365.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Müller, J. J., Lunina, N. L., Urzhumtsev, A., Weckert, E., Heinemann, U. & Lunin, V. Y. (2006). *Acta Cryst.* D**62**, 533–540.
Petrova, T. E., Lunin, V. Y. & Podjarny, A. D. (2000). *Acta Cryst.* D**56**, 1245–1252.
Podjarny, A. D., Schevitz, R. W. & Sigler, P. B. (1981). *Acta Cryst.* A**37**, 662–668.
Strelkov, S. V., Schumacher, J., Burkhard, P., Aebi, U. & Herrmann, H. (2004). *J. Mol. Biol.* **343**, 1067–1080.
Strelkov, S. V., Tao, Y., Rossmann, M. G., Kurochkina, L. P., Shneider, M. M. & Mesyanzhinov, V. V. (1996). *Virology*, **219**, 190–194.
Urzhumtsev, A. G. (1991). *Acta Cryst.* A**47**, 794–801.
Urzhumtsev, A. & Urzhumtseva, L. (2002). *J. Appl. Cryst.* **35**, 750.

3.3 Secondary and Tertiary Structure Elasticity of I65-I70

The crystal structure of I65-I70 enables us for the first time to apply MD simulations to a poly-domain construct directly relevant to I-band elasticity. Specifically, we were interested to explore the secondary and tertiary structure elasticity as coined by Lee and coworkers (Lee et al., 2007) in their study of Z1Z2 dynamics (see chapter 3.4). Analogous to the work presented in that paper we were interested in characterizing the mechanical properties of I65-I70 under weak stretching forces, *i.e.* without unfolding of the Ig domains (which is unlikely to happen under physiological conditions). Of particular interest were the characterization of hinge and torsion motion for the domain pairs and the calculation of the change in free energy this involved with regard to our proposed model of elasticity. There, the long linkers act as flexors along the chain during elongation, possibly by rotating around a bond analogous to residue A99 in Z1Z2, while the straight sections come into play when the chain is stretched, as damping of their libration (*i.e.* internal chain dynamics) would oppose further extension of the chain. Thus we would expect the longer linkers connecting I65-I66-I67 to be more permissive to hinge opening and closing as well as torsion, whereas I67-I68-I69-I70 would be expected to show steric hindrance to hinge motions, due to their tight connections (zero residue linkers). Ideally, hinge and torsion motions would have been probed concomitantly, since they are not independent and most likely interplay, but for calculation of potential mean force, the system had to be decomposed into the individual motions for the domain pairs.

I65-I70 was also subjected to SMD simulation probing the extension of the molecule, in a first step without unfolding of domains, i.e. probing the tertiary structure elasticity stemming from domain rearrangements during straightening. This was followed by complete unfolding for the characterization of the secondary structure elasticity.

MD simulation of the relaxation of I65-I70 after full extension (without unfolding) would have been of particular interest to see if the molecules contracted to the starting position again, but such simulation calculations are at the moment unfeasible since they lie beyond the time frames currently accessible by MD.

Personal contribution to the work: Experimental design and critical evaluation of intermediary and final calculations and results.

Manuscript submitted to Biophys J.

Tertiary and Secondary Structure Elasticity of a Six-Ig Titin Chain

Eric H. Lee[*†‡], Jen Hsin[§‡], Eleonore von Castelmur[¶‖], Olga Mayans[¶‖] and Klaus Schulten[*§**]

September 19, 2009

[*] Center for Biophysics and Computational Biology and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, USA.

[†] College of Medicine, University of Illinois at Urbana-Champaign, Urbana, USA.

[§] Department of Physics and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, USA.

[¶] Division of Structural Biology, Biozentrum, Klingelbergstrasse 70, University of Basel, CH-4056 Basel, Switzerland

[‖] School of Biological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK

[‡]These authors contributed equally to this work.

[**] To whom correspondence should be addressed. Email: kschulte@ks.uiuc.edu. Phone: 217-244-1604. Fax: 217-244-6078.

Summary

The protein titin functions as a mechanical spring conferring passive elasticity to muscle. Force spectroscopy studies have shown that titin exhibits several regimes of elasticity. Disordered segments bring about a soft, entropic spring-type elasticity; secondary structures of titin's immunoglobulin-like (Ig-) and fibronectin type III-like (FN-III) domains provide a stiff elasticity. In this study we demonstrate a third type of elasticity due to tertiary structure and involving domain-domain interaction and reorganization along the titin chain. Through altogether 870 ns of molecular dynamics (MD) simulations involving 29,000 – 635,000 atom systems, the mechanical properties of a six-Ig domain of titin (I65-I70), for which a crystallographic structure is available, are probed. The results reveal a soft tertiary structure elasticity. A remarkably accurate statistical mechanical description for this elasticity is derived and applied. Simulations studied also the stiff, secondary structure elasticity of the I65-I70 chain due to the unraveling of its domains and revealed how force propagates along the chain during the secondary structure elasticity response.

Introduction

Mechanical proteins confer structural support and mechanical compliance to biological cells and tissues. A key example where such support and compliance are needed is muscular contraction. The protein titin provides the passive elasticity required to restore muscle to its resting length after contraction. Titin is the largest known protein in nature (1); its human variant made of 34,350 amino acids is a long linear chain composed of ∼300 modular domains, 90% of which are immunoglobulin-like (Ig-) or fibronectin-type III-like (FN-III) folds (1; 2), the remainder is composed of flexible segments such as N2B and PEVK (rich in proline, glutamate, valine, and lysine) regions, and a catalytic kinase domain (3). Through its elasticity, titin protects muscle fibers also from mechanical injury (4; 5).

The current understanding of titin's mechanical properties arose from single-molecule force spectroscopy investigations of isolated, native titin as well as recombinant fragments (6–13). A picture began to form from these experiments describing how titin reacts to mechanical stretching forces. Upon stretching, titin's chain of domains first straightens without unfolding; this is followed by elongation of disordered segments; finally, at strong forces, the secondary structure of titin's Ig- and FN-III domains unravels, a process referred to as rupture even though the covalent bonds along the domain's back-bone remain intact. Thus, in addition to the entropic elasticity conferred by the protein's disordered domains, titin's mechanical elasticity can be further classified, as illustrated in Figure 1A and B, into two distinct regimes: tertiary structure elasticity (14) due to domain-domain straightening, and secondary structure elasticity (15; 16) due to the unraveling of the domains.

Titin's secondary structure elasticity has been extensively characterized through molecular dynamics (MD) and steered molecular dynamics (SMD) simulations that cast light on how the terminal β-strands of its Ig- and FN-III domains stabilize the protein against rupture (17–22). The molecular basis for titin's tertiary structure elasticity, though, is less well characterized, in part due to the lack of atomic resolution structures for multidomain titin constructs, and also due to the high computational cost associated with performing simulations on large systems involving multiple protein domains. So far, simulations were employed to study the tertiary structure elasticity associated with the tandem Z1 and Z2 domains of titin (14). Tertiary structure elasticity was also probed via simulations for relatively short mechanical repeat proteins such as ankyrin (23) and cadherin (24), the architecture of which differs from that of titin domains. In the case of titin's Z1 and Z2 tandem, the domains exist in their in vivo context anchored at the sarcomeric Z-disc to telethonin (25; 26) such that it is doubtful if the elastic properties of the Z1Z2 tandem are representative of other functional regions of the titin chain.

The recent availability of the crystal structure of a six-Ig fragment I65-I70 (27) from the I-

band of titin offers an opportunity to computationally study, at atomic resolution, the overall flexibility of titin. Inspection of crystal structure and sequence shows that the linkers between the connected Ig-domains involve long (3 residue length) and short (0 residue length) links that alternate according to a conserved pattern (27). A schematic for titin I65-I70 (referred to in the following as "Ig6") is shown in Figure 1C. The compact linear arrangements of the Ig-domains connected by short linkers (I67-68, I68-69, and I69-70) can be recognized in the surface representation of Ig6 in Figure 1D. Differences in linker lengths have also been observed previously in EM micrographs of titin (27). It remains to be understood, however, how these domain-domain interactions control the overall tertiary structure elasticity of titin.

In the present study, we employ MD simulations and statistical mechanical theory to characterize tertiary and secondary structure elasticity of titin Ig6. The free-energy changes associated with bending and twisting motions involving domain pairs were calculated using the adaptive biasing force (ABF) method (28–30). The resulting potentials of mean force were cast into a mathematical description for the force-extension curve characterizing multidomain tertiary structure elasticity. SMD (31) simulations were then carried out to probe the secondary structure elasticity of the entire Ig6 chain by stretching it until all six domains became unfolded. The simulations and theoretical calculations present evidence that the tertiary structure elasticity associated with flexible I-band tandem Ig-domains comprises a soft entropic-like energy barrier to structural deformation, resembling in this respect the elasticity contributed by titin's disordered domains. The simulations also resolves the Ig6 secondary structure elasticity at the atomic level and shows how tension among the connected Ig-domains is relieved every time a single domain unfolds.

Methods

Here we describe the molecular models and methods employed in our simulations as well as the statistical mechanical framework for titin's multidomain tertiary structure elasticity. Further details are found in Supporting Material where the question of timescale adequacy of simulations is discussed (see also (32)).

Simulated systems. Atomic coordinates of titin Ig-domains I65-70 were taken from the crystal structure (PDB entry code 3B43) (27). The Ig6 structure contains six Ig-domains connected in series. The simulation topology of the protein and missing hydrogen atoms were generated using psfgen (33) with the CHARMM27 topology file (34). All simulation systems were solvated using the VMD (35) plugin Solvate in a periodic box of explicit water, large enough to accommodate the protein and the space required for stretching Ig6 out

completely. The long axis of the protein was aligned along the y-axis of the water box. Neutralizing counter ions were added using the VMD plugin Autoionize to reach an ionic concentration of 150 mM NaCl. Acidic and basic residues such as Asp, Glu, Lys, and Arg were modeled in their charged form. The protonation state of His residues was chosen to favor hydrogen bonding where possible. Disulfide bonds in the case of simulations *simEQ-str2* and *simSTR2*, defined in Table 1, were modeled for cysteine pairs CYS23:CYS74 (I65), CYS119:CYS170 (I66), CYS215:CYS277 (I67), and CYS402:CYS464 (I69).

Nine systems were investigated. The first four systems involved Ig6 in water boxes of different sizes: plain equilibration (*simEQ*), extending Ig6 without unfolding its domains (*simEQ-ext* and *simEXT*), fully unfolding all Ig6 domains without disulfide bonds (*simEQ-str1* and *simSTR1*), and fully unfolding Ig6 with disulfide bonds (*simEQ-str2* and *simSTR2*). The final five systems model the individual Ig-pairs from Ig6. The systems are listed in Table 1 and discussed further in Supporting Material. Altogether, 870 ns of simulations were carried out on systems involving 29,000 – 635,000 atoms.

Molecular dynamics simulations. All MD simulations were performed using NAMD 2.6 (33) and the CHARMM27 (34) force field with the CMAP correction (36; 37) for the ϕ-ψ angle backbone dynamics and the TIP3P (38) model for water molecules. Van der Waals interactions cutoff distances were set at 12 Å (the smooth switching function beginning at 10 Å) and long-range electrostatic forces were computed using the particle-mesh Ewald (PME) summation method with a grid size of less than 1 Å, along with the pencil decomposition protocol where applicable. For equilibrium simulations, constant temperature ($T = 300$ K) was enforced using Langevin dynamics with a damping coefficient of 1 ps^{-1}. In both equilibrium and SMD simulations constant pressure (p = 1 atm) was enforced through the Nosé-Hoover Langevin piston method with a decay period of 100 fs and a damping time constant of 50 fs. Equilibrium and adaptive biasing force (ABF) simulations were performed using a multiple time-stepping algorithm (39; 40) which computed covalent bonds every time step, short-range non-bonded interactions every second timestep, and long range electrostatic forces every fourth timestep (1-2-4 stepping). For SMD simulations, 1-1-1 stepping was employed in order to minimize energy drift (32). Equilibrium and SMD simulations employed an integration timestep of 1 fs, while ABF simulations employed 2 fs.

SMD simulations (31; 41–43) fixed the α-carbon at the N-terminus of I65 and applied a force to the α-carbon at the C-terminus of I70. The constant velocity stretching protocol was employed, with stretching velocities of 10 Å/ns in *simEXT* and 25 Å/ns for *simSTR1* and *simSTR2*. The force was applied along the y-axis. Constant force SMD simulations applied a time-independent potential of $V = kd$ to the specified atom(s), where d is the

C_α-C_α distance between the two termini. For the SMD spring constant (44; 45), we chose $k_{SMD} = 3k_B T/\text{Å}^2$ which corresponds to an RMSD value of $\sqrt{k_B T/k_{SMD}} \approx 0.6\,\text{Å}$.

The ABF method (28; 29), adapted into NAMD (derived for the NVT ensemble) (30), was employed to calculate the reversible work, or potential of mean force (PMF), for domain-domain hinge bending and hinge twisting motions. The ABF method uses a continuously updated estimate of a free energy profile along an a priori selected reaction coordinate to apply a bias that overcomes energetic barriers. For each Ig-pair, the reaction coordinate was the separation of two centers of mass located at the opposing tips of the two Ig-domains corresponding to the N-terminus of one domain and to the C-terminus of the other. Further details are provided in Supporting Material.

Theory of multidomain tertiary structure elasticity. The "spring-like" behavior of titin's tertiary structure elasticity, which arises before unfolding of secondary structure occurs, comes about from multiple protein domains connected through linkers. Our simulations measured the PMF to "open" the "hinges" (connections through the linkers) between adjacent domain pairs. With the PMF, one can describe qualitatively the tertiary structure-based elastic behavior of titin I65-70.

For this purpose, we used the "multidomain chain model" in Lee et al. (14). A (planar) multidomain chain is depicted in Figure 2A, in which, as an example, six domains are connected into a chain, and the overall length of the chain is determined by the five hinge angles (θ_{AB} to θ_{EF}). Our ABF calculations treat each Ig pair as an individual unit, uncoupled from its neighbors. For this reason the present model takes the schematic form shown in Figure 2B, where a simplified chain is made of connected independent linkers. This description is assumed actually only to describe the relative extension of Ig6. Twisting and other degrees of freedom are not considered at this point, but are incorporated into the model in Results.

As shown in Figure 2B, a multidomain Ig chain is made of connected domain pairs, each pair j described by a hinge-opening potential function $\tilde{V}_j(\theta_j)$ determined by the ABF method. The angle dependence of $\tilde{V}_j(\theta_j)$ is first replaced by a length dependence via the geometric relation

$$x_j(\theta_j) \;=\; \ell \sin(\theta_j/2), \tag{1}$$

where x_j is the end-to-end distance of the domain pair, θ_j is the hinge angle, and ℓ is the length of the domain pair when it is fully opened (i.e., for $\theta_j = 180°$). In the following, ℓ is set to 90 Å, approximately the end-to-end distance of an open Ig pair. The inverse of Eq. 1 reads $\theta_j = \theta_j(x_j)$.

Given each PMF, $V_j(x_j) = \tilde{V}_j(\theta_j(x_j))$, the length distribution, $p_j(x_j)$, can be computed

via the Boltzmann relation

$$p_j(x_j) \;=\; Z_j^{-1} \exp[-V_j(x_j)/k_BT], \tag{2}$$

where $Z_j \;=\; \int_{-\infty}^{+\infty} \exp(-V_j(x_j)/k_BT)dx_j$ is the partition function.

Since the overall length of the connected chain, X, is the sum of the length of N linker pairs (i.e. X = $\sum_{j=1}^{N} x_j$, illustrated in Figure 2B), the overall length distribution of the multidomain chain is

$$P(X) \;=\; \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \ldots \int_{-\infty}^{+\infty} dx_N \; p_1(x_1)p_2(x_2)\ldots p_N(x_N) \;\; \delta\left(\sum_{j=1}^{N} x_j - X\right), \tag{3}$$

which can be expressed (14)

$$P(X) \;=\; (2\pi)^{(N-2)/2} \int_{-\infty}^{+\infty} \exp[-ikX]dk \;\; \prod_{j=1}^{N} \hat{p}_j(k)\;, \tag{4}$$

where $\hat{p}_j(k)$ is the Fourier transform of $p_j(x_j)$, namely,

$$\tilde{p}_j(k) \;=\; (2\pi)^{-1/2} \int_{-\infty}^{+\infty} p_j(x_j) \exp[ikx_j]dx_j\;. \tag{5}$$

To compute $P(X)$ using Eq. 4, $p_j(x_j)$ needs to be extracted first from ABF data. For this purpose, the ABF data are fitted to a simple mathematical expression for $p_j(x_j)$, such that taking the Fourier transform of $p_j(x_j)$ (Eq. 5) and the subsequent integration (Eq. 4) are feasible. We choose to employ a sum of two Gaussians, non-vanishing only for $x_{j,min} \leq x_j \leq x_{j,max}$, namely

$$p_j(x_j) \;=\; a_1 \exp[-c_1(x_j - b_1)^2] + a_2 \exp[-c_2(x_j - b_2)^2], \tag{6}$$

with fitting parameters a_1, b_1, c_1, a_2, b_2, and c_2, and $x_{j,min}$, $x_{j,max}$ determined through ABF data. This function captures the ABF data quite well as demonstrated in Results.

The central limit theorem (46) states that for large N, $P(X)$ assumes the form of a Gaussian with average $\bar{X}$ and mean square deviation Σ^2, i.e.,

$$P(X) = (2\pi\Sigma^2)^{-1/2} \exp[-(X - \bar{X})^2/2\Sigma^2]. \tag{7}$$

Here $\bar{X}$ is the sum of the averages of x_j ($\bar{X} = \sum_{j=1}^{N} \bar{x}_j$) and Σ^2 the sum of the mean square

deviations of $p_j(x_j)$ ($\Sigma^2 = \sum_{j=1}^{N} \sigma_j^2$).

So far we have considered only the length distribution of the multidomain chain when there is no external force. However, the property of interest is how the length of the multidomain chain changes when a (constant) force f is applied. In this case, the potential for the bending motion of each hinge j is

$$\hat{V}_j(x_j) \;=\; V_j(x_j) - f x_j. \tag{8}$$

As a consequence, the length distribution of each domain pair, $p_j(x_j)$, becomes

$$\hat{p}_j(x_j) \;=\; \exp[-\hat{V}_j(x_j)/k_BT] / \int_{-\infty}^{+\infty} \exp[-\hat{V}_j(x_j)/k_BT] dx_j. \tag{9}$$

$\hat{p}_j(x_j)$ allows one to determine the average domain pair length $\langle x_j \rangle_{\hat{V}_j}$ (the subscript denotes that the average is performed over all configurations weighted by the Boltzmann factor corresponding to $\hat{V}_j$)

$$\langle x_j \rangle_{\hat{V}_j} \;=\; \int_{-\infty}^{+\infty} x_j \hat{p}_j(x_j) dx_j \;=\; \int_{-\infty}^{+\infty} x_j \exp[-\hat{V}_j(x_j)/k_BT] dx_j / \int_{-\infty}^{+\infty} \exp[-\hat{V}_j(x_j)/k_BT] dx_j. \tag{10}$$

The integral in the numerator can be expressed, using Eq. 8,

$$\int_{-\infty}^{+\infty} x_j \exp[-\hat{V}_j(x_j)/k_BT] dx_j \;=\; Z_j \langle x_j \exp[f x_j/k_BT] \rangle_{V_j}. \tag{11}$$

where Z_j is defined in Eq. 2 and where we define the average $\langle f(x) \rangle_{V_j} = \int_{-\infty}^{+\infty} f(x) \exp[-V_j(x_j)/k_BT] dx_j$. One can similarly rewrite the integral in the denominator of Eq. 10, which leads to

$$\langle x_j \rangle_{\hat{V}_j} \;=\; \langle x_j \exp[f x_j/k_BT] \rangle_{V_j} / \langle \exp[f x_j/k_BT] \rangle_{V_j}. \tag{12}$$

The average overall end-to-end distance, $\langle X \rangle = \sum_{j=1}^{N} \langle x_j \rangle_{\hat{V}_j}$, can then be written

$$\langle X \rangle \;=\; \sum_{j=1}^{N} \left[\langle x_j \exp[f x_j/k_BT] \rangle_{V_j} / \langle \exp[f x_j/k_BT] \rangle_{V_j} \right] \equiv g(f). \tag{13}$$

From this one obtains the force-extension curve $f = g^{-1}(\langle X \rangle)$, which is well-defined since g is a monotonic function of f as shown in Supporting Material.

In case that the applied force, f, is small, Taylor expansion of the exponential terms in

Eq. 13 yields

$$\langle X \rangle \approx \sum_{j=1}^{N} \left[\langle x_j \rangle_{V_j} + (f/k_B T)\sigma_j^2 \right] \;=\; \bar{X} + (f/k_B T) \sum_{j=1}^{N} \sigma_j^2, \tag{14}$$

where σ_j^2 is the mean square deviation of the length distribution $p_j(x_j)$ and $\bar{X}$ is defined as $\sum_{j=1}^{N} \langle x_j \rangle_{V_j}$. It is then obvious that the chain behaves as a spring of resting length $\bar{X}$ and overall spring constant

$$k_{chain} \;=\; k_B T / \sum_{j=1}^{N} \sigma_j^2. \tag{15}$$

This behavior is the one that also characterizes the statistical mechanics of the potential in Eq. 7. Eqs. 14 and 15 hold only for small forces, i.e. $fx \ll k_B T$; in general one needs to use Eq. 13.

Results

The following results are based on the simulations of titin I65-I70 (titin "Ig6" for short) summarized in Table 1. The individual simulations will be referred to by the designations listed under the "name" column of Table 1.

Equilibration of titin I65-70 reveals interdomain flexibility. The crystal structure for titin I65-70 was solvated in a water box under physiological ionic conditions and free dynamics were performed for 20 ns in *simEQ* (see Table 1). Analysis of the root mean squared deviation (RMSD) of the protein revealed that the individual domains of titin Ig6 remained stable. The bending and twisting angles between Ig-domains along the crescent shaped chain were observed to fluctuate during relaxation, suggesting that such interdomain motions represent a source of elasticity. Equilibrium simulations alone, however, are not sufficient to quantitatively describe this elasticity. Therefore, we employed a combination of SMD and ABF simulations to fully characterize the underlying energetics of this interdomain-based, so-called tertiary structure elasticity.

Overall tertiary structure elasticity of titin Ig6. In order to assess the soft elasticity arising from extending Ig6 from a crescent-shaped chain into a linear chain, SMD simulations were carried out as described below; further details are provided in Supporting Material. Such simulations (43) had successfully characterized the elasticity of titin

I91 (8; 18), fibronectin (19), ankyrin (23), and cadherin (24). In *simEXT*, the equilibrated Ig6 structure from *simEQ* had its N-terminus α-carbon fixed while a force was applied to the C-terminus α-carbon at a stretching velocity of 10 Å/ns. The direction of stretching was chosen to lengthen the Ig6 chain, and force was applied until the chain was completely straightened, but avoiding the individual domains' secondary structure becoming disrupted. The structural transition is illustrated in Figure 3A. The extension vs. time, $x(t)$, curve is provided in Figure S3 in Supporting Material. $x(t)$ is governed by the Langevin equation in the strong friction limit, which can be written

$$\gamma\,\dot{x} \;=\; f_{\rm chain}(x) - \;k_s\,(x \;- v\,t) \;+\; \sigma\xi(t) \tag{16}$$

where $f_{\rm chain}(x)$ is the force due to the tertiary elasticity of the Ig6 chain, k_s is the SMD spring constant ($k_s = 3k_BT/\text{Å}^2$; see Methods), v is the stretching velocity ($v = 10$ Å/ns), and the last term describes (thermal) Gaussian white noise with RMSD denoted by σ and $\langle\xi(t)\rangle = 0$. According to the fluctuation-dissipation theorem holds $\sigma^2 = 2k_BT\gamma$. As long as $f_{\rm chain}(x)$ is negligible compared to $\gamma\dot{x}$, i.e., for $|f_{\rm chain}(x)| \ll |\gamma\dot{x}|$, one can write for the average extension, $\langle x(t)\rangle$,

$$\gamma\,\langle\dot{x}(t)\rangle \;=\; -\,k_s\,[\langle x(t)\rangle \;- v\,t], \tag{17}$$

the solution of which is

$$\langle x(t)\rangle \;\approx\; vt \;-\; (v\gamma/k_s)[1-\exp(-k_st/\gamma)]. \tag{18}$$

One can recognize that, after $\exp(-k_st/\gamma)$ has decayed to zero, the average extension is $\langle x(t)\rangle \;\sim\; vt \;-\; \Delta x$, where $\Delta x = v\gamma/k_s$ is the extension of the SMD spring. γ can be estimated from the value $D \approx 1.5\times10^{-6}$ cm^2/s of a typical protein diffusion coefficient (47) using $D = k_BT/\gamma$. From this one estimates $\Delta x = 0.2$ Å which, indeed, agrees closely with the simulated extension as shown in Figure S3 in Supporting Material. $F_0 = k_s\Delta x$ is the force that the SMD spring exerts on Ig6. One can readily show $F_0 = v\gamma$, i.e., the force arising in the spring is just the frictional force that resists the tip of Ig6 being dragged with velocity v. Using the expressions above one obtains $F_0 = 28$ pN.

The force-extension curve from our simulation, covering a maximum extension of 100 Å, is shown in Figure 3B. The considerable noise in the force values ($\Sigma_{\rm force} = 91$ pN) seen in Figure 3B can be attributed largely to thermal fluctuations in the SMD spring. Using the known result for the position RMSD of a harmonic spring, $\sigma_{\rm pos} = \sqrt{k_BT/k_s}$, one can estimate the force RMSD through $\sigma_{\rm force} = k_s\,\sigma_{\rm pos} = \sqrt{k_s\,k_BT}$. One finds $\sigma_{\rm force} = 70$ pN, which is 80% of the overall noise value $\Sigma_{\rm force}$ seen in Figure 3B; other degrees of freedom constitute the remainder 20% of the noise. The black trace in Figure 3B shows the average

force value, which is constant during the first half of the simulation period as suggested by the deliberations above and, indeed, matches the estimated value of 28 pN closely. Movie S1, in Supporting Material, illustrates the forced straightening of Ig6, correlating the interdomain rearrangement with the precise point on the force-extension curve in Figure 3B.

So far, the information gained from Figure 3B does not reveal anything about $f_{\text{chain}}(x)$ characterizing the tertiary structure elasticity of Ig6. However, the force trace (black) in Figure 3B exhibits an increase above $\gamma v = 28$ pN beyond 60 Å extension, 60 Å corresponding to the x value for which $f_{\text{chain}}(x)$ begins to rise above the hydrodynamic drag of 28 pN. This motion of Ig6 is then characterized by

$$\gamma \, \langle \dot{x}(t) \rangle \; = \; f_{\text{chain}}[\langle x(t) \rangle] - \; k_s \, (x \; - v \, t), \tag{19}$$

which can be solved numerically. $f_{\text{chain}}[\langle x(t) \rangle]$ is due to bending and twisting motions as well as due to reversible elastic extension of individual Ig domains. These contributions to $f_{\text{chain}}(x)$ will be discussed now.

Local tertiary structure elasticity of titin I65-70. To learn how the hinge-bending and twisting motions contribute to the overall tertiary structure elasticity of the Ig-chain as seen in Figure 3, we employed ABF simulations (28–30) that determined the corresponding PMFs, $V_j(x_j)$. Each Ig-pair of Ig6, j, was described individually as the coupling between the neighboring domain-domain bending and twisting should be small ($V_j(x_j)$ arises from, e.g., the linker made of domains i and $(i+1)$, and is separated by an entire protein domain from the nearest two linkers made of domains $(i-1)$ and i, and $(i+1)$ and $(i+2)$). Consequently, simulations were performed separately on the five pairs of neighboring Ig-domains: I65-66, I66-67, I67-68, I68-69, and I69-70. The ABF simulations listed in Table 1 are named according to the specific Ig-pair and type of motion with A, B, C, D,... corresponding to I65, I66, I67, I68,..., respectively, and "*b*" and "*t*" corresponding to hinge bending and hinge twisting, respectively.

The first set of simulations (*simAB-b* to *simEF-b*) sampled the bending motion (illustrated in Figure 4A) in which the domains bend at the linker towards and away from each other. The simulations reveal how easy or difficult it is to bend the paired Ig-domains, like two adjoining pages of a book, by producing a free energy profile as a function of the bending angle. Figure 4A shows the PMF as a function of bending angle for each of the two-Ig pairs, with the initial conformation observed in the respective equilibrium simulation of two Ig-domains denoted by a green diamond. The PMFs shown in Figure 4A show that the energetic cost of altering the bending angle between domains of Ig6, i.e., flexing them open and closed, is actually quite low, on the order of several k_BT. In the case of bending

the hinge between I67-68 simulated in *simCD-b*, some crowding between the domains due to the short linker did occur as the bending angle was closed. One would expect the same behavior for the bending motions between I68-69 (also with a short linker) simulated in *simDE-b*; however, the equilibrium structure for the I68-69 pair from *simEQ-de* reveals that the domains are offset slightly so that crowding does not pose a significant barrier towards closing this bending angle. The PMFs for all five Ig-pairs show, therefore, soft barriers to domain-domain extension as a result of the low energy cost of opening and closing the individual domain hinges via bending motions.

The second set of ABF simulations (*simAB-t* through *simEF-t*) sampled the twisting motions between adjacent domains, and are illustrated in Figure 4B. These simulations address whether twisting along the Ig6 chain represents a significant source of tertiary structure-based elasticity. Beginning from the structures of each Ig-pair derived from equilibrium simulations (*simEQ-ab* through *simEQ-ef*), one domain was twisted away from the other. Figure 4B depicts the potential of mean force as a function of the twisting angle, α, for the five Ig pairs. The twisting motion PMFs reveal that small angular deviations (approximately $\alpha \leq 40°$) from the equilibrium configurations actually face very little mechanical resistance as a result of domain-domain interactions. However, continued rotation towards larger twisting angles commands a significant energy cost. In the case of the I67-68 pair (*simCD-t*), steric crowding appears to come into play at lower degrees of torsion angle (approximately 37° away from equilibrium structure) compared with the other Ig pairs, the latter not imposing a large penalty to additional twisting until angles between 50-60° are reached. The overall trend of PMFs suggest that there exists heterogeneity along an Ig chain with respect to twisting angles as long as the domains adopt moderate interdomain torsion angles, and that significant elasticity can derive from twisting motion.

Hinge motions in Ig pairs produce an elastic multidomain chain. The free-energy profiles for the hinge bending motions of the tandem Ig-pairs (Figure 4A) permit one to describe the collective elastic behavior of Ig6. In Lee et al. (14), a rudimentary multidomain chain model had been constructed by replicating properties of Z1Z2 hinges into a chain. Here we adopt a similar methodology, described in Methods, with the following modifications: (1) the free-energy profile of each domain pair opening is not assumed to be harmonic, i.e., the length distribution of domain pair j, $p_j(x_j)$, is not necessarily Gaussian, and (2) the free energy of the bending motions of the domain pairs, and consequently their length distributions, do not necessarily have to be uniform, i.e., each $p_j(x_j)$ is different, to reflect the heterogeneity of linker behavior.

Taking the PMF results from Figure 4A ($V_j(x_j)$, where j denotes the five domain pairs

AB, BC, CD, DE, and EF), the individual length distributions $p_j(x_j)$ were computed via Eq. 1 and plotted in Figure 5A. The data were then fitted to a sum of two Gaussians (Eq. 6). As seen in Figure 5A, the $p_j(x_j)$'s are non-identical and non-Gaussian. The mean end-to-end distance of the domain pair j, $\bar{x_j} = \langle x \rangle_{V_j}$, and the root mean square deviation of the length distribution, $\sigma_j = [\langle (x_j - \bar{x}_j)^2 \rangle_{V_j}]^{1/2}$, were computed and are shown also in Figure 5A. The free parameters used for these calculations are listed in Supporting Material in Table S1.

The five bending angles were then connected together to form a hypothetical multidomain chain as depicted in Figure 2B, and the overall length distribution of this chain, $P(X)$, was computed via Eq. 4 and plotted in Figure 5B (black curve). Although each $p_j(x_j)$ is non-Gaussian (Figure 5A), the final $P(X)$ resembles closely a Gaussian distribution (black curve in Figure 5B) as expected from the central limit theorem (Eq. 7, see Methods), with average $\bar{X} = \sum_{j=1}^{5} \langle x_j \rangle_{V_j}$, and overall root mean square deviation Σ given by $\Sigma^2 = \sum_{j=1}^{5} \sigma_j^2$. The Gaussian fit is shown in Figure 5B (gray curve). $\bar{X}$ agrees well with $\langle X \rangle = \int dX P(X) X$ (352.3 Å vs. 358.1 Å), and likewise Σ agrees well with $[\int dX (X - \langle X \rangle)^2 P(X)]^{1/2}$ (14.9 Å vs. 16.8 Å). The close agreement is quite remarkable as the central limit theorem holds strictly only in the limit $N \to \infty$. This result implies that repeat proteins behave overall like harmonic elastic springs in the limit of weak force (see Methods). As numerous other elastic proteins are also made of repeat domains, e.g., ankyrin, cadherin and fibrin (23; 24; 32; 48; 49), this result, i.e., that repeat proteins in obeying the central limit theorem act as Brownian springs, is of general importance, though holds only for small extension.

The relationship between mechanical force and arbitrary chain extension was computed using Eq. 13, which, in terms of the probability distributions $p_j(x_j)$, is

$$\langle X \rangle = \sum_{j=1}^{5} \left[\int_{-\infty}^{+\infty} x_j p_j(x_j) \exp[f x_j / k_B T] dx_j \Big/ \int_{-\infty}^{+\infty} p_j(x_j) \exp[f x_j / k_B T] dx_j \right]. \qquad (20)$$

The chain extension, i.e. $\langle X \rangle$ - $\langle X \rangle_{f=0}$, versus applied force f is plotted in Figure 5C (dashed trace). At low forces ($f <$ 5 pN), the force-extension relation displays the linear behavior (Figure 5C, inset) given by Eqs. 14 and 15, derived in the small force limit, with effective spring constant $k_c = k_B T / \Sigma^2_{soft/stiff}$, $\Sigma^2_{soft/stiff} = \sum_j \sigma_j^2$, where j includes all five domain pairs, both soft and stiff. One can compute the value of k_c, and obtains $k_c \approx 0.005\ k_B T/\text{Å}^2$. When force increases, the chain departs from the linear regime, becoming stiffer, as represented by an increased slope in the force-extension curve, corresponding to opening of the stiffer hinges. At around ~40 Å extension when both soft hinges (BC and DE) are maximally opened (note $\Delta x_{soft,max} = \Delta x_{BC,max} + \Delta x_{DE,max} \approx 40$ Å), the slope of the force-extension

curve increases to $k_BT/\sum_j \sigma_j^2$, where now $j = AB$, CD, and EF, i.e., j counts now mainly the three stiff hinges. To determine the appropriate $f_{\text{chain}}(x)$ one employs again Eq. 20 to obtain $\langle X \rangle = g(f)$. $g(f)$ being a monotonic function of f, i.e., $\partial g/\partial f > 0$ as demonstrated in Supporting Material, one can determine ($x = \langle X \rangle$) $f(x) = g^{-1}(x)$. $f_{\text{chain}}(x)$ was then plugged into Eq. 19 and the applied force as a function of Ig6 extension during the SMD simulation was computed numerically. The result is plotted as a blue trace in Figure 3B.

At most, a ∼70 Å extension ($\Delta x_{soft,max} + \Delta x_{stiff,max} = \sum_{j=1}^{5} \Delta x_{j,max} \approx 70$ Å) can be reached through forces of a few tens of pN, i.e., for extensions deriving purely from domain-domain bending (Figure 5C). Beyond 70 Å extension, the tertiary structure elasticity due to domain-domain bending is exhausted. In order to describe the tertiary structure elasticity over a wider range, i.e., over the interval [0 Å, 100 Å], one should account for all other degrees of freedom that permit stretching of up to 100 Å. An obvious choice is domain twisting neglected so far. In this case the maximum extension is calculated to be $\sum_{j=1}^{10} \Delta_{j,\max} = 93$ Å (see Figure 5C, gray trace; details on the calculation presented in Supporting Material, Figure S1, and Table S2). The Ig6 extension under SMD pulling, including both bending and twisting motions in the chain, was calculated again via Eq. 19 and plotted as a red trace in Figure 3B. It is noteworthy that $f_{\text{chain}}(x)$ accounting for bending and twisting renders the overall chain softer than it does if only bending is accounted for (Figure 5C, inset). Every further degree of freedom accounted for renders a chain overall softer, a stiff degree of freedom adds less so than a soft degree of freedom.

Extension beyond 93 Å involves stretching of individual Ig domains. Significant further extension would lead to rupture of domain secondary structure, but small extension involving reversible (on a ns time scale) domain stretching is permitted without rupturing of secondary structure. To determine the underlying force-extension characteristic of individual domains we determined $U_j(x_j)$ for single domain extension by simply monitoring $x_j(t)$ for the individual domains as described in Supporting Material to obtain the RMSD value of $x_j(t)$ for each domain (Figure S2). The resulting value measured ∼0.6 Å, which corresponds to a force-extension curve for the overall stretching of I65-I70 of $f_0(x) = \sqrt{k_{\text{B}}T/(6 \cdot (0.6\ \text{Å})^2)}(x - x_{\text{eq}})$ where $x_{\text{eq}} = 93$ Å is the equilibrium length of Ig6 after straightening it through domain-domain bending and twisting. Figure 5C shows the force-extension curve for Ig6 determined following the procedure above, adding to domain-domain bending and twisting (black solid curve). As shown by Figure 3B (orange trace), inclusion of the stretching degree of freedom reproduces the simulated force response of the Ig6 chain during SMD pulling very clearly. At extension $x = 100$ Å this force assumes a value of 200 pN which is sufficient to rupture the secondary structure of individual Ig domains, i.e., at this extension the secondary structure elasticity regime of Ig6 sets in.

Secondary structure elasticity of titin I65-70. Whereas low stretching forces straighten the Ig-chain by rearranging its tertiary structure, high stretching forces induce unfolding of the Ig-domain's secondary structure. The secondary structure elasticity regime for titin Ig-domains has been well characterized through force-spectroscopy experiments and computational simulations (6–10; 14; 17; 18; 20; 50; 51). In the case of atomic force microscopy (AFM) experiments, unfolding the long polyprotein chain produces a distinct sawtooth force-extension trace interpreted as the sequential rupture of individual Ig-domains (52). All-atom MD simulations up to this point have been limited by computational resources and by the availability of relevant structures to simulating the unfolding of only single Ig-domains (17; 18; 20; 51). Recent strides in computational efficiency (53) permit us now to carry out SMD simulations to completely extend and unfold titin Ig6 in order to probe the secondary structure elasticity of a natural Ig-chain.

After equilibrating the Ig6 structure (*simEQ-str1*), the N-terminal α-carbon of titin I65 was fixed and a stretching force applied to the α-carbon of the C-terminus of titin I70 in *simSTR1* employing a constant velocity protocol (31) with $v = 25$ Å/ns until all six domains had ruptured and fully extended. The force-extension curve resulting from *simSTR1* is shown in Figure 6A, showing clearly individual force peaks. These peaks, labeled (*ii*) through (*vii*), are correlated with the unraveling of individual Ig-domains, producing a sawtooth-like profile similar to those seen in experiments mentioned above. The subsequent drop in force following each force peak, i.e., the domain unraveling event, indicates a relief in stress along the Ig-chain. A detailed view of the structural dynamics reveals that the domain unraveling in every case is initiated by the separation of the terminal β-strands that are adjacent in each domain forming between them 7 to 9 hydrogen bonds. This strand separation has been described in detail previously (18; 32; 54). As the internal β-strands, less stable than the terminal β-strands, readily unravel, the β-strands of unruptured domains are permitted to relax and stabilize their interstrand hydrogen bonding. Peak (*ii*) corresponds to the rupture of I65, (*iii*) to I70, (*iv*) to I66, (*v*) to I67, (*vi*) to I69, and (*vii*) to I68. Thus, the order of domain unraveling is I65 $\rightarrow$ I70 $\rightarrow$ I66 $\rightarrow$ I67 $\rightarrow$ I69 $\rightarrow$ I68, with the terminal domains rupturing before the ones in the middle of the Ig chain. A schematic of the rupture sequence is provided in Figure S4 in Supporting Material along with a further discussion of rupture events. Figure 6B shows snapshots of the individual domain rupturing events at the timepoints labeled (*i*) through (*vii*). The full unfolding trajectory for *simSTR1* is available as Movies S2 and S3 in Supporting Material.

Additionally, a short simulation (*sim-RF*) exploring the refolding of Ig-domains was performed, starting from partially unfolded I65 and I70 domains. While I70 was observed to refold as it had just crossed the terminal β-strand rupture barrier when the force was

released for relaxation, I65, which was more extended at the beginning of the equilibrium simulation, was not observed to refold on the timescale investigated (see Figure S6 along with further details in Supporting Material).

Simulation *simSTR2* was carried out to stretch Ig6 containing internal disulfide bonds across CYS residues in domains I65, I66, I67, and I69 and to test whether altering the mechanical stability of individual Ig-domains alters the sequence of domain rupture. It turned out that the sequence of rupture remained identical to that observed in *simSTR1* without crosslinked cysteines. Details for these simulations are presented in Supporting Material (Figure S5); the trajectory of the simulation is shown in Movies S5 and S6.

Discussion

The titin I-band, which includes the segment Ig65-Ig70, has been experimentally characterized as an exceptionally flexible region of titin, having elastic properties derived from its multidomain architecture of Ig-like, PEVK, N2B, and other disordered domains (4). The elasticity of titin I-band exhibits three complementary extension regimes (14), a regime of soft entropic elasticity due to the disordered N2B and PEVK regions, a second regime of soft, so-called tertiary structure elasticity due to domain-domain bending and twisting as well as minor domain stretching, and a stiff, so-called secondary structure elasticity due to the rupture of the β-strand structure of individual Ig-domains. The latter two regimes have been the focus of the present theoretical-computational study made possible through the availability of the structure of the I65-I70 segment.

Tertiary structure elasticity of titin I-band segment I65-I70. Tertiary structure elasticity is due to bending and twisting motions involving the hinges between neighboring titin domains. Motions give rise to change, by Δx, in the overall segment length x. Five bending and five twisting degrees of freedom (labelled $j = 1, 2, \ldots 10$) contribute to an increase of x by Δx_j along with weak stretching of each of the six domains (labelled $j = 11, 12, \ldots, 16$). An overall elastic extension of segment I65-I70, $\Delta x = \sum_{j=1}^{16} \Delta x_j$ is assumed to arise in quasi-equilibrium with the segment's remaining degrees of freedom and are governed by additive potentials of mean force $V_j(x_j)$. An external force, f, applied adds a potential $-f\,x_j$ to each contributing degree of freedom and has been shown to lead to the average extension

$$x = \langle X \rangle = \sum_{j=1}^{16} (\langle x_j \exp[f x_j/k_B T]\rangle_{V_j} / \langle \exp[f x_j/k_B T]\rangle_{V_j} = g(f)\,. \qquad (21)$$

$g(f)$ is a fundamental function characterizing the force-extension relationship, $f(x)$, i.e., the elastic behavior, of biopolymers in general as it can be inverted to yield $f(x) = g^{-1}(x)$. This succinct and straightforward description has been shown in the extensive molecular dynamics simulations presented above to be highly accurate, to a degree seldom found in the theory of biopolymers. The analysis outlined reveals that titin segment I65-I70 upon application of forces in the range [0, 200 pN] extends up to 100 Å. The analysis, based on $g(f)$, can be applied to any biopolymer system in the reversible stretching regime; the extension function $g(f)$, accordingly, deserves further investigation.

Stiff secondary structure elasticity. Steered molecular dynamics simulations captured in many previous studies the force-induced unraveling of individual protein domains. Examples include unraveling of individual fibronectin type-III domains (19; 55; 56), spectrin repeats (57), fibrinogen coiled-coils (48), and titin domains such as I1 (54; 58), and I91 (8; 17; 18). In the present case, the size of the simulations was increased to include stretching and unraveling of six connected Ig domains, allowing direct comparison to experimental results, namely the "sawtooth" force-extension trace (52). The ability for SMD simulations to record the sequentially ordered rupture of Ig6 reveals how force propagates between individual domains along the Ig-chain. Specifically, one observes that once a domain ruptures, tension is immediately relieved for all other domains along the chain, illustrating how each Ig-domain function as a "shock absorber" for protecting other domains in the chain from forced rupture.

Putting the pieces of the picture together into an elasticity hierarchy. It is now understood that titin's elastic response to forced stretching stems from the mechanical properties of its many constituent domains. The flexibility of these domains contributes to two distinct regimes of elasticity, a soft regime characterized by the rearrangement of protein tertiary structure and unraveling of disordered segments during protein elongation, and at physiologically extreme forces a regime characterized by the rupture of secondary structure folds of individual domains. Computational studies of single titin domains have contributed to this understanding by revealing that it is the network of hydrogen bonds that span across the terminal β-strands of individual Ig-domains which governs titin's secondary structure elasticity. The MD simulations on titin I65-70 reported here demonstrate additionally how the domain-domain arrangements and motions give rise to tertiary structure elasticity of titin's flexible I-band. Combining these insights along with clues from prior experimental and computational studies, a picture of titin's mechanical properties emerges as a complex molecular spring. Many of nature's other mechanical proteins likely share mechanisms that

employ multiple regimes of elasticity for bearing and transforming forces in cells.

Supporting Material

Five movies (S1-S5), six figures, two tables, and further explanations are available in Supporting Material.

Acknowledgments

The authors thank Christophe Chipot, Ioan Kosztin, Hei-Chi Chan, and Johan Strumpfer for insightful discussions. This work was supported by the National Institutes of Health (NIH P41-RR05969 and R01-GM073655). Computer time was provided through the National Resource Allocation Committee grant (NRAC MCA93S028) from the National Science Foundation. EL was supported in part by the Hazel I. Craig Fellowship and EvC by the Roche Research Foundation.

References

1. Bang, M. L., T. Centner, F. Fornoff, A. J. Geach, M. Gotthardt, M. McNabb, C. C. Witt, D. Labeit, C. C. Gregorio, H. Granzier, and S. Labeit. 2001. The complete gene sequence of titin, expression of an unusual approximate to 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circulation Res.* 89:1065–1072.

2. Labeit, S. and B. Kolmerer. 1995. Titins, giant proteins in charge of muscle ultrastructure and elasticity. *Science*. 270:293–296.

3. Linke, W. A., M. Ivemeyer, P. Mundel, M. R. Stockmeier, and B. Kolmerer. 1998. Nature of PEVK-titin elasticity in skeletal muscle. *Proc. Natl. Acad. Sci. USA*. 95:8052–8057.

4. Tskhovrebova, L. and J. Trinick. 2003. Titin: Properties and family relationships. *Nat. Rev. Mol. Cell Biol.* 4:679–689.

5. Granzier, H. L. and S. Labeit. 2004. The giant protein titin a major player in myocardial mechanics, signaling and disease. *Circulation Res.* 94:284–295.

6. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*. 276:1109–1112.

7. Carrion-Vazquez, M., A. Oberhauser, S. Fowler, P. Marszalek, S. Broedel, J. Clarke, and J. Fernandez. 1999. Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl. Acad. Sci. USA*. 96:3694–3699.

8. Marszalek, P. E., H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. 1999. Mechanical unfolding intermediates in titin modules. *Nature*. 402:100–103.

9. Li, H., W. Linke, A. F. Oberhauser, M. Carrion-Vazquez, J. G. Kerkvliet, H. Lu, P. E. Marszalek, and J. M. Fernandez. 2002. Reverse engineering of the giant muscle protein titin. *Nature*. 418:998–1002.

10. Fowler, S. B., R. B. Best, J. L. T. Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke. 2002. Mechanical unfolding of a titin Ig domain: Structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J. Mol. Biol.* 322:841–849.

11. Watanabe, K., C. Muhle-Gol, M. S. Kellermayer, S. Labeit, and H. Granzier. 2002. Different molecular mechanics displayed by titin's constitutively and diferentially expressed tandem ig segments. *J. Struct. Biol.* 137:248–258.

12. Oberhauser, A. F., P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez. 2001. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. USA*. 98:468–472.

13. Grutzner, A., S. Garcia-Manyes, S. Kotter, C. L. Badilla, J. M. Fernandez, and W. A. Linke. 2009. Modulation of titin-based stiffness by disulfide bonding in the cardiac titin n2-b unique sequence. *Biophys. J.* 97:825–824.

14. Lee, E. H., J. Hsin, O. Mayans, and K. Schulten. 2007. Secondary and tertiary structure elasticity of titin Z1Z2 and a titin chain model. *Biophys. J.* 93:1719–1735.

15. Lee, E. H., M. Gao, N. Pinotsis, M. Wilmanns, and K. Schulten. 2006. Mechanical strength of the titin Z1Z2/telethonin complex. *Structure*. 14:497–509.

16. Linke, W. and A. Grutzner. 2008. Pulling single molecules of titin by AFM - recent advances and physiological implications. *Eur. J. Physiol.* 456:101–115.

17. Lu, H., B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* 75:662–671.

18. Lu, H. and K. Schulten. 2000. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.* 79:51–65.

19. Gao, M., D. Craig, O. Lequin, I. D. Campbell, V. Vogel, and K. Schulten. 2003. Structure and functional significance of mechanically unfolded fibronectin type III1 intermediates. *Proc. Natl. Acad. Sci. USA*. 100:14784–14789.

20. Best, R. B., S. B. Fowler, J. L. T. Herrera, A. Steward, E. Paci, and J. Clarke. 2003. Mechanical unfolding of a titin Ig domain: Structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J. Mol. Biol.* 330:867–877.

21. Cieplak, M., T. Hoang, and M. Robbins. 2004. Thermal effects in stretching of go-like models of titin and secondary structures. *Proteins*. 56:285–297.

22. Keten, S. and M. Buehler. 2008. Strength limit of entropic elasticity in beta-sheet protein domains. *Phys. Rev. E.* 78:1–7.

23. Sotomayor, M., D. P. Corey, and K. Schulten. 2005. In search of the hair-cell gating spring: Elastic properties of ankyrin and cadherin repeats. *Structure.* 13:669–682.

24. Sotomayor, M. and K. Schulten. 2008. The allosteric role of the Ca^{++} switch in adhesion and elasticity of C-cadherin. *Biophys. J.* 94:4621–4633.

25. Valle, G., G. Faulkner, A. DeAntoni, P. Pacchioni, A. Pallavicini, D. Pandolfo, N. Tiso, S. Toppo, S. Trevisan, and G. Lanfranchi. 1997. Telethonin, a novel sarcomeric protein of heart and skeletal muscle. *FEBS Lett.* 415:163–168.

26. Gregorio, C. C., K. Trombitas, T. Centner, B. Kolmerer, G. Stier, K. Kunke, K. Suzuki, F. Obermayr, B. Herrmann, H. Granzier, H. Sorimachi, and S. Labeit. 1998. The NH2 terminus of titin spans the Z-disc: Its interaction with a novel 19-kD ligand (T-cap) is required for sarcomeric integrity. *J. Cell Biol.* 143:1013–1027.

27. von Castelmur, E., M. Marino, D. I. Svergun, L. Kreplak, Z. Ucurum-Fotiadis, P. V. Konarev, A. Urzhumtsev, D. Labeit, S. Labeit, and O. Mayans. 2008. A regular pattern of ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. *Proc. Natl. Acad. Sci. USA.* 105:1186–1191.

28. Darve, E., D. Wilson, and A. Pohorille. 2002. Calculating free energies using a scaled-force molecular dynamics algorithm. *Mol. Sim.* 28:113–144.

29. Rodriguez-Gomez, D., E. Darve, and A. Pohorille. 2004. Assessing the efficiency of free energy calculation methods. *J. Chem. Phys.* 120:3563–3578.

30. Hénin, J. and C. Chipot. 2004. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* 121:2904–2914.

31. Isralewitz, B., M. Gao, and K. Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* 11:224–230.

32. Lee, E. H., J. Hsin, G. Comellas, M. Sotomayor, and K. Schulten. 2009. Discovery through the computational microscope. *Structure.* In press.

33. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comp. Chem.* 26:1781–1802.

34. MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

35. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD – Visual Molecular Dynamics. *J. Mol. Graphics.* 14:33–38.

36. MacKerell, A. D., Jr., M. Feig, and C. L. Brooks III. 2004. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.* 25:1400–1415.

37. Buck, M., S. Bouguet-Bonnet, R. W. Pastor, and A. D. MacKerell. 2006. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophys. J.* 90:L36–38.

38. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

39. Grubmüller, H., H. Heller, A. Windemuth, and K. Schulten. 1991. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.* 6:121–142.

40. Schlick, T., R. Skeel, A. Brünger, L. Kalé, J. A. Board Jr., J. Hermans, and K. Schulten. 1999. Algorithmic challenges in computational molecular biophysics. *J. Comp. Phys.* 151:9–48.

41. Izrailev, S., S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten. 1998. Steered molecular dynamics. In P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, pages 39–65.

42. Gao, M., M. Sotomayor, E. Villa, E. Lee, and K. Schulten. 2006. Molecular mechanisms of cellular mechanics. *Phys. Chem. Chem. Phys.* 8:3692–3706.

43. Sotomayor, M., V. Vasquez, E. Perozo, and K. Schulten. 2007. Ion conduction through MscS as determined by electrophysiology and simulation. *Biophys. J.* 92:886–902.

44. Evans, E. and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. *Biophys. J.* 72:1541–1555.

45. Izrailev, S., S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. 1997. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* 72:1568–1581.

46. Sornette, D. 2006. *Critical Phenomena in Natural Sciences*. Springer, Heidelberg.

47. Park, S. and K. Schulten. 2004. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* 120:5946–5961.

48. Lim, B., E. H. Lee, M. Sotomayor, and K. Schulten. 2008. Molecular basis of fibrin clot elasticity. *Structure*. 16:449–459.

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

50. Williams, P. M., S. B. Fowler, R. B. Best, J. L. Toca-Herrera, K. A. Scott, A. Steward, and J. Clarke. 2003. Hidden complexity in the mechanical properties of titin. *Nature*. 422:446–449.

51. Lu, H. and K. Schulten. 1999. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins: Struct., Func., Gen.* 35:453–463.

52. Sarkar, A., S. Caamano, and J. M. Fernandez. 2007. The mechanical fingerprint of a parallel polyprotein dimer. *Biophys. J.* 92:36–38.

53. Kumar, S., C. Huang, G. Zheng, E. Bohm, A. Bhatele, J. C. Phillips, H. Yu, and L. V. Kale. 2008. Scalable molecular dynamics with NAMD on the IBM Blue Gene/L system. *IBM Journal of Research and Development*. In press.

54. Gao, M., H. Lu, and K. Schulten. 2002. Unfolding of titin domains studied by molecular dynamics simulations. *J. Muscle Res. Cell Mot.* 23:513–521.

55. Craig, D., A. Krammer, K. Schulten, and V. Vogel. 2001. Comparison of the early stages of forced unfolding of fibronectin type III modules. *Proc. Natl. Acad. Sci. USA*. 98:5590–5595.

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

57. Altmann, S. M., R. G. Grünberg, P.-F. Lenne, A. R. Ylänne, K. Herbert, M. Saraste, M. Nilges, and J. K. H. Hörber. 2002. Pathways and intermediates in froced unfolding of spectrin repeats. *Structure*. 10:1085–1096.

58. Gao, M., M. Wilmanns, and K. Schulten. 2002. Steered molecular dynamics studies of titin I1 domain unfolding. *Biophys. J.* 83:3435–3445.

Legends to figures

Figure 1. Schematic representation of the soft tertiary structure and stiff secondary structure elasticity for the six-domain titin segment I65-70 ("Ig6"). (A) depicts how the arrangement of domains in Ig6 manifests itself in tertiary structure elasticity. (B) shows how strong forces lead to the rupture of individual domains, an example of secondary structure elasticity. Shown in (C) is the structure of I65-70 in cartoon representation. The individual domains are color-coded the same way in all subsequent figures. (D) shows the structure of I65-70 in surface representation; the short linkers between I67-68, I68-69, and I69-70 can be recognized clearly.

Figure 2. Schematics of the multidomain chain model. (A) The overall length of a six-domain chain is described by five hinge angles. (B) The present multidomain chain model uses a simplified representation, in which domain pairs are connected together, each domain pair contributing an independent hinge angle not coupled to other domain pairs.

Figure 3. Steered molecular dynamics (SMD) simulations probing the tertiary structure elasticity of titin Ig6. Shown in (A) are snapshots from *simEXT* depicting the extension of Ig6 without unfolding the individual domains. The resulting force-extension curve is shown in (B), with the black trace corresponding to the average force measured in simulation. The force fluctuation can be attributed to the SMD spring as discussed in the text. The quantities f_{ave} and Σ_{force}, discussed in the text, were computed over the extension range of 10-60 Å, i.e., after the initial relaxation and before the increase of the stretching force above 28 pN (value shown as a black dashed line). The blue and red traces were computed by using the multidomain chain model to determine $f_{\text{chain}}(x)$ and solving Eq. 19 as described in the text, with the blue trace computed by considering only the hinge bending motions, and the red

trace considering both the bending and twisting motions. The orange trace describes the last ∼10 Å of Ig6 extension stemming from the intrinsic stretching of the individual domains; the elasticity characterizing this motion was measured from an equilibrium simulation described in the text and in Supporting Material. The agreement between simulation (black) and theoretical description (red and orange) does not involve any fitting parameters.

Figure 4. Adaptive biasing force (ABF) simulations probing the tertiary structure elasticity of titin Ig6. ABF simulations were carried out on each of the connected Ig-domain pairs to probe the energetics of two types of motions: a hinge bending motion (A) in which the domains bend away from each other at the flexible linker, and a hinge twisting motion (B) corresponding to a twisting of the chain. In (A), the potential of mean force (PMF) is shown as a function of the bending angle, with the position observed in respective equilibrium simulations marked by green diamonds. In (B), the PMF is plotted as a function of twisting angle, measured also against the equilibrium position.

Figure 5. Tertiary structure elasticity of a multidomain chain. (A) The PMF resulting from ABF simulations can be related to length distribution for an individual Ig linker pair, represented by open circles. Best-fit expressions $p_j(x_j)$ (Eq. 6) are plotted as solid lines. Also shown in the plots are the mean end-to-end length of each domain pair, $\bar{x}_j$, the root mean square deviation of each $p_j(x_j)$, σ_j, and the approximate maximum extension of each domain pair, $\Delta x_{j,max}$. The hinges are labeled as stiff or soft depending on their σ_j values. (B) The overall length distribution for the chain of five Ig domain pairs, computed using Eq. 4 is shown (black), the Gaussian fit (Eq. 7) being depicted in gray. (C) Shown is the overall extension of the modeled multidomain chain in response to mechanical force. At small forces ($f < 5$ pN), the force-extension curve is linear (inset). The dashed curve is calculated from considering only the bending motion in the multidomain chain; the gray curve includes both bending and twisting motions in the calculation; the black solid curve incorporates also the stretching degree of freedom in each domain.

Figure 6. SMD simulations for full unfolding of titin Ig6. Shown in (A) is a force-extension curve from *sim-STR1*, in which the entire I65-70 Ig-domain tandem was stretched until completely unfolded. Rather than exhibiting simultaneously rupturing across all domains, domains unfold one-by-one, producing a sawtooth pattern in the force extension profile. The force peaks corresponding to ruptures of individual domains are denoted by numerals, corresponding to the close up views of domain rupture shown in (B). The domains rupture in the order I65 → I70 → I66 → I67 → I69 → I68.

Name	Structure	Type	Ensemble	Atoms $\times 1000$	Size Å^3	Special parameters	Time ns
simEQ	I65-70	EQ	NpT	224	$110 \times 339 \times 63$	–	20.0
simEQ-ab	I65-66	EQ	NpT	78	$131 \times 82 \times 77$	–	10.0
simEQ-bc	I66-67	EQ	NpT	72	$138 \times 71 \times 77$	–	10.0
simEQ-cd	I67-68	EQ	NpT	73	$136 \times 74 \times 77$	–	10.0
simEQ-de	I68-69	EQ	NpT	72	$129 \times 81 \times 74$	–	10.0
simEQ-ef	I69-70	EQ	NpT	77	$131 \times 83 \times 75$	–	10.0
simAB-b	I65-66	ABF	NVT	78	$131 \times 82 \times 77$	bending	25.0
simBC-b	I66-67	ABF	NVT	72	$138 \times 71 \times 77$	bending	27.0
simCD-b	I67-68	ABF	NVT	73	$136 \times 74 \times 77$	bending	50.0
simDE-b	I68-69	ABF	NVT	72	$129 \times 81 \times 74$	bending	44.0
simEF-b	I69-70	ABF	NVT	77	$131 \times 83 \times 75$	bending	41.0
simAB-t	I65-66	ABF	NVT	78	$131 \times 82 \times 77$	twisting	43.0
simBC-t	I66-67	ABF	NVT	72	$138 \times 71 \times 77$	twisting	69.0
simCD-t	I67-68	ABF	NVT	73	$136 \times 74 \times 77$	twisting	79.0
simDE-t	I68-69	ABF	NVT	72	$129 \times 81 \times 74$	twisting	125.0
simEF-t	I69-70	ABF	NVT	77	$131 \times 83 \times 75$	twisting	132.0
simEQ-ext	I65-70	EQ	NpT	227	$114 \times 353 \times 72$	[a]	5.0
simEXT	I65-70	SCV	NV	277	$114 \times 353 \times 72$	10Å /ns[b]	10.0
simEQ-str1	I65-70	EQ	NpT	635	$2045 \times 60 \times 53$	[a]	5.0
simSTR1	I65-70	SCV	NV	635	$2045 \times 60 \times 53$	25Å /ns[b]	66.0
simEQ-str2	I65-70	EQ	NpT	425	$1197 \times 64 \times 58$	[a]	5.0
simSTR2	I65-70	SCV	NV	425	$1197 \times 64 \times 58$	25Å /ns[b]	35.0
sim-RF	I65-70	EQ	NpT	225	$430 \times 68 \times 62$		18.6
sim-EQ-I65	I65	EQ	NpT	29	$75 \times 66 \times 62$		18.9

Table 1: Summary of simulations. Under "type", EQ denotes equilibration; ABF denotes adaptive biasing force simulations, and SCV denotes constant velocity SMD simulations. The "ensemble" column lists the variables held constant during the simulations; N, V, p, and T correspond to number of atoms, volume, pressure, and temperature, respectively. Footnotes under special parameters describe the motion sampled in the ABF simulation, and, in the case of SMD simulations, which atoms were fixed and which stretching velocity was employed.

[a] Pre-equilibrated I65-70 from *simEQ* was re-solvated in a large water box to accommodate the SMD simulation.

[b] α-carbon of N-terminus I65 fixed and force applied to the C-terminus α-carbon of I70.

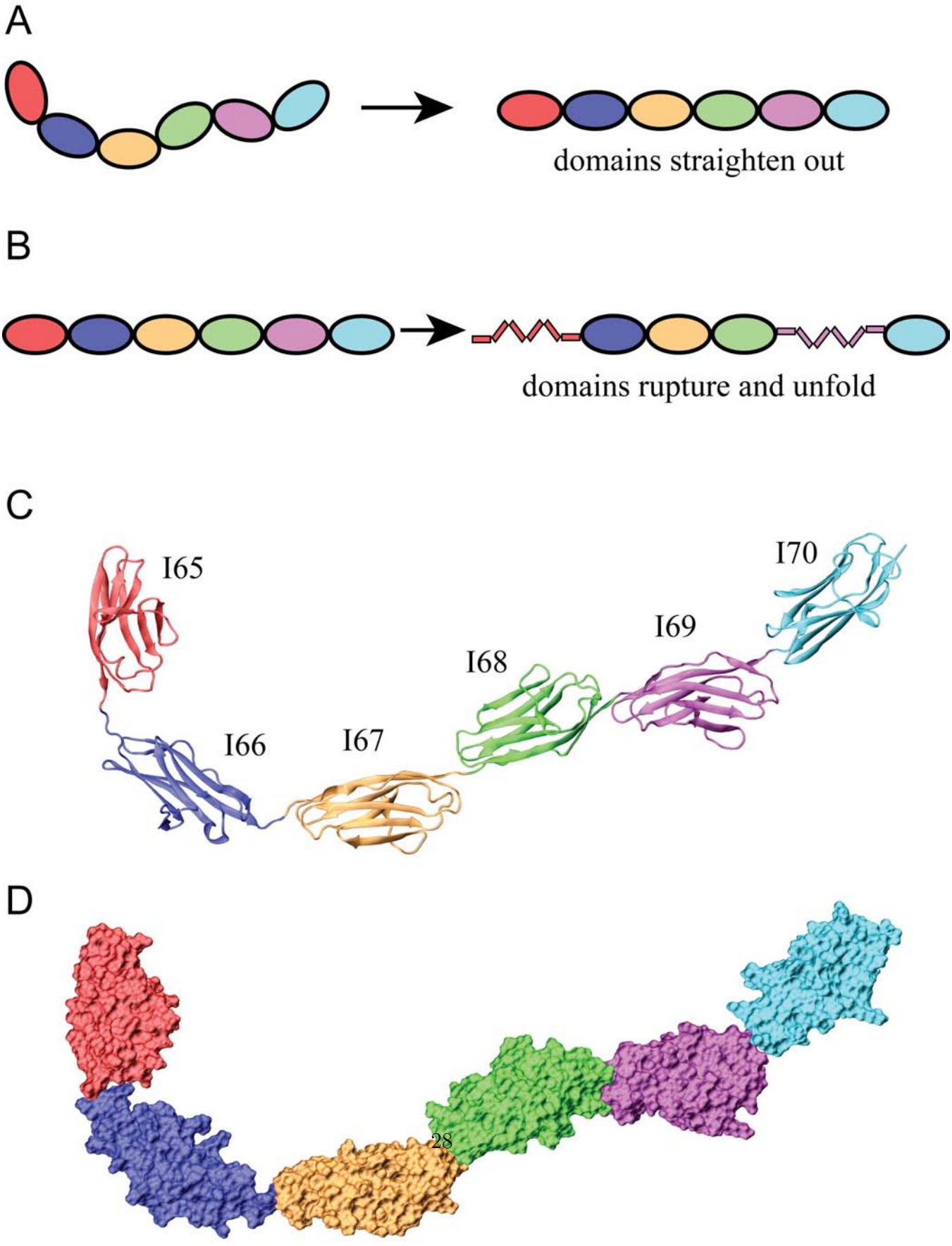

A
domains straighten out
B
domains rupture and unfold
C
I65
I66
I67
I68
I69
I70
D

Figure 1:

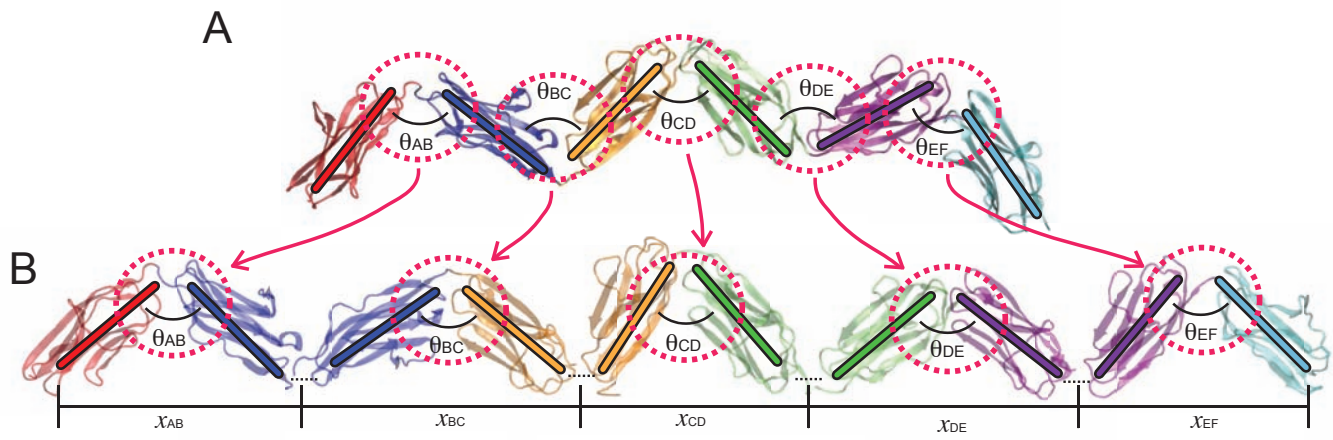


Figure 2:

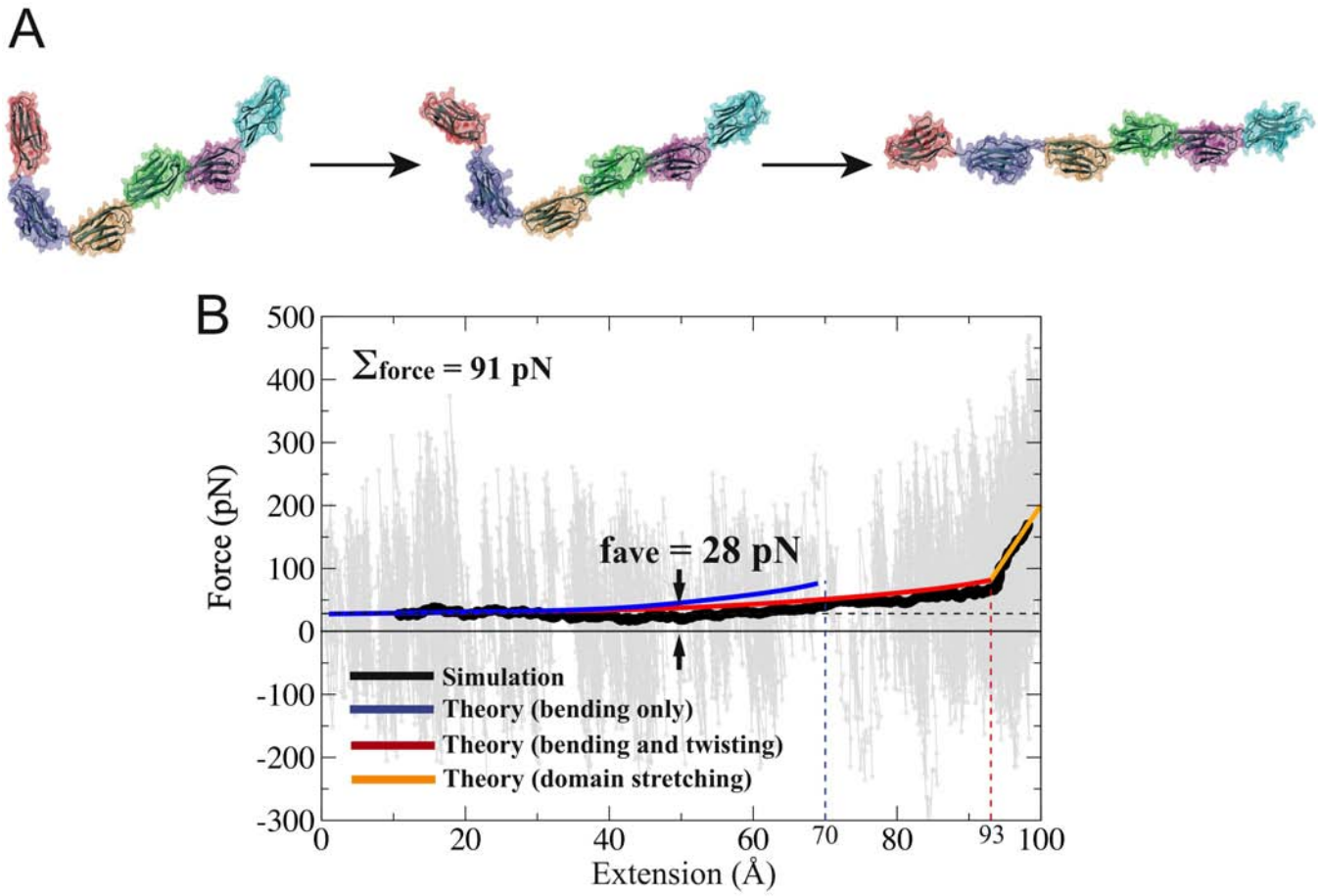

A
B
Σforce = 91 pN
fave = 28 pN
Force (pN)
Extension (Å)
Simulation
Theory (bending only)
Theory (bending and twisting)
Theory (domain stretching)
500
400
300
200
100
0
-100
-200
-300
0
20
40
60
70
80
93
100

Figure 3:

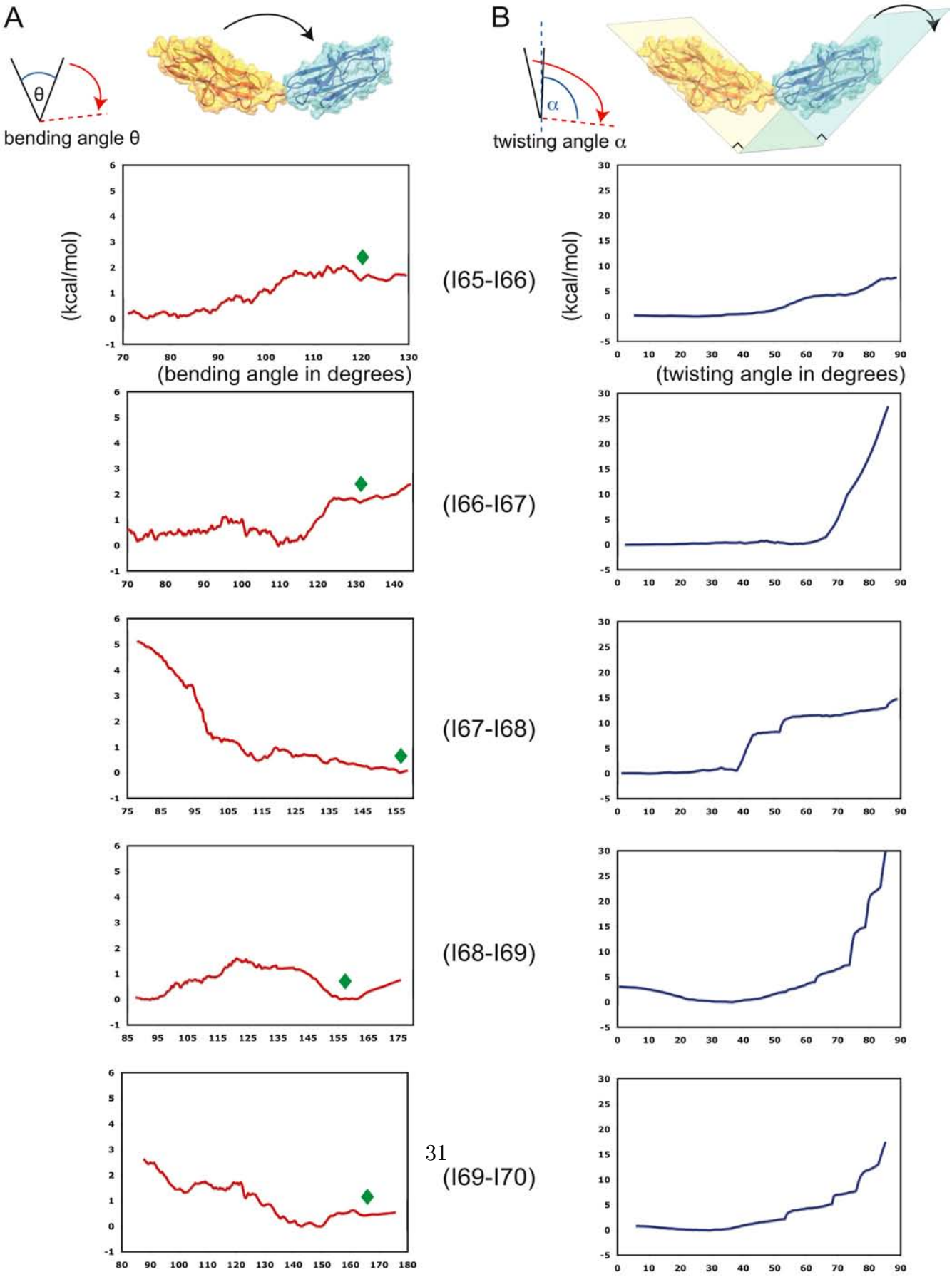

A
θ
bending angle θ
B
α
twisting angle α
(kcal/mol)
(bending angle in degrees)
(I65-I66)
(kcal/mol)
(twisting angle in degrees)
(I66-I67)
(I67-I68)
(I68-I69)
(I69-I70)

Figure 4:

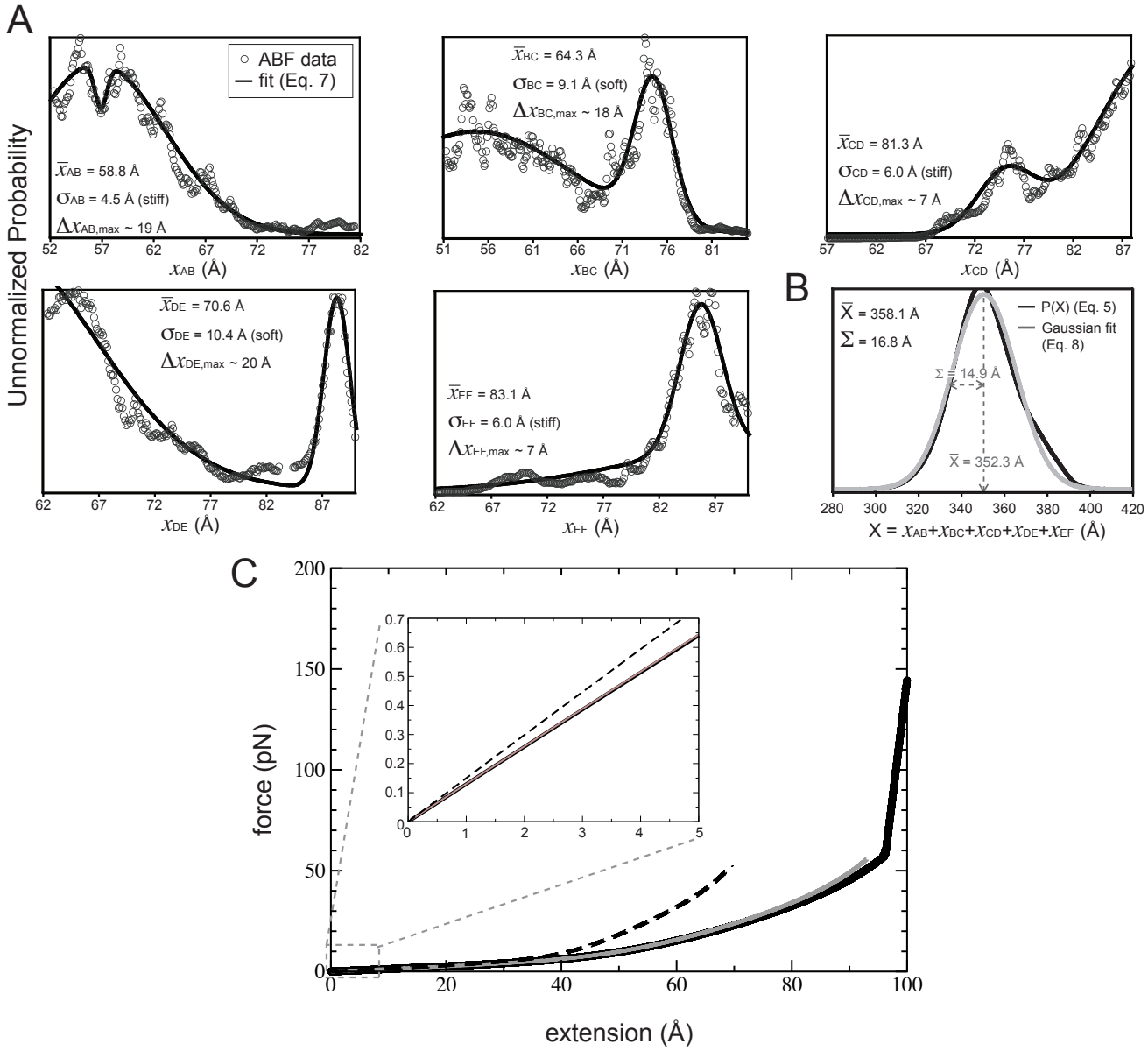

A
Unnormalized Probability
ABF data
fit (Eq. 7)
$\bar{x}_{AB}$ = 58.8 Å
σ_{AB} = 4.5 Å (stiff)
$\Delta x_{AB,max}$ ~ 19 Å
x_{AB} (Å)
$\bar{x}_{BC}$ = 64.3 Å
σ_{BC} = 9.1 Å (soft)
$\Delta x_{BC,max}$ ~ 18 Å
x_{BC} (Å)
$\bar{x}_{CD}$ = 81.3 Å
σ_{CD} = 6.0 Å (stiff)
$\Delta x_{CD,max}$ ~ 7 Å
x_{CD} (Å)
$\bar{x}_{DE}$ = 70.6 Å
σ_{DE} = 10.4 Å (soft)
$\Delta x_{DE,max}$ ~ 20 Å
x_{DE} (Å)
$\bar{x}_{EF}$ = 83.1 Å
σ_{EF} = 6.0 Å (stiff)
$\Delta x_{EF,max}$ ~ 7 Å
x_{EF} (Å)
B
$\bar{X}$ = 358.1 Å
Σ = 16.8 Å
P(X) (Eq. 5)
Gaussian fit (Eq. 8)
Σ = 14.9 Å
$\bar{X}$ = 352.3 Å
X = x_{AB}+x_{BC}+x_{CD}+x_{DE}+x_{EF} (Å)
C
force (pN)
extension (Å)

Figure 5:

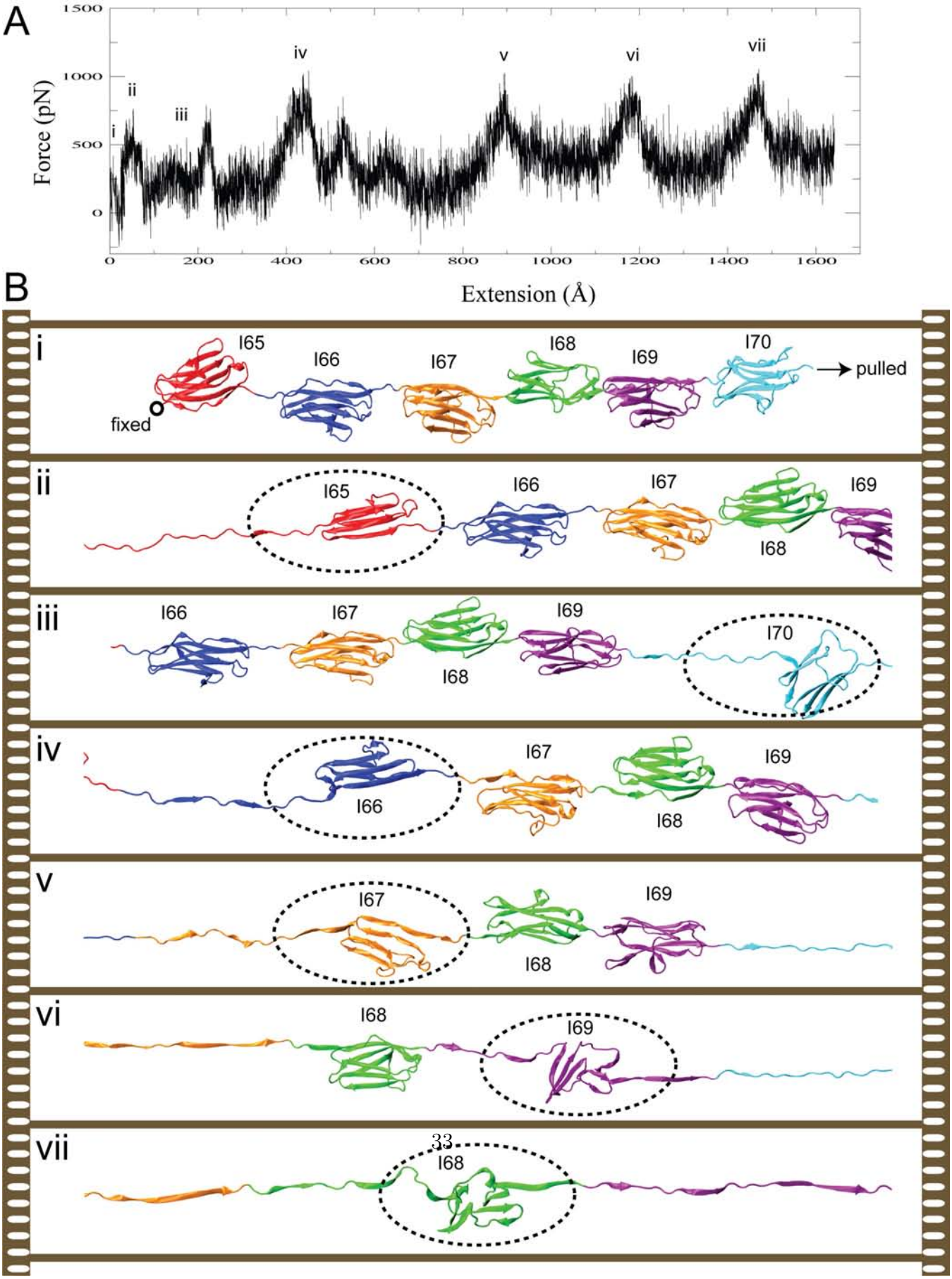

A
Force (pN)
Extension (Å)
i
ii
iii
iv
v
vi
vii
B
I65
I66
I67
I68
I69
I70
fixed
pulled

Figure 6:

3.4 Understanding the mechanics of the titin chain beyond its component modules – a critical review

3.4.1 Models of polymer elasticity

When interpreting data obtained by *in situ* measurements of stretched myofibrils or from single molecule force spectrometry measurements, statistical models adapted from polymer physics are applied (Flory, 1969). The more commonly used **Worm-Like Chain (WLC)** formalism describes homogeneous, semi-flexible polymers of continuous flexibility adopting a random-coil conformation. The WLC model needs only one parameter in its formalism, the persistence length (λ), which is the length over which the polymer starts to show some curvature (i.e. loses its correlation). For the WLC to be flexible the persistence length has to be much smaller than the contour length (L) of the polymer; otherwise, the WLC approaches the appearance of a rod, i.e. flexibility relates to L/λ (see DNA – short stretches of DNA behave like a rod, while long stretches of DNA behave like random coils). It should be mentioned that λ has the dimension of length, but does not necessarily describe a physiological building block of the polymer. Also, in dealing with polymers rather than ideal chains, we have to take into account excluded volume interactions, i.e. no two segments of the chain can occupy the same position in space. This severely reduces the number of conformations available to the polymer and thus increases its mean square end-to-end distance.

The alternative model from polymer physics applied to titin, the **freely-jointed chain (FJC)**, assumes that the chain is not continuously flexible but rather consists of stretches of straight segments, called the Kühn-length (l_K), only bending at discrete points along the chain with no restrictions on the angles of motion. If the segment length in the FJC model is chosen close to the persistence length of the WLC model, no big difference can be seen statistically between the two models, but when the segment length is longer, i.e. non-negligible compared to the contour length of the molecule, the FJC model is more appropriate and gives different results since it still allows for flexibility through bending at discreet points.

Entropic spring force is generated by the fact that the chain resists extension since this would reduce the number of conformations accessible to the random coil chain as it undergoes thermal fluctuations. Entropic contributions come e.g. from the exposure of hydrophobic residues normally hidden in the hydrophobic core, which

become solvated by low-entropy water forming a clathrate structure. Further entropic spring force comes from the damping of internal chain dynamics (librations) by extension, termed the Librational Entropy Mechanism (Urry, 1988). Enthalpic contributions on the other hand come from e.g. bond deformations, electrostatic stiffening, i.e. the interactions between like charges in close proximity (seen in PEVK segment), conformational changes, hydrogen bond breakages, etc.

Further important factors are the polarity and ionic strength of the **solvent**, where a good solvent favors interactions between the solvent and the polymer and thus leads to polymer swelling, while in poor solvents intra-molecular interactions of the polymer predominate and lead to a shrinking of the polymer. The ionic strength also plays an important role when the polymer is a polyelectrolyte, as screening of like charges along the polymer often increase its stability. As mentioned above solvent restructuring during hydrophobic hydration from bulk water to ordered water in clathrate networks is another source of change in entropy (lower entropy in case of the solvent).

Compared to their biological importance, the influence and consequences of lateral **confinement**, e.g. of titin in the sarcomere, are not well explored. Due to the complexity of the system, most research focuses on ideal polymers in dilute solution. While confinement has been explored in relation to DNA confinement in nanotubes, the field of peptidic polymers in confinement is poorly developed and is focused on stiff polymers like microtubules or actin (Cifra et al., 2008; Marko and Siggia, 1995; Wagner et al., 2006). Lateral confinement greatly influences the degrees of freedom and thus the number of conformations the polymer can adopt. Important parameters are the contour length and flexibility (i.e. persistence length) of the chain, and their relation to the length and diameter (d) of the confining lattice, where the stiffness of the walls is another important factor. If the confinement is weak, simplified $\lambda << d$, and the chain is flexible, not much influence can be seen on the end-to-end distribution of the chain. If, on the other hand, the persistence length is close to or even exceeds the diameter of the confining tube ($\lambda \sim d$), the polymer will no longer adopt a random coil but rather swell and wind through the tube, deflecting from the walls upon collision. (Brochard-Wyart et al., 2005; Odijk, 1983; Wagner et al., 2006).

If, furthermore the ends are fixed in space and orientation, as is the case for titin[3], the internal motions become describable in terms of librations, where the change in torsion of one angle, is compensated by an equivalent change in torsion of the opposite sign in one or more other torsion angles (Urry, 1988). The case of lateral confinement in conjunction with fixed end orientations have to our knowledge never been rigorously treated and all conclusions must therefore be qualitative estimations.

3.4.2 Modeling titin elasticity *in vivo* and *in vitro*

There are several aspects that have to be taken into consideration when modeling titin elasticity *in vivo* based on single molecule experiments. Firstly, on the sarcomere level titin is expressed simultaneously in multiple isoforms (Freiburg et al., 2000), and during extraction from muscle tissue titin is very sensitive to proteolytic degradation, which is another source for an inhomogeneous population (Maruyama, 1994; Nave et al., 1989). The use of recombinant fragments on the other hand is limited in the size-range of constructs that can viably be expressed and purified (Figure 3.13). Furthermore is the *in vitro* environment used for the varying experimental techniques, where individual molecules are measured in solution or absorbed to a surface, vastly removed from the conditions *in vivo*, as e.g. the hydrophobic collapse seen in EM is not possible in the sarcomere. There, titin is anchored at either end to sarcomeric structures in the Z-disc and M-band and stretches between the highly organized filament systems of the thin and thick filaments with all associated proteins, leading to a very crowded and confined space within the sarcomeric lattice (Fürst et al., 1988; Gregorio et al., 1999; Labeit and Kolmerer, 1995; Tskhovrebova and Trinick, 2002). The part that the A-band region of titin plays in elasticity measurements *in vitro* cannot be neglected either. *In vivo*, this part of the molecule is constrained by interactions with thick filaments, rendering it more or less inextensible (Fürst et al., 1988; Granzier et al., 1996). In extracted titin fractions most titin molecules are degraded through protease activity to the T2 fragment, also called β-connectin, which mostly consists of the A-band region of titin and only marginally extends into the I-band region (Granzier and Wang, 1993;

[3] Titin is anchored in the sarcomere with its N- and C-termini embedded in the Z-disc and M-line respectively. Also, the interaction with the thick filaments will impose a certain defined orientation on this part of the molecule.

Maruyama et al., 1984; Wang et al., 1984). Thus, measurements performed and conclusions drawn from these samples most likely are of little relevance to titin elasticity *in vivo*. All of this has to be considered when interpreting single molecule experiments and comparing them with *in situ* measurements.

It has been shown that the tandem Ig, PEVK and N2B react as three serially linked, independent spring elements, that have been modeled as WLC or FJC and are characterized by individual persistence or segment lengths (Granzier et al., 2002; Linke et al., 1999). Depending on the titin isoform and the technique used the derived persistence lengths (see Table 3.2) show a wide spread – 21 to 0.4 nm for the Ig tandem and 2.7 to 0.15 nm for the PEVK (N2B ~0.6nm) – but this can partially be explained by grouping the values into the persistence lengths of folded and unfolded as well as native, extracted or recombinant fragments (Figure 3.13). Depending on the size of the fragments the application of the WLC formalism is not entirely appropriate as the length of the individual segments is not insignificant compared to the contour length and the application of the FJC would be more appropriate.

Figure 3.13: Overview of experiments performed on titin
A) Domain structure of the skeletal soleus isoform of titin. B) All atomic-resolution structures of titin fragments are indicated by black boxes. The corresponding structures are displayed as ribbon diagrams above the titin filament. C) Location of antibodies used in single molecule and myofibril experiments. D) Recombinant constructs explored in single molecule AFM and OT experiments. E) Fragments visualized by EM. F) Structures explored by molecular dynamics simulations. G) Constructs employed in SAXS measurements. H) Domains, whose stability has been characterized by CD spectroscopy in conjunction with denaturation.

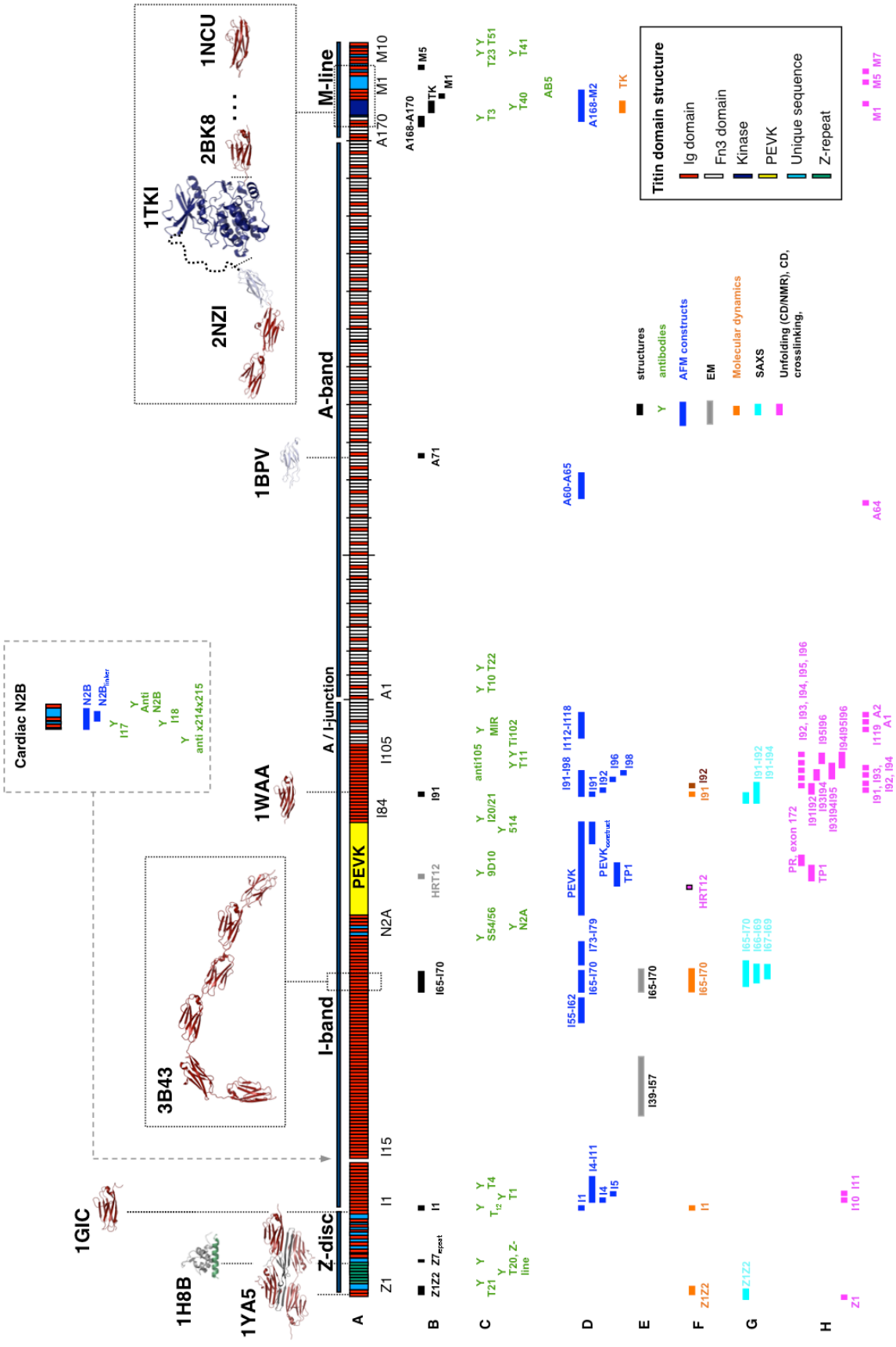

1GIC
1H8B
1YA5
Z-disc
I-band
PEVK
N2A
A / I-junction
1WAA
1BPV
A-band
M-line
Cardiac N2B
N2B
N2B$_{unique}$
Y I17
Y Anti N2B
Y I18
Y anti x214x215
3B43
1TKI
2BK8
1NCU
2NZI
Z1
I1
I15
I84
I105
A1
A170
M1
M10
A168-A170
M5
TK
M1
A71
I91
HRT12
I65-I70
S54/56
9D10
I20/21
anti105
MIR
T10 T22
514
T11
N2A
T21
T20, Z-line
Z1Z2
Z7$_{epsat}$
T$_{12}$
T1
T4
I39-I57
I55-I62
I73-I79
PEVK
PEVK$_{construct}$
TP1
I91-I98
I112-I118
I91
I92
I96
I98
A60-A65
A168-M2
TK
AB5
T3
T23 T51
T40
T41
I4-I11
I5
I4
I92
I91-I92
I91-I94
PR, exon 172
I92, I93, I94, I95, I96
I91I92
I93I94
I95I96
I93I94I95
I94I95I96
I91, I93, I92, I94
I119, A2
A1
A64
M1
M5
M7
I10
I11
structures
antibodies
AFM constructs
EM
Molecular dynamics
SAXS
Unfolding (CD/NMR), CD, crosslinking,
Titin domain structure
Ig domain
Fn3 domain
Kinase
PEVK
Unique sequence
Z-repeat

Table 3.2: Persistence lengths of titin

Component	Source	Method	PL	CL	Reference
PEVK	human soleus muscle fibres	IEM	2.0nm	820nm	(Trombitás et al., 1998)
tandem Ig	human soleus muscle fibres	IEM	15.0nm	340nm proximal 140nm distal	(Trombitás et al., 1998)
tandem Ig	rat psoas myofibrils	IEM +IFM	42nm/21nm	225nm	(Linke et al., 1998b)
whole chain	rabbit long dorsi	EM	13.5nm	0.9-1.0μm	(Tskhovrebova and Trinick, 2001)
whole chain	rabbit long dorsi	DLS	15nm	n.s.	(Higuchi et al., 1993)
PEVK	rat psoas myofibrils	OT + AB	~0.6nm	476nm	(Linke et al., 1998a)
PEVK region	rabbit long dorsi, single fast fibre	OT + AB	0.91nm / 2.7nm [a]	533nm / 579nm [a]	(Leake et al., 2004)
chain I, whole titin	rabbit long dorsi, single fast fibre	OT + AB	0.5nm / 1.7nm [a]	723nm / 505nm [a]	(Leake et al., 2004)
chain II, whole titin	rabbit long dorsi, single fast fibre	OT + AB	3.0nm / 3.6nm [a]	966nm / 906nm [a]	(Leake et al., 2004)
A-band	rabbit long dorsi, single fast fibre	OT + AB	2.7nm / 2.5nm [a]	1039nm / 1055nm [a]	(Leake et al., 2004)
tandem Ig	rabbit long dorsi, single fast fibre	OT + AB	3.3nm / 3.4nm [a]	125nm / 118nm [a]	(Leake et al., 2004)
I-band region chainI	rabbit long dorsi, single fast fibre	OT + AB	0.8nm / 1.3nm [a]	507nm / 421nm [a]	(Leake et al., 2004)
I-band region chainII	rabbit long dorsi, single fast fibre	OT + AB	3.1nm / 3.3nm [a]	374nm / 415nm [a]	(Leake et al., 2004)
titin II	extracted chicken breast muscle	DLS	15nm	~900nm	(Higuchi et al., 1993)
titin II	extracted rabbit psoas	SLS	11nm	1μm	(Di Cola et al., 2005)
titin II	extracted rabbit psoas	DLS	16nm	1μm	(Di Cola et al., 2005)
titin II	extracted rabbit psoas	SANS	10nm	1μm	(Di Cola et al., 2005)
titin II	extracted rabbit psoas	microrheology	16nm	1μm	(Di Cola et al., 2005)
titin I-band fragment	extracted rabbit psoas	DLS	10nm	100nm	(Di Cola et al., 2005)
titin I-band fragment	extracted rabbit psoas	SANS	9nm	100nm	(Di Cola et al., 2005)
Chain I (PEVK) WLC	extracted rabbit dors. Long.	OT + AB	0.15nm	416nm	(Tskhovrebova et al., 1997)
Chain I (PEVK) FJC	extracted rabbit dors. Long.	OT + AB	0.35nm	619nm	(Tskhovrebova et al., 1997)
Chain II (Ig/Fn) WLC	extracted rabbit dors. Long.	OT + AB	4.6nm	1020nm	(Tskhovrebova et al., 1997)
Chain II (Ig/Fn) FJC	extracted rabbit dors. Long.	OT + AB	5.3nm	1070nm	(Tskhovrebova et al., 1997)
titin chain	extracted rabbit dors. Longi	OT + AB	1.5-2.0nm	~4μm	(Kellermayer et al., 1997)
titin chain	extracted rabbit dors. Longi	OT + AB	0.7nm	3,4μm	(Kellermayer et al., 1998)
titin chain (folded)	extracted (various sources)	OT + AB	10nm	1μm	(Kellermayer et al., 2001)
titin chain (unfolded)	extracted (various sources)	OT + AB	1.6nm	10μm	(Kellermayer et al., 2001)
titin chain	extracted rabbit dors long.	AFM	1.5nm	~0.87μm	(Kellermayer et al., 2003)
PEVK (cardiac)	recombinant construct	AFM	0.4-2.5nm	~70nm	(Li et al., 2001)
PEVK (526aa)	recombinant construct	OT + AB	1.8nm / 4.3nm [a]	201nm / 198nm [a]	(Leake et al., 2004)
PEVK fetal TP1	recombinant construct	AFM	~0.15-0.6nm	30-150nm	(Forbes et al., 2005)
PEVK	recombinant constructs	AFM	~0.9nm	31-34nm	(Sarkar et al., 2005)
N2B	recombinant construct	AFM	0.66nm	232nm	(Li et al., 2002)
N2B	recombinant constructs	AFM	0.65nm	~220nm	(Watanabe et al., 2002b)
N2B	recombinant construct	AFM	0.4-0.6nm	120nm	(Leake et al., 2006)
I91-I94, I91-I98	recombinant constructs	AFM	0.4nm	n.s.	(Rief et al., 1997)
I91-I98, I112-I118	recombinant constructs	AFM	0.4nm	n.s.	(Rief et al., 1998)
proximal Ig (I4-I11)	recombinant construct	AFM	10nm / 0.66nm [b]	n.s.	(Li et al., 2002)
distal Ig (I91-I98)	recombinant construct	AFM	10nm / 0.66nm [b]	n.s.	(Li et al., 2002)
I65-I70	recombinant construct	AFM	~0.6nm	191nm	(Watanabe et al., 2002a)
I91-I98	recombinant construct	AFM	~0.6nm	172nm	(Watanabe et al., 2002a)

Abbreviations: PL – persistence length; CL – contour length; IEM – immuno electron microscopy; IFM – immunofluorescence microscopy; DLS – dynamic light scattering; SLS-static light scattering; SANS – small angle neutron scattering; OT – optical tweezers; AB – antibody; AFM – atomic force microscopy; n.s. – not specified. Footnotes: a) measurements performed at 150mM / 300mM ionic strength; b) folded / unfolded chain

In light of the recent availability of the first structures of representative poly-domain constructs it is time to reassess the applicability of these two formalisms, since the available models clearly challenge the view of titin as a chain of beads on a string with no defined inter-domain structure. Especially the WLC suffers in this regard, since it is based on the assumption of a homogeneous polymer of continuous bending rigidity.

3.4.3 Polydomain structures of titin

Recently, the first two-domain structure of the N-terminal 2 Ig, Z1Z2 located in the Z-disc, was solved both in complex with their native partner, telethonin (Figure 3.14B) (Zou et al., 2006) as well as in isolation in a comprehensive study combining X-ray crystallography, SAXS, NMR relaxation data and residual dipolar couplings (RDC) to get a mechanistic insight into its modular dynamics (Figure 3.14A) (Marino et al., 2006). Both Z1 and Z2 belong to the N-conserved group of Ig and are connected by a three-residue linker. They preferentially adopt a semi-extended conformation in solution, as shown by NMR RDC measurements in combination with SAXS analysis. Linker length and a lack of specific interactions between Ig except for a small hydrophobic component would indicate that inter-domain rearrangements should be possible on a large amplitude. Relaxation data show that this linker does not undergo fast fluctuations and that Z1Z2 appears to be an elongated shape in solution with restricted dynamics. That Z1Z2 is indeed able to undergo large-scale rearrangements is seen in the crystal structure. There the three copies in the asymmetric unit adopt either an extended or closed conformation, induced by metal ligation and stabilized by the crystal lattice. Analysis of the differences in conformation shows that closing of the two domains is based on torsion around the backbone of residue A99 in the linker, which acts as a hinge. Thus, Z1Z2 adopts a preferential, semi-extended conformation in solution, which can adapt to environmental stimuli, like the induced straightening by binding to telethonin.

The crystal structure of the complex with telethonin reveals a unique palindromic assembly, where two anti-parallel copies of Z1Z2 in extended conformation bind one telethonin molecule through large-scale backbone hydrogen bonding, leading to an extended β-sheet formation across the molecules (Figure 3.14B). This complex provides a stable anchoring scaffold for titin in the Z-disc. The mechanical strength of this complex was also confirmed in Steered Molecular Dynamics (SMD) simulations (Lee et al., 2006), which showed that complexation of Z1Z2 with telethonin helps distribute the force through the β-sheets to both domains under stretch.

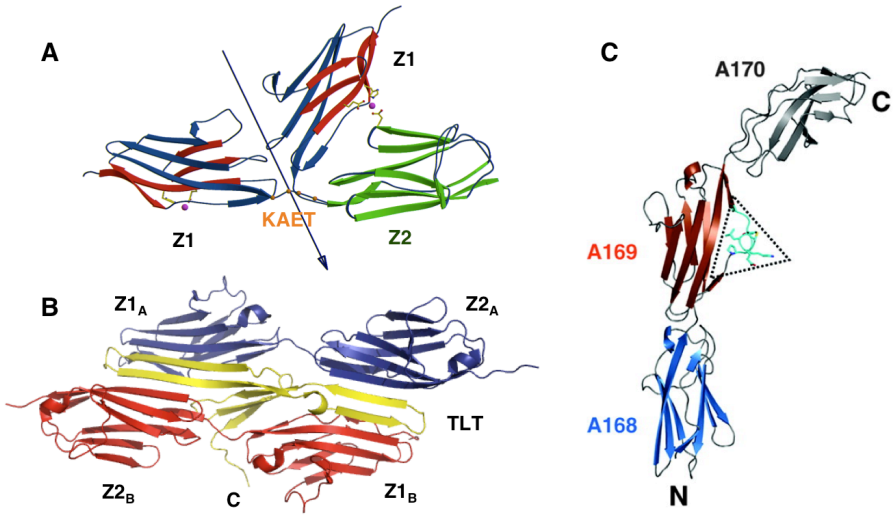


Figure 3.14: Structures of polydomain constructs of titin
A) Superposition of the two conformations seen in the crystal structure of Z1Z2. For easier visualization of the domain torsion, the β-sheets of Z1 are colored differently. Metal ions and coordinating protein residues are displayed. The rotation axis that relates the Z1 domains is shown. (From Marino et al., 2006)
B) Crystal structure of Z1Z2 in complex with telethonin. The extended β-sheet formation across molecules can be seen (adapted from Zou et al., 2006)
C) Crystal structure of A168-A170, with the unique loop insertion in A169 highlighted in cyan and marked by a triangle. A168-A169 lie almost coaxial, while the FnIII domain A170 tilts away from the molecule axis (taken from Mrosek et al., 2007)

Subsequently two related structures, A168-A170 (Mrosek et al., 2007) and A168-A169 (Müller et al., 2007), were solved of the A-band domains directly preceding the titin kinase that are involved in binding of MuRF-1. A168-A170 (Figure 3.14C) is composed of two Ig belonging to the N-conserved group followed by an FnIII domain. Domain A169 contains a unique sequence insertion shown to be important for binding to MuRF1 (see chapter 4.3). The three domains exhibit an extended arrangement and are connected by one-residue linkers that result in extension of secondary structure across the domains and restricted mobility around the linker. While both A168-A169 lie almost coaxial, A170 is slightly bent away from the molecular axis. A feature of this extended domain arrangement is the unusually small interaction surface, which is dominated by hydrophobic clusters. In this it resembles Z1Z2, which still displays a different domain orientation and a certain flexibility around its three-residue linker (Marino et al., 2006). Thus, the presence of the residues of the N-conserved fold alone is no determinant of domain orientation in these multi-domain constructs, but linker length and composition come into play as well. The structure of A168-A70 for the first time also encompasses an Ig-FnIII domain interface that,

based on analysis of sequence conservation of key residues, might be representative for equivalent Ig-FnIII pairs along the A-band.

The crystal structure of I65-I70 offers now the first atomic model directly related to the elastic I-band of skeletal-muscle titin and shows this region to be characterized by a distinctive pattern of regular domain spacings, hinge openings and torsional angles (Figure 3.15; covered in detail in chapter 3.2).

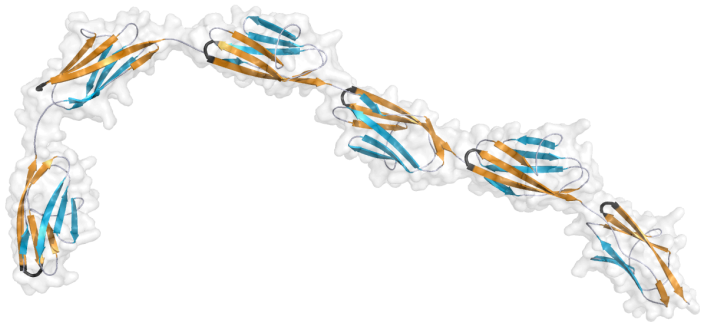


Figure 3.15: Crystal structure of I65-I70
Crystal structure adopts an extended conformation, where domains I67-I70 lie nearly coaxial, while the two N-terminal domains I65 and I66 are slightly bent away from the molecular axis. The β-sheets are colored differently to illustrate the distinctive torsional pattern along the chain.

In summary, the structures of the poly-domain fragments solved to date, which all comprise Ig of the N-conserved group, show that this fold seems to favor extended conformation, probably being oriented through interactions of residues in the BC and FG-loops with linker residues. Domain interfaces were shown to be small and often hydrophobic, lacking any specific inter-domain interactions. The linkers connecting the chain on the other hand are heterogeneous in length and composition, as well as mechanical/dynamic properties, most likely correlating with the functional regions of the chain they are located in. With structural data available, such heterogeneity can be accounted for in a model that incorporates the chemical features of the chain, replacing the homogeneous statistical models available. In this regard, the crystal structure of I65-I70 allows the formulation of the first structure-based model for titin elasticity of the skeletal Ig tandem, incorporating the concept of tertiary structure elasticity with the regular features observed in the skeletal tandem.

3.4.4 The concept of tertiary structure elasticity

Following the publication of the crystal structure of Z1Z2 in combination with data on the conformation and dynamics in solution using NMR and SAXS techniques (Marino et al., 2006), a study combined Steered Molecular Dynamics (SMD) and Adaptive Biasing Force (ABF) simulations to explore the conformational flexibility, dynamics and energetics of higher-order structure in titin as seen in the interdomain arrangement of Z1Z2 (Lee et al., 2007). The relative orientation of the domains with regard to each other is determined by the linker as well as the domain-domain interface. This well defined tertiary structure should oppose forces stretching domains apart through linker reorientation and hinge opening or twisting, giving rise to an elastic force that has been termed tertiary structure elasticity by the authors.

Z1Z2 was chosen as a model since the three structures available, i.e. the closed, semi-extended and open state, are indicative of inter-domain dynamics. Indeed, during equilibration of the crystal structure Lee and coworkers (Lee et al., 2007) saw a spontaneous change in conformation, where Z1Z2 approached the semi-extended state calculated from the RDC and SAXS measurements. Low force SMD simulations explored the forces required to open the doublet without inducing domain unfolding and found that 50pN was not enough to open the doublet within 2ns, whereas forces of 94pN and 164.5pN were sufficient for hinge opening without unfolding. These forces are higher than the forces used to unravel titin chain in the most sensitive force spectroscopy experiments, using optical tweezers (Kellermayer et al., 2001; Leake et al., 2004), but are made necessary by the limited time scales available to MD.

Subsequent ABF experiments were performed to calculate the energy needed for the conformational changes seen in Z1Z2. For this, force is applied along a predetermined reaction coordinate to open Z1Z2 through rotation about the linker into an open state. The calculated potential of mean force (PMF = free energy profile) was found to increase by 20kcal/mol upon forcing Z1Z2 into the open state, indicating that domain rearrangements need a measurable energy input and thus can buffer external forces.

Based on their simulation results they constructed a basic model of a titin chain composed of 300 domains, using as a building block Z1Z2 and its linker and assuming that every angle is contributing to the chain extension. Several simplifications were applied, namely that the system is planar, i.e. there is no twisting out of the plane, and that the angles only assume two states, the open and closed

conformations seen in the crystal and the solution structure. If a force of 50pN, characteristic for *in vitro* experiments, was used to stretch this model, this would lead to an estimated extension of the chain by 600Å, corresponding to ~9% of its length. This extension can be achieved by the consecutive slight opening of each Z1Z2 doublet by a few Å without having to rely on the (full) extension of individual Ig doublets, for which the applied forces would be too low. Thus they showed that Z1Z2 is able to generate elastic force by domain rearrangements and that even small rearrangements in individual domains can add up since propagation of the force through the chain leads to consecutive domain rearrangements along the chain.

3.4.5 The new structure-based carpenter ruler model

The regular arrangement of domain and linker features in I65-I70 in conjunction with the conservation pattern observed in the skeletal I-band tandem allowed for the first time the construction of a model representing a possible conformation of the tandem. The resulting model suggests that conserved Ig-Ig transition motifs generate high-order in the structure of the filament in keeping with the super-repeat pattern identified in this region. In the model, straight stretches of tightly connected domains are interspersed with domains connected by longer linkers. Tightly connected modules are likely to be conformationally restrained, as steric hindrance prevents bending. Ig pairs joined by longer linkers can be expected to allow long-range inter-domain motions as seen in Z1Z2 and, thus, act as flexor points of the chain. Stretch and recoil would be expected to first induce conformational changes in the permissive linkers, with their opening sufficient for full extension of the chain. This led us to formulate a new model of titin elasticity, termed the carpenter ruler model, which is based on the segregation of domains with different bending stiffnesses into two classes that correspond to rods and flexor points in the chain. Experimental evidence that Ig might indeed be correlated along the length of 5-6 domains, corresponding to the length of the super-repeats was found in a study by Linke and colleagues (Linke et al., 1998), where they reported a persistence length of 21nm for the tandem Ig (immunofluorescence, immuno EM and single myofibril mechanics experiments performed on isolated rat psoas myofibrils).

This is clearly a divergence from the WLC model assuming a homogeneous polymer of continuous bending rigidity and therefore should be expressed by the discrete formalism of the FJC. Our application of the FJC formalism to model skeletal

I-band elasticity has been reported previously (see chapter 3.2; von Castelmur et al., 2008), Briefly, based on conserved linker features, the skeletal Ig were grouped into 15 straight segments with an average length of ~ 15.6nm. From the mean square end-to-end distance of this FJC chain, the corresponding persistence length, λ, was derived for comparison with published values. This gave rise to an estimation of 8 nm, in agreement with experimental values for this section of titin (~9 nm; Di Cola et al., 2005), showing that our model explains statistical parameters currently available for titin. It also illustrates, as mentioned earlier, that contrary to the WLC a FJC can be composed of fairly long stiff segments and still be flexible, as long as freedom of motion is given at discrete hinge points.

In I65-I70 individual domain pairs could be shown to adopt a preferred conformation, which is independent of the long-range conformation of the overall chain, but instead based on the constituent linker length and composition. Application of external force, either to stretch or compress the filament, will induce a temporary and reversible conformational change in the chain. Upon release of the external stimulus, individual domain pairs can be expected to return to their preferred local conformation. Thus, titin can be said to have a molecular shape memory, in analogy to the shape memory effect seen in polymers and alloys (Ratna and Karger-Kocsis, 2008). This molecular shape memory can also be seen as the underlying driving force of tertiary elasticity.

In terms of a mechanical model, this would translate to the flexible linkers acting like hinge points that allow rearrangement of the titin I-band under stretch or compression. Since this model has fewer accessible conformations than the continuously bending WLC model, this leads to a lower conformational entropy component and instead an enthalpic component due to the conformational changes in the linker has to be included.

The stiff segments on the other hand appear to allow only minor conformational rearrangements. Instead, their mechanical contribution might be an essential factor in resisting domain unfolding at higher forces. Any stretch beyond full extension of the tandem, would lead to dampening of their librational motions, as seen in elastin (Urry, 1988), and would function as an energy barrier against domain unfolding, by instead triggering the unraveling of the more compliant PEVK spring segments. As a result, the interplay of poly-Ig and PEVK serial springs causes a non-linear response to stretch that defines the mechanical properties of titin.

The results of our exploration of the secondary and tertiary elasticity of I65-I70 (see chapter 3.3) showed that the energetic cost of hinge opening of this well-defined structure lies relatively low, on the order of several k_BT. Together with the plateau-like profile observed for the force necessary for extension of the molecule, this indicates that structural rearrangements of I65-I70 under force indeed represent a source of tertiary structure elasticity that would allow the extension of the Ig tandem at low force through well-defined hinge rearrangements. This is indeed consistent with our carpenter ruler model.

Rigid treatment of the attempt to put our newly proposed model into the context *in vivo* is well beyond the scope of this work, not least since the field of macromolecular polymers and their elasticity in confinement is not well explored. Thus we will focus on a qualitative assessment of the relevant parameters and their likely influence. In striated muscle, the I-band of titin is confined within the acto-myosin lattice, whose lateral spacing is dynamic and depends on the state of extension or contraction. A tube with a diameter of 17-20 nm can approximate the lateral space available to each chain (Schroeter et al., 1996; Tskhovrebova and Trinick, 2002). The conformation of a WLC in this confined space depends on its persistence length, the diameter of the tube as well as the stiffness of the walls. When modeling the skeletal tandem as a WLC composed of 50 domains, each 4.5nm in length, and with a persistence length λ = 10nm, a recent theoretical analysis (Brochard-Wyart et al., 2005) predicts[4] that the chain will occupy a length around 100nm within the tube, and thus assume a rather extended state (contour length L = 225nm)[5]. Thus, the entropy derived from conformational sampling of space is minimized.

In the case of our carpenter ruler model, the longest straight segment contains a stretch of 4-5 domains, which corresponds to ~22.5nm. This is in the same range as the expected diameter of the tube and thus the chain will normally tend to adopt an

---

[4] The optimal length ($R_{||}$) of a polymer in a tube is given by $R_{\|} = N l_K \left(\frac{l_K}{d}\right)^{2/3} \left(\frac{\lambda}{d}\right)^{1/3}$

[5] A simplified modeling of T2 titin as a flexible chain confined in the hexagonal lattice of actin filaments (assuming a single titin chain per tube) by Di Cola et al. came to the conclusion that upon confinement in the tube, the end-to-end distance of the chain would increase ~10fold, i.e. to ~600nm compared to the radius of gyration R_G~59nm they had measured for the chain in solution (Di Cola et al., 2005).

extended conformation along the tube, since this segment won't readily fit transversely in a relaxed state. This is in agreement with the conformation observed in our model of the skeletal I-band (Figure 3.12). The behavior of the titin chain during stretch, as predicted by our carpenter ruler model, should not be affected by this lateral confinement, since the straightening of the titin chain should not be affected by the narrowing of the lateral acto-myosin lattice spacing. During contraction on the other hand the lattice spacing will widen, which should enable the straight segments to fit transversely and thus allow compression of the I-band region and even stretch in the opposite direction, as has been observed in myofibrils shortened to below their slack length (Granzier et al., 1996; Helmes et al., 1996). Upon release of force, the conformational energy stored in the (compressed or straightened) linker segments can be expected to provide the restoring force necessary to bring the sarcomere back to slack length.

Clearly, extensive further experimental work will be required to explore the validity of our model. To our knowledge, though, it does not contradict experimental results obtained to date. Thus, it represents a valuable improvement on currently available models, since for the first time structural information on the higher-order organization of the tandem Ig have been integrated.

3.5 Distal Ig tandem

3.5.1 Introduction

Currently available models of titin Ig-tandems and our recently proposed mechanistic model do not explain the behavior of the constitutive Ig-arrays dominating the cardiac forms of titin, since constitutive Ig exhibit different loop features and Ig-Ig transition sequences that do not appear compatible with the existent model.

We therefore initiated a study to elucidate the crystal structure of a representative fragment of the distal Ig tandem. The work focused on the purification and crystallization of, I101-I106 from human titin. This construct localizes at the distal I-band adjacent to the A/I-band junction and contains 5Ig domains (1 N-conserved, 4 N-variable) and one FnIII domain. The structure should allow us to gain an insight into the architecture of constitutive Ig-tandems of I-band titin relevant to the mechanics of cardiac isoforms. The expectation was that the resulting model should enable the interpretation of the mechanical differences between constitutive and spliced Ig-tandems and, combined with the existent knowledge for the spliced arrays, allow the formulation of an integrated mechanistic model for the diverse cardiac and skeletal poly-Ig chains.

3.5.2 Experimental work

3.5.2.1 Cloning:

The expression plasmid for I101-I106 was provided by Prof. Siegfried Labeit and PD Dr. Dietmar Labeit (Institut für Anästhesiologie und Operative Intensivmedizin, Universitätsklinikum Mannheim, Germany). Coding sequences for domains I101-I106 from human titin (residues 13566-14120, Uniprot entry Q8WZ42) had been cloned into the pETM-11 vector (EMBL vector collection). pETM-11 is a variant of pET-24d (Novagen) including an N-terminal His_6-tag and a TEV (tobacco etch virus) protease cleavage site prior to the inserted gene.

3.5.2.2 Protein production:

Overexpression of I101-106 for structural analysis was carried out in *E.coli* BL21(DE3) Rosetta (Novagen). Cultures were grown at 30°C up to an OD_{600} of 0.6 in Luria Bertani medium supplemented with 25 μg/ml kanamycin and 34 μg/ml chloramphenicol. Expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cultures were further grown at 20°C for approximately 18 additional hours. Cells were harvested by centrifugation at 2800g and 4°C. Bacterial pellet was resuspended in 25mM MOPS pH 7.4, 150mM KCl, 1mM DTT, containing a protease inhibitor cocktail (Roche) and DNAse I and lysed by French press treatment (3 passes at 1000psi). The homogenate was clarified by centrifugation at 15000g and 4°C. Affinity purification used a HiTrap column (GE Healthcare) equilibrated in lysis buffer in conjunction with a stepwise imidazole gradient employed during column wash steps to remove impurities. Elution used 50mM Tris pH 8.0, 100mM NaCl, 2mM βME, 200mM imidazol. Tag removal was by incubation with TEV protease overnight at 4°C during dialysis against 25 mM MOPS pH 7.4, 150 mM KCl, 1 mM DTT. Subtractive purification for protease (His_6-tagged at C-terminus) and non-digested protein was carried out on a subsequent chelating step. The eluted protein fractions were further purified to homogeneity by ion exchange chromatography using a MonoQ column and gel filtration on a Superdex 200 HiLoad 16/60 column (both from GE Healthcare) in 25 mM MOPS 50mM KCl pH 7.4. Protein samples were then concentrated as required and stored at 4°C. The yield of pure protein was 20 mg per 1L of *E. coli* expression culture. Mass spectrometry was used to confirm the identity of the target protein by tryptic digest and peptide mapping.

Analysis of the sequence shows that this construct contains a very large number of charged residues (101 Asp + Glu, 88 Arg + Lys). Some issue with persistent impurities encountered in purification can most likely be attributed to electrostatic interaction and would probably be ameliorated by more extensive salt washes during affinity purification.

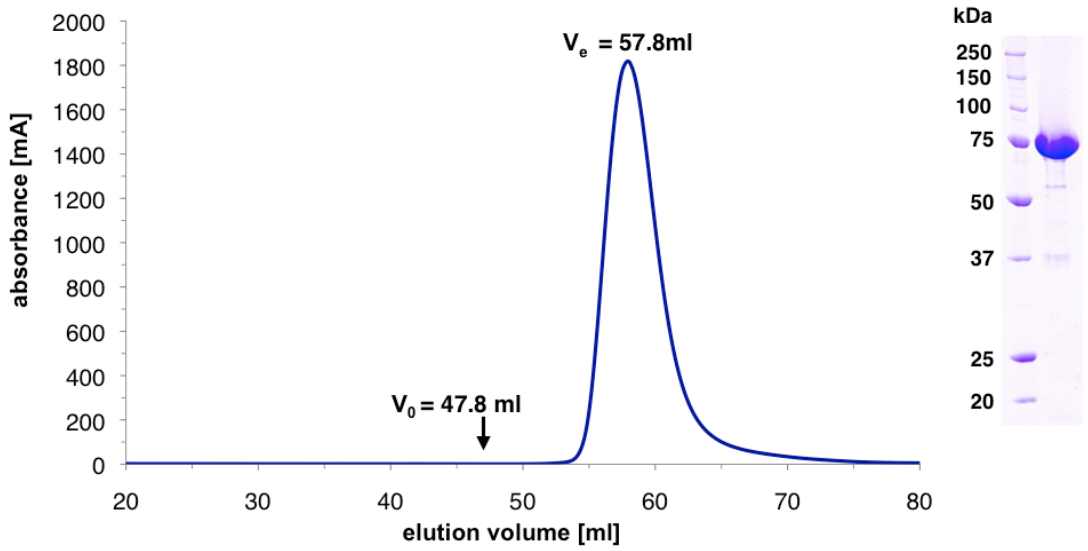


Figure 3.16: Size-exclusion chromatogram of I101-I106
Size exclusion was performed in 25 mM MOPS pH7.4, 50mM KCl using a Superdex200 HiLoad 16l60 column (GE Healthcare). The signal shown corresponds to A_{280nm}.Comparison of elution volumes with I65-I70 suggest that I101-I106 is monomeric in solution and most likely adopts a rather extended conformation (estimated molecular weight for a globular protein would be >600kDa, see calibration curve in appendix 6.2). SDS-PAGE on the right shows final sample purity. Marker used is Precision Protein Plus Standard (Biorad).

3.5.2.3 Crystallization

Initial sparse matrix screening was performed at 20°C and used protein retaining the His_6-tag, as provided by D. & S. Labeit. Trials were set using commercial screen kits, namely the Hampton screens I & II, Quikscreen (Hampton Research), Wizard I+II and Cryo I+II (both Emerald Biosystems) in 96-well plates (Greiner Crystalquick) using the sitting drop vapor diffusion method. 1μl of protein solution at 10 and 20 mg/ml was mixed with 1μl of mother liquor and equilibrated against a reservoir of 70μl. Screening was later repeated with untagged protein using 48-well VDX plates. 1μl of protein solution at 10, 20 and 40 mg/ml was mixed with 1μl of mother liquor and equilibrated against 100μl of reservoir solution (performed by Zoehre Ucurum). Several conditions that gave rise to micro-crystals and small needles growing in the habit of clusters were selected for refinement. Based on the chemical composition of the main precipitant, these conditions could be divided into two groups: *i)* based on ethylene glycol or PEG of varying size or *ii)* 1,4-butanediol.

Table 3.3: Crystallization conditions for titin I101-I106 identified in sparse matrix screens

Condition	Chemical composition
CSI-37	8% PEG 4000, 0.1 M Na Acetate pH 4.6
CSII-37	10% PEG 8000, 0.1 M Hepes pH 7.5, 8% Ethylene Glycol
WizI-12	20% (w/v) PEG-1000, 0.1M imidazole pH 8.0, 0.2M $Ca(OAc)_2$
WizI-35	20% (v/v) 1,4-butanediol 0.1M Acetate pH 4.5
WizII-44	0.2M MgCl2, 0.1M Cacodylate pH6.5, 20% PEG 1000
CryoI-4	40% (v/v) PEG-300, 0.1M HEPES pH 7.5, 0.2 M NaCl
CryoI-12	50% (v/v) ethylene glycol, 0.1M acetate pH 4.5, 5% (w/v) PEG-1000
CryoI-46	40% (v/v) PEG-400, 0.1M MES pH 6.0, 5% (w/v) PEG-3000

CS – Crystal Screen I + II, Hampton Research; Wiz – Wizard I + II, Emerald Biosystems; Cryo I + II – Cryo Screen, Emerald Biosystems

Thus, a refined search concentrated on screening the precipitants, i.e. PEGs of different size and in a range of concentrations as well as 1,4-butanediol, against buffers of varying composition and pH. This led to growth of crystals of improved size. Further screening used additives, ranging from organic solvents (glycerol, alcohols, MPD, DMSO, etc) to sugars, divalent metal salts and amino acids, as well as layering of the reservoir with a mix of silicon and paraffin oil that helped to slow down vapor diffusion and equilibration. Optimized conditions gave rise to single crystals of larger size and improved morphology, but still growing in the habit of thin plates. The most promising crystals grew in the space of a few days and in the habit of large, but very thin, plates of approximate dimensions 300x100x5 μm^3. The reservoir contained 0.1M HEPES pH 7.0, 10% PEG-8000, 10% ethylene glycol, 3% methanol, with a drop ratio 1.25μl of protein sample to 0.75 μl of reservoir solution (Figure 3.17A).

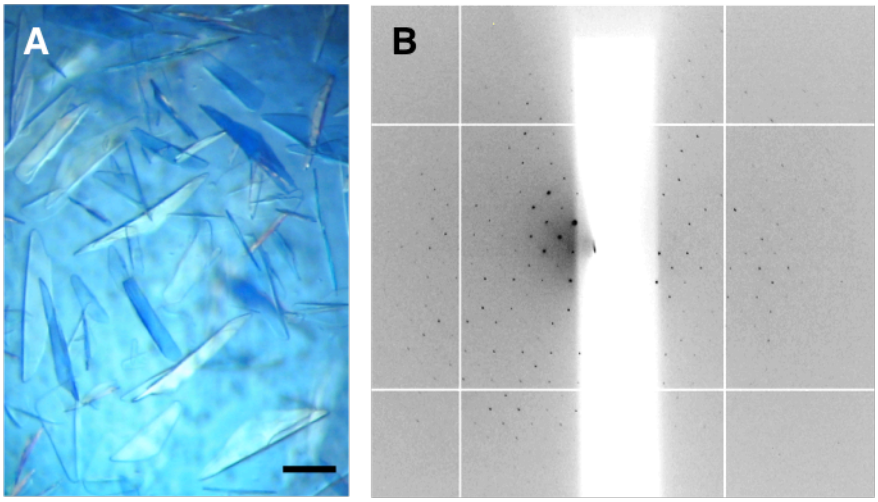


Figure 3.17: I101-I106 Crystals and resulting X-ray diffraction pattern
A) Crystals with ultra-thin plate morphology, representative of those used for diffraction tests, scale bar, 0.1μm. B) Diffraction image from crystals of I101-I106. Crystals showed only weak diffraction and needed prolonged exposure times of 8s for diffraction to an approximate resolution of 8Å. (I04, Diamond, UK)

3.5.2.4 Diffraction test

For diffraction tests, crystals were mounted in litholoops (Molecular Dimensions) and flash-frozen in liquid nitrogen, using mother-liquor supplemented with either 7% [v/v] glycerol or PEG-400 as cryoprotectant. Tests were conducted at beamline I04 at Diamond (Didcot, UK) under cryoconditions. Crystals showed interpretable diffraction patterns to resolutions of approximately 8Å (Figure 3.17B). The low-resolution and overall weakness of the diffraction is predictably due to the thin-plate morphology of these crystals. Furthermore, the thin plate morphology posed difficulties in handling and freezing of the crystals, where the drop surface tension and other mechanical stresses of the manipulation led to breakages or to the buckling of the plate plane. As a result, diffraction images showed split reflections and reflections with distorted profiles. Clearly, availability of crystals with improved morphology, namely increased thickness of the plate, should facilitate handling and result in improved diffraction quality.

3.5.3 Discussion

We could establish protein production and purification of I101-I106 with a yield of pure protein of ~20mg/L of culture. Size-exclusion chromatography indicates that I101-I106, similar to I65-I70, is monomeric in solution and adopts an extended shape. Crystallization conditions have been identified and refined to yield single crystals.

Sadly, for reasons listed above, the crystals obtained to date are not of sufficient quality for collection of diffraction data suitable for structure elucidation. For a successful conclusion of this work, issues in reproducibility and diffraction quality of I101-I106 will have to be resolved.

Our laboratory has an established track record regarding the crystallization and structure elucidation of titin poly-domain fragments (Z1Z2, A168-A170, I65-I70, I67-I69, I67-I69^{E93A}). The experience gained during crystallization of these domains from various regions of the molecule, but especially of I65-I70 and its subfragments, should help in designing further experiments to successfully conclude this project. For I65-I70, since structure solution proved slow and complicated, a range of subfragments was designed and crystallized, which led to the successful structure elucidation of I67-I69 at a resolution of 1.6 Å. A further construct spanning I66-I69 also readily crystallized and single crystals in the shape of rods or needles in a size range of 250-600μm in their longest dimension could be optimized from small needle-clusters. In diffraction tests (at beamline X06SA, SLS, Switzerland) crystals showed poor diffraction, limited to 4.5-5.5Å resolution. In addition, diffraction patterns often showed streaky spot shapes and a high background of diffuse scattering, indicating crystal disorder. No diffraction data of sufficient quality for structure elucidation could be obtained. The subsequent success in solving the structure of I65-I70 shows that it is not necessarily only the size of the molecule that is decisive for the successful crystallization and structure solution, but also the domain composition. This observation is further confirmed by the crystallization of I65-I70$_{\Delta}^{566E}$, where the removal of a single residue at the C-terminus led to a complete change in crystal habit as well as loss of any diffraction (unpublished results).

Yields of soluble protein from expression in bacterial culture are usually not an issue, often exceeding 10mg/L of culture, probably due to the fact that these intracellular Ig do not require disulfide bridge formation for proper folding and stability. Protein purification following standard protocols yields pure, homogeneous proteins that are stable over months. Rather, it is the construct length and boundaries that seem to be determinant for successful growth of diffraction quality crystals. With this in mind we have recently designed the subfragments I101-I105 and I102-I106 and cloning is in progress. This work is still in the initial stages but we expect that these constructs will prove more amenable to the growth of diffraction quality crystals and subsequent structure elucidation.

4 The M-line Signalosome

In this chapter we present the structural and biophysical characterization of MuRF1 and its binding to titin domains A168-A170, adjacent to TK, necessary for its recruitment to the signalosome. Furthermore, the structure elucidation of the twitchin construct covering FnIII-Kin-Ig, corresponding to titin domains A170-TK-M1, for the first time gives us a structural model integrating the kinase within the context of the filament.

4.1 Introduction

4.1.1 Titin kinase

TK is an autoinhibited Ser/Thr kinase of the CAMK-family that shows close homology both in sequence and structure (Mayans et al., 1998) to other giant protein kinases like myosin light chain kinase (MLCK), and invertebrate titin homologs like projectin and twitchin (Bullard et al., 2002) (see Figure 1.5). Remarkably, even the arrangement of the neighboring domains is conserved, giving rise to a common pattern of Ig-Ig-FnIII--Kin-reg-(Ig). The kinase domain is linked on either side to its neighboring domains with unique sequences of varying length, usually between 20-60 residues. The C-terminal linker contains an auto-regulatory sequence (reg) that blocks the active site and keeps the kinase inactive by preventing substrate as well as ATP-binding (Heierhorst et al., 1996a; Heierhorst et al., 1994; Hu et al., 1994b; Kobe et al., 1996; Mayans et al., 1998).

TK exhibits an unusual two-step activation mechanism (Mayans et al., 1998). First, the auto-regulatory sequence that blocks the active site of the kinase has to be removed, which can be achieved by Ca^{2+}/calmodulin binding (Mayans et al., 1998). In a second step the Tyr residue Y170 in the P+1 loop must be phosphorylated (Mayans et al., 1998). TK was shown to be activated by extracts from developing myocytes, but not mature muscle, indicating that TK is phosphorylated by a developmentally regulated kinase, which has yet to be identified. Expression of a constitutively active form of TK in C2C12 myocytes demonstrated that titin phosphorylates telethonin, which in fully differentiated myofibrils is >1μm away at the Z-disc. In developing myocytes though the C-terminus of titin and telethonin can be detected in dot like aggregates clustered on stress-fiber like structures (Mayans et al., 1998).

Recently it has been proposed that TK might act as a stress sensor, activated by stretch-induced structural rearrangements of the autoregulatory tail (Gräter et al., 2005). Molecular dynamics simulations of stretch based TK unfolding suggested that the auto-regulatory tail indeed undergoes a number of conformational rearrangements leading to the release of inhibition by opening the catalytic cleft, while the catalytic core of the molecule retains its structural integrity necessary for activity (Gräter et al., 2005).

Yeast two-hybrid (Y2H) screening of a TK construct designed to mimic this mechanically activated kinase identified the signalosome, composed of nbr1, p62 and MuRF2, which specifically assembles around this potentially semi-opened state of TK (Lange et al., 2005) and appears to function as a link between sarcomere activity and the regulation of gene expression and protein turnover.

4.1.2 MuRFs

MuRF1 was identified in a Y2H screen as a protein that interacts with the titin kinase region (Centner et al., 2001). Two further family members, MuRF2 and MuRF-3, were found to hetero-oligomerize with MuRF-1 through coiled-coil domains (Centner et al., 2001). MuRF3 (originally termed muscle ring finger protein) had previously been identified as a protein whose expression is necessary for skeletal myoblast differentiation and myotube fusion (Spencer et al., 2000). MuRF1 had also been simultaneously described as Striated Muscle Ring Zinc finger (SMRZ, Dai and Liew, 2001) and shown to bind to SUMO3. To date, members of this family have only been reported in vertebrates (Centner et al., 2001).

4.1.2.1 Conserved features of the MuRF family

The MuRF protein family belongs to the RBCC/TRIM protein superfamily, whose members share an evolutionarily conserved tripartite motif organization consisting of up to three conserved Zn-binding domains, typically a RING finger, a B-box type1 (B1) and a B-box type 2 (B2), followed by a coiled-coil (CC) fraction (reviewed in Meroni and Diez-Roux, 2005). In the case of the MuRFs an N-terminal RING finger is followed by a unique MuRF-family specific motif (MFC), a single B2 box, a CC domain and an acidic tail (see Figure 4.1). While RING domains can be found in a range of protein families, the B-Box is the critical determinant of the TRIM/RBCC family. Strikingly, the combination and order of the domains is strictly conserved from N-terminus to C-terminus, as is the spacing between adjacent domains (Reymond et al., 2001; Torok and Etkin, 2001). This suggests that this structure is being selectively maintained as it is optimized to carry out a basic function common to all members of the family.

MuRFs are encoded by three different genes, located in distinct loci in the mouse and human genomes, that exhibit very high conservation. MuRF-1 shares 62% sequence identity (seq. id.) with MuRF-2, and 77% with MuRF-3, while MuRF-2 and

MuRF-3 show 65% identity (Centner et al., 2001). The conservation is especially high in the N-terminal region, where RING, MFC and B2 share approximately 81% seq. id. while the CC fraction is reasonably well conserved (36% seq. id.). The most divergent areas are the N- and C-termini that vary in length and lack any clear similarity (see Figure 4.1; values for conservation calculated by us). Three additional spliceoforms of MuRF2 and one of MuRF-3 have been identified and are characterized mostly by the presence of C-terminal extensions of unrelated structural nature (Centner et al., 2001; Pizon et al., 2002). The small, cardiac specific isoform of MuRF-2, MuRF-2p27 lacks the coiled-coil regions in its C-terminus (Pizon et al., 2002).

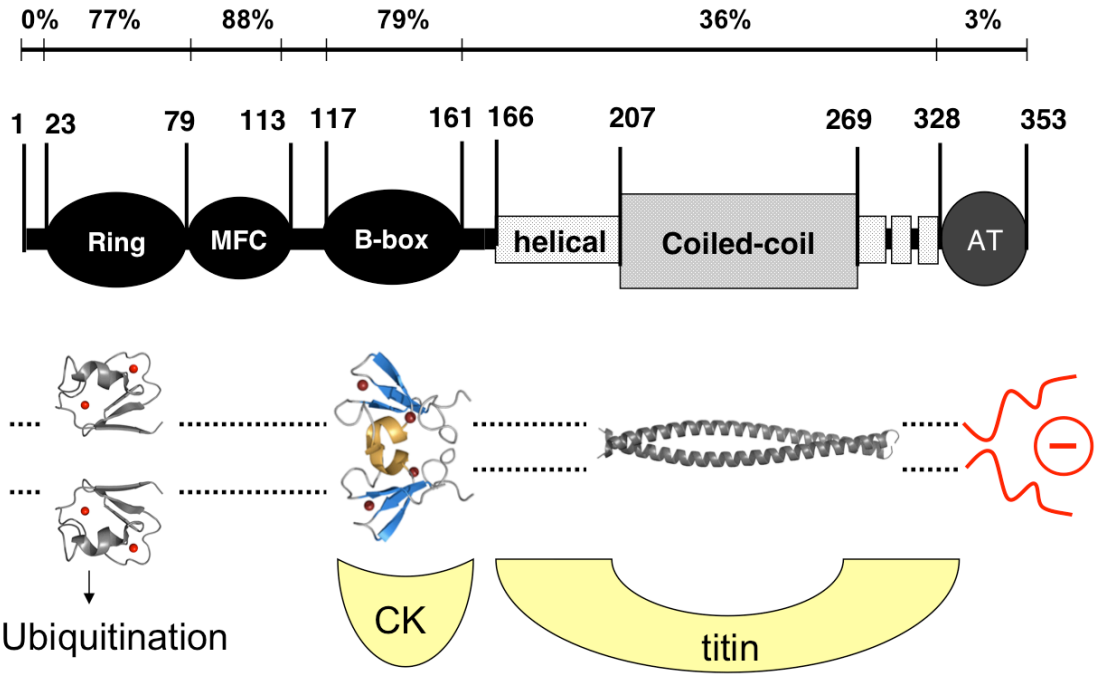


Figure 4.1: Conservation in the MuRF family
Domain structure of MuRF1, with domain boundaries indicated. Above, the sequence identity of the individual domains among the family members is indicated (Mayans & Labeit, 2009, "TRIM proteins". Landes Biosciences, Ed. G. Meroni, *in press*). Underneath, the structure of the B-Box is shown (3DTT; Mrosek et al., 2008). A representative model is presented in grey for the RING domain and coiled coil fraction. The B-Box domain represents a binding site for creatine kinase (CK, muscle isoform) (Koyama et al., 2008), while the interaction with titin has been mapped to the helical fraction (Centner et al., 2001). (Figure prepared by O. Mayans)

In TRIMs, RING motifs are often associated with ubiquitin ligase (E3) activity (Bodine et al., 2001; Kedar et al., 2004; Lorick et al., 1999), B-Box domains appear involved in establishing heterologous protein interactions (Koyama et al., 2008; Peng et al., 2000; Short et al., 2002) and the CC fraction is generally assumed to regulate self-association (Javanbakht et al., 2005; Reymond et al., 2001; Spencer et al., 2000).

However, this functional demarcation is not strict. CC domains also recognize and bind cellular targets, as in MuRF-3 where it targets the protein to microtubules (Spencer et al., 2000) or in MuRF-1 where it binds to titin (Centner et al., 2001; Mrosek et al., 2007). Similarly, B-Boxes and RING fingers can self-associate, as seen for the B2 of MuRF-1 (Mrosek et al., 2008). As for the MuRF-family conserved (MFC) motif, it has been implicated in helping to direct protein interaction (Witt et al., 2005), but neither its fold nor function are known to date, but the very high sequence conservation hints at evolutionary constraints within this family.

4.1.2.2 Physiological roles and cellular targets of MuRFs

MuRF1 is constitutively expressed during development, and found in the cytoplasm, as part of the myofibril, co-localizing with TK in the M-line and in the Z-disc, (see Figure 1.7), (Centner et al., 2001) and in the cell nucleus (Mcelhinny et al., 2002). Its transcription in skeletal muscle is upregulated by a range of signals, all associated with myopathy, namely diabetes, microgravity, denervation, immobilization, TNF-α and dexamethasone (Adams et al., 2008; Bodine et al., 2001; Hirner et al., 2008). Metabolic factors, like starvation or amino acid deprivation also induce MuRF-1 transcription in skeletal muscle (Koyama et al., 2008).

MuRF1 possesses intrinsic ubiquitin ligase activity (Bodine et al., 2001). In addition, MuRF1 null mice are resistant to atrophy and MuRF1 levels are upregulated in skeletal muscle during the atrophic state. Hence, MuRF1 is thought to be involved in ubiquitin-depended degradation of muscle protein (Bodine et al., 2001). Depending on the system used (cultured tissue or live mice) and the mode of induction (e.g. creatine kinase promoter driven expression, non-physiological cytomegalovirus driven overexpression, induction by denervation, amino acid deprivation or stress-induction by TNFα injection) different preferential targets have been reported for MuRF1 (Adams et al., 2008; Hirner et al., 2008; Kedar et al., 2004; Koyama et al., 2008). Further, Y2H studies screened cardiac (Centner et al., 2001; Kedar et al., 2004) and skeletal muscle cDNA libraries (Witt et al., 2005) to find potential ubiquitination targets of MuRF1.

This identified eight myofibrillar proteins as binding partners: **nebulin, titin, NRAP, troponin-I (TnI), troponin-T (TnT), myosin light chain 2 (MLC-2), myotilin** and **telethonin**. Mating studies indicated that all proteins are targeted redundantly by both MuRF1 and MuRF2 The ubiquitination of titin (in its A-band region), nebulin,

myosin light chain-2 (MLC-2), and troponin could be detected in Western Blots of skeletal muscle extracts (Witt et al., 2005). In addition, a second group of proteins targeted by MuRFs included metabolic enzymes.

A summary of proteins found to interact with the MuRFs *in vivo* and *in vitro* is given in Table 4.1.

Table 4.1: MuRF proteins and their interaction partners

Protein	Expression	Localization	Interactions	Proposed function
MuRF-1	constitutive during development	Myofibril (Z-disc, M-line) Nucleus	titin A168-A169 MuRF-1,2,3	Ubiquitin ligase activity
	up-regulated in response to atrophy		SUMO-3, GMEB-1 Ubc-9, ISOT-3	Degradation of muscle proteins during atrophy
			ATP Synthase, creatine kinase	Thick filament assembly
			Rack1 Nebulin	M-line integrity
			N-RAP TnT	myofibril maintenance/turnover
			MLC-2 Myotilin Telethonin (T-cap) cardiac troponin I (cTpl) SRF UbcH5a,b,c	Regulator of energy metabolism
MuRF-2	down-regulated during development	Stable microtubules (Glu-MTs) Myofibril (M-line) Nucleus (upon mechanical inactivity)	titin A164-A169 MuRF-1,2,3	Stability of stable microtubules and intermediate filaments
			p62/SQSTM1 Nebulin	adaptor between titin and microtubules
			N-RAP TnT MLC-2	critical for myoblast fusion and myofibrillogenesis
			Myotilin Telethonin (T-cap)	M-line integrity
				Component of M-line stretch sensor
MuRF-3	up-regulated during development	Stable microtubules (Glu-MTs) Myofibril (Z-disc, M-line)	MuRF-1,2,3	establish/maintain stable Glu-MT array
			Microtubules UbcH5a,b,c	required for myoblast fusion and myogenesis, gene expression and differentiation ubiquitin ligase activity

adapted from (Gregorio et al., 2005)

MuRF2 and MuRF3 show a different expression pattern and subcellular localization than MuRF1 (See Table 4.1). They have both been found associated to glutaminated microtubules (Pizon et al., 2002; Spencer et al., 2000) and their expression appears to be necessary for differentiation. Indeed, knock-down of MuRF2 in skeletal myoblasts delayed myoblast fusion and myofibrillogenesis (Mcelhinny et al., 2004), while ablation of MuRF3 in cell cultures inhibits myoblast differentiation and myoblast fusion by inhibiting the expression of the early myogenic regulatory factors, MyoD and myogenin (Spencer et al., 2000).

Based on their high similarity to MuRF1, it could be expected that MuRF2 and MuRF-3 also act as ubiquitin ligases. Indeed, both MuRF1 and MuRF3 have been shown to be involved in the controlled degradation of β/slow MHC and MHCIIa proteins in striated muscle through ubiquitination (Fielitz et al., 2007). On the other

hand, while MuRF1 and MuRF2 seem to target many proteins redundantly (Witt et al., 2005), there has been no direct evidence so far for MuRF2 ubiquitin ligase activity.

The interaction of MuRF1 with **SUMO-3** (Dai and Liew, 2001), an ubiquitin-related protein implicated in subcellular targeting and nuclear import (for a review see Melchior, 2000), **Isopeptidase T3** (Mcelhinny et al., 2002), a member of an enzyme family responsible for isopeptide bond cleavage, and **UBC9** (Mcelhinny et al., 2002), an enzyme involved in SUMO modifications that catalyzes the formation of an isopeptide bond between SUMO and the target protein, hint at a possible use of sumoylation to regulate the subcellular location, nuclear import and their influence on gene expression.

Immunolocalization studies show that MuRF1 is primarily located in the sarcomeric M-line (Hirner et al., 2008; Mcelhinny et al., 2002), suggesting that anchoring of MuRF1 to M-line titin is a predominant event *in vivo*. Concomitantly, expression of MuRF1-GFP in transfected cardiac myocytes or its central region disrupted titin M-line integrity and also affected thick filament components, but neither the N-terminus or I-band regions of titin, the Z-disc lattice or thin filaments (Mcelhinny et al., 2002). Surprisingly, knockout of MuRF1 in mice showed no primary effect in myofibril integrity, but animals were resistant to muscle atrophy (Bodine et al., 2001).

Thus, while the interaction of titin and MuRF1 clearly plays an important role in the maintenance and regulation of sarcomeric integrity, many questions regarding the exact nature, regulation and outcome of this interaction remain unclear. So far, no evidence has been found for post-translational modification of MuRF1 by titin kinase. Preliminary phosphorylation studies *in vitro* with recombinant TK and MuRF1 proteins were unsuccessful in demonstrating specific phosphorylation of MuRF1 by titin (unpublished results, mentioned in Centner et al., 2001). A second possibility is that titin might be an ubiquitination target of MuRF1, which either labels it for degradation or modulates its biomechanical signaling properties. Again, no evidence that support this view has been gathered to date. While it could be shown that titin is ubiquitinated *in vivo* (Mcelhinny et al., 2002), no change in the ubiquitination levels of titin could be detected after MuRF1 knockout (Adams et al., 2008; Mcelhinny et al., 2002). Alternatively, titin might sequester MuRF1 from the cytoplasm, preventing it from acting on its cellular targets in unchallenged muscles. Future studies will be required to unravel the relation between mechano-signaling, titin kinase and the role of MuRF1 in muscle remodeling.

4.2 Biophysical characterization of MuRF1

4.2.1 Introduction

The MuRF family members mediate a wide variety of protein-protein interactions (see section 4.1), but it is unknown which specific domains or motifs in these proteins are involved in contact formation. This is complicated by the challenge to identify how the similarities and differences in their molecular targeting and function correlate with the pattern of conservation exhibited by the different domains (see section 4.1).

To date, structural data have only become available for the B2 of MuRF1 (Mrosek et al., 2008); work performed in our laboratory before the initiation of this study). The structure reveals that this domain, ~ 40 residues in length, is a representative of the CHC_2C_2HC B2 subtype and a member of the RING-like family. In this domain, an α-helix (α1) packs against a small three stranded β-sheet (β1-β3) that together support the three loops (L1-L3) that are involved in the ligation of the two Zn^{2+} atoms (see Figure 4.2). Zinc coordination follows the classical cross-brace topology of RING domains, where binding site I (ZnI) is formed by the first and third protein ligand pairs and site II (ZnII) by the second and fourth ligand pairs.

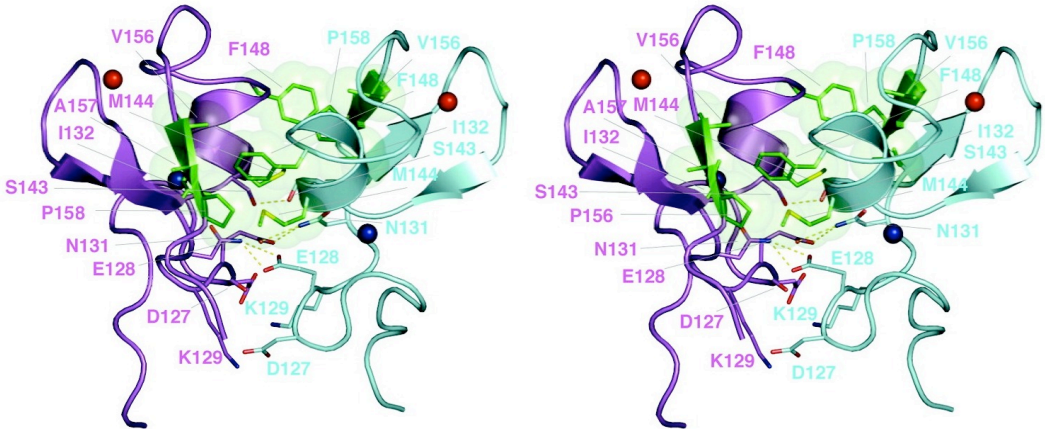


Figure 4.2: Stereo image of B-Box dimer
Stereoview of the interface contacts in the crystal structure of dimeric MuRF1 B2. ZnI is represented as blue, ZnII as a red sphere. Hydrophobic groups are colored green and highlighted using van der Waals spheres. Polar interactions are represented as dashed lines (Mrosek et al., 2008)

MuRF1 B2 forms high-affinity dimers as shown by crystallographic, NMR and ultracentrifugation data (see Figure 4.2) (Mrosek et al., 2008). The dimer interface is formed by the α-helix of one subunit docking into concave depression that is formed between the α-helix and β-sheet of the second subunit. The interface contains clusters

of both polar and hydrophobic interactions. So far, this represents a unique association pattern between RING-like folds, but the interacting groups are relatively well conserved in the B2 family, suggesting a possible similar association of other B2 Boxes.

Analysis of the surface topography and comparison with RING domains with known ligase activity reveal that MuRF1 B2 is unlikely to act as an E3 ligase.

In MuRFs, the B2 domain is followed by a long stretch that is predicted to be α-helical and in its middle part form a dimeric parallel coiled-coil based on structure prediction analysis (Jpred software, Cuff et al., 1998; COILS, Lupas et al., 1991). This is supported by CD data on a correspondingly truncated MuRF-1 variant, MuRF1$^{176-327}$. An estimation of the secondary structure content of this sample based on the obtained spectra suggested an approximate α-helical content of 70% and 30% random coil (fitting parameter R^2=99.8%) (Mrosek et al., 2007).

This domain is not only responsible for self-association, but also mediates the binding of MuRF1 to titin (Centner et al., 2001; Mrosek et al., 2007; Witt et al., 2005) (see section 4.3). In MuRF3 on the other hand the α-helical region has been shown to be necessary for targeting to microtubules. For MuRF2 no such data is available. Compared to the N-terminal region comprising the RING, MFC and B2 domains, the α-helical region exhibits lower sequence conservation. In this region, most likely, the important residues for the structural stability of the CC were maintained while other residues were free to diverge in sequence to accommodate the differing binding affinities exhibit by the MuRFs to their targets.

This study focused on the biophysical and structural characterization of MuRF1, especially in relation to its interaction with titin A168-A170. Since full length MuRF1 has proven challenging to express in soluble form in bacterial systems in quantities necessary for structural and biophysical characterization, we concentrated on fragments of the α-helical region and the neighboring B-Box, which is also expected to be involved in protein targeting.

4.2.2 Experimental work

4.2.2.1 Constructs:

The coding sequence for full length MuRF1 (Swiss-Prot Q969Q1) cloned into pET-9d (Novagen) was provided by Prof. Siegfried Labeit and PD Dr. Dietmar Labeit (Institut für Anästhesiologie und Operative Intensivmedizin, Universitätsklinikum Mannheim, Germany). A range of subfragments was designed, cloned by PCR amplification (list of primers used is given in appendix 6.1) and inserted into pETM-11 (EMBL vector collection), a variant of pET-24d (Novagen) including an N-terminal His_6-tag and a TEV (tobacco etch virus) protease cleavage site prior to the inserted gene. Constructs were chosen to cover the region of the protein interaction interface, i.e. the B-Box and α-helical fraction. Constructs MuRF1$^{176-327}$ and MuRF1$^{176-353}$ had been cloned for a previous study characterizing the binding of MuRF1 to A168-A170 *in vitro* (Mrosek et al., 2007) and were available in our group. They contained a C293S mutation to prevent oxidation during sample storage, which was replicated for other constructs. For ease of purification (using affinity techniques), the original clone of full length MuRF1 was recloned into pETM-11 and truncated short of the acidic tail, to improve crystallizability. Also, it had been shown that the acidic tail did not have an influence on the binding affinity between A168-A170 and MuRF1 (Mrosek et al., 2007). The MuRF1$^{117-327}$ construct was chosen to cover both the previously characterized B-Box domain (Mrosek et al., 2008) with the α-helical region covered in MuRF1$^{176-327}$. This construct juxtaposed two oligomerization motifs that, individually, were shown to be dimeric in solution (Mrosek et al., 2007; Mrosek et al., 2008) but in combination led to the formation of higher order assemblies. MuRF1$^{117-271}$ covers the B-Box and the predicted CC in the middle of the α-helical region, stopping just after a helix-breaking Pro (P271), with the hope that this would still provide binding to titin, but remediate some of the issues seen in crystallization (see section 4.2.2.6) due to its more compact shape and predictably well defined termini without any flexible loops. MuRF1$^{214-271}$ finally covers only the part predicted to be coiled-coil and was cloned with the expectation to avoid issues in crystallization seen in larger fragments and be small enough for NMR study/titration of its binding (work in progress) to titin A168-A170. Figure 4.3 gives an overview of the constructs and Table 4.2 summarizes outcome of experimental work broken down by construct.

A construct for MuRF2$^{118-329}$, cloned into pETM-11, was again provided by D. and S. Labeit. This construct shows 54% overall sequence identity with the corresponding

construct MuRF1$^{117-327}$ (78% for B-Box and 49% for the α-helical fraction). Thus, it is expected to adopt a very similar structure, where sequence variations will be determining the differential protein targeting of the two paralogs. In order for the different MuRFs to hetero-oligomerize through mixed coiled formation (see Centner et al., 2001), these regions would need to have a compatible heptad repeat organization. Indeed, in sequence alignments, the residues in positions **a** and **d** of the heptad repeat involved in hydrophobic core formation are conserved between all three isoforms.

Also, sometimes even only a few different residues can lead to great variation in behavior during crystallization trials and thus it seemed to a good strategy to pursue both constructs for structure elucidation. Given the structures of the two constructs, or even the structure of one in combination with threading algorithms it should be possible to map the sequence variations and correlate them to the differential binding affinities reported for titin. Furthermore, given a protocol for complex formation and purification for MuRF-1$^{117-327}$ and A168-A170 (see chapter 4.3), MuRF2$^{118-329}$ could have been assayed for binding to A168-A170.

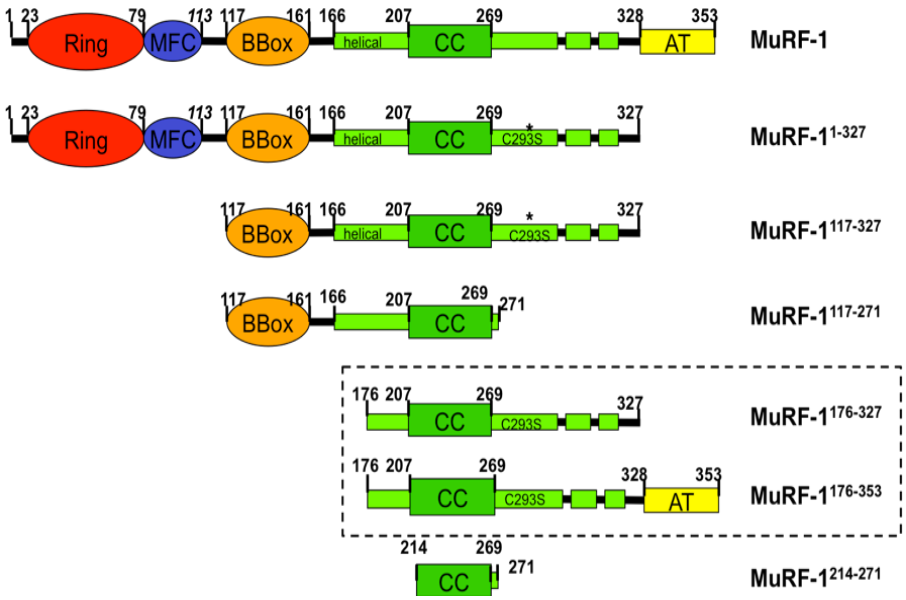


Figure 4.3 Overview of MuRF1 constructs employed in this study
MuRF1 can be divided into several domains, which was exploited in the design of sub-constructs. All sub-constructs were cloned into pETM-11. The two framed constructs had been cloned for a previous study and included the mutation C293S to prevent oxidation *in vitro* (Mrosek et al., 2007).

Table 4.2: Overview of experimental outcome for MuRF1 constructs

Construct	Expression	Purification, yield of pure protein	Crystallization	Conclusions
MuRF1$^{1-327}$	Good 100% insoluble	Purification under denaturing conditions, refolding not successful	Not screened	No soluble protein
MuRF1$^{176-327}$	Good 5-10% insoluble	~ 15mg/L of culture	Screened No crystals gelatinous precipitate	Dimeric in solution, probably parallel coiled-coil formation
MuRF1$^{176-353}$	good, 5-10% insoluble	~ 15-20mg/L of culture	Screened No crystals gelatinous precipitate acidic tail leads to increased solubility	Acidic tail unnecessary for titin binding, increases solubility, potential issues with disorder
MuRF1$^{117-327}$	Good 10-30% insoluble	5-10mg/L of culture, good purity but multiple oligomeric species	Screened Single crystals gelatinous precipitate no diffraction	Multiple higher order oligomeric species in solution Crystals disordered, no diffraction
MuRF1$^{117-271}$	Good 20-30% insoluble	5-10mg/L of culture	Screened no crystals gelatinous precipitate	Very similar in behavior to MuRF1$^{117-327}$ no improvement by truncation of a-helical part to C-ter of CC

4.2.2.2 Protein production:

Production of MuRF1 and MuRF2 constructs employed *E. coli* BL21(DE3) Rosetta2 cells and followed the procedure in (Mrosek et al., 2007; Mrosek et al., 2008). For protein expression of MuRF1 and MuRF2 fragments, the relevant recombinant plasmid was transformed into *E. coli* BL21(DE3) Rosetta2 (Novagen). Cultures were grown at 30°C up to an OD_{600} of 0.6 in Luria Bertani medium supplemented with 25 μg/ml kanamycin and 34 μg/ml chloramphenicol. Expression was induced by addition of isopropyl-β-D-thiogalacto-pyranoside (IPTG) to a final concentration of 0.5 mM. Where constructs contained the B-Box Zn^{2+}-finger domain, 50μM Zn_2Cl was added to the bacterial culture at time of induction to obtain soluble protein. Cultures were further grown at 20°C for approximately 18 additional hours. Cells were harvested by centrifugation at 2800g and 4°C. Bacterial pellet was resuspended in 50mM Tris pH 8.0, 300mM NaCl, 2mM β-ME, containing a protease inhibitor cocktail (Roche) and DNAse I and lysed by sonication or French press. The homogenate was clarified by centrifugation at 15000g and 4°C. Affinity purification used a HistrapHP column (GE Healthcare) equilibrated in lysis buffer in conjunction

with a stepwise imidazole gradient employed during column wash steps to remove impurities. Elution used 50mM Tris pH 8.0, 100mM NaCl, 2mM βME, 200mM imidazol. Tag removal was by incubation with TEV protease overnight at 4°C during dialysis against 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM βME. Subtractive purification for protease (containing uncleavable His_6-tag at C-terminus) and non-digested protein was carried out on a subsequent chelating step. As a final polishing step to remove aggregates the protein was subjected to gel filtration on a Superdex 200 HiLoad 16/60PG column (GE Healthcare) equilibrated in dialysis buffer (Figure 4.4). Protein purity was assessed by SDS-PAGE.

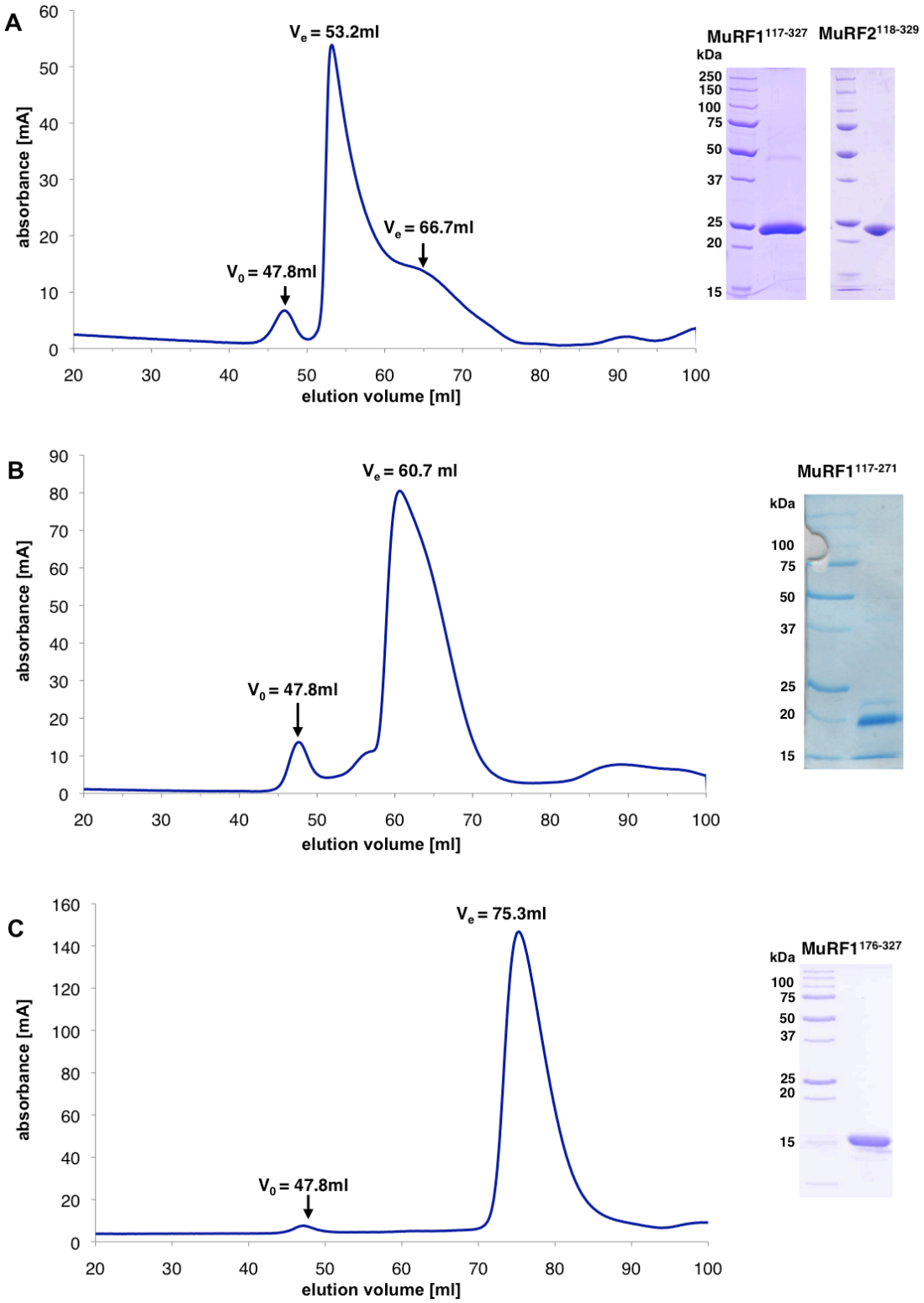


Figure 4.4 Size exclusion chromatograms of MuRF1 constructs

Representative size-exclusion chromatograms of MuRF1 constructs, performed using a Superdex200 16l60 column (GE Healthcare) equilibrated in 50 mM Tris pH8.0, 100mM NaCl. The signal shown corresponds to A_{280nm}. A) Chromatogram of MuRF1$^{117-327}$ (monomeric theoretical Mr=24051 Da); B) MuRF1$^{117-271}$ (monomeric theoretical Mr=17965 Da); C) MuRF1$^{176-327}$ (monomeric theoretical Mr=17350 Da). The final purity of corresponding samples according to SDS-PAGE is shown on the right; as a Mr standard Precision Plus Protein standard (Biorad) was used.

With the exception of MuRF1$^{1-327}$, all constructs were expressed as soluble products, with some percentage (5-40%) of the protein going into inclusion bodies. The amount of pure soluble protein obtained per 1L of E. coli culture showed some considerable variation between the different constructs, ranging from 5-20mg (generally, constructs containing only the helical part expressed better than fragments that contained the B-Box in). One problem encountered with this protein is its apparent tendency to bind to membranes used in devices for protein concentration (Amicon concentrator with Ultracel regenerated cellulose). Generally the amount of protein recovered from the concentrator after careful flushing of the membranes was 10-25% lower than the protein that was "applied". Thus, care was taken not to dilute the sample unnecessarily, to keep concentration steps as brief as possible with regular resuspension of samples, and to use concentrators with small membrane surfaces. Further, during concentration samples containing the B_2-Box together with the α-helical region became viscous at relatively low concentrations (in the range of 5mg/ml as estimated by A_{280})

In size exclusion chromatography MuRF1$^{117-327}$ eluted at 53.2ml. Assuming a globular shape for the molecule this would correspond to a 30mer according to the calibration curve for this column (see appendix 6.2). Since the molecule contains a sizeable stretch predicted to form a dimeric CC, it is likely though that the protein adopts an extended, rather than globular shape. This has to be taken into account in molecular weight estimation since the hydrodynamic radius will be far larger. Moreover, the asymmetric shape of the elution peak is caused by underlying minor species. Based on sample purity in SDS-PAGE this peaks are indicative of multiple oligomeric states rather than co-eluting impurities.

MuRF1$^{117-271}$ eluted at 60.7ml, which would be indicative of a 30mer if assuming a globular shape. Again the peak profile is asymmetric, but less pronounced than for the longer construct. The peak elution volume of 75.3ml for MuRF1$^{176-327}$ corresponds to a 15mer. The peak shape of MuRF1$^{176-327}$ is quite symmetrical, indicative that this construct exists as a single species in solution, most likely as a dimer, induced by the central coiled-coil domain. Thus, in order to assess the oligomeric state accurately size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) was employed, since this method gives accurate molecular weight measurements independently of hydrodynamic radius.

4.2.2.3 Determination of quaternary structure of MuRF1 constructs

Samples used in SEC-MALS measurements had previously been purified by size-exclusion chromatography to remove any impurities and aggregates. This is important since the light scattering signal is directly proportional to the weight-averaged molar mass (Mr) of the sample being measured, multiplied by the concentration, and thus is very sensitive to the presence of aggregates.

Measurements were performed on an ÄKTA basic system (GE Healthcare) connected to a tri-angle static light scattering detector and a differential refractometer (miniDAWN Tristar and Optilab, respectively). A Superdex 200 HR 10/300 GL column (GE Healthcare) was used in 50 mM Tris pH8.0, 100 mM NaCl (chromatography buffer) at a flow rate of 0.7 ml/min. Sample volumes of 100 µl were injected at a concentration of 2 mg/ml. A specific refractive index increment (dn/dc) of 0.185 ml/g was used. The data were recorded and processed using ASTRA software (Wyatt Technology). To determine the detector delay volumes and normalization coefficients for the light scattering detector, a BSA sample (Pierce) was used as reference.

Following the protocol outlined above, SEC-MALS measurements for MuRF1$^{176-327}$ yielded a Mr of 32.3 kDa (protein concentration estimated from refractive index), which is in excellent agreement with the calculated Mr of 34.7 kDa for a dimer of this sample (Figure 4.5). Combined with the CD data showing the high α-helical content (70%) of this fragment, this data suggests that there is a very high chance that this fragment really forms a dimeric coiled coil in solution (see Mrosek et al., 2007). Measurements of the molecular weight of MuRF-1$^{117-327}$ was hindered by the fact that *i)* size-exclusion of this sample is not successful to separate the individual species in the polydisperse sample, *ii)* the elution volume is close to, but not identical with, the exclusion volume of the column and measurements therefore show interference from the big signal of even minor amounts of contaminating aggregates. Still, the data allowed the estimation that the sample forms hexameric or octameric assemblies (Figure 4.5). Thus, it appears that the combination of B-Box and CC, both dimeric in isolation (this study and Mrosek et al., 2008), leads to a higher order of assembly.

Thus, in order to prove that this higher order oligomeric state is indeed induced by the presence of the B-Box, we assayed a sample of MuRF-1$^{117-327}$, where the B-Box had been unfolded by metal depletion. For this, the sample was incubated for 18h on a rotary mixer at 4°C in chromatography buffer supplemented with 50mM EDTA pH8.0. In order to prevent disulfide formation of freed cysteines (i.e. residues normally involved in Zn^{2+} coordination in the B-Box) and consequently sample aggregation

10mM β-mercaptoethanol was added to the mixture. The buffer used for the SEC-MALS measurement of the oligomeric state of this sample was modified to contain 1mM EDTA and 2 mM β-mercaptoethanol. The data thus obtained revealed a Mr of 48.7 kDa, in very good agreement with the calculated molecular weight for a dimer, Mr=of 48.1 kDa (see Figure 4.5).

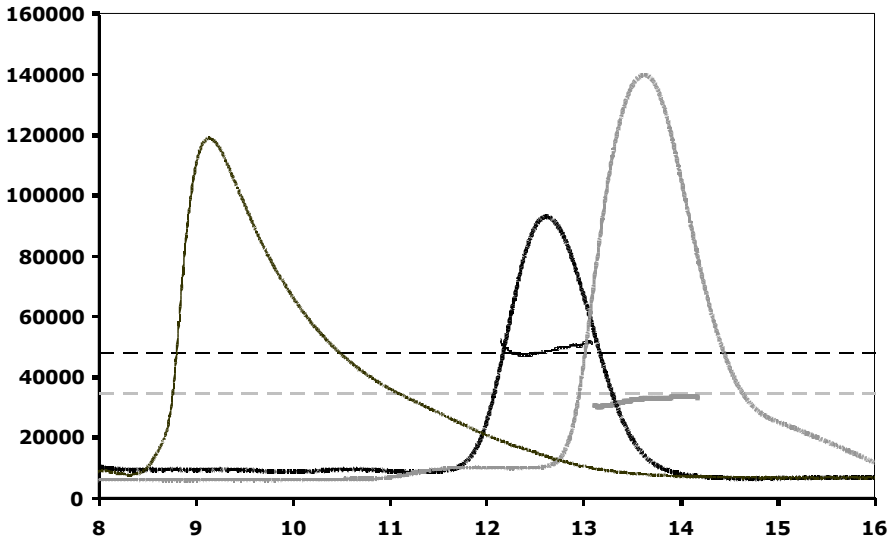


Figure 4.5: Oligomeric states of MuRF1
SEC–MALS measurements of MuRF-1$^{117-327}$ and MuRF-1$^{176-327}$ truncated variants. The average Mr per volume unit (dots) and the normalized refractive index (solid lines) are shown. The theoretical Mr values of the dimeric states of MuRF-1$^{117-327}$ and MuRF-1$^{176-327}$ constructs are represented as dashed horizontal lines. Unfolding of B_2 in the MuRF-1$^{117-327}$ variant was achieved by mixing with 50 mM EDTA and 10 mM β-mercaptoethanol for 16 h at 4 °C.

4.2.2.4 Anionexchange chromatography of MuRF-1$^{117-327}$

In an attempt to separate the multiple species underlying the distorted peak in size exclusion chromatograms MuRF1$^{117-327}$ was subjected to ion-exchange chromatography. This was in the hope that the assembly into different oligomeric species might also influence the polarity of the assemblies.

For anion exchange chromatography MuRF1$^{117-327}$ was taken after tag-removal and reverse purification and dialyzed against 50mM Tris pH 8.0, 50mM NaCl. Anion exchange chromatography used a MonoQ 5l50GL column (GE Healthcare) and a NaCl gradient from 50-525mM NaCl for sample elution. In order to improve resolution, a shallow gradient was applied over 25 column volumes. This succeeded in separating two distinct peaks at salt concentrations of 433 mM and 444 mM, indicating that the underlying species are very similar in their charge profile (Figure 4.6A).

Nevertheless, when the two samples were subsequently assayed by size-exclusion chromatography, they exhibited distinct elution profiles, where the first peak is double the weight of the second peak according to the calibration curve (assuming globular domains, the volumes would correspond to a 28mer and a 14mer). This confirmed that ion exchange was indeed successful at separating two distinct species, which do not interconvert in a dynamic equilibrium (Figure 4.6C). Still, the two species did not display appreciably differing behavior in either crystallization experiments or binding experiments with titin (see chapter 4.3) when compared with each other or with samples that had not been subjected ion chromatography. Thus, anion-exchange chromatography was not further pursued.

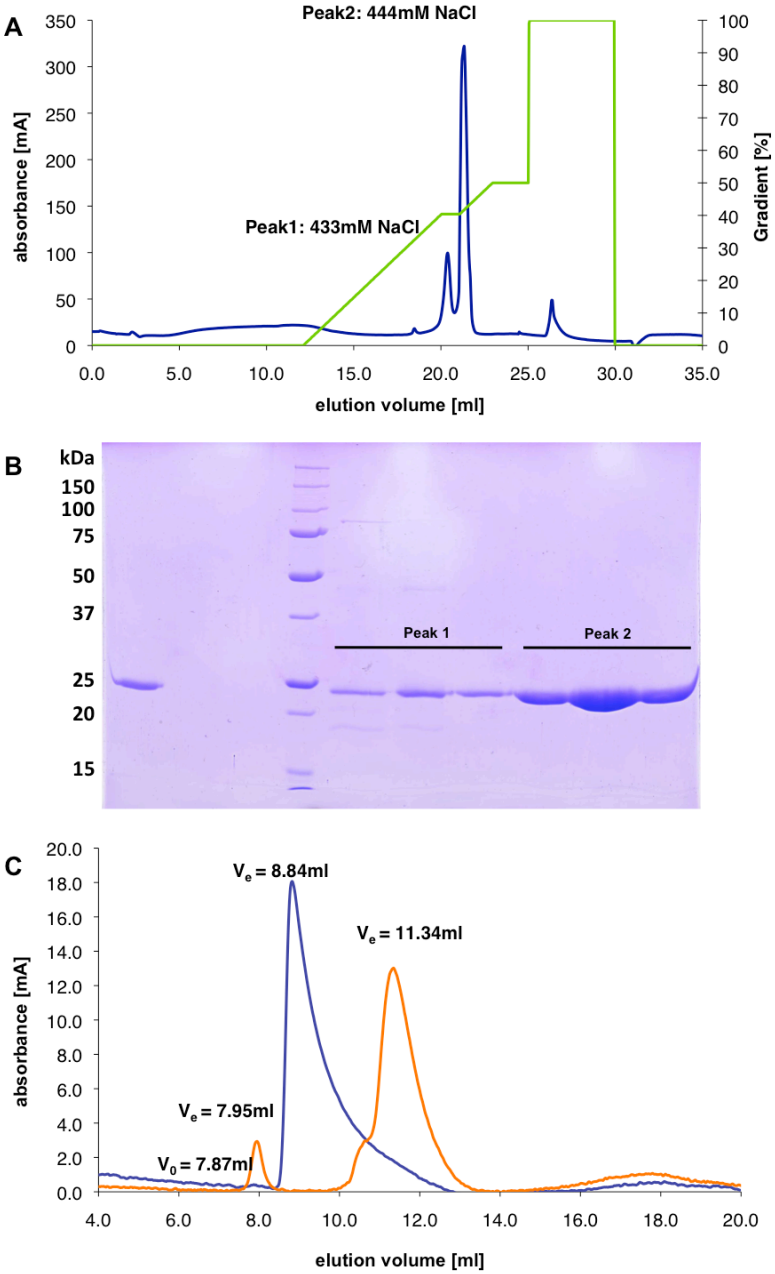


Figure 4.6: Resolution of MuRF1$^{117-327}$ molecular species
A) Anionexchange of MuRF1$^{117-327}$ was performed in Tris pH 8.0 with a NaCl gradient ranging from 50-525mM for protein elution (indicated by the green line) on a MonoQ 5I50GL column (GE Healthcare). The salt gradient was held during peak elution to improve resolution. The protein eluted in two distinct peaks, at 433mM and 444mM NaCl respectively. The signal shown corresponds to A_{280nm}
B) SDS-PAGE of anion exchange samples. The left lane shows the sample as loaded onto the ion exchange column. The first peak shows some impurities that co-elute with MuRF1$^{117-327}$, probably bound to the protein, while the second peak seems to contain only MuRF1$^{117-327}$
C) Size exclusion chromatography of the two fractionated samples obtained in ion exchange chromatography using a Superdex 200 HR 10I300 GL column (GE Healthcare). The orange curve corresponds to A_{280nm} of the peak 1 sample from ion exchange, while the blue curve corresponds the second peak.

4.2.2.5 Expression and purification of MuRF1$^{1-327}$

We attempted to express and purify MuRF1$^{1-327}$ as a representative for full-length MuRF1, since experiments had shown that the acidic tail had no influence on binding affinity of MuRF1 for titin A168-A170 (Mrosek et al., 2007). Also this stretch of highly charged residues could potentially introduce inhomogeneity into sample populations since it is unlikely to adopt a well-defined conformation.

MuRF1$^{1-327}$ was invariably expressed in inclusion bodies. For purification, bacterial pellet was resuspended and lysed according to the established protocol (see above). Inclusion bodies were harvested by centrifugation and resuspended in lysis buffer supplemented with 8M Urea for 30 min by agitation on a stir plate at room temperature. Affinity purification under denaturing condition again followed the protocol with the exception that all buffers were supplemented with 8M Urea (later reduced to 6M urea; Figure 4.7A). Refolding was attempted under a range of experimental procedures, *i)* in dialysis at room temperature or at 4 °C using either the refolding buffer employed for the B-Box fragment (Mrosek et al., 2008), 50mM Tris or HEPES, pH 8.0 150mM NaCl, 2mM $ZnCl_2$, 10mM β-mercaptoethanol, or *ii)* using an "on column" refolding protocol for denatured samples immobilized onto the Ni-agarose beads (Oganesyan et al., 2005), washing the beads first with refolding buffer (50mM Tris, 150mM NaCl, 2mM $ZnCl_2$, 10mM β-mercaptoethanol, pH 7.5) containing 6M Urea, followed by washes with refolding buffer supplemented with 5% glycerol as stabilizing agent, and then eluted using refolding buffer supplemented with 200mM imidazole. Neither refolding method yielded soluble protein, but consistently led to misfolded, precipitated protein (Figure 4.7B). Since this precipitate is often of higher purity than the original sample employed for refolding, samples were solubilized in fresh unfolding buffer and subjected to a second round of refolding, supplementing refolding buffer with an increased concentration of glycerol (up to 10%).

The establishment of a successful refolding protocol often requires extensive screening of buffer compounds, pH, additives as well as refolding method (slow dilution of denaturing agent in dialysis, fast dilution methods or on column refolding). Given that we were successful in purifying and characterizing smaller subfragments that could be proven to be sufficient for high-affinity binding to titin, we decided not to pursue this construct any further.

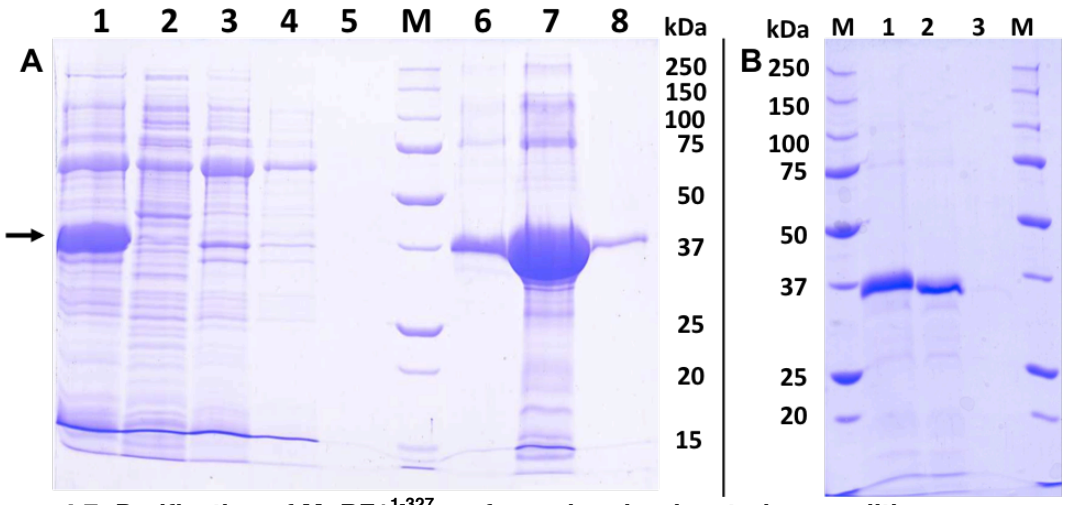


Figure 4.7: Purification of MuRF1$^{1-327}$ performed under denaturing conditions
A), SDS-PAGE of MuRF1$^{1-327}$ metal affinity purification performed under denaturing condition on a Histrap (GE Healthcare), with 6M Urea in all buffers. Lanes: 1) pellet of lysate, 2) Supernatant of lysate, 3) Flowthrough of unbound sample (resolubilized pellet), 4-5) washes, 6-8) elution fractions. The lane marked with M contains the Precision Plus Protein Marker (Biorad). The arrow indicates MuRF1$^{1-327}$.
B) SDS-PAGE of refolded samples is shown. Lanes 1) original sample in denaturing buffer used for refolding dialysis, 2) Pellet of precipitated protein after refolding dialysis, 3) Supernatant of sample dialyzed against refolding buffer.

4.2.2.6 Crystallization

For crystallization protein samples were concentrated to 7-12mg/ml (samples becoming viscous). All crystallization trials were performed using vapor diffusion, where as a standard, unless otherwise mentioned, 1µl of protein was mixed with 1µl of reservoir solution to form a 2µl hanging drop inverted over 1ml reservoir solution in VDX plates (Hampton research).

Screening of **MuRF1$^{176-329}$** and **MuRF1$^{176-353}$** variants (see Figure 4.3) was performed simultaneously for easier comparison in 24 well VDX plates (Hampton Research) and used the sparse matrix kits Hampton Research (HR) CS-I & II and Emerald Wizard I & II (Emerald Biosystems). Screening did not result in any hits for either MuRF1$^{176-329}$ or MuRF1$^{176-353}$ but revealed that the acidic tail increases protein solubility. Further screening concentrated on other constructs MuRF-1$^{117-327}$ and MuRF-1$^{117-271}$, both of which come with the B_2-Box included and with a truncated C-terminus, since ITC binding studies (Mrosek et al., 2007) showed that it does not have an influence on affinity for titin binding.

For initial screening of **MuRF1$^{117-327}$** the sparse matrix kits Hampton Research (HR) CS-I & II, Emerald Wizard I & II and Emerald Cryo I & II, were used. Small

needle-like crystals grew within minutes from HR II-1, containing 2M NaCl and 10% PEG-6000. Further conditions (see Table 4.3) gave rise to small microcrystals, intergrown needles or spherulites over the course of a few weeks and were integrated into refinement of crystallization conditions.

Table 4.3 Crystallization conditions for MuRF1$^{117-327}$

Condition	**Reservoir composition**	**Morphology**
CS II-1	2M NaCl, 10% PEG-6000	Needles
CS II-37	10% PEG-8000, 0.1M HEPES pH7.5, 8% ethylene glycol	Spherulites
Wiz I-10	20% (w/v) PEG-2000 MME, 0.1M Tris pH 7.0	Spherulite
Wiz I-40	10% (v/v) 2-propanol, 0.1M MES pH 6.0, 0.2M $Ca(OAc)_2$	Small plates / needles
Wiz II-8	10% PEG-8000, 0.1M Na/K phosphate pH6.2, 0.2M NaCl	Needles
Wiz II-15	1.26M $(NH_4)_2SO_4$, 0.1M HEPES pH7.5	Spherulites, microcrystals
Wiz II-43	10% PEG-8000, 0.1M Tris pH7.0, 0.2M $MgCl_2$	Needles
Cryo I-7	0.05M $MgCl_2$, 0.1M Tris pH8.5, 40% ethanol	Needles
Cryo II-16	1.0M $(NH_4)_2SO_4$, 0.1M Imidazol pH8.0, 20% PEG-300, 10% glycerol	Microcrystals

CS – Crystal Screen I + II, Hampton Research; Wiz – Wizard I + II, Emerald Biosystems; Cryo I + II – Cryo Screen, Emerald Biosystems

In order to screen as much chemical space with as little sample as possible, protein was shipped to EMBL Hamburg for screening at the crystallization robotics facility. Screening included the screens HR Index, HR Salt RX (both Hampton Research), Jena 1-10 (Jena Biosystems) and the Nextal Classic (Nextal, now Qiagen) in 96-well sitting drop plates (Greiner low profile plates) with a drop size of 200nl protein + 200nl reservoir volume. Unfortunately this did not lead to the identification of any new conditions.

Thus, efforts concentrated on the refinement of the small, fast growing, needle like crystals obtained from condition HR CS II-1, containing 2M NaCl and 10% PEG-6000, and also took into account conditions CSII-37 and Wiz II-43, that all shared PEG as main precipitant. Optimization focused on screening PEG size versus concentration, buffer composition as well as pH, temperature (4°C, 20°C, 25°C, 30°C) and additives (organic solvents, divalent metal salts).

This led to larger needles in the range of 100 μm length (see Figure 4.8C) growing from e.g. 400mM NaCl, 9-11% PEG-6000 at 4°C or 250mM NaCl, 100mM HEPES pH8.0, 8% PEG-6000, 1% butanol, set up at 20°C (Figure 4.8D). Due to the

very fast equilibration and crystallization rate, attempts at seeding were unsuccessful in improving crystal size or habit. Also, crystals were repeatedly not of sufficient quality to yield suitable stock for seeding. (The actual problem here was that the crystals were "chewy" and "sticky" and it was not possible to break them to produce seeds). Unfortunately, crystals were often unexpectedly flexible during handling, bending rather than breaking when applying (moderate) pressure, usually indicative of incorporation of partially disordered proteins and imperfect crystal lattice formation. Further, crystals did not diffract during test exposures at either home source or synchrotron, confirming the suspicion that the crystal did not contain proper order.

Screening of **MuRF2$^{118-329}$** was set up as for MuRF1$^{117-327}$ with the exception that 48-well VDX trays (Hampton research) were used with a reservoir of 200μl. Hits coincided with conditions identified for MuRF1$^{117-327}$ (seq. id. is 54% between the two constructs). Thus, refinement followed largely the same procedure but focused on refinement of the ammonium sulphate condition (Wiz II-15). Refinement of MuRF1$^{117-329}$ microcrystals and spherulites grown from ammonium sulphate led to improved crystal size, but no change in morphology. To date, no single, diffraction quality crystals could be grown for either construct.

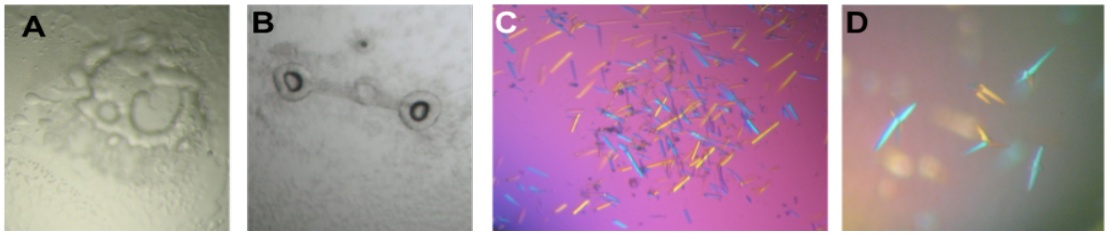


Figure 4.8: Crystallization results for MuRF1 constructs
A) gelatinous/oily protein aggregate of MuRF1$^{117-327}$ 0.5M NaCl, 0.01M Hexadecyltrimethylammonium bromide, 0.01M $MgCl_2$
B) gelatinous protein of MuRF1$^{117-271}$ in 0,1M MES pH6.5, 20% PEG-20'000
C) needles of MuRF1$^{117-327}$ grown at 4C in 400mM NaCl, 10% PEG-6000
D) needles of MuRF1$^{117-327}$ grown in 250mM NaCl, 0.1M HEPES pH8.0, 8% PEG-6000, 1% butanol

In the hope that a shorter construct that stopped right at the end of the predicted coiled coil would potentially be less flexible and better behaved in crystallization, trials for **MuRF1$^{117-271}$** were initiated. Crystallization trials of MuRF1$^{117-271}$ made use of the robotics facility in Daresbury to maximize screening of chemical space and included the Qiagen crystallization screens Classics, ClassicsII, JCSG+, MbClass, MbClassII, MPD, PEGs, PEGsII, pHClearII. Screening was performed in 96 well MRC sitting drop plates at a protein concentration of 12mg/ml, setting drops of 200nl + 200nl over a

reservoir of 70µl. Sadly, screening did not produce any viable leads for optimization. Inspection of crystallization trials showed a clear tendency of the protein to form gelatinous aggregates rather than granular precipitates (see Figure 4.8A and B), an observation that holds true for all constructs assayed to date, and might also be in correlation with the tendency of the samples to become viscous during concentration.

4.2.3 Discussion

MURF1 is one of three paralogs that make up the MuRF family together with MuRF2 and MuRF3. We initiated this study to characterize MuRF1 with a special focus on this protein interaction interface and its targeting of titin A168-A170.

We could show that MuRF-1$^{176-327}$ is present as a dimer in solution, which together with the CD data showing a predominantly α-helical fold for this construct (M. Mrosek), strongly supports the fact that this region assembles into a parallel coiled-coil in its middle region (published in Mrosek et al., 2007). Combination of the dimeric B-Box with the dimeric α-helical region, led to formation of higher-order assemblies.

Unfolding of the B-Box by metal depletion led to reversion to the dimeric state as induced by the α-helical fragment, proving that the B-Box was the determinant of the higher order assembly seen for MuRF1$^{117-327}$. Lacking structural data we cannot resolve the order or mode of association for this higher-order assembly. This case illustrates that it can be precipitous to draw conclusions about the oligomeric state of full-length protein from the assembly characteristics of individual domains. We were hindered in our efforts to determine the oligomeric state of full-length protein by the fact that we could not produce soluble protein recombinantly. It would be interesting to see how the oligomeric state of MuRF1$^{117-327}$ compares to the one of the full-length protein, taking into account that the RING finger also has a propensity to form homo-oligomeric interactions (Borden, 2000).

Crystallization was attempted for all subfragments in a wide range of crystallization conditions, and yielded very few hits. A large fraction of the conditions instead exhibited formation of gelatinous aggregates, possibly correlating with the tendency of samples to become viscous at relatively low concentrations (~5mg/ml) and preventing the formation of ordered interactions necessary for lattice formation and crystal growth. For MuRF1$^{117-327}$ and MuRF2$^{118-329}$ a range of hits was identified and pursued in order to optimize crystal growth. The very fast equilibration and nucleation rate of the crystallization trials prevented successful application of seeding

techniques. To date we have not identified any diffraction quality crystals in test exposures.

While this study did not succeed in its objective to obtain a crystal structure of the titin binding domain of MuRF1, our characterization of oligomeric assembly and the identification of potential regions of flexibility in the acidic tail have led us to design and clone a new construct MuRF1$^{214-271}$, covering the presumed dimeric coiled coil region. Recent work (performed together with B. Franke) has established the purification of this construct and diffraction quality crystals have been obtained (Figure 4.9A). Recently, diffraction data (Figure 4.9B) of native crystals have been recorded to 2.35Å resolution at beamline MX14.1 at BESSY (Berlin, Germany). Currently, heavy atom derivatization for experimental phasing is ongoing.

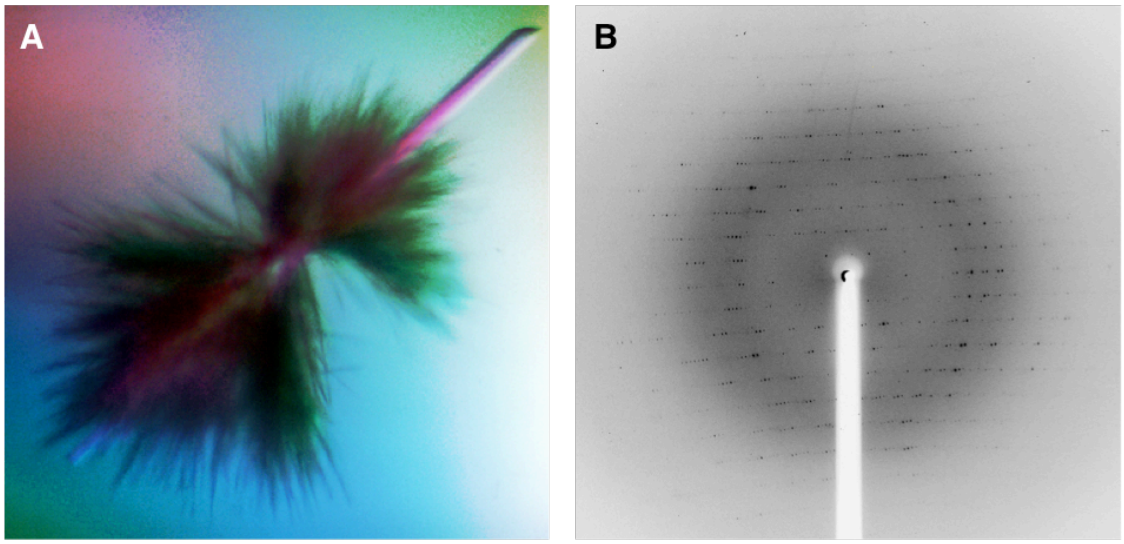


Figure 4.9: Crystal and diffraction pattern for MuRF-1$^{214-271}$
A) Crystal of MuRF-1$^{214-271}$ (grown by B. Franke) used for data collection
B) representative diffraction pattern of MuRF-1$^{214-271}$ collected under cryoconditions

4.2.4 Published material

Personal contribution to both publications: Design and cloning of MuRF-1$^{117-327}$; protein purification for SEC-MALS and CD spectroscopy; SEC-MALS measurements and interpretation

Molecular determinants for the recruitment of the ubiquitin-ligase MuRF-1 onto M-line titin

Michael Mrosek,* Dietmar Labeit,‡ Stephanie Witt,‡ Heiko Heerklotz,† Eleonore von Castelmur,* Siegfried Labeit,‡ and Olga Mayans*,[1]

*Division of Structural Biology and †Chemical Biophysics, Biozentrum, University of Basel, Klingelbergstr 70, Basel, Switzerland; ‡Institut für Anästhesiologie und Operative Intensivmedizin, Universitätsklinikum Mannheim, Mannheim, Germany

ABSTRACT **Titin forms an intrasarcomeric filament system in vertebrate striated muscle, which has elastic and signaling properties and is thereby central to mechanotransduction. Near its C-terminus and directly preceding a kinase domain, titin contains a conserved pattern of Ig and FnIII modules (Ig^{A168}-Ig^{A169}-$FnIII^{A170}$, hereby A168-A170) that recruits the E3 ubiquitin-ligase MuRF-1 to the filament. This interaction is thought to regulate myofibril turnover and the trophic state of muscle. We have elucidated the crystal structure of A168-A170, characterized MuRF-1 variants by circular dichroism (CD) and SEC-MALS, and studied the interaction of both components by isothermal calorimetry, SPOTS blots, and pull-down assays. This has led to the identification of the molecular determinants of the binding. A168-A170 shows an extended, rigid architecture, which is characterized by a shallow surface groove that spans its full length and a distinct loop protrusion in its middle point. In MuRF-1, a C-terminal helical domain is sufficient to bind A168-A170 with high affinity. This helical region predictably docks into the surface groove of A168-A170. Furthermore, pull-down assays demonstrate that the loop protrusion in A168-A170 is a key mediator of MuRF-1 recognition. Our findings indicate that this region of titin could serve as a target to attempt therapeutic inhibition of MuRF-1-mediated muscle turnover, where binding of small molecules to its distinctive structural features could block MuRF-1 access.—Mrosek, M., Labeit, D., Witt, S., Heerklotz, H., von Castelmur, E., Labeit, S., Mayans, O. Molecular determinants for the recruitment of the ubiquitin-ligase MuRF-1 onto M-line titin *FASEB J.* 21, 1383–1392 (2007)**

Key Words: elastic filament titin · muscle atrophy · X-ray crystallography · binding studies

The striated muscle tissue from vertebrates undergoes constant remodeling in adaptation to its mechanical load (reviewed in 1). Adaptive remodeling influences muscle tissue architecture, fiber-type composition, enzymatic apparatus, and innervation. Accordingly, numerous signaling pathways have been identified recently, which are controlled by myofibrillar stretch-strain. A prime candidate for the mediation of mechanical signaling in myocytes is the giant protein titin, the only sarcomeric component that forms a continuous filament system spanning full myofibrils, which can measure their tension and contains signaling domains (reviewed in 2).

Titin is composed of up to 33000 aminoacids, spanning *in situ* ~2 μm, from the Z-disc to the central M-line of the sarcomere (2). It folds into a total of 195 immunoglobulin (Ig) and 132 fibronectin-III (FnIII) domains, which together amount to over 90% of its mass. To date, structures of polydomain tandems of titin are limited to the Ig-doublet Z1Z2 from the N terminus of the filament, which has been elucidated in its free (3) and telethonin-bound states (4). Titin contains one single catalytic domain, a Ser/Thr kinase (TK), located near its C-terminus at the sarcomeric M-line. TK has been implicated in myofibrillogenesis (5) as well as in control of gene expression and protein turnover in mature muscle (6). The deformation of the M-line during sarcomere stretch has lead to suggest that TK could become activated via mechanically-induced conformational changes (7), thereby exerting its action on vicinal titin-associated proteins (6). The latter are thought to shuttle to the nucleus, where they interact further with regulators of gene expression (6, 8). According to this hypothesis, titin mediates a net translation of mechanical strain into chemical signaling.

Proteins associated to TK include MURF-1, MURF-2 (as shown *in situ* and *in vitro*; 8, 9), and nbr1 and p62 (*in vitro* data; 6). MuRF-1 binds to titin domains Ig^{A168}-Ig^{A169}-$FnIII^{A170}$ (hereby A168-A170), located just prior to TK (**Fig. 1*A***) (9). MuRF-1 can heteromerize with MuRF-2, allegedly through coiled-coil segments (9). *In vitro*, MuRF-2 binds p62, and the latter nbr1, which interacts with TK (6). However, with the exception of MuRF-1, it is not currently known to what extent these

[1] Correspondence: Division of Structural Biology Biozentrum, University of Basel, Klingelbergstr. 70, CH-4056 Basel, Switzerland. E-mail: olga.mayans@unibas.ch

doi: 10.1096/fj.06-7644com

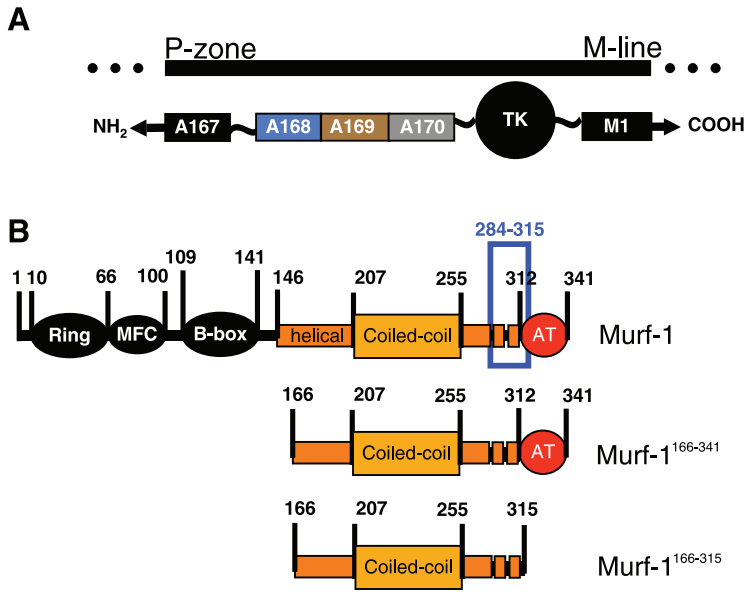


Figure 1. Domain composition of M-line titin and MuRF-1. *A)* M-line titin. Ig and FnIII domains are represented by filled boxes, TK by a globe. The construct under study, A168-A170, is colored; (*B*) MuRF-1 domain composition. The extent of ring finger, MuRF-specific (MFC) and B-box domains is as reported (9). The helical fraction, including an expected coiled-coil motif, has been predicted from sequence data and is in orange. The sequence previously suggested as primary titin binding site (26) is boxed in blue. The C-terminal acidic tail (AT) is in red.

proteins could form stable assemblies on the titin filament. MuRFs are muscle-specific proteins with a RING-finger/B-box/coiled-coil tripartite fold known as TRIM (Fig. 1*B*) (10). Although three MuRF isoforms (MuRF-1, -2, -3; encoded by different genes; 9) have been identified to date, only MuRF-1 has been shown to interact directly with titin. MuRF-1 is an E3 ubiquitin-ligase that becomes up-regulated during muscle loss and has been proposed as a molecular mediator of muscle atrophy (11). Muscle atrophy is characterized by an accelerated protein degradation via the ubiquitin proteasome system and it is an associated condition to numerous diseases, including cancer, diabetes and AIDS, as well as physiological states such as long-term immobilization and microgravity. A role of MuRF-1 in muscle turnover has been confirmed by a knockout mouse model that showed increased resistance to muscle atrophy (11), which suggests the potential pharmacological relevance of this protein. Yet, the cellular function of MuRF-1 might be complex to modulate, since it interacts with a broad range of cellular factors, including the ubiquitin-like modifier SUMO-3 (12), isopeptidase T-3, and the SUMO E2 ligase Ubc9 (8), which thereby might exert their influence on the titin cytoskeleton. Furthermore, nuclear forms of MuRF-1 interact with the transcription regulator GMEB-1 (8). Taken together, current functional data on MuRF-1 suggest its involvement in turnover and muscle gene expression, possibly forming part of mechanotransduction pathways in control of the titin composition of the sarcomere which, in turn, determines mechanical stiffness. Thus, in order to gain an insight into the structure of M-line titin and its recruitment of MuRF-1, we have elucidated the crystal structure of A168-A170 and interpreted it in the context of its binding to MuRF-1. Our findings aid the understanding of the molecular determinants of the interaction and point to a therapeutic interest of this region of titin.

MATERIALS AND METHODS

Cloning

For crystallographic studies, A168-A170 from human titin (residues 24431–24731, EMBL X90568) was cloned into the pETM-11 vector (EMBL collection) at restriction sites *Nco*I and MIuI. pETM-11 is a variant of pET-24d (Novagen, Madison, WI, USA), which includes an N-terminal His_6-tag and a TEV (tobacco etch virus) protease cleavage site prior to the inserted gene. Pull-down assays used A168-A170 cloned in pETM-44 (EMBL collection), which provided the N-terminal fusion of maltose binding protein (MBP) to the titin fragment. A mutated version of this construct, where the 102KTLE105 motif had been replaced by AAAA, was obtained using the QuikChange protocol (Stratagen, La Jolla, CA, USA).

Fragments from MuRF-1 (Swiss-Prot Q969Q1), MuRF-1(166–341) and MuRF-1(166–315), were inserted in pETM-11 via *Nco*I and *Kpn*I restriction sites. Both constructs correspond to a C280S mutated variant designed to prevent unspecific aggregation of the samples due to oxidation during storage.

Protein production

Overexpression of A168-A170 for structural analysis was carried out in *E. coli* BL21(DE3) Rosetta (Novagen). Cultures were grown at 37°C up to an OD_{600} of 0.6 in Luria Bertani medium supplemented with 25 μg/ml kanamycin and 34 μg/ml chloramphenicol. Expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cultures were further grown at 25°C for ~18 additional hours. Cells were harvested by centrifugation at 2800 *g* and 4°C. Bacterial pellet was resuspended in 50 mM Tris-HCl, pH 8.0; 150 mM NaCl; containing a protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany) and DNase I. Lysis was by addition of lysozyme and sonication. The homogenate was clarified by centrifugation at 15000 g and 4°C. The supernatant was applied to a Ni^{2+}-chelating HisTrap column (GE Healthcare) equilibrated in lysis buffer and eluted using 250 mM imidazole. The eluent was dialyzed against 50 mM Tris-HCl, pH 8.0; 150 mM NaCl; and 2 mM DTT in the presence of TEV protease. Given that A168-A170 still interacted with the Ni^{2+}-chelating resin after tag removal, the digested mixture was further purified by gel filtration on

a Superdex 200 Hiload 16/60PG column (GE Healthcare) equilibrated in dialysis buffer. The samples were finally dialyzed against 25mM MOPS, pH 8.0; 150 mM KCl; 1 mM DTT, 1 mM EGTA; and stored at 4°C for further use.

MuRF-1 fragments were prepared as above, but purification of the protease digest used subtractive affinity chromatography prior to gel filtration.

Structure elucidation

Crystals grew in hanging drops at 20°C from solutions containing 2.1 M NaCl; 0.1 M MES, pH 5.5; and 0.6 M Li_2SO_4, mixed in a 1:1 ratio with a 12 mg/ml protein solution. Crystals were in the habit of long rods, hexagonal in cross section with approximate dimensions of 600 × 100 × 100 μm^3 and reached their final size over a 2 month period.

For X-ray data collection, crystals were frozen at 100 K in mother liquor solutions supplemented with 20% [v/v] glycerol as cryoprotectant. X-ray data processing was in XDS (13), and data statistics and crystal parameters are given in **Table 1**. Crystals contained two copies of A168-A170 per asymmetric unit, which were related by an imperfect two-fold axis approximately colocalized with the crystallographic *z*-axis. Since attempts at phasing by molecular replacement were unsuccessful, experimental phases were obtained by SIR combined with a 3λ MAD experiment on a crystal derivatized with $[Ta_6Br_{12}]^{2+}$ by soaking for 30 min at saturation. Derivative data statistics are given in **Table 3**. Anomalous scatterers (2 sites) were located using CNS (14) and phases calculated and refined in SHARP (15) (Table 3). Density modification used solvent flattening, histogram matching, and NCS domain averaging in DM (16). Manual model building was in O (17). Model refinement applied overall anisotropic B-factor scaling, bulk solvent correction, and grouped B-factor minimization as implemented in CNS (14). NCS restraints were applied across individual domains throughout refinement and moderately loosened in the final cycle of refinement (resulting rmsd values were 0.33 Å for Cα atoms across copies of domain Ig^{A168}, 0.25 Å for Ig^{A169}, and 0.24 Å for Fn^{A170}). For refinement, experimental data were divided into a working and a test set using FREERFLAG (16). Solvent molecules were built using the water-pick routine in CNS and validated visually in electron density maps. The final model includes all protein residues of A168-A170, at exception of the 12 C-terminal residues that were disordered in both NCS copies. Two N-terminal residues in positions –1 and –2, remnants of the TEV cleavage site, could be observed and are part of the final model.

Model coordinates and experimental X-ray data have been deposited at the Protein Data Bank with accession code 2NZI.

Circular dichroism

MuRF-1(166–315) was extensively dialyzed against 10 mM sodium tetraborate, pH 8.5; 20 mM NaCl; and concentrated up to 0.1 mg/ml as determined by BCA (Pierce, Rockford, IL, USA) assay. Data were acquired at 25°C on a 62A DS circular dichroism spectropolarimeter (AVIV) equipped with a temperature-controlled quartz cell of 1 mm optical path. The spectrum was recorded at a scan speed of 50 nm/min with a response time of 5 s and averaged over 8 scans. Protein conformation was calculated by fitting the experimental data with a combination of reference spectra for known secondary structure elements (18).

Size exclusion chromatography combined with multiangle light scattering (SEC-MALS)

The oligomeric state of MuRF-1(166–315) in solution was determined via SEC-MALS measurements performed on an ÄKTA basic system (GE Healthcare) connected to a triangle static light-scattering detector and a differential refractometer (miniDAWN Tristar and Optilab, respectively). A Superdex 200 HR 10/300 GL column (GE Healthcare) was used in 50 mM Tris, pH 8.0, and 100 mM NaCl at a flow rate of 0.7

TABLE 1. *Native X-ray data and model refinement statistics*

Spacegroup	$P3_121$
Unit cell dimensions	$a = b = 125.90$ Å, $c = 134.07$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$
Solvent content	72%
X-ray Data:	
X-ray source	ID29–1 (ESRF)
Detector	ADSC Quantum Q210
Wavelength (Å)	0.9792
Resolution (Å)	15–2.9 (3.0–2.9)
Unique reflections	26792 2627
R_{sym}(I)	7.5 53.1
Multiplicity	6.3 6.3
Completeness	96.6 98.6
I/σ (I)	22.7 3.9
Refinement:	
Number of reflections in working/free set	25645/1147
Number of protein residues/solvent molecules	584/45
R-factor / R-free (%)	21.8 / 27.6
Model:	
rmsd bond length (Å) / bond angle (°)	0.010 / 1.59
Average B (Å^2):	
Main chain / side chains / solvent	52.6 / 58.9 / 28.9
Ramachandran plot (%)[a]:	
Core + allowed / generous / disallowed	99.2 / 0.6 / 0.2[b]

[a]Calculated using PROCHECK (32). [b]The one residue in disallowed conformation is His146B.

TABLE 2. *Domain arrangement in poly-domain fragments from titin*

	Opening angle[a]	Torsion[b]	Distance[c]	Buried surface area (Å^2)[d]
$Ig^{A168}Ig^{A169}$	176°	+89°	43.3Å	212 (~1.4%)
$Ig^{A169}Fn^{A170}$	127°	—	40.2Å	308 (~2.0%)
$Ig^{Z1}Ig^{Z2}{}_{XTAL}$[e]	146°	+75°	49.4Å	165 (~1.4%)
$Ig^{Z1}Ig^{Z2}{}_{RDC/SAXS}$[e]	136°	+85°	44.5Å	235 (~1.4%)
$Ig^{Z1}Ig^{Z2}{}_{Tele}$[e]	167°	+48°	52.4Å	160 (~2.0%)

[a] Angle defined by the primary axes of inertia (longitudinal) of consecutive domains. [b] Angle defined by the projection of secondary axes of inertia (cross sectional) of consecutive domains onto a common plane. [c] Distance between centers of mass of individual domains. [d] Calculated with AREAIMOL (33). [e] Crystallographic coordinates for free Z1Z2 ($Ig^{Z1}Ig^{Z2}{}_{XTAL}$) derive from PDB entry 2A38 (3); domain arrangement of Z1Z2 in solution was calculated from SAXS data and NMR residual dipolar couplings ($Ig^{Z1}Ig^{Z2}{}_{RDC/SAXS}$) as reported (3) and crystal coordinates of Z1Z2 complexed to telethonin ($Ig^{Z1}Ig^{Z2}{}_{Tele}$) derive from 1AY5 (4).

ml/min. Sample volumes of 100 μl were injected at a concentration of 2 mg/ml. A specific refractive index increment (dn/dc) of 0.185 ml/g was used. The data were recorded and processed using ASTRA software (Wyatt Technology, Santa Barbara, CA, USA). To determine the detector delay volumes and normalization coefficients for the light scattering detector, a BSA sample (Pierce) was used as reference.

Isothermal titration calorimetry

Data were recorded with a VP-ITC calorimeter (Microcal LLC, Northhampton, MA, USA) at 25°C running Origin v7 software. All samples were previously dialyzed against 50 mM Tris-HCl, pH 8.0, and 20 mM NaCl. Titrations consisted of 5–7 μl injections of A168-A170 concentrated to 1.4 mg/ml into MuRF-1(166–341) or MuRF-1(166–315) solutions at 1 mg/ml and 0.7 mg/ml, respectively (as determined by BCA assay; 19). The time interval between injections was 200–300 s. The data were corrected for the heat of dilution by subtraction of the small constant heat obtained at the end of the titration, where no further binding occurs.

Binding assays

Titin residues mediating MuRF-1 binding were revealed by using a SPOTS blot membrane (JPT, Berlin, Germany) that displayed A168-A170 as a series of 32 overlapping peptides of 20-aa length. Peptides were acetylated at their N terminus to enhance stability. The membrane was initially washed with ethanol and abundantly with TBS, followed by an overnight blocking step in TBS supplemented with 5% (w/v) milk powder. For binding, MuRF-1(166–315) was added at a final concentration of 50 μg/ml and reacted overnight. Unbound proteins were removed by washes of TBS supplemented with 0.05% (v/v) Tween (Sigma, St. Louis, MO, USA). Bound fragments were reacted for 3 h with specific antibodies against MuRF-1 (rabbit polyclonal, 0.5 μg/ml). Bound antibodies were then detected using the Elite Kit (Vectastain) employing biotinylated anti-rabbit IgG (PK6101), and the detection kit LumiLight Plus (Roche Diagnostics) used as indicated by the manufacturer.

Pull-down assays used MBP-A168–170 fusion products, both wild-type and as a KTLE to AAAA mutated variant, which had been immobilized onto amylose resin (New England Biolabs, Beverly, MA, USA). Purified MuRF-1(166–315) was flowed through the column and incubated for 1 h. MBP in isolation was tested as control. Mixtures were washed with 20 column volumes of buffer solutions containing 20 mM Tris-HCl, pH 7.4; 200 mM NaCl; 1 mM EDTA; and 10 mM β-mercaptoethanol. Elution used 10 mM maltose.

RESULTS

Crystal structure of A168-A170

The structure of A168-A170 from human titin has been elucidated at 2.9 Å resolution using X-ray crystallography. Data statistics and model parameters are given in Table 1. A168-A170 comprises two Ig domains (Ig^{A168}-Ig^{A169}) and one FnIII (Fn^{A170}). The molecule adopts an extended conformation with an end-to-end distance of ~115 Å, where the two N-terminal Ig lie almost per-

TABLE 3. $[Ta_6Br_{12}]^{2+}$ *X-ray data and phasing statistics*

Spacegroup/unit cell dimensions	$P3_121/a = b = 125.94$ Å, $c = 134.60$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$					
Xray source	ESRF-ID-23-1					
Detector	MAR-CCD 225					
Set	Peak		Inflecion		Remote	
Energy (eV)	9882.3		9878.8		12299.9	
Resolution (Outer shell) (Å)	20.0–3.35 (3.4–3.35)		15.0–3.45 (3.5–3.45)		15.0–3.5 (3.55–3.5)	
No. Bijvoet reflections	33994	1556	31059	1347	29580	1216
R_{sym}(I) (%)	11.1	42.6	10.6	41.9	12.9	47.5
Multiplicity	3.9	3.9	3.7	3.7	3.9	3.9
Completeness	98.8	99.7	98.2	99.4	98.2	98.9
$<I/\sigma$ (I)$>$	9.8	3.5	10.1	3.5	9.2	3.1
Phasing SIR/MAD (<4.5 Å):						
Number of heavy atom sites	2					
Phasing power(iso/ano)	1.46/1.58		1.34/0.96		1.13/0.87	
R_{cullis}(iso/ano)	0.66/0.71		0.67/0.86		0.71/0.86	
Figure of merit (prior to density modification)	0.66					

fectly coaxial but the FnIII domain is bent away from the molecular axis (**Fig. 2*A***). Domain arrangements are described in **Table 2**. The crystal form used in this study contains two A168-A170 copies in its asymmetric unit. These are essentially identical in conformation (rmsd 0.68 Å for 288 Cα atoms of the whole molecule, calculated with SPDBV; 20), suggesting that this fragment of titin has a well-defined long-range order, and a high degree of stiffness.

Domains Ig^{A168} and Ig^{A169} belong to the I(intermediate)-set of Ig folds (21). They share 20% seq. id. and a high structural similarity (rmsd 0.99 Å for 82 matching Cα atoms, SPDBV). Both belong to the "N-conserved" type of Ig from titin (22), characterized by a conserved N-terminal loop cluster comprising proline residues prior to β-strand A, a PxP motif in the BC loop, and an extended FG β-hairpin hosting an NxxG sequence (**Fig. 3*C***). This Ig type predominates in the Z-disc, the skeletal I-band and the A-band fractions of titin, but not in its constitutive cardiac I-band. Domains Ig^{A168} and Fn^{A170} show no significant deviation from other equivalent modules across titin. However, Ig^{A169} shows a unique 9-residue loop protrusion between β-strands A and A' with sequence PKTLEGMGA, which adopts in part an α-helical conformation (Figs. 2*A* and **Fig. 5*A***). According to structure and sequence data on titin domains, a loop insertion at this position is not detectable in any other Ig of titin.

Domain interfaces

The domains of A168-A170 form a tight tandem. They are connected through one-residue linkers in extended conformation that effectively result in the continuation of structural elements from one domain into the next. The extended arrangement of the molecule is characterized by domain interfaces with unusually small buried areas (Table 2) and engages a minimal number of residues in each domain. Both Ig-Ig and Ig-FnIII interfaces are void of specific interactions and only involve small hydrophobic clusters.

Domains Ig^{A168}-Ig^{A169} interact through hydrophobic groups that surround the linker strand dorsally and ventrally (Fig. 3*A*). Given their coaxial orientation, the interacting residues originate from the linker region (V92, aliphatic portion of E93, V94) and the C-terminus of β-strand A' (aliphatic portion of R13, Y14) in Ig^{A168} and the conserved elements of this Ig type in Ig^{A169}, namely the FG turn (F180) and the PxP motif of the BC loop (P128, K127) (Fig. 3*C*). A comparison of Ig^{A168}-Ig^{A169} and Z1Z2 (the only other Ig doublet of titin structurally characterized to date) (3, 4) reveals certain similarities. Also Z1Z2 exhibit an extended conformation, lack specific interdomain contacts, and include a small hydrophobic cluster between modules. As that of Ig^{A168}-Ig^{A169}, the cluster of Z1Z2 comprises residues C-terminal to β-strand A', the BC loop and linker residues (Fig. 3*C*). Despite, Z1Z2 displays a different domain orientation (Table 2) and has a longer, three-residue linker that allows certain modular dynamics (3). It can then be concluded that the mere presence of elements characteristic of the "N-conserved" Ig type does not determine domain orientations in these doublets. Since the identified interface residues are not conserved in Ig across titin (23), it

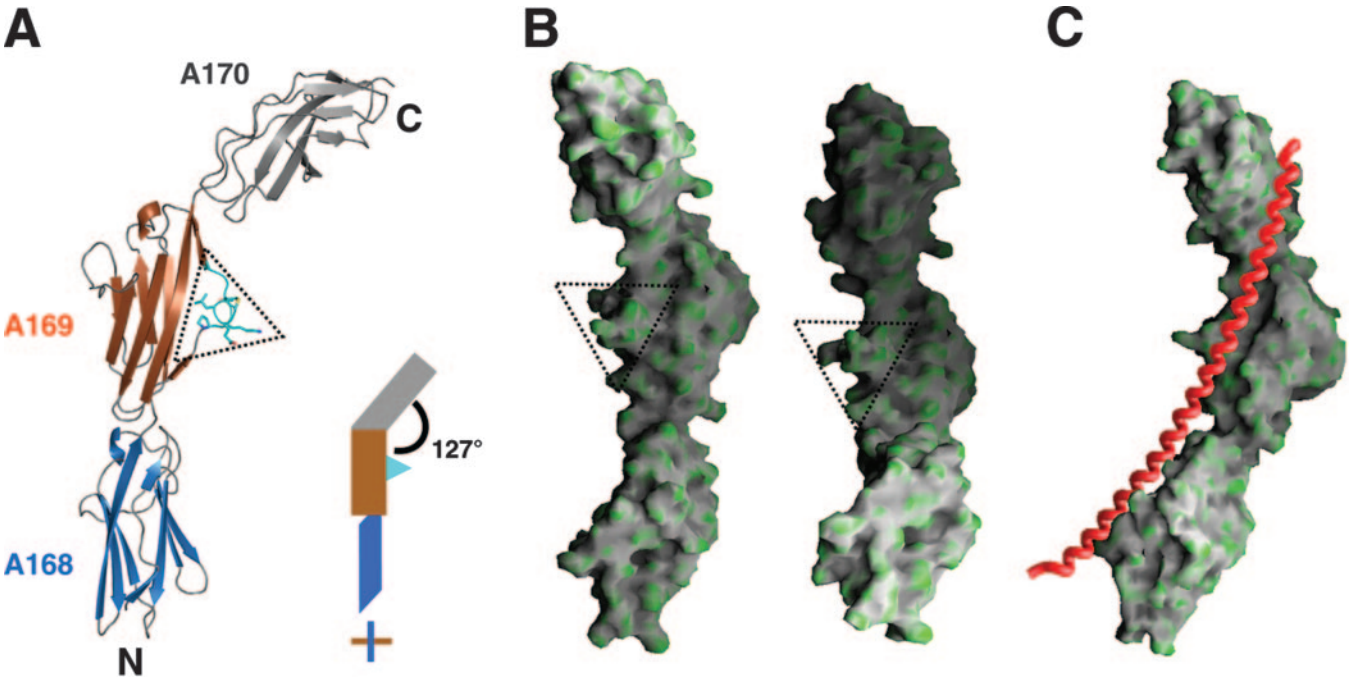


Figure 2. Crystal structure of A168-A170. *A)* Ribbon representation. The insertion loop in Ig^{A169} is displayed in cyan. The inset shows the molecular conformation schematically; (*B*) Surface representation in two related views, where green emphasizes the local curvature. A ridge is present in the concave side of the molecule. The distinct loop of Ig^{A169} is marked; (*C*) Docking of an α-helical peptide in coiled-coil conformation onto the surface of A168-A170 to satisfy shape complementarity. The helical peptide shown derives from lamin A coil 2B (PDB code 1X8Y).

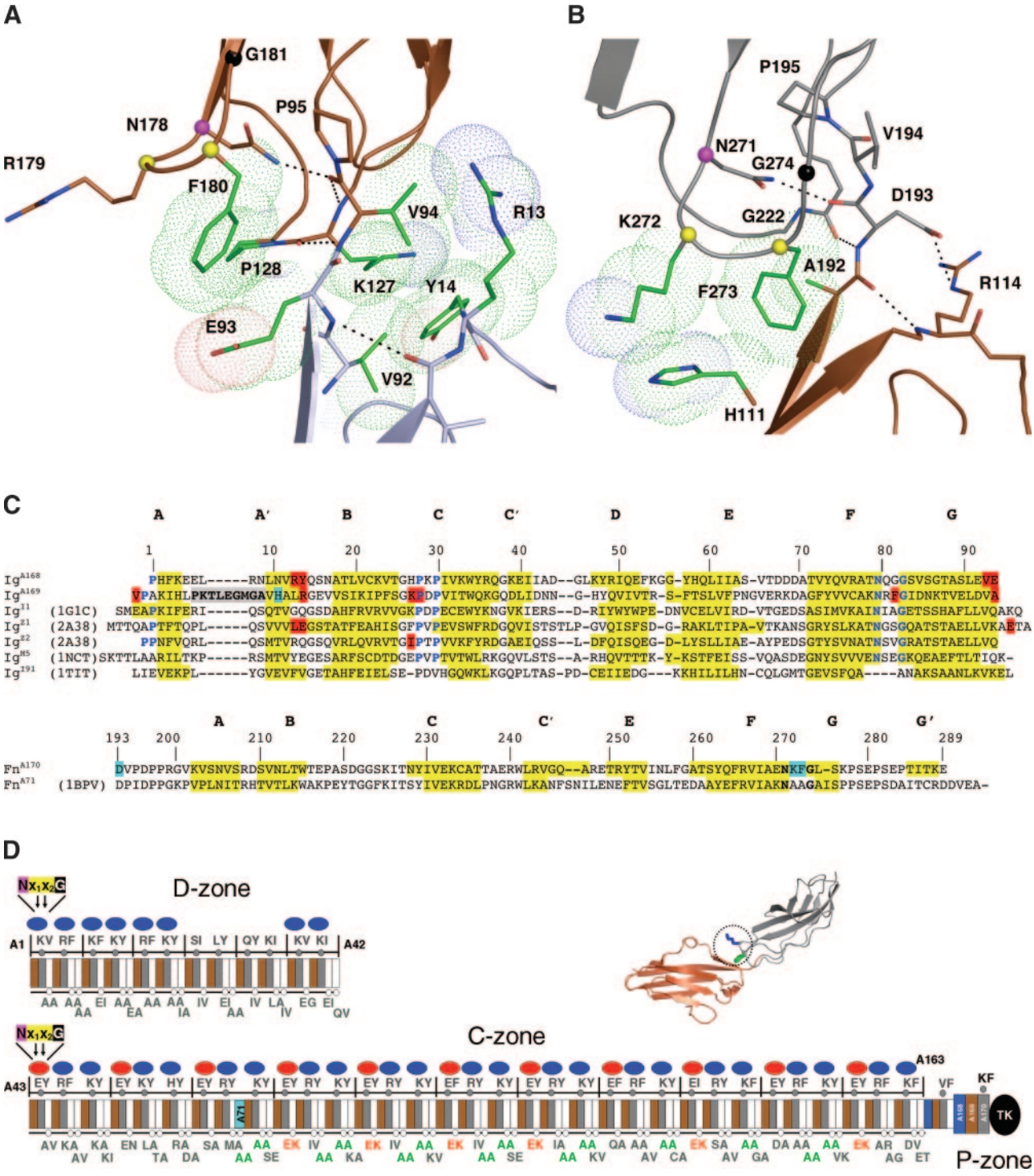


Figure 3. Domain interfaces. *A)* Ig^{A168}-Ig^{A169} and (*B*) Ig^{A169}-Fn^{A170} interfaces. Domain color code as in Fig. 2*A*. Hydrophobic contributions are shown in green. Selected van der Waals surfaces are displayed. The Cα atoms of residues in the NxxG motif of the FG β-turns of both Ig^{A169} and Fn^{A170} are shown as spheres. Hydrogen bonds are indicated by dotted lines; *C)* Structure-based sequence alignment of titin Ig and FnIII of known structure (PDB accession codes are given). β-Strand composition is shown in yellow; residues at Ig-Ig and Ig-FnIII interfaces are highlighted in red and cyan, respectively. Features characteristic of "N-conserved" Ig (A168, A169, I1, Z1, Z2 and M5) are shown in blue. I91 was formerly I27. Strand nomenclature for Ig as in (21) and for FnIII as in (31). The distinct loop of Ig^{A169} is shown in gray; *D)* A-band of titin. Ig and FnIII domains are shown as filled boxes, where Ig domains flanked by FnIII are in brown, FnIII following an N-terminal Ig in gray, other FnIII in white and Ig followed by another C-terminal Ig in blue. Fn^{A71}, the only other FnIII of titin with known structure (1BPV), is in cyan. The two residues in variable positions of the NxxG β-turn of FnIII domains are displayed. Domain super-repeats are indicated. The pattern of alternating charges is displayed, where red and blue indicate negative and positive charges, respectively.

cannot be predicted at this point how closely the observed conformations represent those of other tandems. Yet, the overall structural principles of Ig arraying revealed by these studies might be generic to the titin filament as indicated by the fact that all poly-Ig arrays of titin characterized to date by small angle X-ray scattering have revealed comparable extended arrangements independently of their domain or linker composition (22).

The interface of domains Ig^{A169}-Fn^{A170} exhibits a hydrophobic component as that of Ig^{A168}-Ig^{A169} (Fig. 3*b*). Similarly, contacts involve β-strand A' (H111) in the initial domain Ig^{A169} and the NxxG motif of β-turn FG in Fn^{A170} (K272, F273) (Fig. 3*C*). However, in contrast to Ig, the FG turns in FnIII of titin are highly conserved in sequence, as revealed by the alignment of 123 FnIII sequences of A-band titin (CLUSTALW; 23). This analysis showed that those FnIII domains following an Ig within the 11-domain super-repeats of the central C-zone region of A-band titin, supposedly part of the myosin-based thick filament, consistently host a conserved residue pair in the variable positions of their NxxG motif (Fig. 3*D*). These invariably include a charged residue followed by an aromatic group, most often tyrosine or phenylalanine. Since these residues are part of the modular interface, Ig-FnIII pairs from the C-zone of titin can be expected to exhibit a similar interdomain conformation as that of A169-A170. Strikingly, the charges within this motif are distinctly distributed along the C-zone repeats, where every first FnIII hosts a negative charge and the rest a positive group (Fig. 3*D*). Such segregation is not observed in the 7-domain super-repeats of the preceding D-zone (Fig. 3*D*). In the crystal structure, the lateral amino group of K272 is not involved in interactions, suggesting its availability for binding to other sarcomeric proteins. Myosin-binding protein C (MyBP-C) binds titin by interacting with the first Ig domain of every super-repeat in the C-zone, so that titin possibly aids thereby the regular polymerization of myosin filaments (25). Given that the FG β-turn of Fn^{A170} is located toward the intermodular space, its charge might be speculated to contribute to MyBP-C docking. Interestingly, the FG turns of FnIII domains at other positions within repeats of the D- and C-zones also exhibit certain conservation of their sequence motifs (Fig. 3*D*), suggesting that these might be generically important for FnIII-FnIII arrangements and/or interactions to other A-band components.

MuRF-1 binding

The surface of A168-A170 reveals a shallow groove at the concave side of the molecule that spans its full length (Fig. 2*B*). This could be expected to accommodate the binding counterpart of MuRF-1. Based on SPOTS blots of MuRF-1 peptides, the latter has been proposed to bind titin through a sequence just prior to its acidic C-terminal tail (Fig. 1*B*) (26). According to our own analysis (Jpred software; 27), the C-terminal half of MuRF-1 (residues 166–315, which include the proposed binding sequence) has high tendency to α-helical formation, with at least its middle section predictably following a coiled-coil association (COILS; 28) (Fig. 1*B*). This is supported by CD data on a correspondingly truncated MuRF-1 variant, MuRF-1(166–315), whose spectrum exhibits a maximum at 190 and minima at 208 and 222 nm indicative of a predominant α-helical composition (**Fig. 4*A***). An estimation of the secondary structure content of this sample based on these data suggested an approximate α-helical content of 70% and 30% random coil (fitting parameter R^2=99.8%). This is in excellent agreement with results from both manual examination and docking prediction software (PatchDock; 29), which indicated that the dimensions and geometry of the surface groove in A168-A170 can well accommodate an α-helix along its length. In particular, a best docking is obtained if the helix exhibits a long-range bending of its axis as that of components of a dimeric coiled-coil. For example, as seen in human lamin A coil 2B (30) used as a prototypic structural template to probe the surface of A168-A170 in this study (Fig. 2*C*). To explore further the interaction area between titin and MuRF-1, we investi-

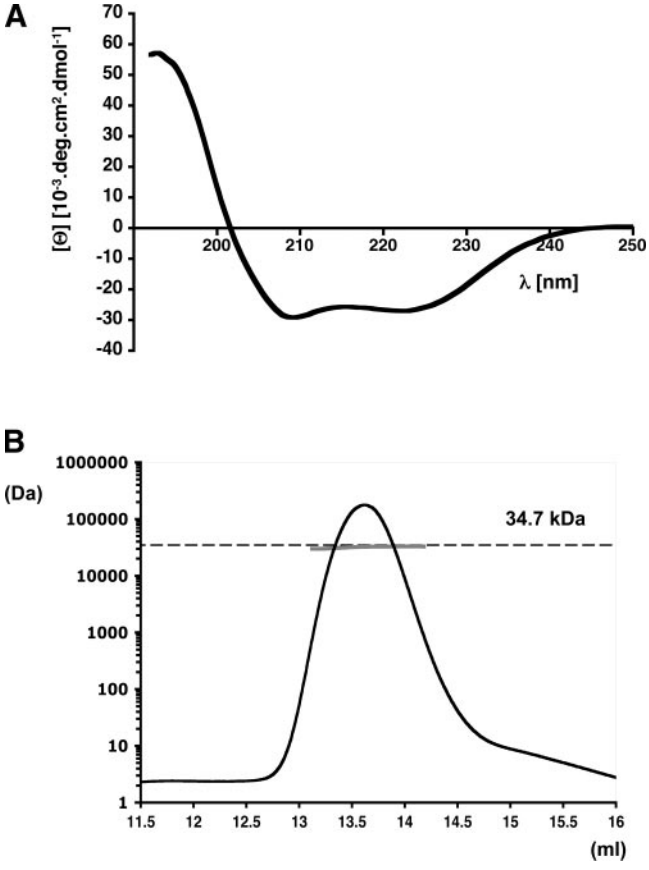


Figure 4. Biophysical characterization of the titin-binding domain of MuRF-1. *A)* CD spectrum of MuRF-1(166–315) exhibiting features characteristic of a sample with a predominantly α-helical content; *B*) SEC-MALS measurements of MuRF-1(166–315). Average molecular mass per vol unit (gray) and normalized refractive index (continuous line) are shown. The theoretical mass of a dimer of this construct is indicated by a dashed horizontal line.

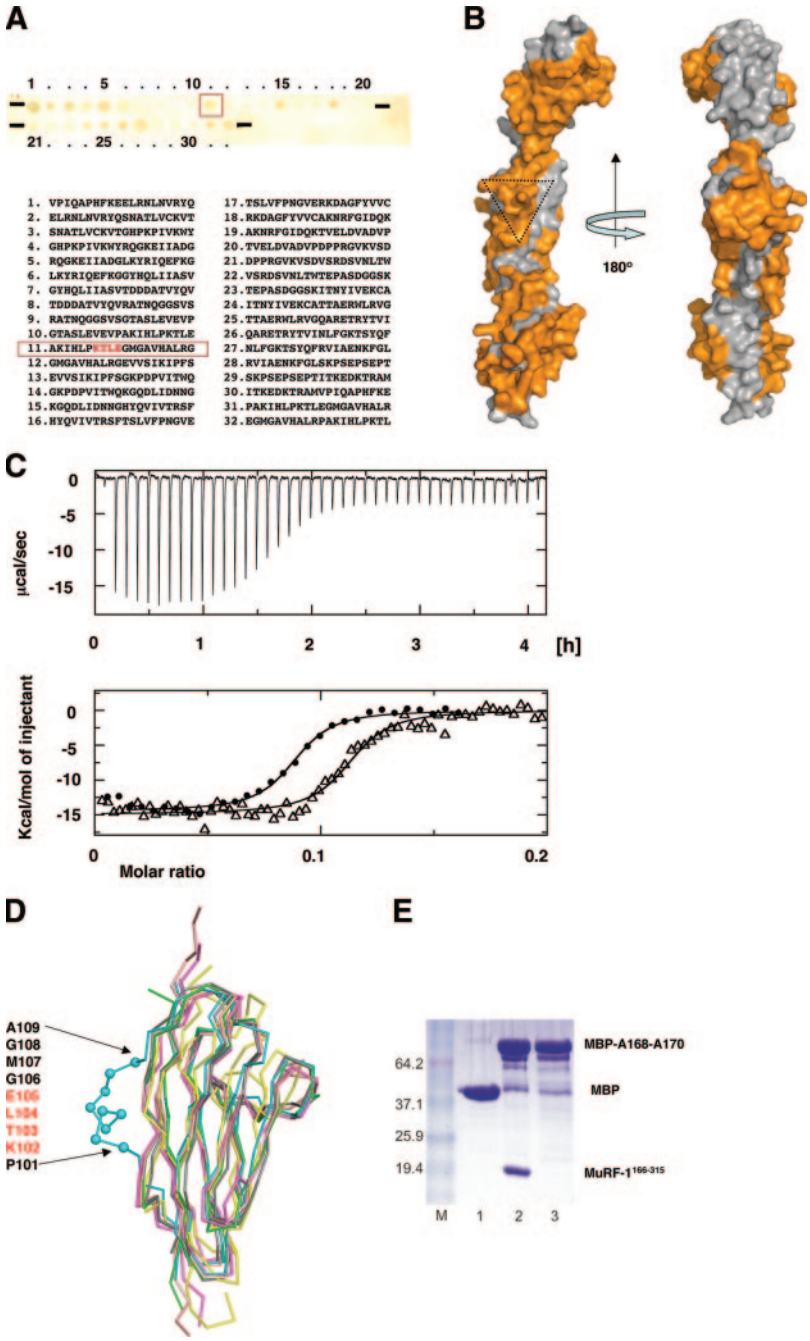


Figure 5. Binding of MuRF-1 to A168-A170. *A)* SPOTS blots of peptide fragments from A168-A170 assayed *vs.* MuRF-1(166–315). Data corresponding to the insertion motif in Ig^{A169} are boxed in red; *B*) Mapping of interacting SPOTS peptides (orange) onto the molecular surface of A168-A170. They cluster at the concave face (left), while only one peptide (peptide 18) can be mapped to the convex side of the molecule (right), which otherwise remains unmatched. The distinct loop insertion of Ig^{A169} is marked; *C*) Binding of MuRF-1 to A169-A170 monitored by ITC. The experimental pattern of injections of MuRF-1(166–341) is shown (upper panel). Heats of binding for titrations of A168-A170 into MuRF-1(166–341) (filled circles) and MuRF-1(166–315) (open triangles) recorded in 50 mM Tris-HCl, pH 8.0; 20 mM NaCl at 25°C is displayed (lower panel). Both curves show the same K_d and ΔH and differ only slightly in the active fraction of MuRF-1; *D*) Superimposition of Ig domains from titin for which a structural model is available, namely Ig^{A169} (cyan), Ig^{A168} (green), I1 (magenta; PDB accession code 1G1C), M5 (yellow; 1NCT), Z1 (salmon; 2A38) and Z2 (gray, 2A38). The unique loop insertion of Ig^{A169} is clearly identifiable. The Cα atoms of its component residues are displayed as solid spheres and the sequence given. I91 (formerly I27), which does not belong to this Ig subgroup and has shorter loop regions, is excluded from the superimposition to ease visualization. As the other Ig, I91 does not comprise loop insertions in its N-terminal β-strand region. The KTLE motif altered by site-site directed mutagenesis is in red; *E*) SDS-PAGE of a pull-down assay where MBP-A168-A170 was immobilized in affinity resin in the presence of MuRF-1(166–315). Lane 1 shows isolated MBP as control; lane 2 reveals the binding of MuRF-1 to wild-type A168-A170; lane 3 employed tetra-mutated A168-A170, where the KTLE motif had been disrupted, with the resulting loss of MuRF-1 binding.

gated their binding by SPOTS blots of peptide fragments of A168-A170 assayed against the central helical fraction of MuRF-1, MuRF-1(166–315). The results confirm that all three modules in A168-A170 are involved in the interaction (Fig. 5*A*). Furthermore, the mapping of the selected peptides onto the crystal structure of A168-A170 showed that, with the exception of one peptide, all interacting fragments colocalized at the concave face of the molecule (Fig. 5*B*). Thus, it could be concluded that this is the primary MuRF-1 binding interface.

Further, we have established the interaction between A168-A170 and the helical fraction of MuRF-1 in solution and quantified it using isothermal calorimetry (ITC). This study used two N-terminally truncated MuRF-1 constructs, MuRF-1(166–341) and MuRF-1(166–315) (Fig. 1*B*), where the latter lacks in addition the acidic C-terminal tail. Binding monitored by ITC showed that both MuRF-1 variants interacted strongly with A168-A170, yielding heat release values of 14 and 28 kcal/mol and K_d affinities of 35 and 37 nM, respectively (Fig. 5*C*). Since K_d values are similar for both constructs, it can be concluded that the acidic tail of MuRF-1 does not influence binding but that this is determined by its helical fraction. It was also found that saturation occurs at a molar ratio of ~0.1, which indicates that ~10 MuRF-1 molecules bind one A168-A170. Such high ratio, however, is unusual and points to either a high oligomerization state of MuRF-1 variants or possible technical problems in the estimation of this parameter by ITC. To gain a further insight into

this issue, we investigated the association state of the helical fraction of MuRF-1 in isolation using size-exclusion chromatography combined with multiangle static light scattering (SEC-MALS). This technique yields an accurate determination of the molecular mass (MM) without being influenced by molecular shape or hydrodynamic parameters. SEC-MALS measurements of MuRF-1(166–315) yielded a MM of 32.3 kDa (protein concentration estimated from refractive index), which is in excellent agreement with the calculated MM of 34.7 kDa for a dimer of this sample (Fig. 4*B*). This supports predictions based on sequence data that suggest a self-association of MuRF-1 via a dimeric coiled-coil segment (COILS software; 28) (Fig. 1*B*). Since A168-A170 is monomeric according to structural and chromatographic data, it could be expected that the basic stoichiometry of the titin/MuRF-1 complex is 1:2 (or a multiple of this ratio). Unfortunately, we have not been able to investigate directly the composition of titin/MuRF-1 assemblies by other biophysical techniques during this study. Mixing and storage of complexed samples at concentrations and in conditions compatible with biophysical analysis led to precipitation that rendered the preparations unsuitable for study. Thus, at this time, we cannot discard the possibility that the ratio measured by ITC might represent a more complex scenario in which, on binding, conformational changes take place in the individual components that result in a higher level of assembly. However, the differing ITC value is more likely to be of technical origin. Since ITC curves were consistent over the several repetitions of the recordings and for both constructs, MuRF-1(166–341) and MuRF-1(166–315), the measured stoichiometry cannot be due to random or variable effects, such as unspecific aggregation of MuRF-1 samples. The latter, for which no experimental evidence exists, would have caused drifts in the saturation point of the different ITC measurements depending on sample batch, storage times and conditions. Instead, ITC data most probably reflect a consistent error in estimating the concentration of MuRF-1 fragments in this study. The MuRF-1 helical domain has a low content of aromatic residues and calculation of its concentration required of the BCA colorimetric method, whose response to different proteins can vary widely. Moderate, cumulative errors in quantifying the amounts of MuRF-1 variants and/or A168-A170 could well explain the deviation of recorded ITC values (~0.1) from the expected (~0.5) titin/MuRF-1 ratio based on the oligomeric state of the individual components.

Structural and biochemical data here presented indicate that MuRF-1 and titin interact through an extensive contact area. However, in A168-A170 the unique loop insertion of Ig^{A169} is located at the midpoint of its surface groove (Fig. 2*B*, *C*) and it could be expected to play a central role in MuRF-1 binding or function. Effectively, SPOTS blots confirmed that this motif is involved in MuRF-1 interaction (Fig. 5*A*). To further assess the relevance of this feature in the binding, we constructed a tetra-mutated variant of A168-A170, where residues 102-KTLE-105 had been exchanged for 102AAAA105, and tested its interaction to MuRF-1(166–315) using pull-down assays (Fig. 5*D*). Results confirmed that alteration of this unique motif abolishes the interaction with MuRF-1 and, therefore, that this is central to its molecular recognition by A168-A170.

DISCUSSION

The muscle filament titin, despite its colossal dimensions, has a simple and repetitive architecture consisting of linear tandems of Ig and FnIII modules. The structure of A168-A170 reveals the molecular details of Ig-Ig and Ig-FnIII interfaces, where the latter are likely to be generic to Ig-FnIII domain pairs along titin. A168-A170 has a pronouncedly extended conformation and exhibits a well-defined long-range order with limited conformational freedom. This seems to result from extremely short linker sequences and the presence of small hydrophobic clusters that restrict the modular orientations. Specific contacts are not found across any of these domain pairs. Hydrophobic interactions include groups from similar structural elements in Ig and FnIII modules, with residues around the linker region and the FG β-turn being involved in both cases (Fig. 3*C*). The observed Ig-Ig contacts are poorly conserved (Fig. 3*C*). In agreement, domain orientations in Ig^{A168}-Ig^{A169} and Z1Z2 differ somewhat (Table 2). On the contrary, sequence conservation at the Ig-FnIII interface indicates that this modular conformation might be common to equivalent pairs from the A-band. It is yet to be established whether the conservation of interface groups in this case, in particular the FG β-turn of FnIII domains, might play an additional role in interactions to other sarcomeric proteins—such as MyBP-C (25).

A168-A170 selectively recruits MuRF-1 to the M-line region of titin. The specificity of the interaction appears achieved through two idiosyncratic features of this fragment of titin: *i*) the topography of a shallow groove spanning the length of its surface and defined by the long-range domain arrangement of this distinct tandem; and *ii*) the unique loop insertion of Ig^{A169}. ITC data show that the titin/MuRF-1 interaction is of high affinity, where the α-helical region of MuRF-1 (residues 166–315) is sufficient for binding. Based on the oligomeric state of individual components, we propose the stoichiometry of the titin/MuRF-1 complex to be 1:2. SPOTS blots support the view that the groove in A168-A170 constitutes the binding interface to MuRF-1 as predicted from the crystal structure. The groove could host up to 70% of the maximal hypothetical coiled-coil length achievable by MuRF-1, if its full helical region was to form such motif. The fact that the interaction involves more than one modular unit in titin suggests that this filament uses long-range steric determinants, provided by its interdomain arrangements, to achieve specificity in the recruitment of other proteins to distinct locations of its otherwise repetitive frame. Finally, we have identified a 102KTLE105 insertion motif in Ig^{A169} as primary molecular determinant of MuRF-1 interaction by pull-down assays on a A168-

A170 mutated variant. So that even if the interacting surface in the MuRF-1/titin complex appears extensive, the binding could be modulated by targeting this key motif. The high affinity of the interaction and the fact that this is primarily governed by a defined sequence motif suggest that the A168-A170 region of titin could serve as a potential therapeutic target against muscle atrophy, where binding of small molecules to its distinctive structural features could block the access of MuRF-1. FJ

Our gratitude goes to B. Sheppard and D. Nurizzo (ESRF) for their support during X-ray data collection and to A. Tsamaloukas (Biozentrum) for help with ITC. M.M. was supported by SNF grant (3100A0–100852) to OM. This project was further financed by DFG (La668/7–2 and La1969/1–1).

REFERENCES

1. Bassel-Duby, R., and Olson, E. N. (2006) Signaling pathways in skeletal muscle remodeling. *Annu. Rev. Biochem.* **75,** 19–37
2. Granzier, H. L., and Labeit, S. (2004) The giant protein titin: a major player in myocardial mechanics, signaling, and disease. *Circ. Res.* **94,** 284–295
3. Marino, M., Zou, P., Svergun, D., Garcia, P., Edlich, C., Simon, B., Wilmanns, M., Muhle-Goll, C., and Mayans, O. (2006) The Ig-doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. *Structure* **14,** 1437–1447
4. Zou, P., Pinotsis, N., Lange, S., Song, Y. H., Popov, A., Mavridis, I., Mayans, O., Gautel, M., and Wilmanns, M. (2006) Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. *Nature* **439,** 229–233
5. Mayans, O., van der Ven, P. F., Wilm, M., Mues, A., Young, P., Furst, D. O., Wilmanns, M., and Gautel, M. (1998) Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* **395,** 863–869
6. Lange, S., Xiang, F., Yakovenko, A., Vihola, A., Hackman, P., Rostkova, E., Kristensen, J., Brandmeier, B., Franzen, G., Hedberg, B., *et al.* (2005) The kinase domain of titin controls muscle gene expression and protein turnover. *Science* **308,** 1599–1603
7. Grater, F., Shen, J., Jiang, H., Gautel, M., and Grubmuller, H. (2005) Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys. J.* **88,** 790–804
8. McElhinny, A. S., Kakinuma, K., Sorimachi, H., Labeit, S., and Gregorio, C. C. (2002) Muscle-specific RING finger-1 interacts with titin to regulate sarcomeric M-line and thick filament structure and may have nuclear functions via its interaction with glucocorticoid modulatory element binding protein-1. *J. Cell Biol.* **157,** 125–136
9. Centner, T., Yano, J., Kimura, E., McElhinny, A. S., Pelin, K., Witt, C. C., Bang, M. L., Trombitas, K., Granzier, H., Gregorio, C.C., *et al.* (2001) Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain. *J. Mol. Biol.* **306,** 717–726
10. Meroni, G., and Diez-Roux, G. (2005) TRIM/RBCC, a novel class of single protein RING finger E3 ubiquitin ligases. *Bioessays* **27,** 1147–1157
11. Bodine, S.C., Latres, E., Baumhueter, S., Lai, V.K., Nunez, L., Clarke, B.A., Poueymirou, W.T., Panaro, F.J., Na, E., Dharmarajan, K., *et al.* (2001) Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science* **294,** 1704–1708
12. Dai, K. S., and Liew, C. C. (2001) A novel human striated muscle RING zinc finger protein, SMRZ, interacts with SMT3b via its RING domain. *J. Biol. Chem.* **276,** 23992–23999
13. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26,** 795–800
14. Brunger, A.T. *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. Allogr.* **D54,** 905–921
15. Fortelle, E. de la, and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472–494
16. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Allogr.* **D50,** 760–763
17. Jones, T.A., Zou, J.Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Allogr.* **A47,** 110–119
18. Yang, J. T. C. S. Wu, and Martinez, H. M. (1986) Calculation of protein conformation from circular dichroism. *Methods Enzymol.* **130,** 208–269
19. Brown, R. E. (1989) Protein measurement using bicinchoninic acid: elimination of interfering substances. *Anal. Biochem.* **180,** 136
20. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723
21. Harpaz, Y., and Chothia, C. (1994) Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238,** 528–539
22. Marino, M., Svergun, D. I., Kreplak, L., Konarev, P. V., Maco, B., Labeit, D., and Mayans, O. (2005) Poly-Ig tandems from I-band titin share extended domain arrangements irrespective of the distinct features of their modular constituents. *J. Muscle. Res. Cell. Motil.* **26,** 355–365
23. Witt, C. C. Olivieri, N., Centner, T., Kolmerer. B., Millevoi. S., Morell. J., Labeit. D., Labeit. S., Jockusch. H., Pastore, A. (1998) A survey of the primary structure and the interspecies conservation of I-band titin's elastic elements in vertebrates. *J. Struct. Biol.* **122,** 206–215
24. Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680
25. Freiburg, A., and Gautel, M. (1996) A molecular map of the interactions between titin and myosin-binding protein C. Implications for sarcomeric assembly in familial hypertrophic cardiomyopathy. *Eur. J. Biochem.* **235,** 317–323
26. Witt, S. H., Granzier, H., Witt, C. C., and Labeit, S. (2005) MURF-1 and MURF-2 target a specific subset of myofibrillar proteins redundantly: towards understanding MURF-dependent muscle ubiquitination. *J. Mol. Biol.* **350,** 713–722
27. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998) Jpred: a consensus secondary structure prediction server. *Bioinformatics* **14,** 892–893
28. Lupas, A., Van Dyke, M., and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252,** 1162–1164
29. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33,** W363–W367
30. Strelkov, S. V., Schumacher, J., Burkhard, P., Aebi, U., and Herrmann, H. (2004) Crystal structure of the human lamin A coil 2B dimer: implications for the head-to-tail association of nuclear lamins. *J. Mol. Biol.* **343,** 1067–1080
31. Leahy, D. J., Hendrickson, W. A., Aukhil, I., and Erickson, H. P. (1992) Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* **258,** 987–991
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26,** 283–291
33. Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55,** 379–400

Received for publication October 31, 2006.
Accepted for publication December 6, 2006.

Structural Analysis of B-Box 2 from MuRF1: Identification of a Novel Self-Association Pattern in a RING-like Fold[†,‡]

Michael Mrosek,[§] Sebastian Meier,[§,||] Zöhre Ucurum-Fotiadis,[§,||] Eleonore von Castelmur,[§] Erik Hedbom,[§] Ariel Lustig,[§] Stephan Grzesiek,[§] Dietmar Labeit,[⊥] Siegfried Labeit,[⊥] and Olga Mayans*[,§,@]

Division of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland, Institut für Anästhesiologie and Operative Intensivmedizin, Universitätsklinikum Mannheim, D-68167 Mannheim, Germany, and School of Biological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, U.K.

Received April 24, 2008; Revised Manuscript Received July 31, 2008

ABSTRACT: The B-box motif is the defining feature of the TRIM family of proteins, characterized by a RING finger−B-box−coiled coil tripartite fold. We have elucidated the crystal structure of B-box 2 (B2) from MuRF1, a TRIM protein that supports a wide variety of protein interactions in the sarcomere and regulates the trophic state of striated muscle tissue. MuRF1 B2 coordinates two zinc ions through a cross-brace α/β-topology typical of members of the RING finger superfamily. However, it self-associates into dimers with high affinity. The dimerization pattern is mediated by the helical component of this fold and is unique among RING-like folds. This B2 reveals a long shallow groove that encircles the C-terminal metal binding site ZnII and appears as the defining protein−protein interaction feature of this domain. A cluster of conserved hydrophobic residues in this groove and, in particular, a highly conserved aromatic residue (Y133 in MuRF1 B2) is likely to be central to this role. We expect these findings to aid the future exploration of the cellular function and therapeutic potential of MuRF1.

Zinc-binding motifs play key physiological and structural roles in a broad range of proteins, including storage and structural proteins, enzymes, and transcription factors. The "B-box" is a small zinc-binding domain (∼40 residues) found exclusively in eukaryotes. It occurs within the context of a tripartite fold known as TRIM[1] (*1*−*3*), which consists of up to three sequentially conserved zinc-binding domains [commonly a RING finger, a B-box type 1 (B1), and a B-box type 2 (B2)] followed by a coiled coil fraction (CC). In those cases where a single B-box is present, it is usually of type B2. The presence of a B-box is the determinant of the TRIM fold, since it is found only in members of this family.

TRIM proteins have been attributed roles in the regulation of gene expression, cell growth, and differentiation, thereby being associated with human proliferative pathologies [e.g., cancer, familiar Mediterranean fever, Opitz/BBB syndrome, promyelocytic leukemia, mulibrey nanism, and thyroid carcinomas (*4*)]. Accumulating evidence suggests that B-boxes mediate protein−protein interactions specific to each TRIM class (e.g., refs *5*−*7*) and are crucial for the efficient self-assembly of this fold (*3* and references therein). Recently, atomic structures of B1 and B2 boxes have become available, namely, the B1 boxes of MID1 (*8*) and TRIM19 (PDB entry 2D8V) and the B2 boxes of ATDC (2CSV), TRIM39 (2DID), TRIM41 (2EGM), TRIM5 (2YRG), MuRF1 (2D8U), MID2 (2DJA), and MID1 (*9*). These structures reveal that the B-box fold exhibits a $\beta\beta\alpha$ RING-like architecture that coordinates two zinc ions in a classical "cross-brace" topology, where zinc-binding site I (ZnI) is formed by the first and third ligand pairs and zinc-binding site II (ZnII) by the second and fourth ligand sets. The recent B-box models differ significantly from a structure previously described for B2 of *Xenopus* XNF7 (*10*), which adopted a topology unique among zinc-binding folds and coordinated one single zinc ion leaving four potential metal ligands unoccupied. For more than a decade, the fold of XNF7 B2 was thought to represent the canonical B-box motif. However, it is now evident that that model, calculated from a refolded synthetic peptide, does not represent a physiologically relevant fold.

To further investigate the characteristics of the B2 motif, we have analyzed the structure and self-association properties of the B2 from the ubiquitin ligase MuRF1 using X-ray crystallography, NMR spectroscopy, light scattering, and analytical ultracentrifugation. MuRF1 contains a RING domain in the N-terminal position, a specific MuRF family domain (MFC), a single B2 box, a CC domain, and a C-terminal acidic tail (Figure 1). We selected MuRF1 for our studies since it contributes to the control of the trophic state of cardiac and skeletal muscle and its inactivation can slow the development of muscle atrophy (*11, 12*). MuRF1 interacts with an intriguingly wide variety of cellular partners,

[†] This work was supported by grants from The Royal Society to O.M. and DFG (La668/11-1) to S.L. M.M. and Z.U.-F. were supported by SNF (3100A0-100852).

[‡] Atomic coordinates and experimental structure factor amplitudes have been deposited in the Protein Data Bank as entry 3DDT.

* To whom correspondence should be addressed. Telephone: +44 (0)151 7954472. Fax: +44 (0)151 7954406. E-mail: Olga.Mayans@liverpool.ac.uk.

[§] University of Basel.

[||] These authors contributed equally to this work.

[⊥] Universitätsklinikum Mannheim.

[@] University of Liverpool.

[1] Abbreviations: B2, B-box type 2; CC, coiled coil; MuRF, muscle-specific RING finger; TRIM, tripartite motif.

10.1021/bi800733z CCC: $40.75 © 2008 American Chemical Society
Published on Web 09/17/2008

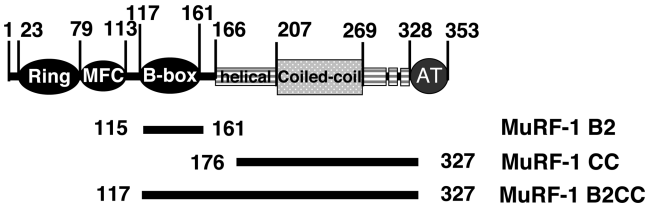


FIGURE 1: Domain structure of MuRF1. MFC refers to a MuRF family specific motif, and AT denotes the C-terminal acidic tail. The residue composition of the protein constructs used in this study is indicated.

namely, the titin filament (*13, 14*), isopeptidase T-3 (*15*), the transcription regulator GMEB-1 (*15*), CARP, SQSTM1, and EEF1G (*12*). Further, MuRF1 also participates in sumoylation pathways (*13, 15*). Although the specific motifs in the multidomain MuRF1 that are responsible for these interactions are still unknown, its B2 box has recently been shown to participate in the binding of creatine kinase (*16*), providing a further link between myofibril turnover and muscle metabolism. Our data show that MuRF1 B2 presents a distinct dimerization pattern and allow us to propose sequence and surface features that might sustain the protein interactions mediated by this motif, aiding future studies of its cellular role and potential therapeutic targeting.

EXPERIMENTAL PROCEDURES

Cloning. Truncated variants of human MuRF1 (Q969Q1) (Figure 1) were cloned into vector pETM-11 (EMBL collection) containing an N-terminal His_6 tag and a TEV protease cleavage site prior to the target gene. The $B2^{C137D}$ variant was engineered in pETM-11 using the QuikChange protocol (Stratagene). The B2CC construct corresponded to a C293S variant designed to prevent sample oxidation.

Protein Production. Expression occurred in *Escherichia coli* BL21(DE3) Rosetta (Novagen). Cultures were grown at 37 °C to an OD_{600} of 0.6 in Luria-Bertani medium supplemented with 25 μg/mL kanamycin and 34 μg/mL chloramphenicol. Expression was induced with 0.2−0.5 mM IPTG, and cultures were grown for a further 18 h at 20−25 °C. To obtain soluble B2 and B2CC samples, media were supplemented with 50 μM $ZnCl_2$ upon induction. Harvesting was accomplished via centrifugation. Pellets were resuspended in 50 mM Tris (pH 8.0), 100 mM NaCl, and 10 mM β-mercaptoethanol containing a protease inhibitor cocktail (Roche). Lysis was by addition of lysozyme and French pressing in the presence of DNase. The homogenate was clarified by centrifugation and the supernatant applied to a Ni^{2+}-chelating HisTrap column (GE Healthcare) equilibrated in lysis buffer containing 20 mM imidazole. Elution used 250 mM imidazole. The eluate was dialyzed against 50 mM Tris (pH 8.0), 150 mM NaCl, and 10 mM β-mercaptoethanol in the presence of His-tagged TEV protease at 4 °C overnight. Subsequent purification involved reverse affinity chromatography on a HisTrap column followed by gel filtration on a Superdex 75 16/60PG column (GE Healthcare) equilibrated in lysis buffer. Pure samples were concentrated to 10 mg/mL (BCA assay, Pierce) and stored until further use.

Expression of $B2^{C137D}$ according to the protocol described above resulted in insoluble sample, possibly reflecting a decrease in zinc affinity. Consequently, the bacterial pellet was resuspended in lysis buffer supplemented with 8 M urea and subjected to affinity purification under denaturing conditions. Eluted fractions were dialyzed overnight against a refolding buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, 10 mM β-mercaptoethanol, and 1 mM $ZnCl_2$. The refolded sample was processed further as the wild type.

Table 1: Experimental Diffusion Parameters for MuRF-1 B2 from ^{15}N Relaxation Data[a]

model	$\tau_{c,eff}$[b] (ns)
isotropic	6.39
axially symmetric	6.39
fully asymmetric	6.40

[a] At 25 °C in water. [b] $\tau_{c,eff}$ is calculated from $(6D)^{-1}$, where $D = (D_x + D_y + D_z)/3$; rotational diffusion was analyzed with TENSOR (*25*).

Crystal Structure Elucidation. Crystals grew from solutions containing 1.8 M ammonium sulfate, 0.1 M Tris (pH 8.5), and 15% glycerol at 4 °C in hanging drops composed of a 1:1 ratio of protein and reservoir solutions. Hexagonal crystals with dimensions of ~200 μm × 200 μm × 100 μm grew within 2 days, typically showing visible macrotwinning. X-ray data were collected from a selected single crystal frozen at 100 K in native mother liquor. Data processing used the XDS package (*17*) (statistics given in Table 2). Phasing was carried out using a 3λ MAD experiment that exploited the signal from endogenous zinc ions. The location of anomalous scatterers and phase calculation were conducted with SOLVE (*18*), while density modification and initial model building were conducted in RESOLVE (*19*). Further automated building used ARP/wARP (*20*). Manual model editing was conducted in O (*21*) and final refinement in PHENIX (*22*). The final model contains all protein residues with the exception of C-terminal residue S161 in chains B and C that was not visible in electron density maps.

Crystals of $B2^{C137D}$ grew from 1.8 M ammonium sulfate, 0.1 M Tris (pH 8.5), and 5% glycerol at 18 °C (protein concentration like that of the wild type). X-ray data were collected on an Elliot GX20 rotating anode up to 3.4 Å resolution. Crystals belonged to space group $P2_12_12_1$ with six B-box protomers (three biological dimers) per asymmetric unit and the following cell dimensions: $a = 60.77$ Å, $b = 66.80$ Å, and $c = 77.40$ Å. Lattice reconstruction was by molecular replacement in PHASER (*23*) using the wild-type monomer lacking zinc ions as a search model. Map calculation was preceded by mild refinement.

Analytical Ultracentrifugation. Sedimentation equilibrium data were recorded using an Optima XL-A analytical ultracentrifuge (Beckman Instruments) equipped with 4 and 12 mm Epon double-sector cells in an An-60 Ti rotor and absorption optics. Runs were performed at 20 °C in 50 mM Tris (pH 8.0) and 150 mM NaCl at 24000, 28000, and 34000 rpm and protein concentrations of 0.025−0.4 mg/mL, where OD values remained linear. A detection wavelength of 232 nm was used because of the low extinction coefficient of the sample. Average molecular masses were evaluated using SEGAL (*24*). A protein partial specific volume of 0.73 mL/g was employed, while the solution density was taken to be 1.003 g/mL and the viscosity to be 1.02 cP.

NMR Experiments. ^{15}N-labeled B2 was expressed in M9 minimal medium using $^{15}NH_4Cl$ as the nitrogen source and

Table 2: X-ray Data and Model Refinement Statistics

		MAD (six zinc atoms)		
		peak	inflection	remote
space group	$P6_522$			
unit cell dimensions	$a = b = 76.22$ Å, $c = 146.93$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$			
solvent content	69%			
X-ray source	X06SA (SLS)	ID23-1 (ESRF)		
detector	MAR-CCD225	ADSC Quantum Q210		
energy (keV)	12.3000	9.8825	9.6624	10.2000
resolution (Å)	18−1.9 (2.0−1.9)	15−3.35 (3.4−3.35)	15−3.45 (3.5−3.45)	15−3.5 (3.55−3.5)
no. of unique reflections/Bijvoets*	20520 (2851)	*33994 (1556)	*31059 (1347)	*29580 (1216)
R-factor (%)	7.1 (38.6)	11.1 (42.6)	10.6 (41.6)	12.9 (47.5)
multiplicity	10.4 (10.7)	3.9 (3.9)	3.7 (3.7)	3.9 (3.9)
completeness	99.7 (100.0)	98.8 (99.7)	98.2 (99.4)	98.2 (98.9)
$I/\sigma(I)$	20.7 (7.8)	9.8 (3.5)	10.1 (3.5)	9.2 (3.1)

no. of reflections in working/free set	19667/853
no. of protein chains	3
no. of protein residues/solvent atoms/zinc ions	139/158/6
R-factor/R_{free} (%)	19.9/25.2
rmsd for bond lengths (Å)/angles (deg)	0.012/1.35
Ramachandran plot, residues in disallowed regions	1 (Q44 in chain A)

following the protocol described above. This resulted in production of an insoluble sample, presumably due to the formation of $Zn_3(PO_4)_2$ sediments that depleted the medium of soluble zinc. Thus, refolding and purification protocols like those described for $B2^{C137D}$ were followed for all labeled samples used in NMR measurements.

NMR data were recorded at a sample concentration of 3 mM in 20 mM Tris (pH 7.0), 75 mM NaCl, and 10 mM β-mercaptoethanol. Protein backbone ^{15}N and ^{1}H resonances were assigned with ^{15}N-separated NOESY and TOCSY spectra. The oligomeric state of MuRF1 B2 was estimated from ^{1}H T_2 spin echo relaxation measurements as well as ^{15}N T_1, ^{15}N T_2, and ^{15}N heteronuclear NOE data. ^{15}N relaxation data were analyzed with TENSOR (*25*). RDCs were measured in a solution of 1 mM protein dissolved in 20 mM Tris (pH 7.0), 75 mM NaCl, and 10 mM β-mercaptoethanol supplemented with Pf1 phage (ASLA Biotech) (*26*) at a concentration of 15 mg/mL. The electrostatic alignment was tuned by addition of NaCl to a final concentration of 150 mM. Intersubunit NOEs to residue F148 were identified using a two-dimensional NOESY spectrum with a mixing time of 100 ms. Spectra were recorded on a BRUKER DRX 600 spectrometer equipped with a TXI probe and on a BRUKER DRX 800 spectrometer equipped with a TCI cryoprobe. Data were processed with NMRPipe (*27*) and analyzed with PIPP (*28*).

Multiangle Light Scattering Coupled to Size Exclusion Chromatography (SEC−MALS). Measurements were performed on an ÄKTA basic system (GE Healthcare) connected to a tri-angle light scattering detector and a differential refractometer (miniDAWN Tristar and Optilab, respectively). A Superdex 200 HR 10/300 GL column (GE Healthcare) was used in 50 mM Tris (pH 8.0) and 100 mM NaCl at a flow rate of 0.7 mL/min. Sample volumes of 100 μL were injected at a concentration of 2 mg/mL. A specific refractive index increment (d*n*/d*c*) value of 0.185 mL/g was used. The data were recorded and processed using ASTRA (Wyatt Technology). To determine the detector delay volumes and normalization coefficients for the MALS detector, a BSA sample (Sigma catalog no. A-8531) was used as a reference.

RESULTS AND DISCUSSION

Crystal Structure. The structure of human MuRF1 B2 has been elucidated by X-ray crystallography to 1.9 Å resolution (Figure 2a). The crystal form used in this study contains three molecular copies in its asymmetric unit that are nearly identical [average rmsd of 0.30 Å for 43 matching Cα atoms, calculated using SPDBV (*29*)]. Crystallographic data show that the B2 domain spans residues H119−L159 in MuRF1. It consists of a structural core formed by an α-helix (α1) that packs against a three-stranded antiparallel β-sheet (β1−β3), supporting three loop regions (L1, L2, and L3) involved in zinc binding, where L2 is, in fact, a conserved β-turn. Anomalous difference Fourier maps revealed that each molecular copy of B2 binds two zinc ions in cross-brace topology at a mutual distance of ∼10.7 Å, in agreement with inter-zinc distances observed in other B2 boxes (10.3−10.9 Å) (Figure 2b). Zinc-binding site ZnI involves residues C122xxH125 and C142xxC145, while site ZnII involves C134xxC137 and H151xxC154. The C-terminal cysteine ligand, C154, is a unique feature of human MuRF proteins. This residue is a conserved histidine in all other B2 boxes [52 sequences analyzed (Figure S1 of the Supporting Information)]. Thus, MuRF1 B2 is a representative of the CHC_2C_2HC B2 subtype.

The crystal structure of the MuRF1 B2 monomer is in good agreement with NMR models of the B2 domains from ATDC (2CSV), TRIM39 (2DID), TRIM41 (2EGM), TRIM5 (2YRG), MID2 (2DJA), MID1 (*9*), and MuRF1 (2D8U) (Figure 2b). A more detailed comparison of the crystal and NMR structures of MuRF1 B2 is given in Figure S2 of the Supporting Information.

In the B2 motif, zinc is coordinated through highly conserved cysteine and histidine residues, with the exception of the fourth ligand (in site ZnII) that is broadly degenerated. This position is often occupied by an aspartate residue (31 occurrences in 52 sequences), but it can also host a cysteine (16 of 52), a glutamate (3 of 52), a histidine (1 of 52), or a serine (1 of 52) (Figure S1 of the Supporting Information). All models of B2 boxes elucidated to date display regular aspartates in this location, with the exception of MuRF1 B2

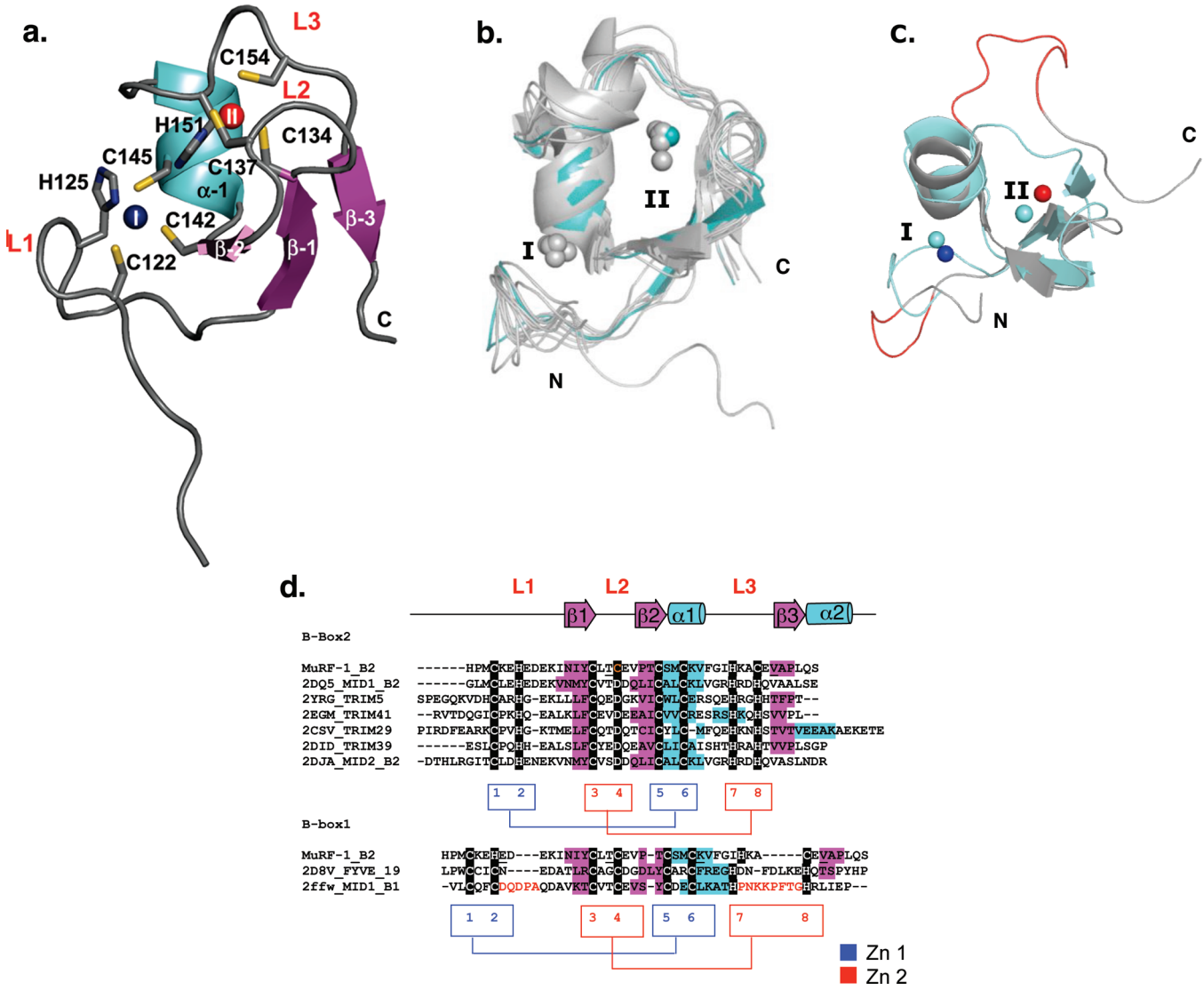


FIGURE 2: Structural characterization of B-boxes. (a) Crystal structure of MuRF1 B2. Metal binding ligands are displayed. The N- and C-terminal zinc ions are colored blue and red, respectively. Secondary structure elements and loops are labeled. (b) Superposition of B2 structures available to date, where MuRF1 B2 is highlighted in cyan. (c) Superposition of MuRF1 B2 (cyan) and MID1 B1 (gray), where loop insertions in the B1 family are colored red. (d) Structure-based sequence alignment of B-boxes with known structure. Secondary structure elements and loop regions are colored as in panel a. Metal ligands are highlighted in black, numbered, and colored according to the metal site they form. The mutated cysteine residue in the B^{C137D} variant is colored orange. The cross-brace topology is sketched.

that contains a cysteine. To investigate whether aspartate and cysteine ligands are interchangeable at this position, we analyzed a C137D mutated variant of MuRF1 B2 ($B2^{C137D}$). Comparative ^{1}H NMR measurements showed that both wild-type B2 and $B2^{C137D}$ shared a high degree of spectral similarity, indicative of common structural features. This was confirmed by a subsequent crystallographic analysis of $B2^{C137D}$, which demonstrated that the amino acid substitution neither influences the fold nor alters its zinc binding properties.

MuRF1 B2 Has a Dimeric Structure. The crystal structure of MuRF1 B2 reveals a dimer, where the α1 helix of one protomer docks into a concave depression formed by the α-helix and β-sheet of the other subunit (Figure 3b). The interface, which comprises an area of 515 Å^2 (calculated using the PISA server), contains clusters of hydrophobic and polar interactions (Figure 3c). Hydrophobic interactions involve residues M144 and F148 in helix α1 of one subunit and residues I132′ in strand β1 and 156′-VAP-158′ in strand β3 of the other subunit. Polar contacts are established by residues E128−N131′ and by the mutual packing of residue S143 from both α1 helices. Electron density maps for interface groups are shown in Figure S3 of the Supporting Information.

The groups involved in dimer formation in MuRF1 B2 are relatively well conserved within the B2 family (Figure S1 of the Supporting Information), in particular the hydrophobic groups in strands β1 and β3 and helix α1. Remarkably, positions S143 and F148 are subject to compensatory sequence variations, where position 143 commonly hosts a hydrophobic group and position 148 is most frequently a serine. The L1 loop also shows a propensity to host charged residues that could form interactions across protomers equivalent to those of MuRF1 B2, although a pattern of conservation is not detected. Thus, one might speculate that other B2 boxes could also dimerize following the pattern of MuRF1 B2. In contrast, B1 boxes seem to be unlikely to adopt this dimerization model. As opposed to the short loops of B2 boxes, which are conserved in length, B1 boxes contain two long variable regions following the second and seventh

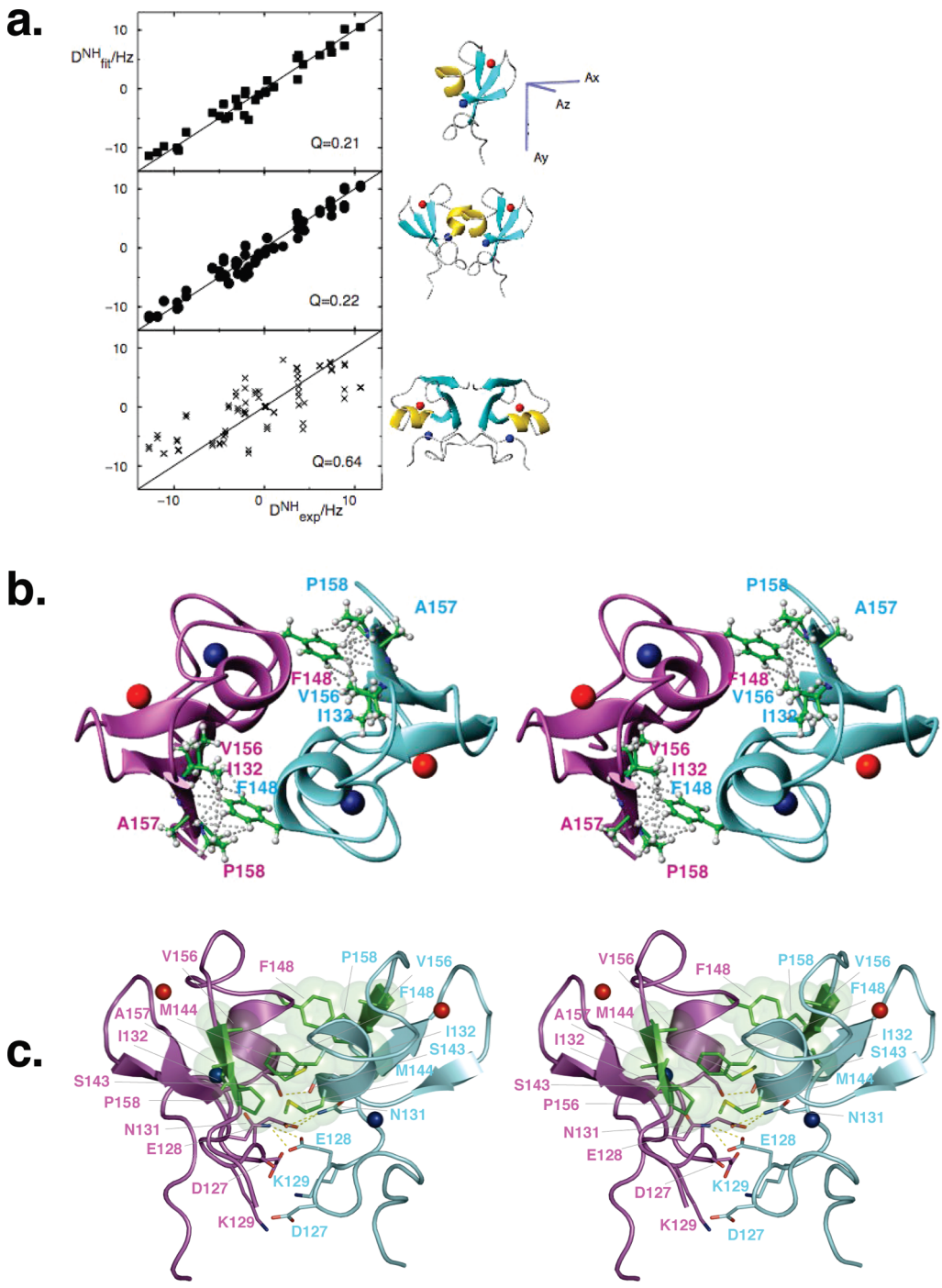


FIGURE 3: Dimeric structure of MuRF1 B2. (a) Fitting of $^{1}H-^{15}N$ RDC data to the crystal structures of the monomer (top), the dimer in the asymmetric unit (middle), and the dimer formed by crystallographic symmetry (bottom). (b) Stereo representation of intersubunit NOEs (15 in total; shown as dashed lines) measured between the protons of F148 and neighboring hydrophobic residues. (c) Stereoview of the interface contacts in the crystal structure of dimeric MuRF1 B2. Hydrophobic groups are colored green and highlighted using van der Waals spheres. Polar interactions are represented as dashed lines.

zinc ligands (*2*) (Figure 2c,d). In MID1 B1 (*12*), the C-terminal insertion adopts a bulky, "lasso-like" structure stabilized by conserved proline residues that can be expected to sterically prevent the formation of a dimeric arrangement like that of MuRF1 B2.

Oligomeric State in Solution. To establish the oligomeric state of MuRF1 B2 in solution, we analyzed it using analytical ultracentrifugation. Sedimentation equilibrium experiments confirmed the dimeric character of the domain and suggested a strong association constant (Figure 4a). Since the sample remained predominantly dimeric at the lowest assayed concentration of 4.8 μM, we concluded that the dissociation constant of the interaction must be below that, possibly in the submicromolar range.

These conclusions were supported by NMR data obtained from ^{15}N-enriched samples. The correlation time (τ_c) at 25 °C derived from ^{1}H T_2 and a complete set of ^{15}N relaxation data (T_1, T_2, and $\{^{1}H\}-^{15}N$ NOEs) was 6.4 $\pm$ 0.1 ns, corresponding to a molecular mass (M_r) of ~12 kDa, approximately twice that of a B2 monomer (5.2 kDa). Experimental diffusion parameters and residue-specific relaxation properties are given in Table 1 and Figure S4 of

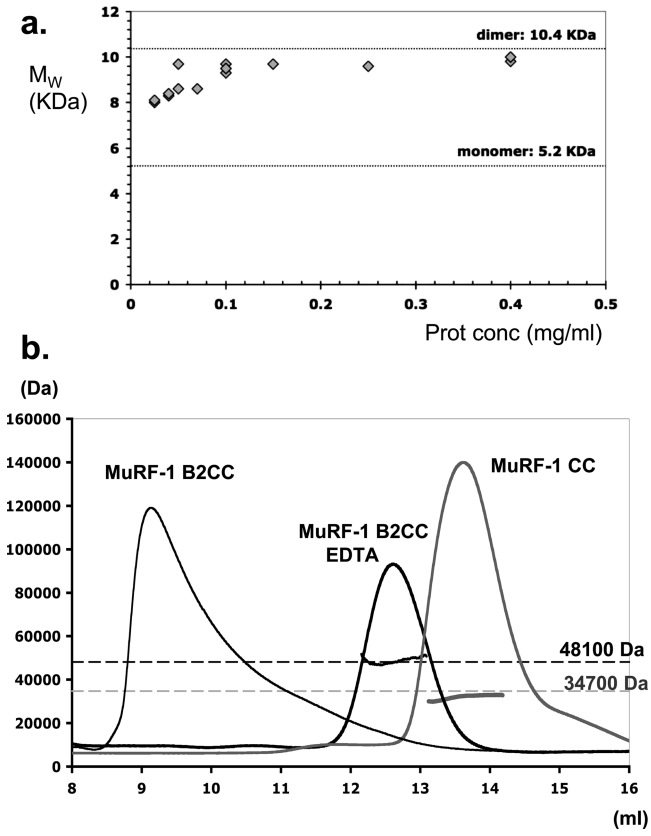


FIGURE 4: Oligomeric states of MuRF1 B2 and CC domains. (a) Molecular masses (M_r) of MuRF1 B2 calculated from sedimentation equilibrium measurements. (b) SEC−MALS measurements of MuRF1 CC and B2CC truncated variants. The average M_r per volume unit (dots) and the normalized refractive index (solid lines) are shown. The theoretical M_r values of the dimeric states of CC and B2CC constructs are represented as dashed horizontal lines. Unfolding of B2 in the B2CC variant was achieved by mixing with 50 mM EDTA and 10 mM β-mercaptoethanol for 16 h at 4 °C.

the Supporting Information. The stability of the dimer was confirmed by monitoring the sample through a dilution series [^{1}H−^{15}N HSQC spectra in the concentration range 3−0.3 mM and ^{1}H spectra for further dilutions down to 20 μM (Figures S5 and S6 of the Supporting Information)]. In agreement with sedimentation equilibrium data, NMR spectra remained virtually unchanged at the concentrations that were tested, indicating that MuRF1 B2 retained its association state through the concentrations that were assayed. Equivalent results were obtained from the NMR analysis of B2^{C137D}, proving that the dimerization properties of this domain are independent of its zinc-binding ligands.

Validation of the Dimer Structure. To assess whether the dimer of MuRF1 B2 in the crystalline state corresponds to that in solution, ^{1}H−^{1}H NOE and ^{15}N−^{1}H residual dipolar couplings (RDCs) recorded from weakly aligned samples in a medium containing Pf1 phage were compared to all possible structural models derived from the crystal lattice, namely, (i) the monomer, (ii) the dimer in the asymmetric unit (described above), and (iii) one additional dimeric form that arises in the lattice by crystallographic symmetry (Figure 3a). A description of the lattice of this crystal form is given in Figure S7 of the Supporting Information. The NMR data (Figure S8 of the Supporting Information) revealed 15 NOEs between the H$^\delta$ and H$^\epsilon$ ring protons of residue F148 and the cluster formed by I132, V156, A157, and P158. These are perfectly satisfied by dimer ii as short intermolecular contacts (<5 Å) (Table S1 of the Supporting Information and Figure 3b) but cannot be satisfied within the monomeric structure (i) or the lattice dimer (iii) [in which the respective proton distances are >10 Å (Table S1 of the Supporting Information)]. It should be noted that dimer ii, but not dimer iii, has its termini arranged in parallel and is consistent with the known composition of the TRIM fold where the B2 box is followed by a parallel coiled coil domain.

This result was in agreement with a subsequent analysis based on RDC data (Figure S9 of the Supporting Information), where alignment tensors were obtained by fitting the RDCs to the respective coordinates (models i−iii) (Figure 3a). In the case of dimeric structures, dipolar couplings for the equivalent internuclear vectors in both protomers are averaged. Thus, one of the principal axes of the alignment tensors for dimers must coincide with the dimer C_2 axis, and monomeric and dimeric forms cannot be distinguished on the basis of the fitting of RDCs to coordinates alone. The RDC data showed excellent agreement with the tertiary structure of the monomer (i) in the crystal ($Q = 0.21$; Q-factor definition as in ref *30*), confirming that the crystal structure of the MuRF1 B2 protomer represents closely the protein in solution. The fitting of RDCs to the noncrystallographic dimer ii yielded an agreement ($Q = 0.22$) similar to that of the monomer, which is expected if the internal symmetry of the dimer in the crystal is identical to that in solution. In contrast, lattice dimer iii fitted the data poorly ($Q = 0.64$), which rules out the existence of this form in solution. Thus, both NOE and RDC data corroborated the initial expectation that the dimer revealed by the crystal structure coincides with that present in solution.

Distinct Dimerization Pattern. The dimeric arrangement of MuRF1 B2 is different from that of other RING-like domains (Figure 5). Even though the dimerization modes of RING folds are highly heterogeneous, involving different secondary structure elements and sequence groups, some general patterns can be observed. The RING finger of RAG1 (*31*), the heterodimeric BRCA1−BARD1 tumor suppressor (*32*), the homodimeric FYVE domain from EEA1 (*33*) and the U-box of CHIP (*34, 35*) self-assemble via additional α-helical extensions. In those cases where direct contacts across the zinc-binding motifs are present, as in the Hdm2 RING (*36*) and the Prp19 U-box (*37*), these commonly involve the β-sheet component and/or its neighboring loop region. Helix α1 invariably lies in the outside, opposite to the dimer interface. A similar arrangement is also observed in the tandem interaction of boxes B1 and B2 in MID1 (*38*). By comparison, the MuRF1 B2 dimer could be regarded as having a "reverse" geometry. Furthermore, in the absence of additional helical extensions [with the exception of Hdm2 RING (*36*)], the self-association of other motifs appears weak (*37*), while the dimerization of MuRF1 B2 is stable and occurs at low concentrations. This suggests that the B2 box might contribute to the overall assembly of MuRF1.

Contribution of B2 and CC Domains to MuRF1 Quaternary Structure. MuRF1 contains a dimeric CC motif C-terminal to its B2 box (*14*). To investigate how the dimerization properties of the B2 motif relate to those imposed on MuRF1 by its CC domain, we compared the M_r values of two MuRF1 variants corresponding to the isolated

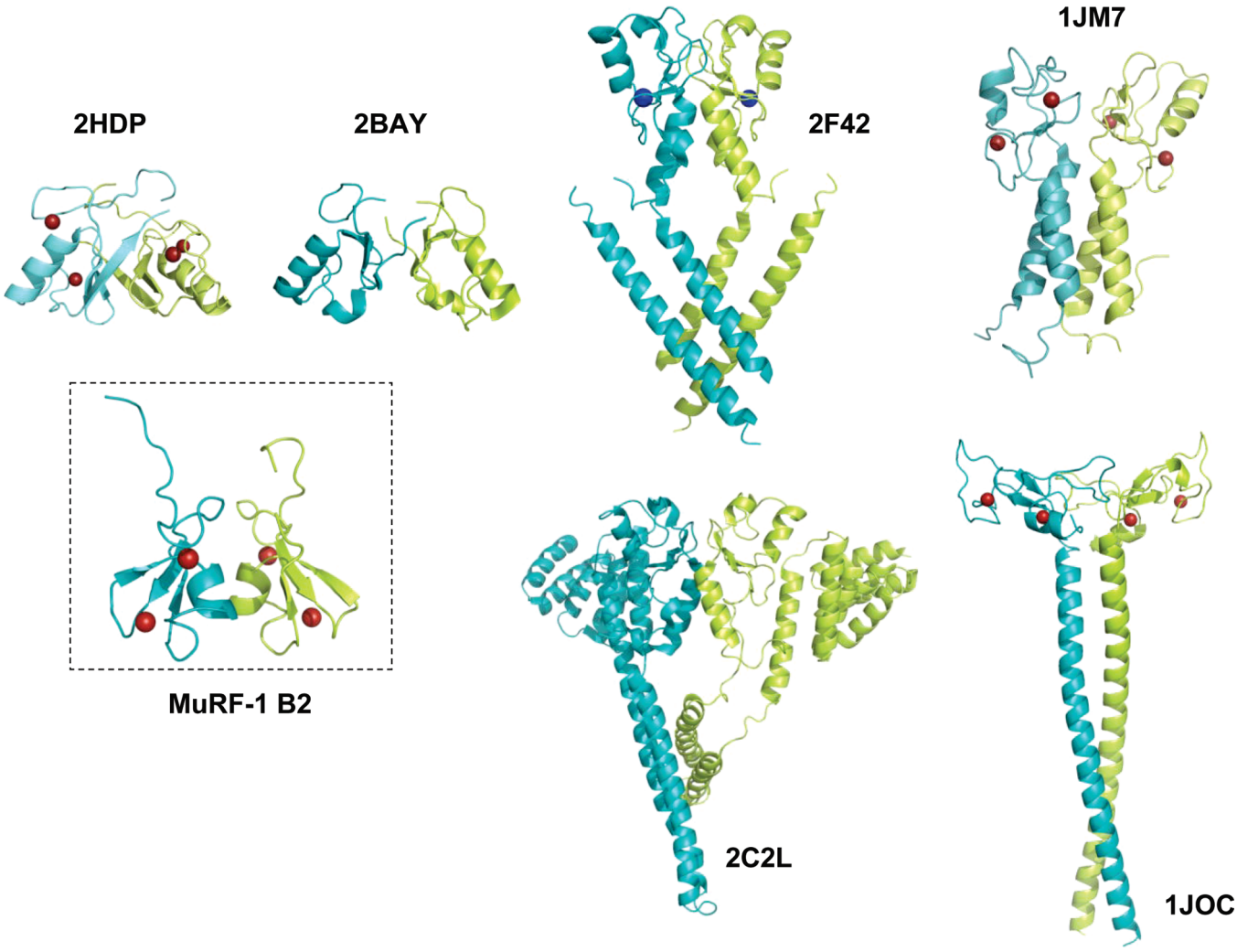


FIGURE 5: Dimeric structure of RING-like motifs. Structures correspond to the Hdm2 RING domain (2HDP), the U-box (2BAY), dimerization and U-box domains of CHIP (2F42), the BRCA1−BARD1 heterodimer (1JM7), the CHIP−Ubc13−Uev1a complex (2C2L), and the EEA1 homodimer (1JOC).

CC motif (MuRF1 CC) and the B2 box followed by the CC motif (MuRF1 B2CC) (construct definition as in Figure 1) using multiangle light scattering coupled to size exclusion chromatography (SEC−MALS) (Figure 4b). The data showed that MuRF1 CC forms dimers as main species in solution but that the presence of B2 in the B2CC variant leads to the formation of higher-order oligomeric species. Even though the polydispersity of the resulting assemblies did not allow calculation of their M_r accurately, this could be approximated to that of hexameric or octameric formations. To verify that the increased level of assembly in B2CC was due to the presence of the B2 box, a B2CC sample in which the B2 had been unfolded by metal depletion (using EDTA) was assayed. The latter formed exclusively dimers mediated by the intact CC fraction (Figure 4b), proving that B2 induces a high-order association. Clearly, the oligomeric state of full-length MuRF1 must become known before the physiological relevance of the B2-induced higher-order assembly can be established. Nonetheless, it is worth noting that these findings correlate with the evolutionary invariability of the B2 domain directly preceding the CC region, which suggests that the B2CC fraction might act as an integrated module in securing the correct quaternary structure of the TRIM fold.

Surface Features. To reveal the possible determinants of protein interaction in MuRF1 B2, we examined its surface topography. Each protomer exhibits a long, shallow groove that encircles the C-terminal metal site ZnII formed by loops L2 and L3. The groove comprises a hydrophobic cluster contributed by the solvent-exposed surface of the β-sheet that includes the highly conserved aromatic residue Y133. This groove does not resemble those defining ubiquitin ligase activity in RING domains (*34*), suggesting that this B2 box is unlikely to be involved in ubiquitination. Instead, one might speculate that the hydrophobic features of this groove, which are conserved in B2 sequences, might mediate the heterologous interactions established by this motif. These conclusions concur with those obtained from MID1 B2, whose role in ubiquitination was also thought to be unlikely and whose hydrophobic surface area was suggested as possible mediators of protein interactions (*9*). Potentially, this feature of the B2 motif might provide specificity when strategies for attempting the therapeutic targeting of TRIM proteins are designed.

Comparative Overview. The B-box fold belongs to the RING finger superfamily of zinc-binding motifs, which comprises RING (*39*), PHD (*40*), FYVE (*41*), ZZ (*42*), U-box (*43*), and B-boxes B1 (*8*) and B2. These folds,

although poorly related in sequence and loop structure, share a common $\beta\beta\alpha$ core that supports the ligation of two zinc ions in a cross-brace fashion (Figure S10 of the Supporting Information) [with the exception of the U-box domain that does not bind zinc but, instead, exhibits a network of hydrogen bonds and salt bridges (*43*)]. The different folds become individualized through the nature of their conserved zinc binders and the length and composition of the spacing sequences, which confer them substantial structural plasticity and allow a broad range of protein interactions and cellular roles to be supported. The B2 box is the most compact of these folds. The short sequences between its metal binders result in an inter-zinc distance of ~10.7 Å, similar to that of the FYVE motif, but shorter than that of ZZ and B1 boxes (~13 Å) and RING and PHD domains (~14 Å). Conserved hydrophobic residues within the respective RING folds mediate the packing of the β-sheet against the α-helix and are primarily located around the fifth metal ligand, in the segment connecting strand β2 to helix α1 (Figure S10 of the Supporting Information). Interestingly, the hydrophobic residues of MuRF1 B2 in this position do not form a motif core but are part of its dimerization interface. In fact, MuRF1 B2 does not have a defined hydrophobic core beyond the interaction of V156 with the aliphatic chain of K146, indicating that it is mostly stabilized by metal binding and self-assembly.

The α/β core of RING folds is most structurally similar at metal site ZnI, while the loops and metal ligands around C-terminal site ZnII display significant variability, both within and across families (Figure S10 of the Supporting Information). Also in the B2 group, variations accumulate around ZnII. All structures available to date correspond to the predominant $CHCDC_2H_2$ form, while MuRF1 B2 is the only representative of a CHC_2C_2HC variant. The difference in metal binders, however, does not translate into noticeable structural differences in the family, which forms a closely related structural group (Figure 2b). Finally, MuRF1 B2 is to serve as a close representative of the B2 domains of the MuRF2 and MuRF3 members of the MuRF family. Since this box contains unique structural features in contrast to other RING-like folds, it might provide specificity for the therapeutical targeting of MuRF1. In this respect, future studies of the molecular interactions established by this B2 domain will shed light on its cellular role and potential pharmacological relevance.

ACKNOWLEDGMENT

Special thanks go to the staff at beamlines ID23-1 (ESRF, Grenoble, France) and X06SA (SLS, Villigen, Switzerland) for support during data collection.

SUPPORTING INFORMATION AVAILABLE

Interproton distances within the MuRF1 B2 monomer and its possible dimeric states in the crystallographic lattice (Table S1), alignment of B2 sequences (Figure S1), comparison of crystallographic and NMR models of MuRF1 B2 (Figure S2), views of the $(2F_{obs} - F_{calc})\alpha_{calc}$ electron density map (Figure S3), NMR relaxation data (Figure S4), assigned HSQC NMR spectrum (Figure S5), comparative HSQC data for diluted B2 samples (Figure S6), description of dimeric formations in the crystalline lattice (Figure S7), NOE data (Figure S8), RDC data (Figure S9), and a comparative overview of RING-like motifs (Figure S10).

This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Meroni, G., and Diez-Roux, G. (2005) TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. *BioEssays 27*, 1147–1157.
2. Short, K. M., and Cox, T. C. (2006) Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *J. Biol. Chem. 281*, 8970–8980.
3. Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., Riganelli, D., Zanaria, E., Messali, S., Cainarca, S., Guffanti, A., Minucci, S., Pelicci, P. G., and Ballabio, A. (2001) The tripartite motif family identifies cell compartments. *EMBO J. 20*, 2140–2151.
4. Torok, M., and Etkin, L. D. (2001) Two B or not two B? Overview of the rapidly expanding B-box family of proteins. *Differentiation 67*, 63–71.
5. Shoham, N., Cohen, L., Gazit, A., and Yaniv, A. (2003) The Tat protein of the caprine arthritis encephalitis virus interacts with the Notch2 EGF-like repeats and the epithelin/granulin precursor. *Intervirology 46*, 239–244.
6. Short, K. M., Hopwood, B., Yi, Z., and Cox, T. C. (2002) MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, α4, to microtubules: Implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders. *BMC Cell Biol. 3*, 1–6.
7. Peng, H., Begg, G. E., Harper, S. L., Friedman, J. R., Speicher, D. W., and Rauscher, F. J., III (2000) Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. *J. Biol. Chem. 275*, 18000–18010.
8. Massiah, M. A., Simmons, B. N., Short, K. M., and Cox, T. C. (2006) Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. *J. Mol. Biol. 358*, 532–545.
9. Massiah, M. A., Matts, J. A., Short, K. M., Simmons, B. N., Singireddy, S., Yi, Z., and Cox, T. C. (2007) Solution structure of the MID1 B-box2 $CHC(D/C)C_2H_2$ zinc-binding domain: Insights into an evolutionarily conserved RING fold. *J. Mol. Biol. 369*, 1–10.
10. Borden, K. L., Lally, J. M., Martin, S. R., O'Reilly, N. J., Etkin, L. D., and Freemont, P. S. (1995) Novel topology of a zinc-binding domain from a protein involved in regulating early *Xenopus* development. *EMBO J. 14*, 5947–5956.
11. Bodine, S. C., Latres, E., Baumhueter, S., Lai, V. K., Nunez, L., Clarke, B. A., Poueymirou, W. T., Panaro, F. J., Na, E., Dharmarajan, K., Pan, Z. Q., Valenzuela, D. M., DeChiara, T. M., Stitt, T. N., Yancopoulos, G. D., and Glass, D. J. (2001) Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science 294*, 1704–1708.
12. Witt, C. C., Witt, S. H., Lerche, S., Labeit, D., Back, W., and Labeit, S. (2008) Cooperative control of striated muscle mass and metabolism by MuRF1 and MuRF2. *EMBO J. 27*, 350–360.
13. Centner, T., Yano, J., Kimura, E., McElhinny, A. S., Pelin, K., Witt, C. C., Bang, M. L., Trombitas, K., Granzier, H., Gregorio, C. C., Sorimachi, H., and Labeit, S. (2001) Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain. *J. Mol. Biol. 306*, 717–726.
14. Mrosek, M., Labeit, D., Witt, S., Heerklotz, H., von Castelmur, E., Labeit, S., and Mayans, O. (2007) Molecular determinants for the recruitment of the ubiquitin-ligase MuRF1 onto M-line titin. *FASEB J. 21*, 1383–1392.
15. McElhinny, A. S., Kakinuma, K., Sorimachi, H., Labeit, S., and Gregorio, C. C. (2002) MuRF1 interacts with titin to regulate sarcomeric M-line and thick filament structure and may have nuclear functions via its interaction with glucocorticoid modulatory element binding protein-1. *J. Cell Biol. 157*, 125–136.
16. Koyama, S., Hata, S., Witt, C. C., Ono, Y., Lerche, S., Ojima, K., Chiba, T., Doi, N., Kitamura, F., Tanaka, K., Abe, K., Witt, S. H., Rybin, V., Gasch, A., Franz, T., Labeit, S., and Sorimachi, H. (2008) Muscle RING-finger protein-1 (MuRF1) as a connector of muscle energy metabolism and protein synthesis. *J. Mol. Biol. 376*, 1224–1236.
17. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr. 26*, 795–800.

18. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D55*, 849–861.
19. Terwilliger, T. C. (2002) Automated main-chain model-building by template-matching and iterative fragment extension. *Acta Crystallogr. D59*, 34–44.
20. Perrakis, A., Morris, R. M., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol. 6*, 458–463.
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110–119.
22. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D58*, 1948–1954.
23. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D60*, 432–438.
24. Machaidze, G., and Lustig, A. (2006) SEGAL, a semi-automatic program for fitting sedimentation equilibrium patterns from analytical ultracentrifugation. *J. Biol. Phys. Chem. 6*, 91–102.
25. Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR 16*, 23–28.
26. Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol. 5*, 1065–1074.
27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) Nmrpipe: A Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR 6*, 277–293.
28. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) A Common-Sense Approach to Peak Picking in 2D, 3D, and 4D Spectra Using Automatic Computer-Analysis of Contour Diagrams. *J. Magn. Reson. 95*, 214–220.
29. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis 18*, 2714–2723.
30. Cornilescu, G., Marquardt, J. L., Ottiger, M., and Bax, A. (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc. 120*, 6836–6837.
31. Bellon, S. F., Rodgers, K. K., Schatz, D. G., Coleman, J. E., and Steitz, T. A. (1997) Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. *Nat. Struct. Biol. 4*, 586–591.
32. Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M. C., and Klevit, R. E. (2001) Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nat. Struct. Biol. 8*, 833–837.
33. Dumas, J. J., Merithew, E., Sudharshan, E., Rajamani, D., Hayes, S., Lawe, D., Corvera, S., and Lambright, D. G. (2001) Multivalent endosome targeting by homodimeric EEA1. *Mol. Cell 8*, 947–958.
34. Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000) Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell 102*, 533–539.
35. Xu, Z., Devlin, K. I., Ford, M. G., Nix, J. C., Qin, J., and Misra, S. (2006) Structure and interactions of the helical and U-box domains of CHIP, the C terminus of HSP70 interacting protein. *Biochemistry 45*, 4749–4759.
36. Kostic, M., Matt, T., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2006) Solution structure of the Hdm2 C2H2C4 RING, a domain critical for ubiquitination of p53. *J. Mol. Biol. 363*, 433–450.
37. Vander Kooi, C. W., Ohi, M. D., Rosenberg, J. A., Oldham, M. L., Newcomer, M. E., Gould, K. L., and Chazin, W. J. (2006) The Prp19 U-box Crystal Structure Suggests a Common Dimeric Architecture for a Class of Oligomeric E3 Ubiquitin Ligases. *Biochemistry 45*, 121–130.
38. Tao, H., Simmons, B. N., Singireddy, S., Jakkidi, M., Short, K. M., Cox, T. C., and Massiah, M. A. (2008) Structure of the MID1 tandem B-boxes reveals an interaction reminiscent of intermolecular ring heterodimers. *Biochemistry 47*, 2450–2457.
39. Barlow, P. N., Luisi, B., Milner, A., Elliott, M., and Everett, R. (1994) Structure of the C3HC4 domain by ^{1}H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. *J. Mol. Biol. 237*, 201–211.
40. Capili, A. D., Schultz, D. C., Rauscher, I. F., and Borden, K. L. (2001) Solution structure of the PHD domain from the KAP-1 corepressor: Structural determinants for PHD, RING and LIM zinc-binding domains. *EMBO J. 20*, 165–177.
41. Misra, S., and Hurley, J. H. (1999) Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. *Cell 97*, 657–666.
42. Legge, G. B., Martinez-Yamout, M. A., Hambly, D. M., Trinh, T., Lee, B. M., Dyson, H. J., and Wright, P. E. (2004) ZZ domain of CBP: An unusual zinc finger fold in a protein interaction module. *J. Mol. Biol. 343*, 1081–1093.
43. Ohi, M. D., Vander Kooi, C. W., Rosenberg, J. A., Chazin, W. J., and Gould, K. L. (2003) Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nat. Struct. Biol. 10*, 250–255.

BI800733Z

4.3 Recruitment of MuRF1 onto titin A168-A170

4.3.1 Introduction

The MuRF1 binding site in titin was mapped to A168-A170 (Centner et al., 2001). The crystal structure of this fragment has been elucidated and its interaction with MuRF1 confirmed by isothermal calorimetry (ITC), (work performed in our group prior to the initiation of this study; Mrosek et al., 2007).

The crystal structure of A168-A170 shows a monomer that adopts a rigid, extended conformation and exhibits a long shallow surface groove that spans its full length. The dimensions and geometry of this groove could well accommodate an α-helix along its length, especially if the helix exhibits axial bending as that of dimeric coiled-coils (Figure 4.10A). ITC confirmed the prediction (from SPOTS blots, Witt et al., 2005) that the helical fraction of MuRF1 is sufficient to induce high affinity interaction with A168-A170 (Kd $\approx$ 35nM).

Domain A169 shows an unusual 9 residue loop insertion between strands A and A' which is unique among all Ig of titin. Its location at the centre of the extended groove along the surface of A168-A170 indicates, that this loop protrusion could be important for binding to MuRF1. Indeed, a mutated A168-A170 variant, where residues $_{102}$-KTLE-$_{105}$ (according to internal numbering of the structure) had been replaced by alanines did no longer bind MuRF1 in pull down experiments (Mrosek et al., 2007).

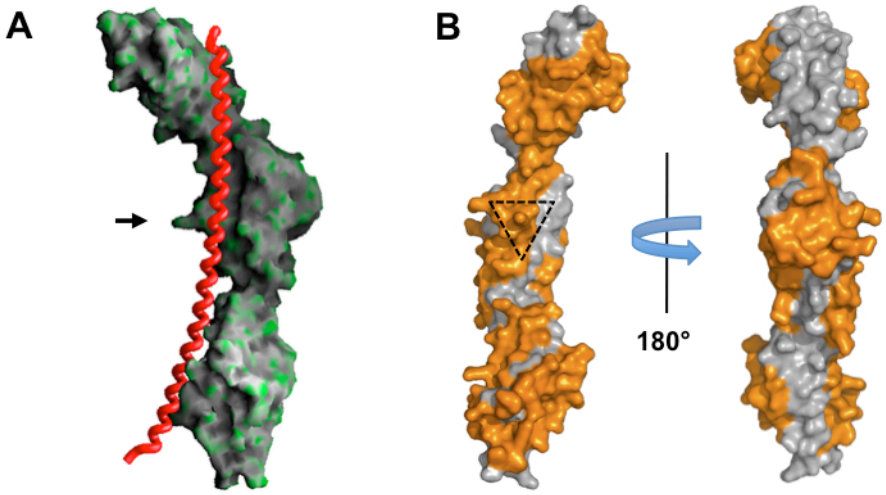


Figure 4.10: Binding of MuRF1 to A168-A170.
A) Surface representation of crystal structure of A168-A170, where green emphasizes local curvature. Docking of α-helical peptide in coiled-coil conformation (model extracted from Lamin A coil 2B, PDB code 1X8Y) according to shape complimentarity shows that this groove could accommodate the extended α-helical region of MuRF1. The arrow marks the unique loop insertion in A169
B) Mapping of SPOTS blots peptides of A168-A170 that interact with MuRF1$^{176-327}$ mapped onto the surface of the structure. The Triangle marks the unique loop insertion in A169.

In this study we set out to establish stable complex formation between A168-A170 and MuRF1 *in vitro* for biophysical and structural characterization. This work builds on the characterization of MuRF1 described in the previous chapter.

4.3.2 Experimental work

4.3.2.1 Constructs:

For details of the expression plasmid for A168-A170 see (Mrosek et al., 2007). Constructs of MuRF1 employed in this study are described in the previous chapter.

4.3.2.2 Protein production:

Production of titin A168-A170 (Figure 4.11) followed the procedure reported in (Mrosek et al., 2007). MuRF1 constructs were expressed and purified as outlined in chapter 4.2.

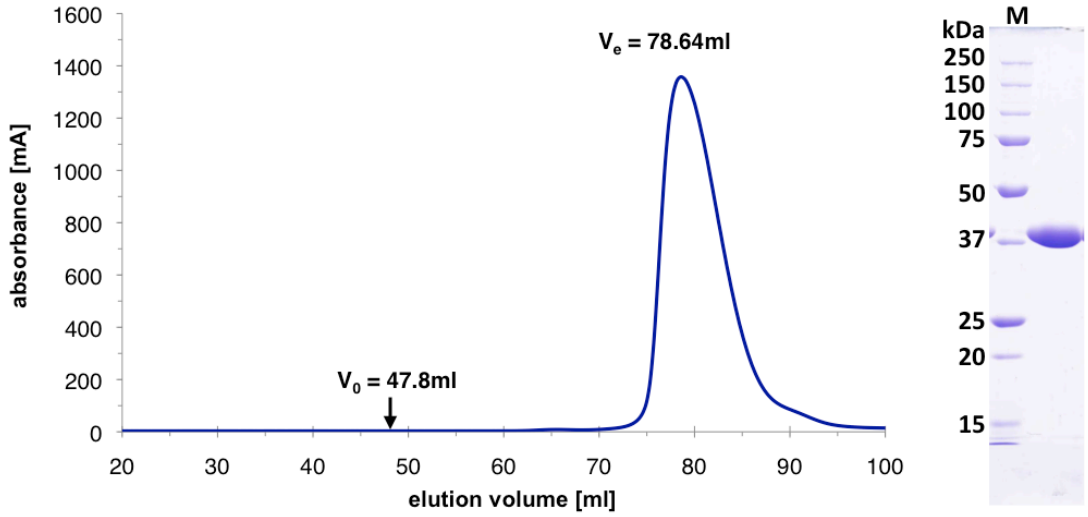


Figure 4.11: Size exclusion chromatogram of titin A168-170
Representative size-exclusion chromatogram of A168-A170 (Mr= 33874 Da), performed on a Superdex200 16l60 column (GE Healthcare) in 50 mM Tris pH8.0, 100mM NaCl as buffer. The signal shown corresponds to A_{280nm}. The final purity of corresponding samples according to SDS-PAGE is shown on the right. The protein standard (lane marked M) is the Precision Plus Protein standard (Biorad).

4.3.2.3 Co-crystallization:

Co-crystallization of the complex was attempted by incubating pure samples of A168-A170 with MuRF1$^{176-327}$ or MuRF1$^{176-353}$. All proteins were at an initial concentration of 10mg/ml and were mixed in a molar ratio of 1.2:2 for A168/MuRF (taking into account dimeric MuRF samples, this corresponds to excess of A168-A170). Subsequently samples were reconcentrated to 10mg/ml.

Crystallization trials were performed using vapor diffusion, where 1µl of protein was mixed with 1µl of reservoir solution to form a 2µl hanging drop. Each droplet was equilibrated against 1ml of reservoir solution in VDX plates (Hampton Research). For initial screening, the sparse matrix kits Hampton Research CS-I & II, and Emerald Wizard I & II, were used. For each condition 4 drops were set up corresponding to MuRF1$^{176-327}$, MuRF1$^{176-353}$, A168-A170/MuRF1$^{176-327}$ or A168-A170/MuRF1$^{176-353}$ samples. This resulted in some small needle like crystals in drops containing the complexation sample but not the MuRF constructs alone. Reproduction with the complexation sample and control drops containing only A168-A170 proved that the crystals contained only A168-A170 without any MuRF1 complexed (similar reservoir composition with regard to the condition reported for A168-A170 crystallization in Mrosek et al., 2007).

Thus, in order to ensure that samples employed in crystallization trials contained the complex, further work focused on isolation of complexed protein samples prior to crystallization trials.

4.3.2.4 Complexation of A168-A170 and MuRF1

Complex isolation used pull-down protocols on His_6-tagged A168-A170 that had been mixed and incubated in solution with untagged MuRF1$^{117-327}$ and MuRF1$^{176-327}$ samples. A168-A170 and MuRF1 samples were incubated either in batch dialysis (e.g. for buffer exchange during incubation) or by step-wise titration followed by incubation on a rotary mixer. Both forms of complex incubations were performed at 4°C and at room temperature, and with protein concentrations ranging from 0.2-7mg/ml. The incubation time varied between two to three hours for standard incubations using step-wise titration to over night incubations in dialysis with concomitant buffer exchange.

Both proteins are soluble and stable at concentrations of 5-10mg/ml in their storage buffer, 50mM Tris pH 8.0, 100mM NaCl. However, mixing of the samples, either in dilute solutions (~0.1-0.2 mg/ml), followed by concentration, or with one or both proteins more highly concentrated (~1-2mg/ml or 5-7mg/ml for titration) led to precipitation in a concentration dependent manner. In incubations at low concentrations solutions remained clear or showed some minor precipitation. During concentration samples precipitated in the form of white floccules. If incubations were performed at higher protein concentrations, in order to avoid the need for concentration, samples precipitated during incubation, often detectable within minutes. Analysis of the sample precipitate by SDS-PAGE showed the presence of both proteins in the pelleted precipitate. A168-A170 could also be detected in solution, presumably in the form of excess protein that was not involved in complex formation (see Figure 4.12).

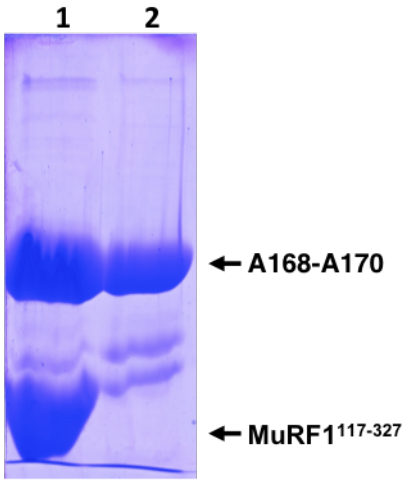


Figure 4.12: Analysis of the insoluble fraction of complexation samples
Complexation of A168-A170 with MuRF1$^{117-327}$, incubated at 4°C for 4h, under agitation. 1) pelleted precipitate and 2) supernatant of sample after complexation incubation. Arrows indicate the two proteins.

Thus, it appeared that the protein complex was formed, but that the protein complex has a different solubility profile. By increasing the salt concentration to 300mM, the complexation sample stayed in solution and could be assayed by size exclusion chromatography. Both proteins eluted as separate peaks with no extra peak indicating complex formation (see Figure 4.13). Thus, it appears that complex formation is dependent on electrostatic contacts. Indeed the two proteins exhibit very different pI (A168-A170 theoretical pI: 8.6; MuRF1$^{117-327}$ theoretical pI: 4.9; MuRF1$^{117-327}$ theoretical pI: 4.7). Thus MuRF1 constructs will carry an overall negative charge in solution, while A168-A170 will most likely be slightly positively charged. Thus, when combined in solution, the complex might form, and bring the two molecules together to form an equivalent of a dipole. It could be speculated that these complexed molecules might further associate into electrostatically bound aggregates.

It cannot be excluded though, that the elution of the proteins as individual peaks could also be due to complex separation by the gel filtration matrix and flow pressure exerted on the sample.

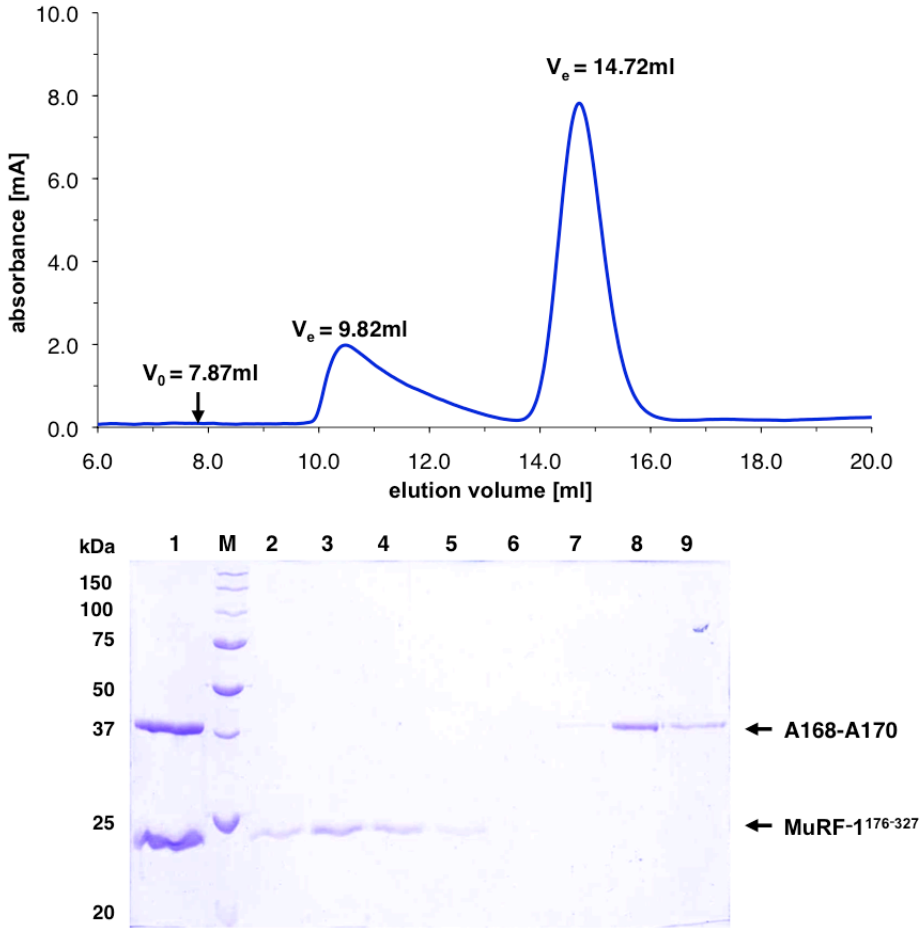


Figure 4.13: Size-exclusion chromatogram of A168-A170/MuRF1$^{117-327}$ mixtures
Size exclusion run of A168-A170:MuRF1$^{117-327}$ complex run on Superdex200 10l300GL Tricorn equilibrated in 50mM Tris 100mM NaCl, pH 8.0. The signal shown corresponds to A_{280nm}. The two peaks of the chromatogram correspond to the normal elution volumes of the individual protein samples, with no extra peak indicating complex formation.
SDS-PAGE of samples collected from size exclusion chromatography. Lane 1 corresponds to the complexation sample loaded onto the column, while lanes 2-9 correspond to consecutive samples taken covering the two peak fractions, from 9-16ml elution volume. Lane M corresponds to the Precision Protein Plus Standard (Biorad)

Assuming that complex formation is dependent on the formation of electrostatic contacts in the interface, the concomitant shielding of these charged residues in the protein interface could be responsible for the different solubility profile of the complex. Since pH is an important determinant of protein solubility by influencing the protonation state of polar residues, we assessed the influence of pH on complex formation and solubility. For this, incubations were set up in dialysis against a range of buffer pH and assayed for soluble complex formation by pull down experiments.

Samples were prepared in 50mM Tris, pH 8.0, 300mM NaCl, since this salt concentration was high enough in previous experiments to prevent complex formation (and therefore precipitation). The samples were mixed and immediately put into dialysis against: 1) 50 mM sodium acetate, pH 4.5, 100mM NaCl; 2) 50mM sodium

acetate, 100mM NaCl, pH 5.0; 3) 50mM MES, 100mM NaCl pH 6.0; 4) 50mM HEPES, 100mM NaCl, pH 7.0; 5) 50mM HEPES, 100mM NaCl, pH 8.0; 6) 50mM CHES, 100mM NaCl, pH 9.0. After dialysis over night at 4°C samples were incubated with Ni-agarose beads (Qiagen) for one hour, washed, and bound sample eluted by supplementing buffers with 200mM imidazole.

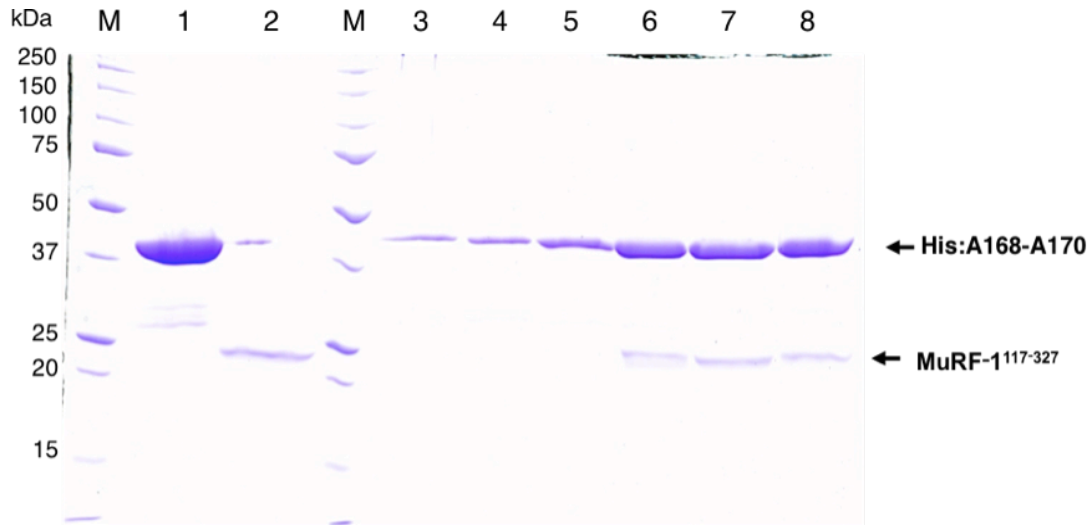


Figure 4.14: Pull-down of the titin A168-170:MuRF1$^{117-327}$ complex
Pull-downs were performed on samples incubated over night and made use of the Histag of A168-A70 samples.
Lanes: 1) A168-A70 sample used for complexation, 2) MuRF1$^{117-327}$ sample prior to mixing
lanes 3-9) elution of pull-downs of complexed samples at varying pH.
3) 50mM sodium acetate, 100mM NaCl, pH 4.5
4) 50mM sodium acetate, 100mM NaCl, pH 5.0
5) 50mM MES, 100mM NaCl pH 6.0
6) 50mM HEPES, 100mM NaCl, pH 7.0
7) 50mM HEPES, 100mM NaCl, pH 8.0
8) 50mM CHES, 100mM NaCl, pH 9.0

As seen in Figure 4.14, complex formation could be detected by pull-down at pH 7.0, 8.0 and 9.0. Complex samples incubated at lower pH might not have bound to the resin, and thus, complex formation at these lower pH cannot be excluded. In order to separate any uncomplexed protein and for confirmation of complex formation samples were concentrated for size-exclusion chromatography. As seen previously for other incubations at dilute concentrations, the samples started to precipitate; attempts to prevent this by the addition of 5% glycerol as a stabilizing agent did not succeed.

4.3.2.5 Co-expression of A168-A170 with MuRF1 constructs MuRF1$^{117-327}$ and MuRF1$^{176-327}$

For co-expression of A168-A170 with MuRF1$^{117-327}$ or MuRF1$^{176-327}$, the corresponding coding sequences were cloned into the bi-cistronic pET-DUET vector (Novagen), which is designed for the simultaneous expression of two target proteins. The vector contains two multiple cloning sites (MCS), each of which is preceded by a T7 promotor/lac operator and a ribosome binding site. The restriction enzymes XbaI and BamHI were used to excise the coding sequence for A168-A170 as well as the preceding His_6-tag and TEV cleavage site. This sequence was then inserted into the MCS1 of pET-DUET. MuRF1$^{176-327}$ or MuRF1$^{117-327}$ were prepared for insertion into MCS2 by PCR amplification, and treated with Nde1 and Xho1 restriction enzymes for ligation into MCS2. This gave rise to two variations of the vector with A168-A170 including a cleavable His6-tag in MCS1 and MCS2 containing either MuRF1$^{176-327}$ or MuRF1$^{176-353}$, both without an affinity tag.

Protein production followed the protocol for the production of the individual proteins, using the E. coli strains BL21(*DE3*) Rosetta, pLysS and RIL. Cultures were grown in Luria Bertani medium supplemented with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol. Lysates were tested for soluble protein expression and complex formation using Ni-NTA agarose beads (Qiagen) and the fractions assayed by SDS-PAGE (see Figure 4.15A).

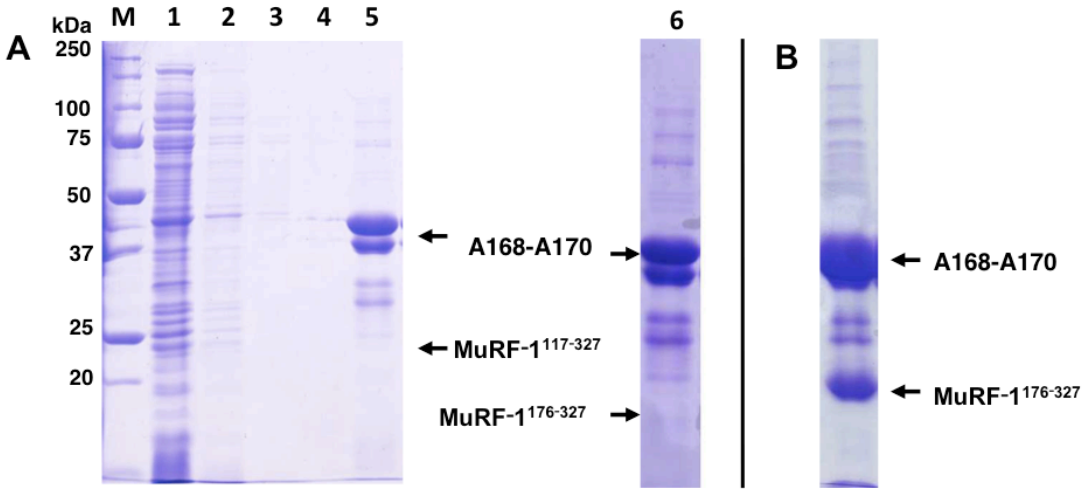


Figure 4.15: Affinity purification of co-expressed A168-A170 and MuRF1 constructs
Clarified cell lysate was incubated with Ni-agarose beads (Qiagen) for batch binding, subsequently washed with lysis buffer and bound protein eluted with lysis buffer supplemented with 250mM imidazole.
A) Affinity purification of A168-A170 coexpressed with MuRF1$^{117-327}$. Lanes: M shows Precision Plus Protein standard (Biorad), 1) soluble fraction of lysate, 2) unbound sample, 3-4) washes, 5) elution of bound protein. Lane 6) elution of corresponding affinity purification of the coexpression of A168-A170 with MuRF1$^{176-327}$. A168-A170 is clearly overexpressed in the soluble fraction and binds to the resin through its His_6-tag, but no expression or copurification of MuRF1 constructs could be detected. The prominent double bands correspond to A168-A170. Addition of extra reducing agent leads to the collapse into a single band.
B) Elution fraction of affinity purification of samples in the soluble fraction from expression in co-transformed *E. coli* cells. Since in this case both constructs carry a His_6-tag, the presence of both proteins in the elution fraction does not necessarily indicate to complex formation

No expression could be detected for either MuRF1$^{117-327}$ or MuRF1$^{176-327}$ when using this bicistronic vector. Therefore, we decided to test co-transformation as a method for coexpression of A168-A170 and MuRF1$^{176-327}$ (C. Stutz, Biozentrum Basel). For this MuRF1$^{176-327}$ in pETM-11 and the pETDuet-1 vector containing only the A168-A170 gene were simultaneously transformed into *E. coli* RIL strain and expression was according to the standard protocol. Lysates were tested for soluble protein expression and complex formation using Ni-NTA agarose beads (Qiagen) and the fractions assayed by SDS-PAGE (see Figure 4.15B). This showed that co-transformation had worked and that both proteins were expressed in soluble form and at similar levels as observed for the individual constructs. Since both constructs carried a His_6-tag, the co-localization of both constructs to the elution fraction could not be taken as an indication for successful complex formation. To assess complex formation size exclusion chromatography was performed. For this, the sample was dialyzed against 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM DTT at 4°C over night and subsequently concentrated for application on a Superdex200 (10l300 Tricorn) column (GE Healthcare).

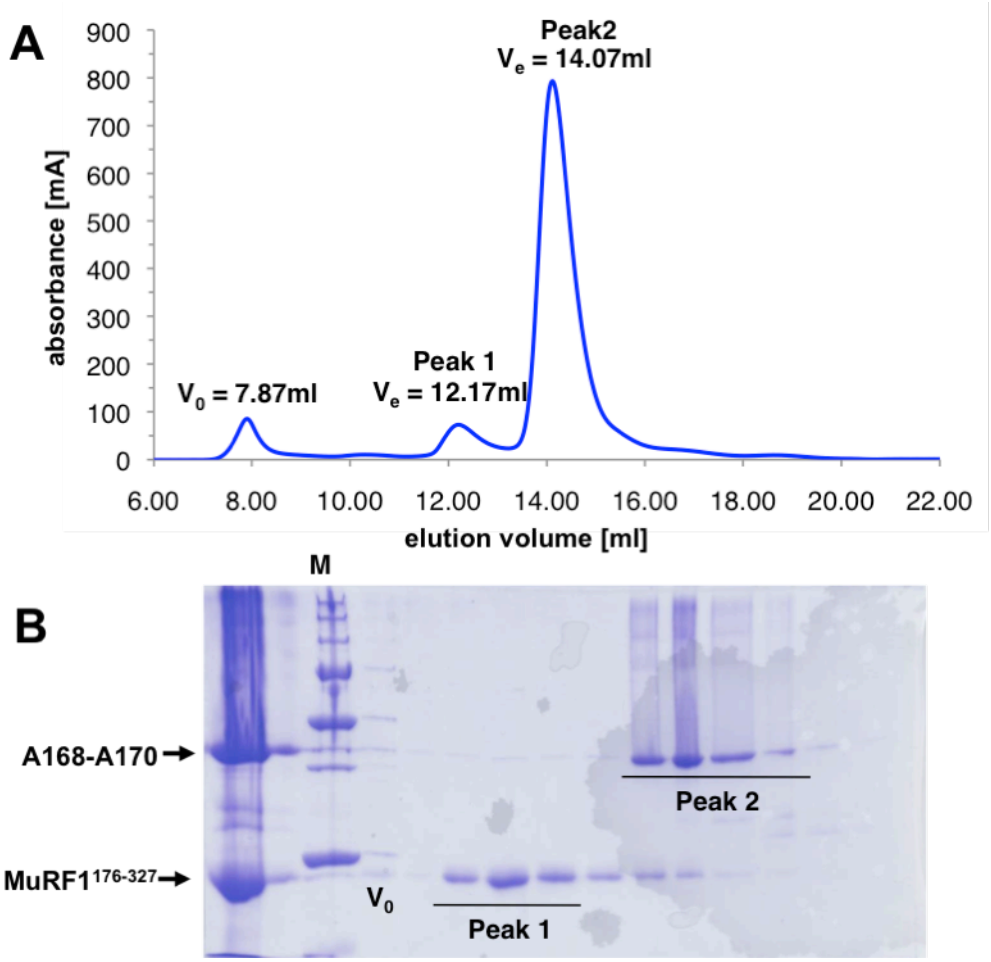


Figure 4.16: Size-exclusion chromatography of A168-A170 and MuRF1$^{176-327}$ coexpression sample.
A) Size exclusion run of the A168-A170/MuRF1$^{176-327}$ coexpressed sample performed on a Superdex200 (10l300 Tricorn) equilibrated in 50mM Tris 100mM NaCl, pH 8.0. The signal shown corresponds to A_{280nm}. The two peaks of the chromatogram correspond to the normal elution volumes of the individual protein samples, with no extra peak indicating complex formation.
B) SDS-PAGE of samples collected from size exclusion chromatography. Lane 1 corresponds to the complexation sample loaded onto the column, while M is the Precision Plus Protein Standard (Biorad). V_0 marks the fraction corresponding to the exclusion volume, the fractions corresponding to the peak 1 and peak 2 are marked correspondingly.

Unfortunately, size exclusion chromatography showed that the samples localized in independent exclusion volumes, with each volume being characteristic of the isolated species, with no extra peak indicating complex formation (Figure 4.16). Thus, it appears that co-expression followed by co-purification of the samples did not lead to a successful production of a complexed state. Again, it cannot be excluded though, that the elution of the proteins as individual peaks could also be due to complex separation by the gel filtration matrix and flow pressure exerted on the sample.

4.3.3 Discussion

The high-affinity interaction between titin A168-A170 and MuRF1 has previously been shown by ITC and pull-downs (Mrosek et al., 2007). Here, the main goal was the further biophysical and structural characterization of the A168-A170/MuRF1 complex. This was contingent on establishing a protocol for successful production of soluble complex samples. While we could indeed establish conditions in which the complex was formed, as confirmed by pull-down experiments, we were unable to obtain the complex at concentrations necessary for biophysical characterization or crystallization screening. Based on the experience gained during this study we would suggest a range of possible experiments to overcome the solubility issues of the complex sample.

As a first suggestion, complex formation should be carried out in the conditions identified to yield complex in dilute samples, with the supplementation of NDSBs (non-detergent sulfo-betaines) to buffers employed. NDSBs are non-denaturing, zwitterionic small compounds that show a good solubility in water without significantly altering the pH or viscosity of biological buffers. The have been successfully employed for solubilization of folding-intermediates during protein refolding (Vuillard et al., 1998) and as additives in crystallization experiments (Vuillard et al., 1996). The latter would be a bonus if soluble complex could be obtained, as the buffer would have not have to be modified for crystallization screening.

Alternatively, samples could be incubated in dilute solutions for complex formation, followed by chemical cross-linking. This would allow increasing the salt concentration during sample concentration, while preserving the complex. This should render possible further biophysical characterization and crystallization of the complex.

Furthermore, recently a new construct of MuRF1, MuRF-1$^{214\text{-}271}$, has become available and structure elucidation is in progress (see above). Assuming that this minimal construct covering the predicted coiled-coil fraction of the α-helical region still interacts with A168-A170, its small size should allow for NMR titration experiments to map the interacting residues between the two proteins, which would allow for guided docking of the two structures.

4.4 Structural studies of C. elegans twitchin kinase

4.4.1 Introduction

While it would be very desirable to have a structural model of the M-line region of titin surrounding the kinase and thus integrating the domain into the context of its position in the filament, human TK is extremely difficult to produce in yields required for structural biology. Therefore, *C. elegans* twitchin is often chosen as a model, due to its high similarity and the ease of recombinant production in the yields necessary for biophysical and structural characterization.

Twitchin, encoded by the gene *unc-22*, is a protein of ~800kDa that interacts with myosin and is associated to thick filaments (Benian et al., 1989; Moerman et al., 1988). Studies of Aplysia (Probst et al., 1994) and Mytilus twitchin (Siegman et al., 1998) showed that twitchin is involved in the inhibition of the rate of relaxation and that cAMP dependent phosphorylation of twitchin relieves this inhibition. Recombinant constructs of twitchin kinase were shown to have Ser/Thr kinase activity, both in autophosphorylation (predominantly on Thr5910, but possibly also on Ser5912 and Ser5913; Lei et al., 1994) and towards peptides derived from vertebrate myosin light chains (Heierhorst et al., 1996b; Lei et al., 1994).

Twitchin encompasses a single MLCK-like kinase (belonging to the CAMK family) that is 52% identical to chicken smooth muscle myosin light chain kinase (39% sequence identity with TK; Lei et al., 1994), as well as 31 FnIII, 30Ig, and some unique insertions. For much of its length twitchin exhibits the equivalent of a minimal titin A-band repeat composed of FnIII-FnIII-Ig that is followed by a conserved Ig-Ig-FnIII---Kin-reg-Ig region towards the C-terminus (see Figure 4.17) (Benian et al., 1989; Flaherty et al., 2002; Labeit et al., 1992).

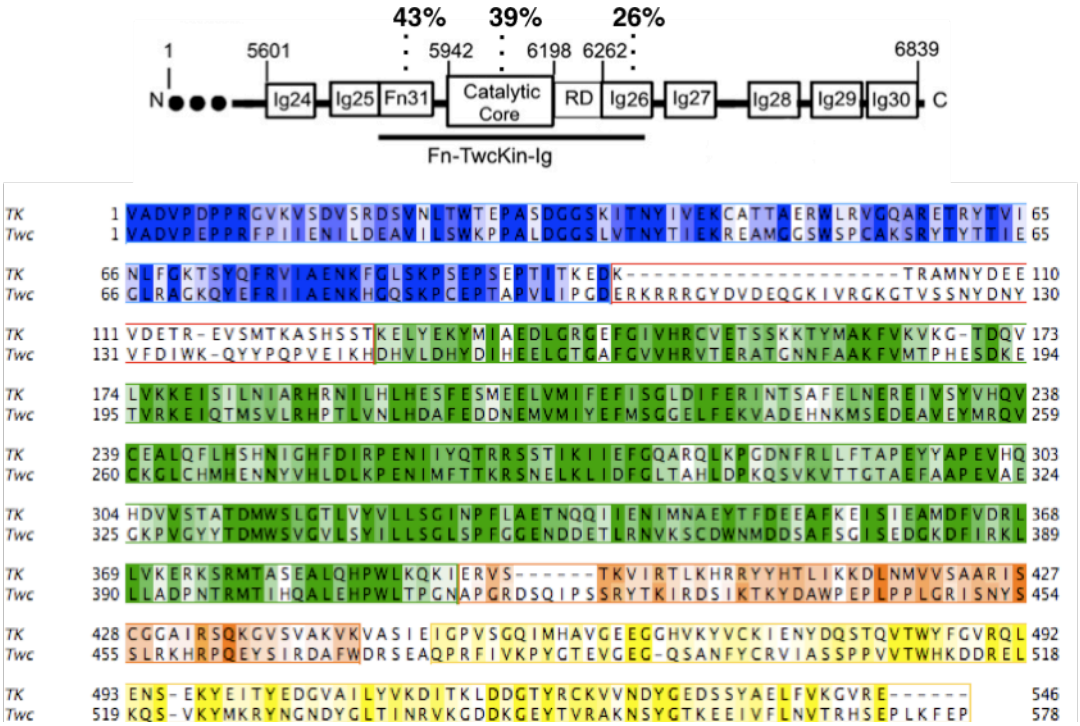


Figure 4.17: Domain composition of twitchin C-terminal region
The domains Ig24-Ig25-Fn31-twcKin-RD-Ig26 follow the conserved domain arrangement seen in many of the giant protein kinases and correspond to human titin domains A168-A169-A170-TK-RD-M1. The construct under investigation is indicated below. RD – regulatory domain (adapted from Greene et al., 2008). The sequence identity between twitchin and titin is given on a per domain basis. Below, the sequence alignment of the twitchin construct and the corresponding region of titin is colored on a per domain basis, to facilitate visualization: FnIII – blue; linker sequence – red; kinase domain – green; regulatory domain – orange; Ig – yellow.

In vivo, twitchin kinase is auto-inhibited by a 60 residue C-terminal regulatory sequence (Lei et al., 1994), immediately following the kinase catalytic core and preceding the Ig domain, that wedges itself between to the two lobes of the kinase (Figure 4.18). The crystal structure of twitchin kinase reveals that the regulatory tail makes extensive contacts with the catalytic core (1KOA; Hu et al., 1994b; Kobe et al., 1996) and exploits practically all features of the active site, blocking binding sites for protein substrate, ATP and Mg^{2+} as well as catalytic residues and, particularly the Thr-rich activation loop (Figure 4.19).

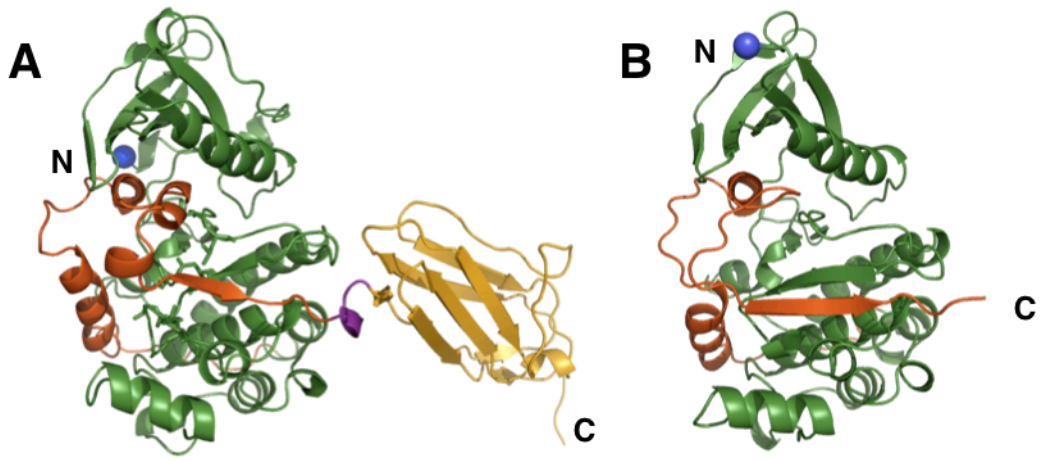


Figure 4.18: Structures of twitchin and titin kinase
A) Ribbon diagram of the structure of twitchin kinase (1KOA).
B) Ribbon diagram of titin kinase (1TKI), in the same orientation as twitchin kinase. The protein kinase domain is colored green, the autoregulatory sequence orange, Ig-26 yellow and the linker between the protein kinase and Ig-26 domain is in purple. The N-terminus is marked with a blue sphere for better visibility. The two structures are very similar, with an rmsd of 1.1 over 275 Cα atoms (calculated using SPDBV; Guex and Peitsch, 1997).

Specifically, the αR2 helix binds to the kinase ATP binding site and specifically interacts with the residues involved in catalysis. Hindsight analysis showed this to be the reason why attempts to soak ATP into the crystals were unsuccessful (Hu et al., 1994b). Furthermore, in the crystal structure the following Ig (26') makes contact with the kinase domain on the opposite side from the active site (Figure 4.18). It has been speculated that this domain might bind to myosin and thus help position the twitchin kinase next to its substrate (Kobe et al., 1996). The model suggests a relatively rigid inter-domain association with around 720 $Å^2$ of surface area buried between the two domains, (if counting the linker as part of the Ig domain this number rises to 1006 Å2; Kobe et al., 1996).

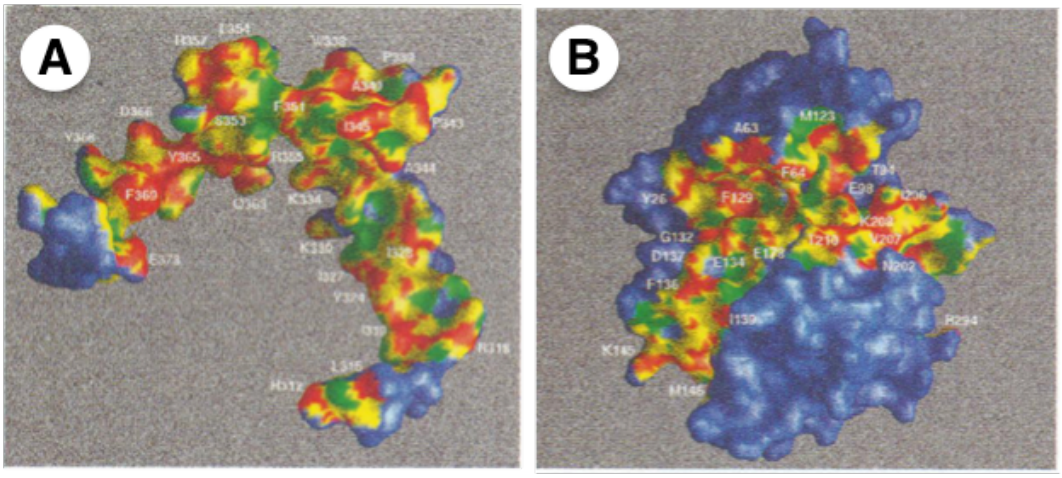


Figure 4.19: Interactions of the autoregulatory tail with the TWK43 kinase domain
(A) Molecular surface of the regulatory region of TWK-43 color-coded according to surface complementarity. Selected residues are labeled. (B) Molecular surface of the catalytic domain of TWK-43 coded-coded according to surface complementarity as in (A). The orientation is related to (A) by a 180° rotation around the y-axis. (Modified from Kobe et al., 1996)

Most members of the MLCK family, including TK, are activated by Ca^{2+}-calmodulin binding to the regulatory tail (Dabrowska et al., 1977; Means et al., 1991). Both nematode and *Aplysia* twitchin can be shown to bind Ca^{2+}-calmodulin, but neither enzyme becomes activated (Heierhorst et al., 1996a). While binding of $Ca^{2+}S100A1_2$ on the other hand was shown to activate the Twitchin kinase >1000-fold (Heierhorst et al., 1996a), this is clearly not the mechanism of kinase activation *in vivo*, since the *C. elegans* genome does not contain any S100 proteins (Ferrara et al., 2005). Also, the *C. elegans* proteome lacks a telethonin homolog, thus the physiological substrates for twitchin remain to be identified (Ferrara et al., 2005).

Possibly, twitchin kinase is regulated by the same stretch-induced mechanism that has been proposed for TK (Gräter et al., 2005). In their study Gräter and colleagues employed force-probe MD simulations to explore the tension-induced activation of TK by partial unfolding. Both kinase termini were subjected to a harmonic pulling potential to mimic the tension in a stretched titin filament (Pulling speeds in MD simulations are 4-5 orders of magnitude larger than for AFM experiments or *in vivo*). This allowed estimation of the forces required for mechanical unfolding and examination of force-induced structural changes. They concluded that mechanical strain leads to an ordered sequence of conformational changes that result in an opening of the catalytic cleft, without compromising the structural integrity of the catalytic core of the enzyme. The stability of the core appears to be a function of the orientation of the molecule with respect to the pulling force, as β-sheets exhibit a differential force resistance depending on their relative orientation to force and

consequential unfolding mechanism. β-sheets parallel to the pulling force rupture at higher loads through shearing (concurrent rupture of several hydrogen bonds), and thus are more force resistant, than the β-sheets vertical to the pulling direction, which are pulled apart in a zipper like fashion (hydrogen bonds ruptured in a sequential fashion). In the context of their simulation[6], the apparent pulling geometry of activation for TK is such that the β-sheets most responsible for the force resistance are located parallel to the pulling force, but the β-sheets responsible for exposing the active site are oriented perpendicular to the force. Thus, the autoregulatory tail can be pulled from the catalytic core while maintaining the integrity of the active site.

In order to test this model, a recent study explored the step-wise unfolding of a twitchin kinase construct (FnIII-TwcKin-Ig) as a model for titin kinase by AFM (see Figure 4.17) (Greene et al., 2008). They found that the twitchin kinase unfolds at 30-150pN, which is similar to the unfolding range of nematode Ig and Fn domains at 40-180pN assessed in the same study. Force extension curves exhibit a sawtooth like pattern with four peaks, indicative of a step-wise unfolding of domains. First, the autoregulatory tail unwinds, probably at very low forces (<10pN), which is indistinguishable from the noise of the AFM. This is followed by the biphasic rupture of the catalytic core, where the smaller N-terminal lobe unfolds first at forces of ~50pN, followed by the unraveling of the larger C-terminal lobe at forces of ~80pN. The last two force peaks are assigned to the rupture of the Ig and FnIII domain. The authors employed a SMD simulation to confirm this model (Greene et al., 2008).

A subsequent study followed up on these results and explored ATP binding of mechanically activated human titin A168-M2 by AFM (Puchner et al., 2008). On the premise that *i)* mechanical activation must uncover the ATP site while leaving the active site intact and *ii)* subsequent ATP binding alters the stability of the enzyme, they explored ATP-binding dependent changes in TK unfolding profiles by AFM and MD simulations. They also reported autophosphorylation[7] of a recombinant construct lacking the autoregulatory tail (thus mimicking mechanically activated state) in an ATP dependent manner as well as its activity on telethonin. Thus, they conclude that TK can indeed function as a biological stress sensor.

---

[6] Currently there is no experimental evidence on the orientation of the kinase within the context of the filament and the directionality of the applied forces.

[7] on a tyrosine residue as detected by antibodies against phosphotyrosine

Our study was initiated with the aim to extend the repertoire of structural data available for this region. While, indeed, the structures of many domains have been solved, namely titin A168-A170 (2NZI; Mrosek et al., 2007), human TK (1TKI; Mayans et al., 1998), twitchin kinase (1KOA; Kobe et al., 1996) and titin M1 (2BK8; Müller *et al.*, unpublished), we currently lack structural data on the linkers that connect these domains (see Figure 4.20).

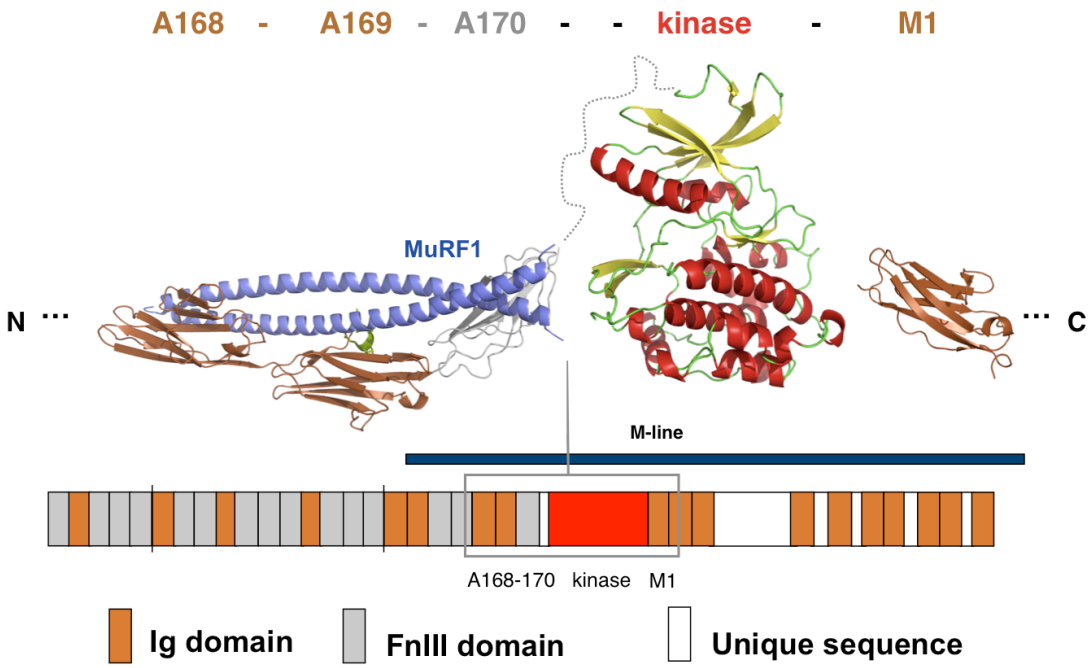


Figure 4.20: Modeling the M-line signalosome
Assembly of the available structures of titin domains involved in the M-line signalosome. In order to build a complete model of the signalosome, structural data is needed for the linkers connecting TK to its neighboring domains on either side as well as for the complex of A168-A170/MuRF1, here modeled by docking of a parallel coiled-coil. (Figure prepared by O. Mayans)

Without this essential information we cannot build an integrated model of the kinase region in the filament. While the structure of twitchin kinase includes the following domain, Ig26, thus linking the kinase to the C-terminal region, we have no data for the N-terminal linker, corresponding to the connection between titin kinase and the MURF1 interacting domains A168-A170. Structural data relating these two fragments to each other in space should help to shed light on how binding to A168-A170 positions MuRF1 in relation to TK and thus how these potentially interact.

A structural model connecting the kinase to its neighboring domains will also provide important information in relation to the model of titin kinase as a stretch

sensor, by revealing the connections that help anchor and orient the kinase with regard to the stretching force transmitted by the filament.

We therefore set out to elucidate the crystal structure of the twitchin fragment Fn31-TwcKin-Ig26, corresponding to A170-TK-M1 in titin.

4.4.2 Experimental work

The work presented as part of this study was performed by myself with the contribution of S. Barbieri in the stages of crystal optimization and data collection.

4.4.2.1 Cloning

The DNA coding for Twitchin FnIII-Kin-Ig inserted in pET-28 was provided by Dina Greene and Dr. Guy Benian (Emory University, School of Medicine, Atlanta). This construct was designed for the AFM experiments reported in (Greene et al., 2008) and thus contained a double Cys tag at the C-terminus for attachment on the gold surface that can be problematic in crystallization trials as they are susceptible to oxidation. Furthermore, the N-terminus contained a long extension between the protease cleavage site for tag removal and the N-terminus of the FnIII domain. Thus a new construct was designed for insertion into pETM-11, minimizing the number of terminal residues not involved in tertiary structure (construct spanning residues 4498-5575 of the Uniprot entry Q7JN85). Because of an internal recognition sequence for NcoI, the primers were designed to use BsmBI and KpnI for subsequent restriction digestion and compatible overhangs for insertion into pETM-11 that had been digested with NcoI and KpnI. Primer sequences are given in appendix 6.1.

4.4.2.2 Protein production:

The recombinant plasmid was transformed into *E. coli* BL21(*DE3*) Rosetta2 (Novagen). Cultures were grown at 30°C up to an OD_{600} of 0.6 in Luria Bertani medium supplemented with 25 *μ*g/ml kanamycin and 34 *μ*g/ml chloramphenicol. Expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cultures were further grown at 20°C for approximately 18 additional hours. Cells were harvested by centrifugation at 2800g and 4°C. Bacterial pellet was resuspended in 50mM Tris pH 7.9, 500mM NaCl, 2mM βME, containing a protease inhibitor cocktail (Roche) and DNAse I and lysed by

French press treatment (2 passes at 1000psi). The homogenate was clarified by centrifugation at 15000g and 4°C. Affinity purification used a HistrapHP column (GE Healthcare) equilibrated in lysis buffer in conjunction with an imidazole gradient. Elution used 50mM Tris pH 7.9, 300mM NaCl, 2mM βME, 200mM imidazol. Tag removal was by incubation with TEV protease overnight at 4°C during dialysis against 50 mM Tris pH 7.9, 200 mM NaCl, 2 mM βME. Subtractive purification for protease (His6-tagged at C-terminus) and non-digested protein was carried out on a subsequent chelating step. This two-step purification was enough to purify the protein to homogeneity. Further chromatography steps did not improve sample homogeneity, as evidenced by a single peak on anion exchange chromatography and the absence of a peak of aggregated protein in size exclusion chromatography and only delayed screening (Figure 4.21). The yield of pure protein was approximately 60mg/L of culture.

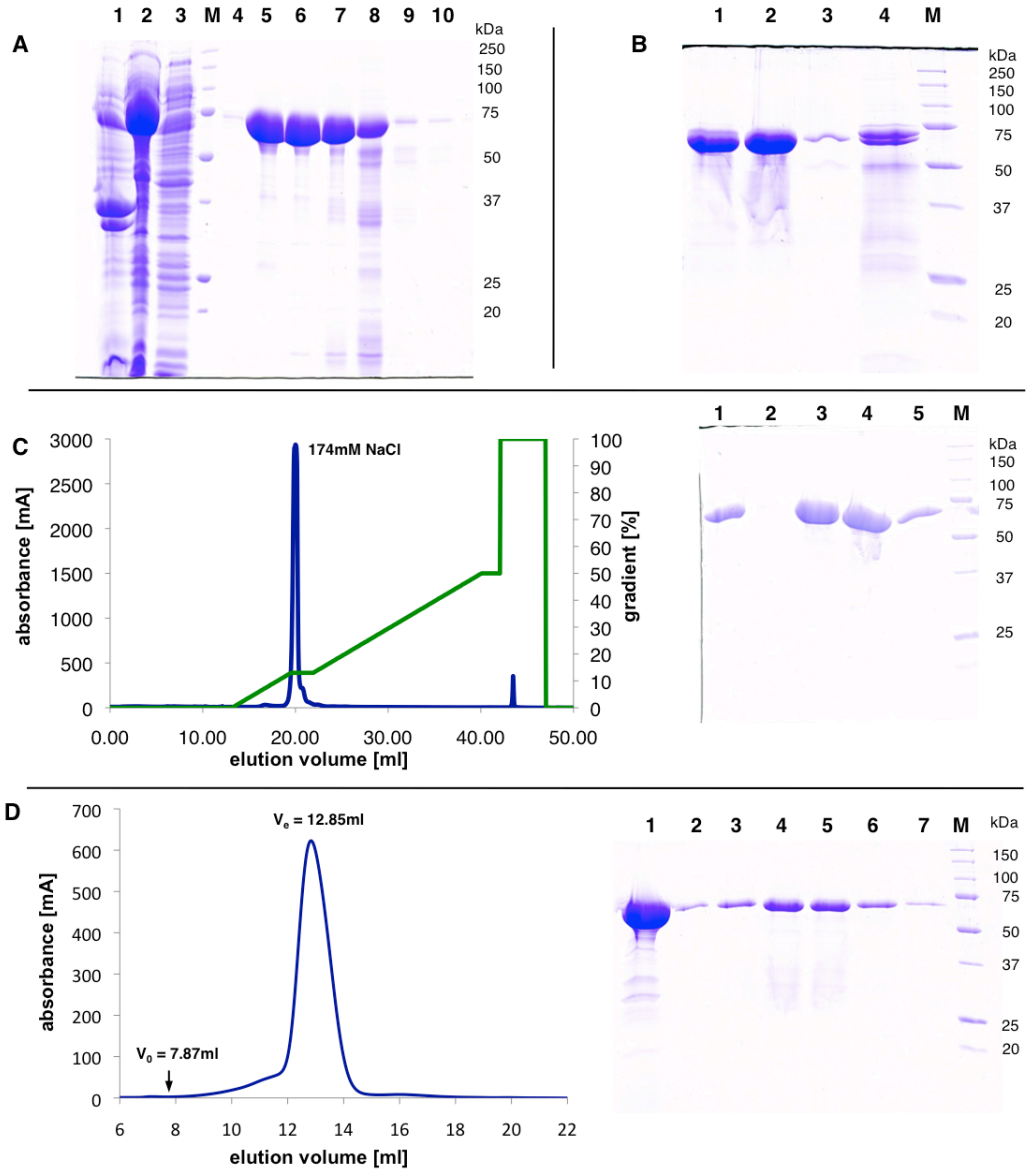


Figure 4.21: Purification of Twitchin FnIII-Kin-Ig

A) SDS-PAGE of affinity purification of Twc Kin. Lanes: 1) Pellet of lysate, 2) supernatant of lysate, 3) unbound sample, 4) wash 5-10) elutions, M) Precision Plus Protein standard (Biorad)

B) SDS-PAGE of reverse affinity purification of Twc Kin after tag removal by incubation with TEV protease. Lanes 1) incubation mixture, 2) unbound Twc Kin, 3) wash, 4) elution of uncleaved Twc Kin as well as TEV protease. M) Precision Plus Protein standard (Biorad)

C) Anion exchange chromatography performed on MonoQ (GE Healthcare). Sample was applied in 50mM Tris, pH 7.9, 50mM NaCl, elution used a salt gradient from 50-525mM NaCl. Signal shown corresponds to A_{280nm}. SDS-PAGE of samples collected from anion exchange. Lanes 1) sample applied to the column, 3-5) fractions of the single peak, M) Precision Plus Protein standard (Biorad)

D) Size exclusion chromatography performed on a Superdex 200 (Tricorn, GE Healthcare) equilibrated in 50mM Tris, pH7.9, 50mM NaCl. SDS-PAGE of samples collected from size exclusion chromatography: 1) Sample applied to the column, 2-7) consecutive fractions collected from ~9-16ml elution volume. M) Precision Plus Protein standard (Biorad).

Streaking observed in SDS-PAGE was due to heat sensitivity of the samples and could be prevented by loading samples without previous boiling.

4.4.2.3 Crystallization

For crystallization the buffer salt concentration was lowered to 50mM Tris pH7.9, 50mM NaCl, 2mM DTT. Sparse matrix screening of crystallization conditions used protein concentrated to 12mg/ml. Trials were set up in MRC sitting drop crystallization plates according to the parameters given in Table 4.4. All experiments were kept at 20°C, with the exception of the Cryo screen, which was set in duplicate and also stored at 4°C.

Table 4.4: Sparse matrix screening of crystallization conditions for Twc Kin

Screen	**Method**	**Drop size** (protein + reservoir)	**Reservoir**	**Temp.**
HR CS I + 2	manual	1μl + 1μl	80μl	20°C
Wiz I + II	manual	1μl + 1μl	80μl	20°C
PEGsII Suite	manual	1μl + 1μl	100μl	20°C
Cryo I + II	Honeybee93	200nl + 200nl	45μl	20°C + 4°C
PACT premier	Honeybee93	200nl + 200nl	45μl	20°C

CS – Crystal Screen I + II, Hampton Research; Wiz – Wizard I + II, Emerald Biosystems; Cryo I + II – Cryo Screen, Emerald Biosystems, PEGsII Suite (Qiagen prefilled easyxtal plate), PACT premier (Molecular Dimensions)

Crystals grew in the habit of small multiple needles within one week in 100mM Citrate pH 5.5, 20% PEG-3000 (Wizard condition I-6, see Figure 4.22B), similar in composition to the original growth conditions for the twitchin kinase crystals (Hu et al., 1994a; Hu et al., 1994b). Crystals in the shape of small plates or cubes grew within 1 month from 0.2M $MgCl_2$, 0.1M Citrate pH5.5, 40% PEG-400 (Cryo I-5, Figure 4.22D). Screening of citrate buffer pH, PEG size as well as concentration, in addition to varying protein concentrations led to crystals growing in rosettes of plates over 2-3 weeks, approximately 300x15x5μm^3 in size, when mixing 1μl of protein at 24mg/ml with 1μl of mother liquor composed of 100mM citrate pH 5.5, 20% PEG-600 (see Figure 4.22E).

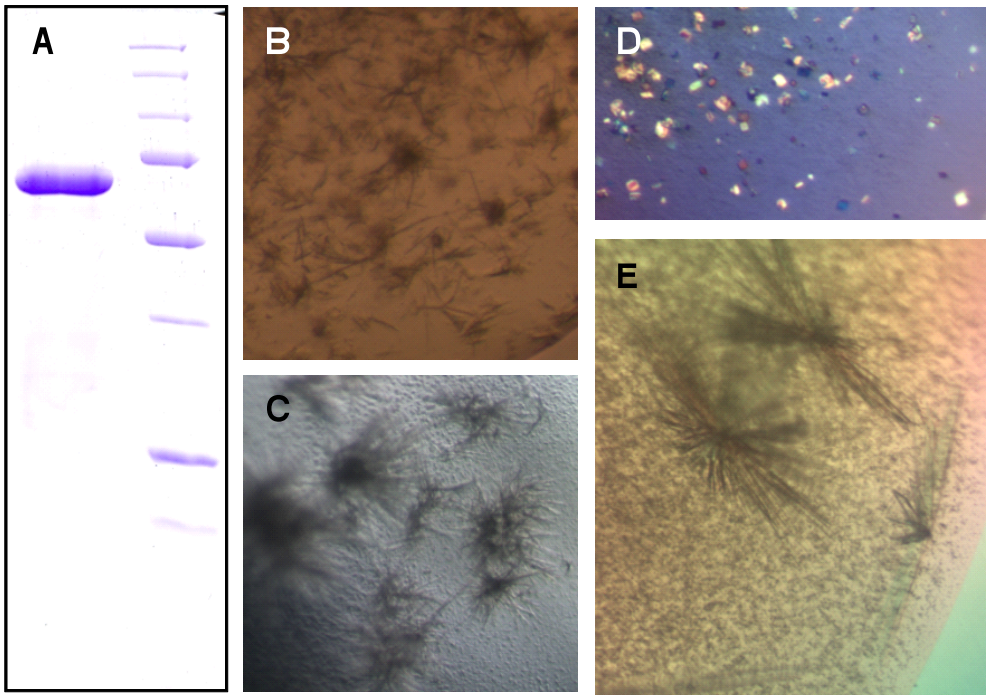
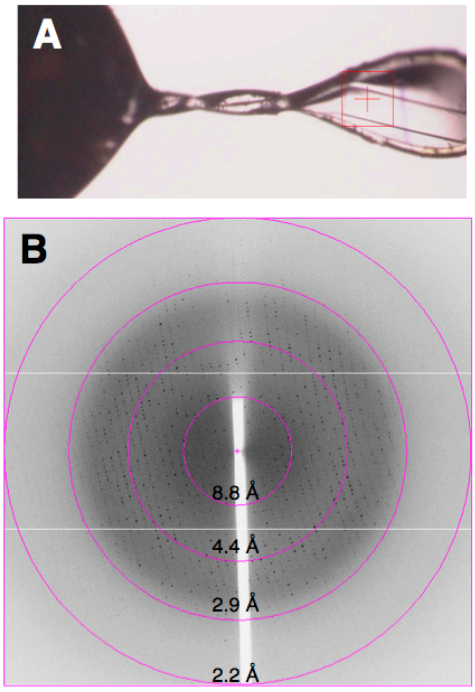


Figure 4.22: Crystallization trials of Twc FnIII-Kin-Ig
A) SDS-PAGE of Twc FnIII-Kin-Ig showing the final purity of the sample used for crystallization.
B) small needles grown in initial screening from condition Wiz I-6, 0.1M citrate pH5.5, 20% PEG-3000.
C) clusters of larger needles, grown from refined conditions, 0.1M citrate pH5.5, 15% PEG-3000
D) small crystals identified from robotic screening in condition Cryo I-5, 0.2M $MgCl_2$, 0.1M Citrate pH5.5, 40% PEG-400
E) whorls of plates grown from 0.1M citrate pH5.5, 20% PEG-600, growing to approximate dimensions of 300x15x5 μm, showing signs of multiple, intergrown crystals

Further refinement of these conditions yielded diffraction quality, single crystals of similar dimensions, but improved thickness. Crystals grew in the space of several days from 100mM citrate pH 5.5, 20% PEG600, 50mM $MgCl_2$. Single crystals (Figure 4.23A) were cryoprotected in mother liquor supplemented with 20% PEG-400 and shock frozen in liquid nitrogen.

4.4.2.4 Data collection and processing

Native data for Twc FnIII-Kin-Ig were collected at beamline I02 at Diamond (Didcot, UK), which is equipped with an ADSC Quantum Q315 detector. Data were collected under cryoconditions in a series of 1°, non-overlapping frames and processed (by S. Barbieri) to 2.4Å resolution using XDS (Kabsch, 1993). A representative diffraction pattern is given in Figure 4.23B and data statistics are given in Figure 4.23C.



C

	Twc FnIII-Kin-Ig	
Spacegroup	$P2_12_12$	
Unit cell dimensions	a = 121.70Å b = 158.21Å c = 60.46 Å α = β = γ = 90°	
Solvent content	44% with 2 mol/asu	
X-ray data		
X-ray source	DIAMOND I02	
Detector	ADSC Quantum Q315	
Wavelength (Å)	0.9796	
Resolution (Å)	18-2.4	(2.45-2.40)
Unique reflections	45313	(2670)
R_{sym} (I)	9.7	(58.0)
Multiplicity	4.86	(4.82)
Completeness (%)	97.2	(96.9)
I/σ (I)	14.43	(3.31)

Figure 4.23: X-ray diffraction pattern and data statistics for Twc FnIII-Kin-Ig
A) Crystal used for data collection. The red square indicates beam position. A representative diffraction pattern is shown in (B) while the table in (C) gives the statistics of the data set.

4.4.2.5 Structure elucidation

Phasing was by molecular replacement in Phaser (McCoy et al., 2007) using 1KOA (Kobe et al., 1996), covering the kinase and Ig, and domain A170 (2NZI; Mrosek et al., 2007) as a search model for the FnIII domain (43% sequence identity). The asymmetric unit contained two molecules, corresponding to 44% solvent content. The solution from Phaser was first subjected to rigid body refinement in Phenix (Adams et al., 2002), followed by ARPwARP (Perrakis et al., 2001) using the "model improvement through free atoms" option. This resulted in maps of sufficient quality for localization and building of the missing residues constituting the linker sequence between the FnIII and kinase domains. Cycles of model building in COOT (Emsley and Cowtan, 2004) were followed by further refinement in Phenix using restrained refinement in combination with TLS refinement of atomic displacement parameters (treating each domain as a separate body). Inspection of electron density maps revealed the presence of ordered components of the crystallization buffer, which were modeled using the add monomer routine in COOT. Building of the solvent layer made use of the ordered solvent picking routine in Phenix and was manually edited in COOT. In total 490 solvent molecules as well as one nonaethyleneglycol, one citrate, two diethylene glycol and one glycerol molecule were added to and co-

refined with the model. At no point during building or refinement were NCS restraints applied to the molecules. Refinement statistics are given in Table 4.5.

Table 4.5: Refinement statistics of Twc FnIII-Kin-Ig

Resolution	19.8-2.4Å
Number of reflections in working / test set	45258 / 1165
Number of protein / non protein atoms	9097 / 551
R-factor / R-free (%)	16.3 / 24.0
Rmsd bond length (Å) / bond angle(°)	0.007 / 1.123

4.4.3 Results: the Structure of Twc FnIII-Kin-Ig

The crystal structure of FnIII-Kin-Ig of *C. elegans* twitchin was solved at 2.4Å resolution (Figure 4.24). This structure allows us for the first time to analyze the structural arrangement of the kinase within the filament.

The asymmetric unit contains two molecular copies that are basically identical (0.66Å rmsd between 559 Cα atoms; SPDBV). The structural arrangement of the kinase domain and the following Ig is as seen in the previous structure, 1KOA, indicating that this arrangement is indeed not induced by preferential crystallization of a specific conformation (0.74Å rmsd between 441 Cα atoms; SPDBV). The FnIII domain preceding the kinase shows high structural similarity with the corresponding domain A170 in titin (0.83Å rmsd over 94 Cα atoms; SPDBV). The linker connecting the kinase domain to the preceding FnIII (and thus integrating it within the filament) exhibits a well-defined conformation. The FnIII domain is positioned in such a way against the side of the kinase domain that it effectively traps the autoregulatory tail in its inhibitory position blocking the active site.

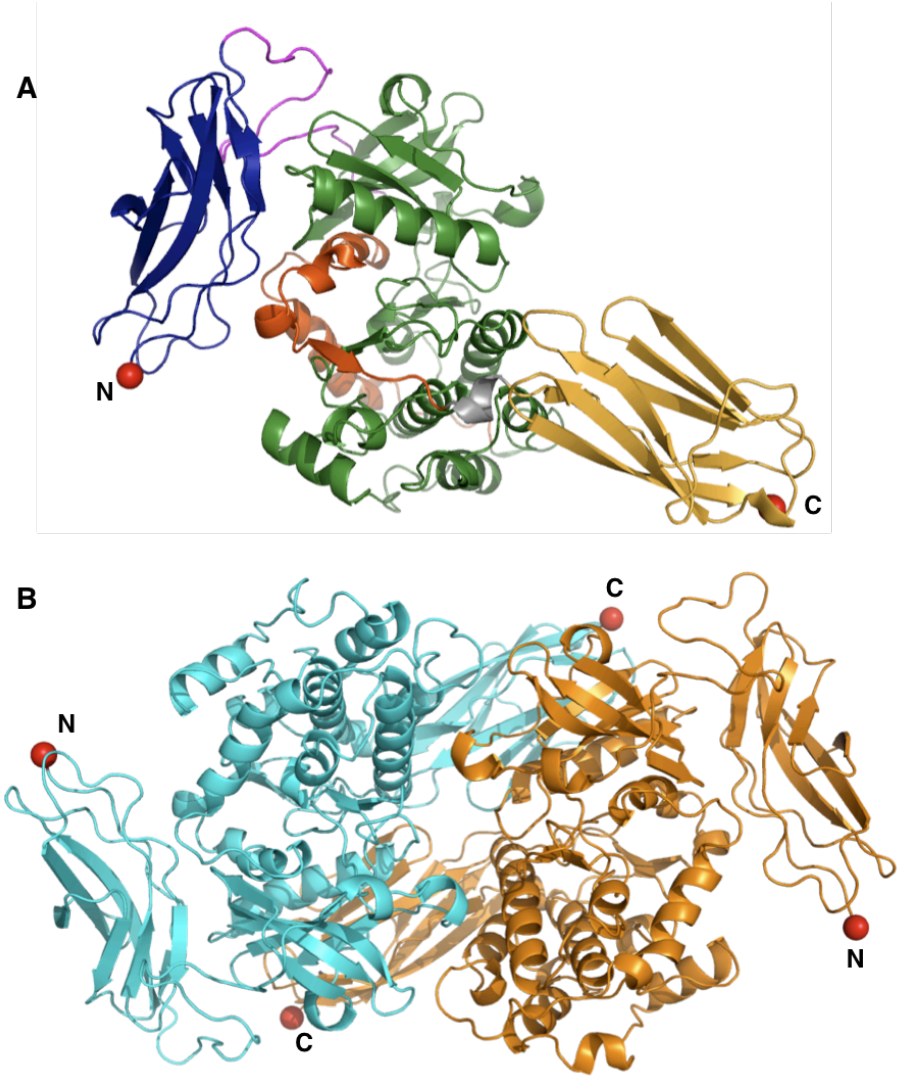


Figure 4.24: Crystal structure of twitchin FnIII-Kin-Ig
A) Ribbon diagram of the structure as seen from the side (see also Figure 4.18A). As can be seen, the regulatory tail wraps around the kinase domain, and is further trapped in place by the FnIII domain, which comes into close contact with it/ both the regulatory tail as well as the core of the kinase domain. The kinase domain is colored green, the autoregulatory sequence orange, Ig-26 in yellow (see also Figure 4.18A), the newly modeled FnIII in blue and the linker sequence connecting the FnIII to the kinase domain in magenta. B) Ribbon diagram of the two molecules in the asymmetric unit, seen from the top. For better visibility, each N- and C-terminus is marked by a red sphere.

Analysis of the domain interfaces made use of PISA (Krissinel and Henrick, 2007). The FnIII domain makes extensive contact with the kinase domain along its side, with an interaction surface of 644.4Å^2. Most of the interactions are non-specific, hydrophobic interactions (listed in Table 4.7) with the exception of a single cluster of hydrogen bonds between residues of the FnIII domain and the autoregulatory tail (Figure 4.26; Table 4.6). The side-chain of R456 is involved in hydrogen bonding with the main-chain atoms of E86 and P87, while the side-chain of E 86 in turn is involved in hydrogen bonding with main-chain atoms of Q458 and E459. All these residues

are conserved in TK (see Figure 4.17), but analysis shows that this part of the regulatory tail adopts a slightly different conformation, which might be due either to *i)* genuinely different orientations of the domains, *ii)* reflect an adaptation in geometry owing to the differences in linker length between the two proteins or *iii)* represent a possible conformation of the regulatory tail in the absence of its binding partner.

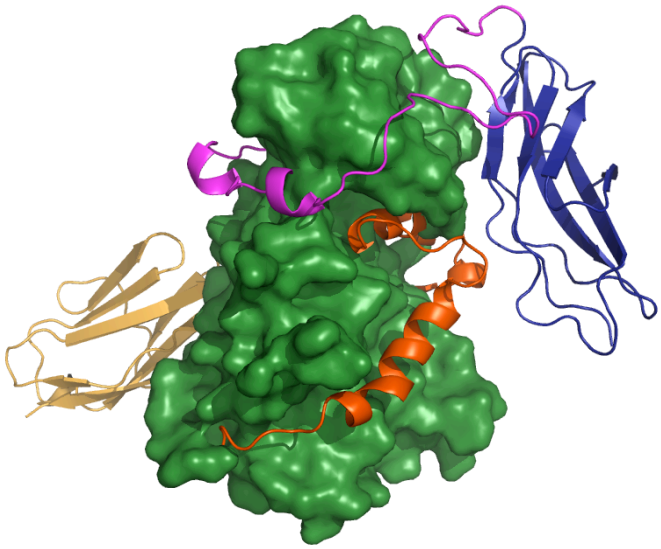


Figure 4.25: The FnIII domain traps the autoregulatory tail in its inhibitory position
The autoregulatory tail (orange) wraps around the kinase (green, displayed in surface representation) and is trapped in this conformation by the FnIII domain (blue). The linker (magenta) wraps around the N-terminal lobe of the kinase.

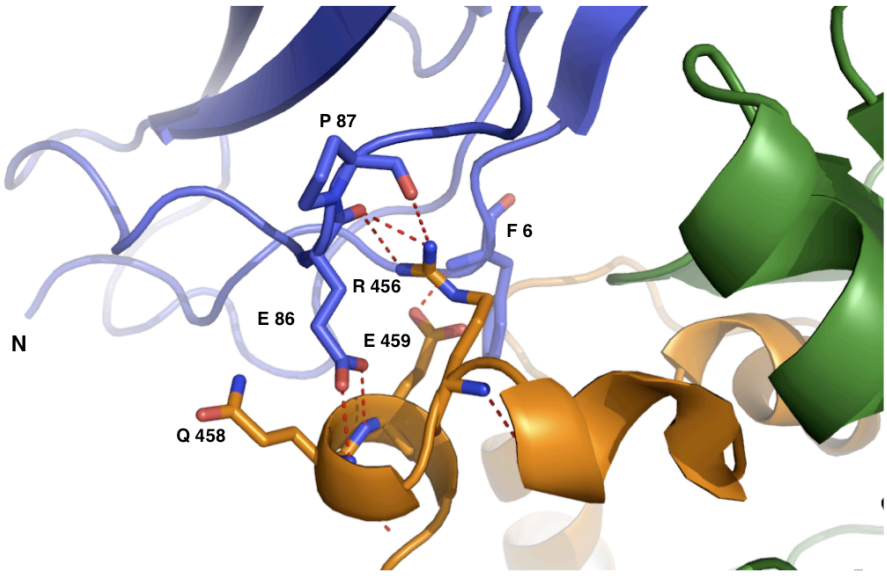


Figure 4.26: Specific interactions between the autoregulatory tail and the FnIII domain
Hydrogen bonding interactions are shown by red dotted lines. The coloring scheme follows the one in Figure 4.24 with the FnIII domain in blue, the kinase domain in green and the autoregulatory tail in orange. Residues involved in hydrogen bonding interactions are displayed and numbered.

Table 4.6: Specific interactions at domain interfaces

Interaction	STRUCTURE 1	Dist. [Å]	STRUCTURE 2
H-bonding	*Linker*		*FnIII*
1	ASP 96 [N]	3.04	PRO 94 [O]
2	GLU 97 [N]	3.07	PRO 94 [O]
3	ARG 98 [N]	3.73	GLY 95 [O]
4	ARG 98 [NH1]	3.08	ASP 14 [OD1]
5	GLY 110 [O]	3.57	ASN 11 [ND2]
6	GLN 109 [NE]	3.36	TYR 57 [OH]
H-bonding	*Kinase*		*FnIII*
1	GLU 459 [N]	2.92	GLU 86 [OE1]
2	GLN 458 [N]	2.99	GLU 86 [OE2]
3	ARG 456 [NH1]	2.89	GLU 86 [O]
4	ARG 456 [NH2]	3.20	GLU 86 [O]
5	ARG 456 [NH1]	3.35	PRO 87 [O]
6	GLU 459 [OE2]	2.75	PHE 6 [N]
H-bonding	*Kinase*		*Linker*
1	ASP 216 [OD1]	2.97	ARG 102 [NH1]
2	ASP 217 [OD2]	3.04	ARG 102 [NH2]
3	GLU 219 [OE2]	2.50	ARG 114 [NH2]
4	HIS 153 [O]	3.70	LYS 116 [N]
Salt bridges	*Linker*		*FnIII*
1	ARG 98 [NE]	3.80	ASP 14 [OD1]
2	ARG 98 [NH1]	3.08	ASP 14 [OD1]
3	LYS 111 [NZ]	3.97	GLU 10 [OE2]
Salt bridges	*Linker*		*Kinase*
1	ASP 216 [OD2]	3.22	ARG 102 [NH1]
2	ASP 217 [OD2]	3.47	ARG 102 [NH1]
3	ASP 216 [OD1]	2.97	ARG 102 [NH1]
4	ASP 217 [OD2]	3.04	ARG 102 [NH2]
5	GLU 219 [OE1]	3.53	ARG 114 [NE]
6	GLU 219 [OE1]	2.89	ARG 114 [NH2]
7	GLU 219 [OE2]	2.50	ARG 114 [NH2]

Only specific interdomain interactions that are observed here for the first time are included. Adapted from PISA (Krissinel and Henrick, 2007)

Table 4.7: Non-specific interface interactions between the FnIII and kinase domains

Residue	ASA [Å²]	BSA [Å²]	buried area	ΔG	Residue	ASA [Å2]	BSA [Å2]	buried area	ΔG
PRO 4	3.56	1.35	IIIII	-0.02	GLY 159	5.86	5.38	IIIIIIIIII	-0.06
ARG 5	91.40	37.43	IIIII	-0.26	ALA 160	48.76	42.16	IIIIIIIIII	0.08
PHE 6	130.19	100.97	IIIIIIIII	1.47	PHE 161	18.09	2.4	II	0.04
PRO 7	38.95	24.62	IIIIIII	0.28	GLY 162	8.47	4.3	IIIIII	0.03
ILE 8	83.95	72.25	IIIIIIIIII	1.16	MET 182	85.06	49.84	IIIIIII	1.25
ILE 9	42.97	31.59	IIIIIIII	0.11	PRO 184	103.07	76.74	IIIIIIII	0.86
GLU 10	90.05	14.26	III	-0.14	HIS 185	113.92	58.7	IIIIII	0.42
ASN 11	74.40	0		0	SER 187	40.5	0.12	I	0
ILE 12	74.84	48.35	IIIIIII	0.77	ASP 188	17.72	10.41	IIIIIII	-0.14
LYS 22	94.65	61.92	IIIIIIII	0.54	ASN 218	94.41	30.31	IIII	-0.07
PRO 23	60.27	3.85	II	0.06	LYS 432	94.07	9.45	II	-0.11
GLN 68	80.02	13.11	III	-0.15	TYR 433	25.9	0		0
PRO 84	46.56	2.45	II	-0.03	ASP 434	107.28	3.07	I	-0.04
CYS 85	10.52	3.68	IIIII	0.06	ALA 435	90.44	65.81	IIIIIIII	0.62
GLU 86	150.51	121.13	IIIIIIIII	-0.45	TRP 436	89.14	66.3	IIIIIIII	0.75
PRO 87	62.06	6.75	II	-0.08	PRO 437	90.13	1.34	I	0.02
THR 88	12.41	1.84	II	0.03	ASN 448	12.97	0.44	I	0
ALA 89	74.58	46.46	IIIIIII	0.72	HIS 455	76.1	0.76	I	-0.01
PRO 90	67.32	40.21	IIIIIII	0.14	ARG 456	131.95	106.47	IIIIIIIII	-1.77
VAL 91	11.55	10.88	IIIIIIIII	0.17	PRO 457	60.65	7.20	II	0.12
LEU 92	106.99	3.06	I	-0.03	GLN 458	171.55	29.35	III	-0.06
					GLU 459	106.42	60.06	IIIIIII	-0.26
					TYR 460	43.52	11.99	IIII	0.06

ASA – accessible surface area; BSA – buried surface area; ΔG – solvation energy effect, [kcal/mol]; buried area given as percentage, with every bar representing 10%. Adapted from PISA

Contrary to the previous crystal structure of twitchin kinase (1KOA), in which part of the linker sequence had been present in the crystal but disordered, and thus could not be modeled, in our structure this region adopts a mostly well-defined conformation. Here, the linker wraps almost entirely around the N-terminal lobe of the kinase and is involved in a range of interactions with the kinase as well as FnIII domain. Part of the linker deviates into a β-hairpin like structure that packs against one of the β-sheets of the FnIII domain and most likely helps to fix the orientation of the domain (Table 4.6; Figure 4.25 Figure 4.26). Since linker sequences and thus most likely linker orientation vary considerably between twitchin and TK, the interactions identified are most likely specific to the individual proteins.

The structure reveals also that T118, implicated to be autophosphorylated by Twc kinase (Lei et al., 1994), is exposed to the solvent and not directly accessible to the catalytic site. Thus it seems unlikely that this residue is autophosphorylated. Instead, our structure reveals Y104 to be a more likely candidate for regulatory phosphorylation. The side chain of Y104 is partially buried in a pocket of the N-terminal lobe of the kinase and its phosphorylation would entail a conformational change. Upon phosphorylation, the neighboring stretch of positively charged Arg and Lys residues could be expected to wrap around the negative charge of the phosphate, which in turn might lead to structural rearrangements that remove the

FnIII from its position blocking the exit of the autoregulatory tail. This would then enable the removal of the autoregulatory tail from its inhibitory position by an as yet unidentified mechanism.

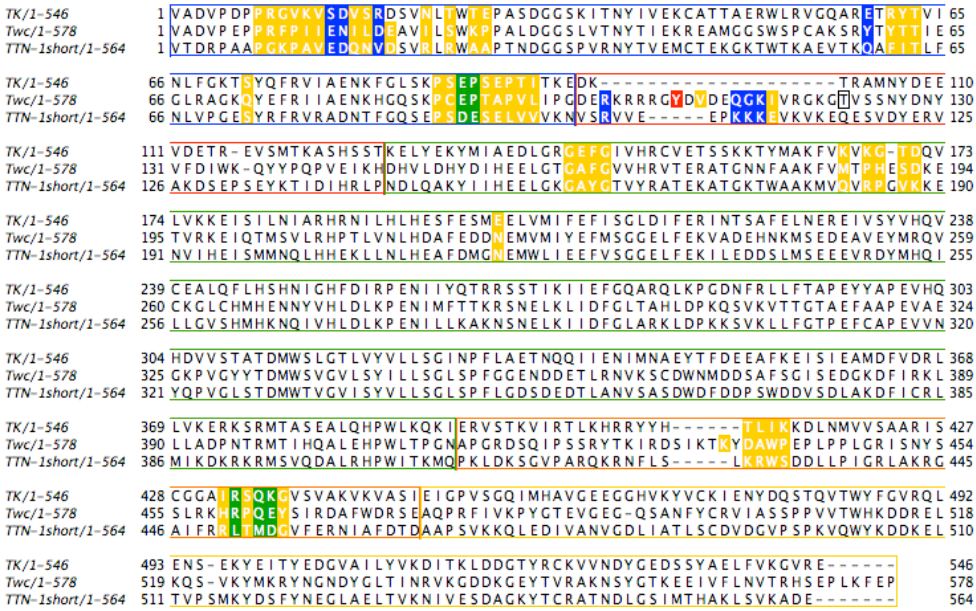


Figure 4.27: Conservation of residues involved in interface interactions
Non-specific interactions are indicated in yellow; specific interactions between the FnIII and autoregulatory tail are indicated in green, while specific interactions between the linker and the FnIII domain are highlighted in blue. The potential phosphorylation site Y104 is highlighted in red. The autophosphorylated T118 is boxed in black. The individual domain boundaries are indicated by colored boxes (as in Figure 4.17).

4.4.4 Discussion

The crystal structure of the twitchin FnIII-Kin-Ig fragment allows us for the first time a glimpse at how the kinase domain is embedded within the filament. Our model reveals that the N-terminal linker wraps around the kinase and positions the preceding FnIII domain in such a way against the kinase, that the autoregulatory tail is blocked in its inhibitory conformation. Furthermore, the geometry of how the kinase is integrated into the filament through the linker sequences, and thus the geometry of force transmission to the kinase indicates that stretch propagated along the filament would induce a shearing or twisting motion of the N-terminal with regard to the C-terminal lobe. This indicates mechanical stretch to be an unlikely activation mechanism. This could be predicted to cause non-native structural disruptions to the kinase fold, affecting its ATP and substrate binding sites. The single molecule AFM experiments exploring the unfolding behavior of this twitchin fragment (Greene et al., 2008) are not sensitive enough to allow the detection of any structural

rearrangements like the unfolding of the linker or regulatory tail or the reorientation of the kinase domains prior to the first domain unfolding peaks. Instead, repeating the SMD dynamics simulations of twitchin kinase unfolding with our structure might be more revealing about how trapping of the autoregulatory tail by the FnIII domain and the geometry of force transmission influence twitchin kinase unfolding.

From the structure it seems more plausible that twitchin kinase is activated through phosphorylation, possibly on T118 (autophosphorylated *in vitro*). Phosphorylation of this exposed residue might be the signal for the recruitment of other proteins to the linker region, where their binding to the linker would induce the conformational changes necessary to enable the removal of the autoregulatory tail from the active site. Alternatively Y104 appears to be a promising candidate for activation by phosphorylation by an upstream kinase. Currently, Y2H screening is ongoing in collaboration with Prof. Guy Benian using a bait construct corresponding to the fragment employed for structure elucidation. The aim is to identify potential binding partners that bind N-terminally of the kinase domain and could act as activators of twitchin kinase, either by phosphorylating twitchin or by inducing conformational changes upon binding that relieve autoinhibition.

While the main residues involved in specific interactions between the autoregulatory tail and the FnIII domain are conserved between twitchin and titin, the autoregulatory tail in titin adopts a slightly different conformation. Also, the linker sequence connecting A170 to the kinase is shorter in titin and thus will most likely adopt a different conformation, either not wrapping around the whole N-terminal domain or without deviating into the β-hairpin like structure. Thus, to be able to conclude about the geometry of the interface of A168-A170/MuRF1 with regard to the kinase domain, more structural data is needed. Furthermore, crystallization and structure elucidation of the corresponding fragment of *C. elegans* TTN-1 (clone available in the laboratory) might also show the structural variability induced by different linker lengths and compositions.

5 Conclusions

Titin is the longest molecule known to date and a central component of the sarcomere, where it fulfils essential roles in the assembly and maintenance of the lattice as well as its continued functioning and remodeling in response to trophic and mechanical demands. Its unique layout spanning half sarcomeres together with its modular domain architecture are a prerequisite by providing the means for regional adaptation of the filament in response to varying demands. Their unique mechanical properties as well as adaptability regarding engineering of protein interaction sites make the Ig and FnIII domains the ideal building blocks. As such they have evolved in the titin I-band to be “non-sticky” in order to be unhindered in their function as spring elements while the other poly-domain structures of titin that have been elucidated, Z1Z2 (Marino et al., 2006; Zou et al., 2006) and A168-A170 (Mrosek et al., 2007) have evolved to provide a highly specific protein interface, targeting a single interaction partner (or possible protein family in case of MuRFs). Analysis of these structures show that in both cases the binding surface stretches across several domains allowing for cooperativity as well as a means to control formation of a specific interface in this modular protein through regulation of the inter-domain orientation. This clearly highlights the need for structural information not only of the individual components but also in their full context within the chain to be able to reach valid conclusions. As a case in point, the crystal structure of I65-I70 allowed us for the first time to formulate a structure-based mechanistic model of titin elasticity. In this, conformationally stiff segments interspersed with pliant hinges form a regular pattern of dynamic super-motifs leading to segmental flexibility in the chain. This mechanistic model, based on freely jointed rigid segments, rationalizes current experimental data on the response to stretch of skeletal titin Ig-tandems. Efforts to solve the crystal structure of I101-I106 are still ongoing and should benefit from the experience gained in the lab in dealing with poly-domain constructs of titin and filamentous proteins.

Our efforts aimed at elucidating the crystal structure of the A168-A170/MuRF1 complex were hindered by the fact that full length MuRF1 could not be produced in soluble form from bacterial culture. In probing the interaction with shorter constructs complex formation could be established but the structural and biophysical characterization were set back by the severely compromised solubility of complexed

sample. In the construct MuRF-1$^{214-271}$ we hope to have overcome these issues and the project is in a promising phase with first diffraction quality crystals recently obtained. Thus, in the near future we expect to solve the structure of the coiled-coil fraction of MuRF1 and use this in docking experiments, which could make use of NMR titration experiments to map the interacting residues involved in complex formation (using the HADDOCK docking software; Dominguez et al., 2003).

The crystal structure of the Twc FnIII-Kin-Ig construct enables us for the first time to analyze the orientation of the kinase domain within the context of the filament. From the structure we can conclude that the orientation of the FnIII domain blocking the regulatory tail in its inhibitory conformation most likely adds an extra layer of inhibition and that the proposed stretch activation model for TK is rather unlikely to be the physiological mechanism. Instead, analysis of the structure has led us to propose Y104 in the linker sequence as a potential phosphorylation site that could be involved in twitchin kinase activation. Y2H screening is ongoing to identify binding partners that recognize the region N-terminal of the kinase domain and might be involved in its activation and possible phosphorylation of the identified tyrosine. Finally, analysis of sequence conservation together with structural comparison of twitchin and titin kinase showed that the main residues involved in hydrogen bonding interactions between the FnIII and autoregulatory tail are conserved but that the conformation of the autoregulatory tail within TK deviates from Twc kinase. Furthermore, due to length and sequence variations in the linker regions between the FnIII and kinase domain, the structure of Twc FnIII-Kin-Ig cannot be directly translated to the situation in titin. Thus, future structural studies on M-line titin and possibly other titin homologs will be necessary to shed light on the variability of these linkers and how they relate the A168-A170/MuRF1 complex with regard to the kinase in the assembly of the M-line signalosome.

6 Appendix

6.1 Primers used in PCR amplification of selected constructs

PCR amplification and cloning into pETM-11 for subfragments of titin I101-I106		
I101-I105	forward	5'- CAT GCC ATG GAA AAA CTC AGG ATC -3'
	reverse	5'- CGG GGT ACC TTA AAC ATC AAT AAC T -3'
I102-I106	forward	5'- CAT GCC ATG GAC CTT AGG ATT GTT G -3'
	reverse	5'- CGG GGT ACC TTA AGG GCC TGG AAC ATC -3'

MuRF1 PCR amplification and cloning into pETM-11		
MuRF1$^{1-327}$	forward	5'- GGA GAT ATA CCA TGG AGA ACT TGG AGA AGC-3'
	reverse	5'- CGG GGT ACC TTA ATC TGT CCC AAA GTC-3'
MuRF1$^{117-327}$	forward	5'- CAT GCC ATG GGC AGT CAC CCC ATG TG-3'
	reverse	5'- CGG GGT ACC TTA ATC TGT CCC AAA GTC-3'
MuRF1$^{117-271}$	forward	5'- CAT GCC ATG GGC AGT CAC CCC ATG TG-3'
	reverse	5'- AGT GGT ACC TTA CCC AGG CTC GTC CAG-3'
MuRF1$^{214-271}$	forward	5'- CAT GCC ATG GAC ACG TTG TAT GC-3'
	reverse	5'- AGT GGT ACC TTA CCC AGG CTC GTC CAG-3'

Primers for insertion of MuRF1 constructs into MCS2 of pET-DUET		
MuRF1$^{176-327}$	forward	5'- GGA ATT CCA TAT GCT GGT GGC-3'
	reverse	5'- GCA TGC TCG AGT TAA TCT GTC CC – 3'
MuRF1$^{117-327}$	forward	5'- GGA ATT CCA TAT GGG CAG TCA CC – 3'
	reverse	5'- GCA TGC TCG AGT TAA TCT GTC CC – 3'

Primers for Twc FnIII-Kin-Ig PCR amplification and cloning into pETM-11		
FnIII-Kin-Ig	forward	5'- CAT GCG TCT CAC ATG GTC GCC GAC GTA CCT GAA-3'
	reverse	5'- CGG GGT ACC TTA TGG CTC GAA TTT GAG TGG-3'

6.2 Calibration curves

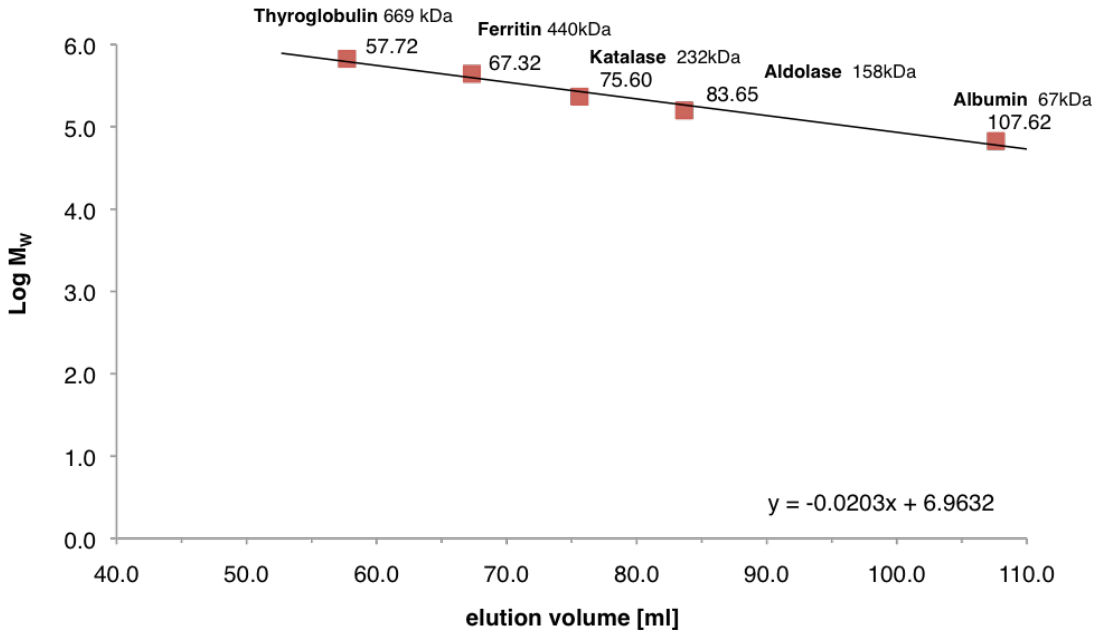


Figure 6.1: Calibration curve for Superdex 200 HiLoad 16l60

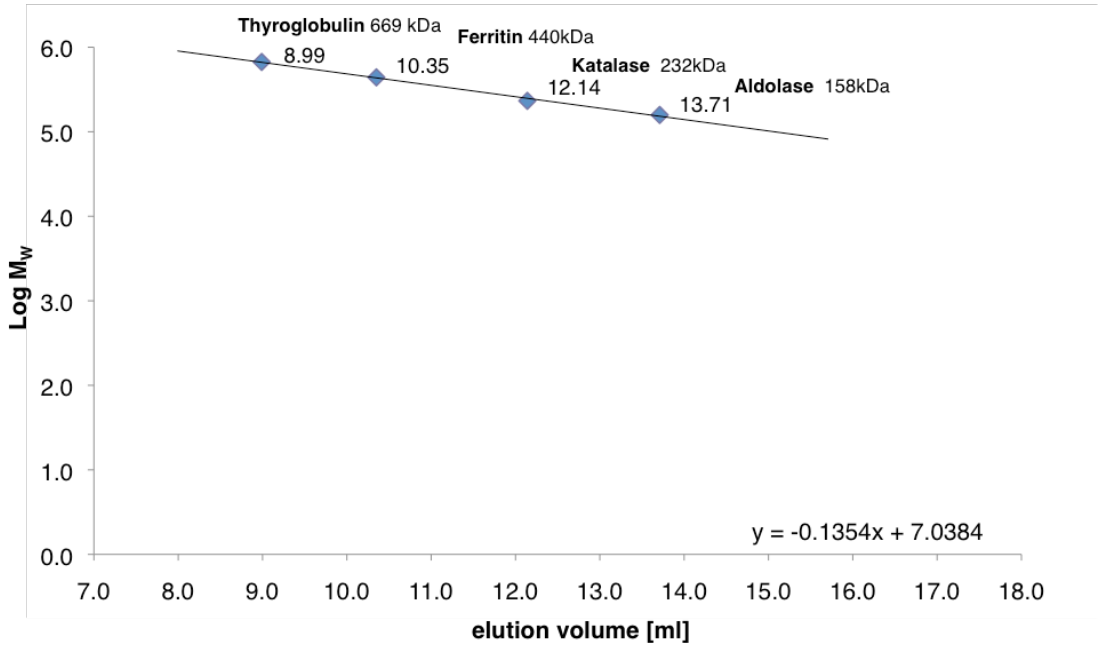


Figure 6.2: Calibration curve for Superdex 200 10l300GL Tricorn

7 References

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr *52*, 30-42.

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., Mccoy, A., Moriarty, N.W., Read, R., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr *58*, 1948-1954.

Adams, V., Linke, A., Gielen, S., Erbs, S., Hambrecht, R., and Schuler, G. (2008). Modulation of Murf-1 and MAFbx expression in the myocardium by physical exercise training. European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology *15*, 293-299.

Agarkova, I., and Perriard, J.C. (2005). The M-band: an elastic web that crosslinks thick filaments in the center of the sarcomere. Trends in Cell Biology *15*, 477-485.

Arber, S., Halder, G., and Caroni, P. (1994). Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differentiation. Cell *79*, 221-231.

Bähler, M., Wallimann, T., and Eppenberger, H.M. (1985). Myofibrillar M-band proteins represent constituents of native thick filaments, frayed filaments and bare zone assemblages. J Muscle Res Cell Motil *6*, 783-800.

Bang, M.L., Centner, T., Fornoff, F., Geach, A.J., Gotthardt, M., Mcnabb, M., Witt, C.C., Labeit, D., Gregorio, C.C., Granzier, H.*, et al.* (2001). The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. Circulation Research *89*, 1065-1072.

Benian, G.M., Kiff, J.E., Neckelmann, N., Moerman, D.G., and Waterston, R.H. (1989). Sequence of an unusually large protein implicated in regulation of myosin activity in C. elegans. Nature *342*, 45-50.

Boateng, S.Y., Belin, R.J., Geenen, D.L., Margulies, K.B., Martin, J.L., Hoshijima, M., de Tombe, P.P., and Russell, B. (2007). Cardiac dysfunction and heart failure are associated with abnormalities in the subcellular distribution and amounts of oligomeric muscle LIM protein. Am J Physiol Heart Circ Physiol *292*, H259-269.

Bodine, S.C., Latres, E., Baumhueter, S., Lai, V.K., Nunez, L., Clarke, B.A., Poueymirou, W.T., Panaro, F.J., Na, E., Dharmarajan, K.*, et al.* (2001). Identification of ubiquitin ligases required for skeletal muscle atrophy. Science *294*, 1704-1708.

Borden, K.L. (2000). RING domains: master builders of molecular scaffolds? Journal of Molecular Biology *295*, 1103-1112.

Brochard-Wyart, F., Tanaka, T., Borghi, N., and de Gennes, P.G. (2005). Semiflexible polymers confined in soft tubes. Langmuir : the ACS journal of surfaces and colloids *21*, 4144-4148.

Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472-475.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Bullard, B., Ferguson, C., Minajeva, A., Leake, M., Gautel, M., Labeit, D., Ding, L., Labeit, S., Horwitz, J., Leonard, K.R.*, et al.* (2004). Association of the chaperone alphaB-crystallin with titin in heart muscle. J Biol Chem *279*, 7917-7924.

Bullard, B., Linke, W.A., and Leonard, K. (2002). Varieties of elastic protein in invertebrate muscles. J Muscle Res Cell Motil *23*, 435-447.

Centner, T., Yano, J., Kimura, E., McElhinny, A.S., Pelin, K., Witt, C.C., Bang, M.L., Trombitas, K., Granzier, H., Gregorio, C.C*., et al.* (2001). Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain. Journal of Molecular Biology *306*, 717-726.

Cifra, P., Benková, Z., and Bleha, T. (2008). Effect of confinement on properties of stiff biological macromolecules. Faraday Discuss *139*, 377-392; discussion 399-417, 419-320.

Clark, K., Mcelhinny, A., Beckerle, M., and Gregorio, C. (2002). Striated muscle cytoarchitecture: an intricate web of form and function. Annu Rev Cell Dev Biol *18*, 637-706.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Craig, R., and Offer, G. (1976). The location of C-protein in rabbit skeletal muscle. Proc R Soc Lond, B, Biol Sci *192*, 451-461.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., and Barton, G.J. (1998). JPred: a consensus secondary structure prediction server. Bioinformatics *14*, 892-893.

Dabrowska, R., Aromatorio, D., Sherry, J.M., and Hartshorne, D.J. (1977). Composition of the myosin light chain kinase from chicken gizzard. Biochemical and Biophysical Research Communications *78*, 1263-1272.

Dai, K.S., and Liew, C.C. (2001). A novel human striated muscle RING zinc finger protein, SMRZ, interacts with SMT3b via its RING domain. J Biol Chem *276*, 23992-23999.

de La Fortelle, E., and Bricogne, G. (1997). SHARP: A Maximum-Likelihood Heavy-Atom Parameter Refinement Program for the MIR and MAD Methods. Methods Enzymol.

Di Cola, E., Waigh, T.A., Trinick, J., Tskhovrebova, L., Houmeida, A., Pyckhout-Hintzen, W., and Dewhurst, C. (2005). Persistence length of titin from rabbit skeletal muscles measured with scattering and microrheology techniques. Biophysical Journal *88*, 4095-4106.

Dominguez, C., Boelens, R., and Bonvin, A.M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc *125*, 1731-1737.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Ferrara, T., Flaherty, D., and Benian, G. (2005). Titin/connectin-related proteins in C. elegans: a review and new findings. J Muscle Res Cell Motil *26*, 435-447.

Fielitz, J., Kim, M., Shelton, J.M., Latif, S., Spencer, J.A., Glass, D., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2007). Myosin accumulation and striated muscle myopathy result from the loss of muscle RING finger 1 and 3. J Clin Invest *117*, 2486-2495.

Fisher, T.E., Marszalek, P.E., and Fernandez, J.M. (2000). Stretching single molecules into novel conformations using the atomic force microscope. NATURE STRUCTURAL BIOLOGY *7*, 719-724.

Flaherty, D., Gernert, K.M., Shmeleva, N., Tang, X., Mercer, K.B., Borodovsky, M., and Benian, G. (2002). Titins in C.elegans with unusual features: coiled-coil domains, novel regulation of kinase activity and two new possible elastic regions. Journal of Molecular Biology *323*, 533-549.

Flory, P.J. (1969). Statistical mechanics of chain molecules (New York,, Interscience Publishers).

Fong, S., Hamill, S.J., Proctor, M., Freund, S.M., Benian, G.M., Chothia, C., Bycroft, M., and Clarke, J. (1996). Structure and stability of an immunoglobulin superfamily domain from twitchin, a muscle protein of the nematode Caenorhabditis elegans. Journal of Molecular Biology *264*, 624-639.

Forbes, J., Jin, A., Ma, K., Gutierrez-Cruz, G., Tsai, W., and Wang, K. (2005). Titin PEVK segment: charge-driven elasticity of the open and flexible polyampholyte. J Muscle Res Cell Motil *26*, 291-301.

Fraternali, F., and Pastore, A. (1999). Modularity and homology: modeling of the type II module family from titin. Journal of Molecular Biology *290*, 581-593.

Freiburg, A., and Gautel, M. (1996). A molecular map of the interactions between titin and myosin-binding protein C. Implications for sarcomeric assembly in familial hypertrophic cardiomyopathy. Eur J Biochem *235*, 317-323.

Freiburg, A., Trombitas, K., Hell, W., Cazorla, O., Fougerousse, F., Centner, T., Kolmerer, B., Witt, C., Beckmann, J.S., Gregorio, C.C.*, et al.* (2000). Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. Circulation Research *86*, 1114-1121.

Fukuda, N., Wu, Y., Farman, G., Irving, T., and Granzier, H. (2005). Titin-based modulation of active tension and interfilament lattice spacing in skinned rat cardiac …. Pflügers Archiv European Journal of Physiology.

Fuller, S.J., Gillespie-Brown, J., and Sugden, P.H. (1998). Oncogenic src, raf, and ras stimulate a hypertrophic pattern of gene expression and increase cell size in neonatal rat ventricular myocytes. J Biol Chem *273*, 18146-18152.

Fürst, D.O., Osborn, M., Nave, R., and Weber, K. (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. The Journal of Cell Biology *106*, 1563-1572.

Gautel, M. (1996). The super-repeats of titin/connectin and their interactions: glimpses at sarcomeric assembly. Adv Biophys *33*, 27-37.

Gautel, M., and Goulding, D. (1996). A molecular map of titin/connectin elasticity reveals two different mechanisms acting in series. FEBS Lett *385*, 11-14.

Gotthardt, M., Hammer, R.E., Hübner, N., Monti, J., Witt, C., Mcnabb, M., Richardson, J.A., Granzier, H., Labeit, S., and Herz, J. (2003). Conditional expression of mutant M-line titins results in cardiomyopathy with altered sarcomere structure. J Biol Chem *278*, 6059-6065.

Granzier, H., Helmes, M., and Trombitás, K. (1996). Nonuniform elasticity of titin in cardiac myocytes: a study using immunoelectron microscopy and cellular mechanics. Biophysical Journal *70*, 430-442.

Granzier, H., Labeit, D., Wu, Y., and Labeit, S. (2002). Titin as a modular spring: emerging mechanisms for elasticity control by titin in cardiac physiology and pathophysiology. J Muscle Res Cell Motil *23*, 457-471.

Granzier, H., and Labeit, S. (2002). Cardiac titin: an adjustable multi-functional spring. The Journal of Physiology *541*, 335-342.

Granzier, H.L., and Irving, T.C. (1995). Passive tension in cardiac muscle: contribution of collagen, titin, microtubules, and intermediate filaments. Biophysical Journal *68*, 1027-1044.

Granzier, H.L., and Labeit, S. (2004). The giant protein titin: a major player in myocardial mechanics, signaling, and disease. Circulation Research *94*, 284-295.

Granzier, H.L., and Wang, K. (1993). Gel electrophoresis of giant proteins: solubilization and silver-staining of titin and nebulin from single muscle fiber segments. Electrophoresis *14*, 56-64.

Gräter, F., Shen, J., Jiang, H., Gautel, M., and Grubmüller, H. (2005). Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. Biophysical Journal *88*, 790-804.

Greaser, M. (2001). Identification of new repeating motifs in titin. Proteins *43*, 145-149.

Greene, D.N., Garcia, T., Sutton, R.B., Gernert, K.M., Benian, G., and Oberhauser, A.F. (2008). Single-molecule force spectroscopy reveals a stepwise unfolding of Caenorhabditis elegans giant protein kinase domains. Biophysical Journal *95*, 1360-1370.

Gregorio, C., Perry, C., and Mcelhinny, A. (2005). Functional properties of the titin/connectin-associated proteins, the muscle-specific RING finger proteins (MURFs), in striated muscle. J Muscle Res Cell Motil *26*, 389-400.

Gregorio, C.C., Granzier, H., Sorimachi, H., and Labeit, S. (1999). Muscle assembly: a titanic achievement? Curr Opin Cell Biol *11*, 18-25.

Gregorio, C.C., Trombitás, K., Centner, T., Kolmerer, B., Stier, G., Kunke, K., Suzuki, K., Obermayr, F., Herrmann, B., Granzier, H.*, et al.* (1998). The NH2 terminus of titin spans the Z-disc: its interaction with a novel 19-kD ligand (T-cap) is required for sarcomeric integrity. The Journal of Cell Biology *143*, 1013-1027.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Gutierrez-Cruz, G., Van Heerden, A.H., and Wang, K. (2001). Modular motif, structural folds and affinity profiles of the PEVK segment of human fetal skeletal muscle titin. J Biol Chem *276*, 7442-7449.

Harpaz, Y., and Chothia, C. (1994). Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. Journal of Molecular Biology *238*, 528-539.

Hayashi, C., Ono, Y., Doi, N., Kitamura, F., Tagami, M., Mineki, R., Arai, T., Taguchi, H., Yanagida, M., Hirner, S.*, et al.* (2008). Multiple molecular interactions implicate the connectin/titin N2A region as a modulating scaffold for p94/calpain 3 activity in skeletal muscle. J Biol Chem *283*, 14801-14814.

Heierhorst, J., Kobe, B., Feil, S.C., Parker, M.W., Benian, G.M., Weiss, K.R., and Kemp, B.E. (1996a). Ca2+/S100 regulation of giant protein kinases. Nature *380*, 636-639.

Heierhorst, J., Probst, W.C., Vilim, F.S., Buku, A., and Weiss, K.R. (1994). Autophosphorylation of molluscan twitchin and interaction of its kinase domain with calcium/calmodulin. J Biol Chem *269*, 21086-21093.

Heierhorst, J., Tang, X., Lei, J.Y., Probst, W.C., Weiss, K.R., Kemp, B.E., and Benian, G.M. (1996b). Substrate specificity and inhibitor sensitivity of Ca2+/S100-dependent twitchin kinases. Eur J Biochem *242*, 454-459.

Helmes, M., Trombitás, K., and Granzier, H. (1996). Titin develops restoring force in rat cardiac myocytes. Circulation Research *79*, 619-626.

Hirner, S., Krohne, C., Schuster, A., Hoffmann, S., Witt, S., Erber, R., Sticht, C., Gasch, A., Labeit, S., and Labeit, D. (2008). MuRF1-dependent regulation of systemic carbohydrate metabolism as revealed from transgenic mouse studies. Journal of Molecular Biology *379*, 666-677.

Holden, H.M., Ito, M., Hartshorne, D.J., and Rayment, I. (1992). X-ray structure determination of telokin, the C-terminal domain of myosin light chain kinase, at 2.8 A resolution. Journal of Molecular Biology *227*, 840-851.

Horowits, R., Kempner, E.S., Bisher, M.E., and Podolsky, R.J. (1986). A physiological role for titin and nebulin in skeletal muscle. Nature *323*, 160-164.

Houmeida, A., Holt, J., Tskhovrebova, L., and Trinick, J. (1995). Studies of the interaction between titin and myosin. The Journal of Cell Biology *131*, 1471-1481.

Hu, S.H., Lei, J.Y., Wilce, M.C., Valenzuela, M.R., Benian, G.M., Parker, M.W., and Kemp, B.E. (1994a). Crystallization and preliminary X-ray analysis of the auto-inhibited twitchin kinase. Journal of Molecular Biology *236*, 1259-1261.

Hu, S.H., Parker, M.W., Lei, J.Y., Wilce, M.C., Benian, G.M., and Kemp, B.E. (1994b). Insights into autoregulation from the crystal structure of twitchin kinase. Nature *369*, 581-584.

Improta, S., Krueger, J.K., Gautel, M., Atkinson, R.A., Lefèvre, J.F., Moulton, S., Trewhella, J., and Pastore, A. (1998). The assembly of immunoglobulin-like modules in titin: implications for muscle elasticity. Journal of Molecular Biology *284*, 761-777.

Improta, S., Politou, A.S., and Pastore, A. (1996). Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure *4*, 323-337.

Javanbakht, H., Diaz-Griffero, F., Stremlau, M., Si, Z., and Sodroski, J. (2005). The contribution of RING and B-box 2 domains to retroviral restriction mediated by monkey TRIM5alpha. J Biol Chem *280*, 26933-26940.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr, A, Found Crystallogr *47 (Pt 2)*, 110-119.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallogr *26*, 795-800.

Kedar, V., McDonough, H., Arya, R., Li, H.H., Rockman, H.A., and Patterson, C. (2004). Muscle-specific RING finger 1 is a bona fide ubiquitin ligase that degrades cardiac troponin I. Proc Natl Acad Sci USA *101*, 18135-18140.

Kellermayer, M.S., Bustamante, C., and Granzier, H.L. (2003). Mechanics and structure of titin oligomers explored with atomic force microscopy. Biochim Biophys Acta *1604*, 105-114.

Kellermayer, M.S., Smith, S.B., Bustamante, C., and Granzier, H.L. (1998). Complete unfolding of the titin molecule under external force. J Struct Biol *122*, 197-205.

Kellermayer, M.S., Smith, S.B., Bustamante, C., and Granzier, H.L. (2001). Mechanical fatigue in repetitively stretched single molecules of titin. Biophysical Journal *80*, 852-863.

Kellermayer, M.S., Smith, S.B., Granzier, H.L., and Bustamante, C. (1997). Folding-unfolding transitions in single titin molecules characterized with laser tweezers. Science *276*, 1112-1116.

Kenny, P.A., Liston, E.M., and Higgins, D.G. (1999). Molecular evolution of immunoglobulin and fibronectin domains in titin and related muscle proteins. Gene *232*, 11-23.

Knupp, C., Luther, P.K., and Squire, J.M. (2002). Titin organization and the 3D architecture of the vertebrate-striated muscle I-band. Journal of Molecular Biology *322*, 731-739.

Kobe, B., Heierhorst, J., Feil, S.C., Parker, M.W., Benian, G.M., Weiss, K.R., and Kemp, B.E. (1996). Giant protein kinases: domain interactions and structural basis of autoregulation. EMBO J *15*, 6810-6821.

Kolmerer, B., Olivieri, N., Witt, C.C., Herrmann, B.G., and Labeit, S. (1996). Genomic organization of M line titin and its tissue-specific expression in two distinct isoforms. Journal of Molecular Biology *256*, 556-563.

Koyama, S., Hata, S., Witt, C., Ono, Y., Lerche, S., Ojima, K., Chiba, T., Doi, N., Kitamura, F., Tanaka, K.*, et al.* (2008). Muscle RING-finger protein-1 (MuRF1) as a connector of muscle energy metabolism and protein synthesis. Journal of Molecular Biology *376*, 1224-1236.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology *372*, 774-797.

Krüger, M., Kötter, S., Grützner, A., Lang, P., Andresen, C., Redfield, M.M., Butt, E., dos Remedios, C.G., and Linke, W.A. (2009). Protein kinase G modulates human myocardial passive stiffness by phosphorylation of the titin springs. Circulation Research *104*, 87-94.

Krüger, M., and Linke, W.A. (2006). Protein kinase-A phosphorylates titin in human heart muscle and reduces myofibrillar passive tension. J Muscle Res Cell Motil *27*, 435-444.

Krüger, M., and Linke, W.A. (2009). Titin-based mechanical signaling in normal and failing myocardium. Journal of Molecular and Cellular Cardiology *46*, 490-498.

Labeit, D., Watanabe, K., Witt, C., Fujita, H., Wu, Y., Lahmers, S., Funck, T., Labeit, S., and Granzier, H. (2003). Calcium-dependent molecular spring elements in the giant protein titin. Proc Natl Acad Sci USA *100*, 13716-13721.

Labeit, S., Barlow, D.P., Gautel, M., Gibson, T., Holt, J., Hsieh, C.L., Francke, U., Leonard, K., Wardale, J., and Whiting, A. (1990). A regular pattern of two types of 100-residue motif in the sequence of titin. Nature *345*, 273-276.

Labeit, S., Gautel, M., Lakey, A., and Trinick, J. (1992). Towards a molecular understanding of titin. EMBO J *11*, 1711-1716.

Labeit, S., and Kolmerer, B. (1995). Titins: giant proteins in charge of muscle ultrastructure and elasticity. Science *270*, 293-296.

Lange, S., Auerbach, D., McLoughlin, P., Perriard, E., Schäfer, B.W., Perriard, J.C., and Ehler, E. (2002). Subcellular targeting of metabolic enzymes to titin in heart muscle may be mediated by DRAL/FHL-2. J Cell Sci *115*, 4925-4936.

Lange, S., Ehler, E., and Gautel, M. (2006). From A to Z and back? Multicompartment proteins in the sarcomere. Trends in Cell Biology *16*, 11-18.

Lange, S., Xiang, F., Yakovenko, A., Vihola, A., Hackman, P., Rostkova, E., Kristensen, J., Brandmeier, B., Franzen, G., Hedberg, B.*, et al.* (2005). The kinase domain of titin controls muscle gene expression and protein turnover. Science *308*, 1599-1603.

Leake, M., Wilson, D., Gautel, M., and Simmons, R.M. (2004). The elasticity of single titin molecules using a two-bead optical tweezers assay. Biophysical Journal *87*, 1112-1135.

Lee, E., Gao, M., Pinotsis, N., Wilmanns, M., and Schulten, K. (2006). Mechanical strength of the titin Z1Z2-telethonin complex. Structure *14*, 497-509.

Lee, E., Hsin, J., Mayans, O., and Schulten, K. (2007). Secondary and tertiary structure elasticity of titin Z1Z2 and a titin chain model. Biophysical Journal *93*, 1719-1735.

Lei, J.Y., Tang, X., Chambers, T.C., Pohl, J., and Benian, G.M. (1994). Protein kinase domain of twitchin has protein kinase activity and an autoinhibitory region. J Biol Chem *269*, 21078-21085.

Linke, W.A., and Grützner, A. (2008). Pulling single molecules of titin by AFM--recent advances and physiological implications. Pflugers Arch *456*, 101-115.

Linke, W.A., Ivemeyer, M., Olivieri, N., Kolmerer, B., Rüegg, J.C., and Labeit, S. (1996). Towards a molecular understanding of the elasticity of titin. Journal of Molecular Biology *261*, 62-71.

Linke, W.A., Rudy, D.E., Centner, T., Gautel, M., Witt, C., Labeit, S., and Gregorio, C.C. (1999). I-band titin in cardiac muscle is a three-element molecular spring and is critical for maintaining thin filament structure. The Journal of Cell Biology *146*, 631-644.

Linke, W.A., Stockmeier, M.R., Ivemeyer, M., Hosser, H., and Mundel, P. (1998). Characterizing titin's I-band Ig domain region as an entropic spring. Journal of Cell Science *111 (Pt 11)*, 1567-1574.

Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M., Hatakeyama, S., and Weissman, A.M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc Natl Acad Sci USA *96*, 11364-11369.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162-1164.

Ma, K., Forbes, J., Gutierrez-Cruz, G., and Wang, K. (2006). Titin as a giant scaffold for integrating stress and Src homology domain 3-mediated signaling pathways: the clustering of novel overlap ligand motifs in the elastic PEVK segment. J Biol Chem *281*, 27539-27556.

Ma, K., Kan, L., and Wang, K. (2001). Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. Biochemistry *40*, 3427-3438.

Ma, K., and Wang, K. (2002). Interaction of nebulin SH3 domain with titin PEVK and myopalladin: implications for the signaling and assembly role of titin and nebulin. FEBS Lett *532*, 273-278.

Ma, K., and Wang, K. (2003). Malleable conformation of the elastic PEVK segment of titin: non-co-operative interconversion of polyproline II helix, β-turn and unordered structures. Biochem J *374*, 687.

Marino, M., Svergun, D., Kreplak, L., Konarev, P., Maco, B., Labeit, D., and Mayans, O. (2005). Poly-Ig tandems from I-band titin share extended domain arrangements irrespective of the distinct features of their modular constituents. J Muscle Res Cell Motil *26*, 355-365.

Marino, M., Zou, P., Svergun, D., Garcia, P., Edlich, C., Simon, B., Wilmanns, M., Muhle-Goll, C., and Mayans, O. (2006). The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. Structure *14*, 1437-1447.

Marko, J.F., and Siggia, E.D. (1995). Statistical mechanics of supercoiled DNA. Physical review E, Statistical physics, plasmas, fluids, and related interdisciplinary topics *52*, 2912-2938.

Maruyama, K. (1994). Connectin, an elastic protein of striated muscle. Biophysical Chemistry *50*, 73-85.

Maruyama, K., Kimura, S., Yoshidomi, H., Sawada, H., and Kikuchi, M. (1984). Molecular size and shape of beta-connectin, an elastic protein of striated muscle. J Biochem *95*, 1423-1433.

Mayans, O., van der Ven, P.F., Wilm, M., Mues, A., Young, P., Fürst, D.O., Wilmanns, M., and Gautel, M. (1998). Structural basis for activation of the titin kinase domain during myofibrillogenesis. Nature *395*, 863-869.

Mayans, O., Wuerges, J., Canela, S., Gautel, M., and Wilmanns, M. (2001). Structural evidence for a possible role of reversible disulphide bridge formation in the elasticity of the muscle protein titin. Structure *9*, 331-340.

Mccoy, A., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr D Biol Crystallogr *61*, 458-464.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Mcelhinny, A., Kakinuma, K., Sorimachi, H., Labeit, S., and Gregorio, C. (2002). Muscle-specific RING finger-1 interacts with titin to regulate sarcomeric M-line and thick filament structure and may have nuclear functions via its interaction with glucocorticoid modulatory element binding protein-1. The Journal of Cell Biology *157*, 125-136.

Mcelhinny, A., Perry, C., Witt, C., Labeit, S., and Gregorio, C. (2004). Muscle-specific RING finger-2 (MURF-2) is important for microtubule, intermediate filament and sarcomeric M-line maintenance in striated muscle development. Journal of Cell Science *117*, 3175-3188.

Means, A.R., Bagchi, I.C., VanBerkum, M.F., and Kemp, B.E. (1991). Regulation of smooth muscle myosin light chain kinase by calmodulin. Adv Exp Med Biol *304*, 11-24.

Melchior, F. (2000). SUMO--nonclassical ubiquitin. Annu Rev Cell Dev Biol *16*, 591-626.

Meroni, G., and Diez-Roux, G. (2005). TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. Bioessays *27*, 1147-1157.

Miller, M., Bang, M., Witt, C., Labeit, D., Trombitas, C., Watanabe, K., Granzier, H., Mcelhinny, A., Gregorio, C., and Labeit, S. (2003). The Muscle Ankyrin Repeat Proteins: CARP, ankrd2/Arpp and DARP as a Family of Titin Filament-based Stress Response Molecules. Journal of Molecular Biology *333*, 951-964.

Miller, M.K., Granzier, H., Ehler, E., and Gregorio, C. (2004). The sensitive giant: the role of titin-based stretch sensing complexes in the heart. Trends in Cell Biology *14*, 119-126.

Moerman, D.G., Benian, G.M., Barstead, R., Schriefer, L., and Waterston, R.H. (1988). Identification and intracellular localization of the unc-22 gene product of Caenorhabditis elegans. Genes & Development *2*, 93-105.

Mrosek, M., Labeit, D., Witt, S., Heerklotz, H., von Castelmur, E., Labeit, S., and Mayans, O. (2007). Molecular determinants for the recruitment of the ubiquitin-ligase MuRF-1 onto M-line titin. The FASEB Journal *21*, 1383-1392.

Mrosek, M., Meier, S., Ucurum-Fotiadis, Z., von Castelmur, E., Hedbom, E., Lustig, A., Grzesiek, S., Labeit, D., Labeit, S., and Mayans, O. (2008). Structural analysis of B-Box 2 from MuRF1: identification of a novel self-association pattern in a RING-like fold. Biochemistry *47*, 10722-10730.

Muhle-Goll, C., Habeck, M., Cazorla, O., Nilges, M., Labeit, S., and Granzier, H. (2001). Structural and functional studies of titin's fn3 modules reveal conserved surface patterns and binding to myosin S1--a possible role in the Frank-Starling mechanism of the heart. Journal of Molecular Biology *313*, 431-447.

Müller, S., Lange, S., Gautel, M., and Wilmanns, M. (2007). Rigid conformation of an immunoglobulin domain tandem repeat in the A-band of the elastic muscle protein titin. Journal of Molecular Biology *371*, 469-480.

Nagy, A., Grama, L., Huber, T., Bianco, P., Trombitás, K., Granzier, H.L., and Kellermayer, M.S. (2005). Hierarchical extensibility in the PEVK domain of skeletal-muscle titin. Biophysical Journal *89*, 329-336.

Navaza, J. (2001). Implementation of molecular replacement in AMoRe. Acta Crystallogr D Biol Crystallogr *57*, 1367-1372.

Nave, R., Fürst, D.O., and Weber, K. (1989). Visualization of the polarity of isolated titin molecules: a single globular head on a long thin rod as the M band anchoring domain? The Journal of Cell Biology *109*, 2177-2187.

Odijk, T. (1983). The statistics and dynamics of confined or entangled stiff polymers. Macromolecules *16*, 1340-1344.

Oganesyan, N., Kim, S., and Kim, R. (2005). On-column protein refolding for crystallization. J Struct Funct Genomics *6*, 177-182.

Ono, Y., Kakinuma, K., Torii, F., Irie, A., Nakagawa, K., Labeit, S., Abe, K., Suzuki, K., and Sorimachi, H. (2004). Possible regulation of the conventional calpain system by skeletal muscle-specific calpain, p94/calpain 3. J Biol Chem *279*, 2761-2771.

Pawson, T., and Scott, J.D. (1997). Signaling Through Scaffold, Anchoring, and Adaptor Proteins. Science *278*, 2075-2080.

Peng, H., Begg, G.E., Harper, S.L., Friedman, J.R., Speicher, D.W., and Rauscher, F.J. (2000). Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. J Biol Chem *275*, 18000-18010.

Perrakis, A., Harkiolaki, M., Wilson, K.S., and Lamzin, V.S. (2001). ARP/wARP and molecular replacement. Acta Crystallogr D Biol Crystallogr *57*, 1445-1450.

Pfuhl, M., and Pastore, A. (1995). Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the I set. Structure *3*, 391-401.

Pizon, V., Iakovenko, A., Van Der Ven, P.F., Kelly, R., Fatu, C., Fürst, D.O., Karsenti, E., and Gautel, M. (2002). Transient association of titin and myosin with microtubules in nascent myofibrils directed by the MURF2 RING-finger protein. Journal of Cell Science *115*, 4469-4482.

Politou, A.S., Gautel, M., Improta, S., Vangelista, L., and Pastore, A. (1996). The elastic I-band region of titin is assembled in a "modular" fashion by weakly interacting Ig-like domains. Journal of Molecular Biology *255*, 604-616.

Politou, A.S., Gautel, M., Pfuhl, M., Labeit, S., and Pastore, A. (1994). Immunoglobulin-type domains of titin: same fold, different stability? Biochemistry *33*, 4730-4737.

Probst, W.C., Cropper, E.C., Heierhorst, J., Hooper, S.L., Jaffe, H., Vilim, F.S., Beushausen, S., Kupfermann, I., and Weiss, K.R. (1994). cAMP-dependent phosphorylation of Aplysia twitchin may mediate modulation of muscle contractions by neuropeptide cotransmitters. Proc Natl Acad Sci USA *91*, 8487-8491.

Puchner, E.M., Alexandrovich, A., Kho, A.L., Hensen, U., Schäfer, L.V., Brandmeier, B., Gräter, F., Grubmüller, H., Gaub, H.E., and Gautel, M. (2008). Mechanoenzymatics of titin kinase. Proc Natl Acad Sci USA *105*, 13385-13390.

Ratna, D., and Karger-Kocsis, J. (2008). Recent advances in shape memory polymers and composites: a review. Journal of Materials Science.

Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., Riganelli, D., Zanaria, E., Messali, S., Cainarca, S.*, et al.* (2001). The tripartite motif family identifies cell compartments. EMBO J *20*, 2140-2151.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M., and Gaub, H.E. (1997). Reversible unfolding of individual titin immunoglobulin domains by AFM. Science *276*, 1109-1112.

Sadoshima, J., Jahn, L., Takahashi, T., Kulik, T.J., and Izumo, S. (1992). Molecular characterization of the stretch-induced adaptation of cultured cardiac cells. An in vitro model of load-induced cardiac hypertrophy. J Biol Chem *267*, 10551-10560.

Schroeter, J.P., Bretaudiere, J.P., Sass, R.L., and Goldstein, M.A. (1996). Three-dimensional structure of the Z band in a normal mammalian skeletal muscle. The Journal of Cell Biology *133*, 571-583.

Scott, K.A., Steward, A., Fowler, S.B., and Clarke, J. (2002). Titin; a multidomain protein that behaves as the sum of its parts. Journal of Molecular Biology *315*, 819-829.

Short, K.M., Hopwood, B., Yi, Z., and Cox, T.C. (2002). MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, alpha 4, to microtubules: implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders. BMC Cell Biol *3*, 1.

Siegman, M.J., Funabara, D., Kinoshita, S., Watabe, S., Hartshorne, D.J., and Butler, T.M. (1998). Phosphorylation of a twitchin-related protein controls catch and calcium sensitivity of force production in invertebrate smooth muscle. Proc Natl Acad Sci USA *95*, 5383-5388.

Sjöström, M., and Squire, J.M. (1977). Fine structure of the A-band in cryo-sections. The structure of the A-band of human skeletal muscle fibres from ultra-thin cryo-sections negatively stained. Journal of Molecular Biology *109*, 49-68.

Sorimachi, H., Kinbara, K., Kimura, S., Takahashi, M., Ishiura, S., Sasagawa, N., Sorimachi, N., Shimada, H., Tagawa, K., and Maruyama, K. (1995). Muscle-specific calpain, p94, responsible for limb girdle muscular dystrophy type 2A, associates with connectin through IS2, a p94-specific sequence. J Biol Chem *270*, 31158-31162.

Sorimachi, H., Ono, Y., and Suzuki, K. (2000). Skeletal muscle-specific calpain, p94, and connectin/titin: their physiological functions and relationship to limb-girdle muscular dystrophy type 2A. Adv Exp Med Biol *481*, 383-395; discussion 395-387.

Spencer, J.A., Eliazer, S., Ilaria, R.L., Richardson, J.A., and Olson, E.N. (2000). Regulation of microtubule dynamics and myogenic differentiation by MURF, a striated muscle RING-finger protein. The Journal of Cell Biology *150*, 771-784.

Storoni, L.C., Mccoy, A., and Read, R. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr *60*, 432-438.

Torok, M., and Etkin, L.D. (2001). Two B or not two B? Overview of the rapidly expanding B-box family of proteins. Differentiation *67*, 63-71.

Trinick, J., Knight, P., and Whiting, A. (1984). Purification and properties of native titin. Journal of Molecular Biology *180*, 331-356.

Trombitás, K., Greaser, M., Labeit, S., Jin, J.P., Kellermayer, M., Helmes, M., and Granzier, H. (1998). Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. The Journal of Cell Biology *140*, 853-859.

Trombitás, K., Jin, J.P., and Granzier, H. (1995). The mechanically active domain of titin in cardiac muscle. Circulation Research *77*, 856-861.

Trombitás, K., Redkar, A., Centner, T., Wu, Y., Labeit, S., and Granzier, H. (2000). Extensibility of isoforms of cardiac titin: variation in contour length of molecular subsegments provides a basis for cellular passive stiffness diversity. Biophysical Journal *79*, 3226-3234.

Trombitás, K., Wu, Y., Mcnabb, M., Greaser, M., Kellermayer, M.S., Labeit, S., and Granzier, H. (2003). Molecular basis of passive stress relaxation in human soleus fibers: assessment of the role of immunoglobulin-like domain unfolding. Biophysical Journal *85*, 3142-3153.

Tskhovrebova, L., and Trinick, J. (2001). Flexibility and extensibility in the titin molecule: analysis of electron microscope data. Journal of Molecular Biology *310*, 755-771.

Tskhovrebova, L., and Trinick, J. (2002). Role of titin in vertebrate striated muscle. Philos Trans R Soc Lond, B, Biol Sci *357*, 199-206.

Tskhovrebova, L., and Trinick, J. (2003). Titin: properties and family relationships. Nat Rev Mol Cell Biol *4*, 679-689.

Tskhovrebova, L., Trinick, J., Sleep, J.A., and Simmons, R.M. (1997). Elasticity and unfolding of single molecules of the giant muscle protein titin. Nature *387*, 308-312.

Urry, D.W. (1988). Entropic elastic processes in protein mechanisms. I. Elastic structure due to an inverse temperature transition and elasticity due to internal chain dynamics. J Protein Chem *7*, 1-34.

Vinkemeier, U., Obermann, W., Weber, K., and Fürst, D.O. (1993). The globular head domain of titin extends into the center of the sarcomeric M band. cDNA cloning, epitope mapping and immunoelectron microscopy of two titin-associated proteins. J Cell Sci *106 (Pt 1)*, 319-330.

von Castelmur, E., Marino, M., Svergun, D., Kreplak, L., Ucurum-Fotiadis, Z., Konarev, P.V., Urzhumtsev, A., Labeit, D., Labeit, S., and Mayans, O. (2008). A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. Proc Natl Acad Sci USA *105*, 1186-1191.

Vondriska, T.M., Pass, J.M., and Ping, P. (2004). Scaffold proteins and assembly of multiprotein signaling complexes. J Mol Cell Cardiol *37*, 391-397.

Vuillard, L., Baalbaki, B., Lehmann, M., Nørager, S., Legrand, P., and Roth, M. (1996). Protein crystallography with non-detergent sulfobetaines. Journal of crystal growth *168*, 150-154.

Vuillard, L., Rabilloud, T., and Goldberg, M.E. (1998). Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. Eur J Biochem *256*, 128-135.

Wagner, F., Lattanzi, G., and Frey, E. (2006). Conformations of confined biopolymers. Macromolecules Phys Rev E *75*, 050902.

Wang, K., Ramirez-Mitchell, R., and Palter, D. (1984). Titin is an extraordinarily long, flexible, and slender myofibrillar protein. Proc Natl Acad Sci USA *81*, 3685-3689.

Watanabe, K., Muhle-Goll, C., Kellermayer, M.S., Labeit, S., and Granzier, H. (2002a). Different molecular mechanics displayed by titin's constitutively and differentially expressed tandem Ig segments. J Struct Biol *137*, 248-258.

Watanabe, K., Nair, P., Labeit, D., Kellermayer, M.S., Greaser, M., Labeit, S., and Granzier, H. (2002b). Molecular mechanics of cardiac titin's PEVK and N2B spring elements. J Biol Chem *277*, 11549-11558.

Weinert, S., Bergmann, N., Luo, X., Erdmann, B., and Gotthardt, M. (2006). M line-deficient titin causes cardiac lethality through impaired maturation of the sarcomere. The Journal of Cell Biology *173*, 559-570.

Whiting, A., Wardale, J., and Trinick, J. (1989). Does titin regulate the length of muscle thick filaments? Journal of Molecular Biology *205*, 263-268.

Williams, A.F. (1987). A year in the life of the immunoglobulin superfamily. Immunology Today *8*, 298 - 303.

Williams, A.F., and Barclay, A.N. (1988). The immunoglobulin superfamily--domains for cell surface recognition. Annu Rev Immunol *6*, 381-405.

Witt, C.C., Olivieri, N., Centner, T., Kolmerer, B., Millevoi, S., Morell, J., Labeit, D., Labeit, S., Jockusch, H., and Pastore, A. (1998). A survey of the primary structure and the interspecies conservation of I-band titin's elastic elements in vertebrates. J Struct Biol *122*, 206-215.

Witt, S., Granzier, H., Witt, C., and Labeit, S. (2005). MURF-1 and MURF-2 target a specific subset of myofibrillar proteins redundantly: towards understanding MURF-dependent muscle ubiquitination. Journal of Molecular Biology *350*, 713-722.

Yamasaki, R., Berri, M., Wu, Y., Trombitás, K., Mcnabb, M., Kellermayer, M.S., Witt, C., Labeit, D., Labeit, S., Greaser, M.*, et al.* (2001). Titin-actin interaction in mouse myocardium: passive tension modulation and its regulation by calcium/S100A1. Biophysical Journal *81*, 2297-2313.

Yamasaki, R., Wu, Y., Mcnabb, M., Greaser, M., Labeit, S., and Granzier, H. (2002). Protein kinase A phosphorylates titin's cardiac-specific N2B domain and reduces passive tension in rat cardiac myocytes. Circulation Research *90*, 1181-1188.

Young, P., Ehler, E., and Gautel, M. (2001). Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly. The Journal of Cell Biology *154*, 123-136.

Zoghbi, M.E., Woodhead, J.L., Moss, R.L., and Craig, R. (2008). Three-dimensional structure of vertebrate cardiac muscle myosin filaments. Proc Natl Acad Sci USA *105*, 2386-2390.

Zou, P., Pinotsis, N., Lange, S., Song, Y., Popov, A., Mavridis, I., Mayans, O., Gautel, M., and Wilmanns, M. (2006). Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature *439*, 229-233.