Structural studies on the association of filamentous proteins in the human M-Bands

DISSERTATION ZUR ERLANGUNG DES

NATURWISSENSCHAFTLICHEN DOKTORGRADES DER

BAYERISCHEN JULIUS-MAXIMILIANS-UNIVERSITÄT

WÜRZBURG

VORGELEGT VON

FLORIAN SAUER

AUS WÜRZBURG

WÜRZBURG 2011

Eingereicht am: _____

Mitglieder der Promotionskommission:

Vorsitzender: Prof. Dr. Wolfgang Rössler

Erstgutachter: Prof. Dr. Hermann Schindelin

Zweitgutachter: Prof. Dr. Jörg Schultz

Tag des Promotionskolloquiums: _____

Doktorurkunde ausgehändigt am: _____

Man sollte auf alles achten, denn man kann alles deuten.

Hermann Hesse

Das Glasperlenspiel

Abstract

Cross-striated muscles enable higher animals to perform directed movements and to create mechanical force. The cells of heart and skeletal muscles consist of myofibrils, serial arrays of the smallest contractile subunits, the sarcomeres. Main components of the sarcomeres are the thin and thick filaments, large protein assemblies consisting of mainly actin (thin filaments) and myosin (thick filaments), whose energy-dependent interaction is responsible for the contraction of sarcomeres and so of the whole muscle.

The thin filaments are anchored in the sarcomere bordering Z-discs, while the thick filaments are anchored in the M-bands, traverse structures in the sarcomere center.

Electron-microscopic studies revealed that the M-bands consist of regular, lattice-like structures that appear to cross-link the thick filaments. A number of proteins could be identified by immune-fluorescence and biochemical binding studies to be present and interact with each other in the M-bands. These data have been integrated into preliminary models of the M-bands. Detailed knowledge of how these proteins interact with each other in the center of the sarcomeres is, however, largely missing.

The current study focuses on the structural characterization of the interactions between the titin, myomesin-1, obscurin and obscurin-like 1 (OBSL1), modular filamentous proteins interacting with each other in the M-bands.

The high-resolution **crystal structure of the titin M10 – OBSL1 Ig1 complex** was solved. The structure and additional biophysical data show that titin and OBSL1 as well as titin and obscurin form stable binary complexes through the formation of a small intermolecular ß-sheet. In contrast to previously characterized intermolecular assemblies of sarcomeric proteins, this sheet is formed between parallel non-homologous ß-strands of the interaction partners. The investigation of disease-related variants of the M10 domain by biophysical methods did not allow to draw unambiguous conclusions on a direct connection between impaired OBSL1/obscurin binding and disease development.

Two out of four known M10 variants have effects on the correct domain folding and so interfere with the ability to bind obscurin/OBSL1. The two other known variants displayed however only minor effects on fold and binding affinities. It should therefore be further elucidated whether a direct connection between impaired complex formation and disease development exists.

A direct interaction between **titin and myomesin-1** could not be confirmed *in vitro*. Possible explanations for the different results are discussed.

While the consequences of the inability of both proteins to interact are unclear, the further characterization of the putative interacting parts of titin and myomesin-1 led to the discovery of **two new potential sites of self-assembly on M-band titin and myomesin-1**.

The **crystal structure of titin M4** showed that this domain can form dimeric assemblies through the formation of a disulfide bridge and an intermolecular metal binding site between residues that are unique to this domain.

On myomesin-1, in addition to the described C-terminal interaction site, a **potential second site of self-assembly** was found in its central Fn3-domain segment. The interacting site was mapped to the predicted Fn3 domain My5. The crystal structure of the domain in its dimeric form showed that the interaction is mediated by a mechanism that has previously not been observed in sarcomeric proteins. Two My5 interact with each other by the mutual exchange of an N-terminal ß-strand which complements the Fn3 fold on the binding partner.

This type of interaction can be interpreted as misfolding. However, the position of the interacting domain and its mode of interaction allowed the postulation of a model of how myomesin-1 could be integrated in the M-bands. This model is in good agreement with the electron-microscopic appearance of the M-bands.

Zusammenfassung

Die quergestreifte Muskulatur befähigt höhere Tiere zur zielgerichteten Bewegung und Ausübung mechanischer Kraft. Herz- und Skelettmuskelzellen bestehen aus Myofibrillen, die wiederum aus aneinandergereihten, kleinen kontrahierenden Untereinheiten, den Sarkomeren aufgebaut sind. Hauptbestandteile der Sarkomere sind die dünnen und dicken Filamente, große Proteinkomplexe die hauptsächlich aus Aktin (dünne Filamente) und Myosin (dicke Filamente) bestehen und deren energieabhängige Interaktion für die Kontraktion der Sarkomere und damit des gesamten Muskels verantwortlich sind.

Die dünnen Filamente sind in den Sarkomer-begrenzenden Z-Scheiben und die dicken Filamente in der M-Bande im Zentrum der Sarkomere verankert.

Elektronenmikroskopische Studien zeigten, dass die M-Banden aus regelmäßigen, gerüstartigen Strukturen bestehen, die die dicken Filamente querzuvernetzen scheinen. Durch Immunfluoreszenz und Bindungstudien konnte eine Anzahl an Proteinen identifiziert werden, die neben Myosin am Aufbau dieses Gerüsts beteiligt sein könnten. Basierend auf diesen Daten wurden vorläufige Modelle des Aufbaus der M-Banden postuliert. Eine detaillierte Charakterisierung der Interaktionen dieser Proteine auf struktureller Ebene hat bisher jedoch nicht stattgefunden.

Die hier präsentierte Arbeit beschäftigt sich mit der strukturellen Charakterisierung der Interaktionen zwischen den Proteinen Titin, Myomesin-1, Obscurin und OBSL1 in den M-Banden von Wirbeltiersarkomeren.

Die hochaufgelöste **Kristallstruktur des Titin M10 – OBSL1 Komplexes** wurde gelöst. Die Struktur und zusätzliche biophysikalische Daten zeigen, dass der C-Terminus von Titin und die N-termini von OBSL1 bzw. Obscurin stabile, binäre Komplexe ausbilden. Im Gegensatz zu schon bekannten Komplexen zwischen Ig-ähnlichen Domänen sarkomerer Proteine, wird die Interaktion hier durch die Ausbildung eines intermolekularen ß-Faltblattes zwischen parallel orientierten ß-Strängen, vermittelt.

Die Untersuchung von Varianten der M10 Domäne, die mit der Entwicklung von erblichen Muskelkrankheiten in Zusammenhang gebracht werden, ließen keine eindeutigen Schlussfolgerungen darüber zu, ob ein direkter Zusammenhang zwischen der Beeinträchtigung der Bindung an Obscurin/OBSL1 und der Entwicklung der

Krankheiten besteht. Zwei der vier bekannten M10 Varianten haben Auswirkungen auf die korrekte Faltung der Domäne, weshalb sie Obscurin und OBSL1 nicht binden können. Die beiden anderen Varianten zeigten jedoch nur geringfügige Auswirkungen auf Faltung und Affinität zu Obscurin und OBSL1. Es sollte daher weiter untersucht werden, ob ein direkter Zusammenhang zwischen der Bindung an Obscurin oder OBSL1 und der Entstehung vor Muskelkrankheiten besteht.

Eine direkte Interaktion zwischen Titin und Myomesin-1 *in vitro* konnte nicht bestätigt werden. Verschiedene Erklärungen die zu den Unterschieden zwischen den hier gezeigten negativen und den an anderer Stelle beschrieben positiven Ergebnissen der Bindungsstudien geführt haben könnten, werden diskutiert. Die Konsequenzen der möglichen 'Unfähigkeit' Titins mit Myomesin-1 zu interagieren sind momentan unklar. Die weitere Charakterisierung der vermeintlichen Bindungspartner führte jedoch zur Entdeckung **zweier neuer Selbstbindungsstellen auf Titin und Myomesin-1**.

Die **Kristallstruktur der Ig-ähnlichen Domäne M4 von Titin** zeigte, dass diese durch einer intermolekularen Disulfidbrücke und einer Zinkkoordinierungsstelle, Dimere bilden kann.

Zusätzlich zu der beschrieben C-terminalen, wurde eine mögliche **zweite Selbstbindungsstelle auf Myomesin-1** im zentralen Fn3-Domänensegment des Proteins entdeckt. Der für die Bindung verantwortliche Bereich konnte auf die Fn3 Domäne My5 eingegrenzt werden. Die Kristallstruktur der Domäne in ihrer dimeren Form zeigte, dass die Interaktion durch einen zuvor bei Muskelproteinen nicht beschriebenen Mechanismus vermittelt wird. Zwei My5-Domänen interagieren durch den gegenseitigen Austausch eines N-terminalen ß-Stranges, der die Faltung des Bindunspartners komplementiert.

Diese Art von Proteininteraktion kann als Resultat der Fehlfaltung der Domäne interpretiert werden. Die Position der interagierenden Domäne und die Art der Interaktion erlaubten es jedoch, ein Modell aufzustellen, das erklären könnte, wie Myomesin-1 in die M-banden eingebaut ist. Dieses Modell stimmt mit dem elektronenmikroskopischen Erscheinungsbild der M-Banden gut überein.

Table of contents

List of Figures

List of tables

Abbreviations

aa	Amino acid(s)
ATP	Adenosine-triphosphate
ADP	Adenosine-diphosphate
AFM	Atomic force microscopy
BME	Beta-mercaptoethanol
CD	Circular dichroism
cDNA	Complementary (reversely transcribed) DNA
cv	Column volume
Cryo EM	Cryo electron microscopy
DESY	Deutsches Elektronensynchrotron
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ELC/RLC	Essential / regulatory light chain
EOFMC	Early onset myopathy with fatal cardiomyopathy
ESRF	European Synchrotron Radiation Facility
FHL2	Four and a half LIM domain containing protein 2
Fn3	Fibronectin type 3 module/domain
fw	Forward
GST	Glutathione S-transferase
HMERF	Hereditary myopathy with early respiratory failure
HPSF	High Purity Salt Free
Ig	Immunoglobulin
IMAC	Immobilized metal affinity chromatography
Is1-Is7	Interspersed segment 1-7
ITC	Isothermal titration calorimetry
LGMD	Limb-girdle muscular dystrophy
M1-M10	Titin Ig-like domains M1-M10
MAD	Multiple wavelength anomalous dispersion
MALDI-TOF	Matrix assisted laser desorption ionization – time of flight
MBP	Maltose-binding protein
MR	Molecular replacement

MURF	Muscle ring finger protein
NBR1	Neighbor of BRCA1 gene 1 protein
NMR	Nuclear magnetic resonance
NTA	Nitrilo triacetic acid
Obs	Obscurin
OBSL1	Obscurin like 1
OMIM	Online mendelian inheritance in men
PAA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Poly ethylene glycol
pI	Isoelectric point
P_i	Inorganic phosphate
rv	Reverse
SAD	Single wavelength anomalous dispersion
SEC	Size exclusion chromatography
SLS	Static light scattering
SR	Sarcoplasmic reticulum
TBE	Tris borate EDTA
Tk	Titin kinase domain
TLS	Translation, libration, screw
TMD	Tibial muscular dystrophy
Ve	Elution volume
Y2H	Yeast Two-Hybrid

1   Introduction

The muscular system consists of contractile organs that are responsible for the generation of nearly all active forces created in higher metazoans. The system is divided into three muscle types that differ in their morphology, mode of force generation and excitability.

Smooth muscles are covering the surface of hollow organs. By application of involuntary long lasting force they exhibit functions like the regulation of the blood pressure by vessel contraction or the forward motion of nutrients by contraction of the gastrointestinal tract.

Striated muscle tissue is the major constituent of the heart muscle, skeletal muscle, tongue and larynx. In contrast to smooth muscle it creates strong directed force over short periods of time. Striated muscles are divided into two types.

Skeletal muscles are under voluntary control of the central nervous system. They are connected to the skeleton through tendons and generate the force for movements of the body.

Heart muscle is responsible for the maintenance of blood circulation. Although also consisting of striated muscle tissue, it differs from skeletal muscle in morphology and protein composition and is a self-exciting tissue under indirect control of the autonomous nervous system.

1.1 Vertebrate striated muscle

1.1.1   General architecture

Skeletal musculature is with up to 50% of the human body weight the biggest organ in man. Its force generating tissue, the striated muscle tissue, is hierarchically organized. On top of the hierarchy are the **muscle fibers**, long multinuclear cells with a diameter of 20 to 100 μm [1] and a length of up to several centimeters that run along the whole muscle. These fibers consist of membranous systems, such as the calcium storing sarcoplasmic reticulum, organelles, such as ATP generating mitochondria and the **myofibrils**, small cylindrical structures with a diameter of up to 5 μm [1]. Myofibrils are serial arrays of small subunits: the **sarcomeres**. The well ordered arrangement of the sarcomeres within myofibrils and muscle fibers is the origin of the 'cross striated' appearance as different sarcomeric regions appear with differing brightness under a

polarization or electron microscope.
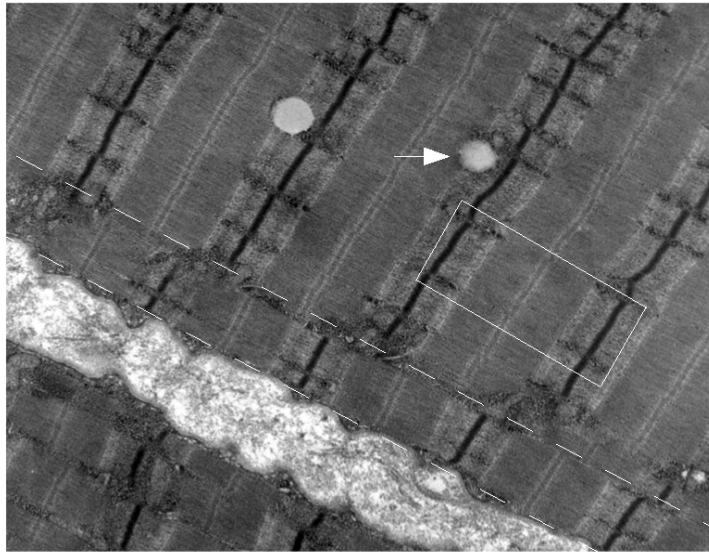
Figure 1-1 Electron micrograph of a muscle fiber. Longitudinal section.

Muscle fibers are composed of bundles of parallel myofibrils (between dashed lines). Myofibrils are serial arrays of sarcomeres (white box). The arrow points to a mitochondrion. The bright area running in parallel to the myofibrils is part of the sarcoplamic reticulum (SR). Image modified from wikicommons.org [2].

1.1.2   The sarcomere

The sarcomeres represent the smallest contractile units of striated muscles. In resting muscle they have a length of 2.2-2.3 μm and are mainly composed of the **thick** and the **thin filaments**.

Thin filaments are polar double-helical assemblies of actin that are covered by tropomyosin and troponin-complexes with a 7:1:1 stoichiometry [3].

Thick filaments are tree shaped bi-polar assemblies of up to several hundred molecules of type II (or skeletal muscle) myosin. Myosin is a hexameric protein composed of two heavy chains and four light chains. Each heavy chain is composed of a long (ca. 1500Å) C-terminal rod domain and an N-terminal globular motor domain, which associates with one regulatory and one essential light chain (RLC/ELC). Although the exact structure of the assembled filaments is not known, electron microscopic and X-ray diffraction studies showed that the thick filaments are helical assemblies with a periodicity of 42.9 nm in which the myosin-heads are found in intervals of ca. 14.3 nm [4].

The striped pattern of the sarcomeres originates from the distribution of both filament

types **(Figure 1-2 A and B)**: Thick filaments are located in the A-bands (German: anisotrop). Due to high density of protein it appears dark on electron micrographs. The thin filaments are anchored in the sarcomere bordering Z-disk (German: Zwischenscheibe) and span the I-band (German: isotrop), a region of low density, to then protrude into the A-band where they overlap with the thick filaments. The center of the A-band contains two bright stripes, the so-called H-zones (Hensen-zone) which are again separated by a dark zone, the M-band (German: Mittelbande). The H-zone appears bright due to the absence of myosin heads (and overlapping thin filaments) at the stem of the thick filaments. The M-band lies in the center of the sarcomere and is the site where the rods of myosin assemble in an anti-parallel manner. Its darker appearance is presumably due to the presence of additional myosin crosslinking proteins (see below). The described filament order creates an inner Z-I-A-H-M-H-A-I-Z symmetry of the sarcomeres with the M-band as inversion center

Cross sections of myofibers reveal differences in the geometrical arrangement of the filaments in different sarcomeric regions. At the level of the I-bands, the thin filaments are arranged as a square lattice. This changes to a hexagonal one as they overlap with the trigonally arranged thick filaments at the level of the A-band. At the H-zone level, only trigonally arranged stems of the thick filaments are visible and additional density of thick filament cross-linking proteins can be observed at the level of the M-bands **(Figure 1-2 C)**.
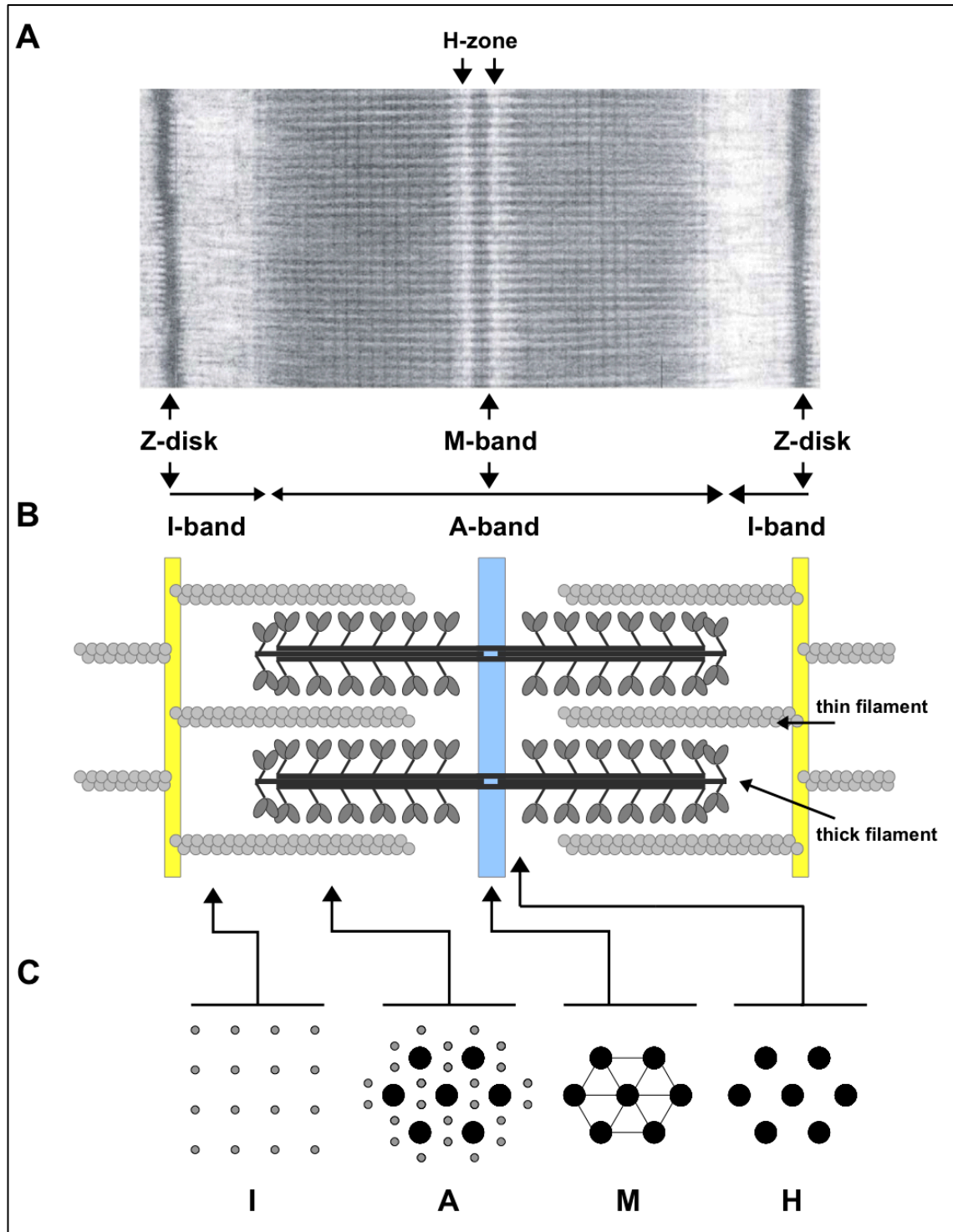
Figure 1-2 Basic architecture of the sarcomere.

(A) Electron micrograph of a sarcomere showing the characteristic cross striated pattern of Z-disks, I-, A-and M-bands and the H-zone. **(B)** Schematic representation of the sarcomere with its main components the thick (dark) and the thin filaments (light). The Z-discs are shown as yellow, the M-band as blue bar. **(C)** Filament geometry at different levels of the sarcomere. The large black dots represent the thick, the smaller grey dots the thin filaments. I: I-band; A: A-band; M: M-band; H: Hensen zone. (Electron micrograph modified from [1])

1.1.3   Sarcomere contraction: the sliding filament model

Contraction is the main function of the sarcomeres. It is triggered by electric stimuli from the nervous system in skeletal muscle or by auto-excitation of electrically coupled cells in heart muscle.

First evidence of how sarcomeres contract came from the observation that I-bands and H-zones shorten during contraction, while the length of the A-bands is preserved. More than 50 years ago this led to the 'sliding filament model of the sarcomere' [5][6] according to which, the Z-disc anchored thin-filaments move during contraction along the thick filaments and thus shorten the sarcomeres. It was found later that this sliding of the filaments along each other is the result of the formation of the actomyosin complex between F-actin and the myosin head domain, the so-called 'cross bridging cycle'. It is basically divided into two steps.

(1) A release of Ca^{2+} into the sarcomere results in a conformational change of the calcium binding protein troponin. Calcium binding by troponin induces a conformational change in tropomyosin. This change uncovers the myosin binding sites on F-actin.

(2) Step 1 allows the binding of myosin-heads to F-actin. Binding to actin is accompanied by a change in the orientation of the myosin head relative to the stem of the thick filament (first described in [7]). This 'power stroke' allows the release of $ADP+P_i$ from myosin and moves it forward on the thin filament. The binding of ATP then dissociates the actomyosin complex and brings the myosin head to its original orientation from where it hydrolyzes the ATP and a new cycle can start.

The contraction of a whole sarcomere is the result of the continuous cross-bridge cycling of thousands of myosin-molecules on the actin filaments and depends on the presence of ATP and calcium ions.

1.2 The M-band: a crosslink for the thick filaments

The understanding of how sarcomeres contract as the result of the simple interaction of two polymeric protein complexes and their appearance as highly ordered subcellular structures allows the conclusion that this degree of order is necessary for their function. A variety of forces act on the sarcomeres. Force production by fast shortening is

inherently connected to an increase in disorder. Muscles do not only contract but are also passively stretched upon contraction of antagonistic muscles and can be deformed due to volume changes by the contraction within the same muscle. This begs the question: How are the intrasarcomeric order and the order of sarcomeres within the muscular hierarchy and with surrounding membranous structures like the calcium storing sarcoplasmic reticulum maintained?

From the appearance of a sarcomere, it seems likely that the two transverse structures that lie in the centers of the force generating structures, the Z-disks (in the center of the I-bands) and the M-bands (in the center of the A-bands) play an important role in the solution of these specific problems.

The **Z-disks** have been (and are) the subject of extensive investigations in the past and their structure and function both as a mechanical linker of the sarcomeres, and as a platform for the force generating thin filaments as well as a large number of regulatory proteins are well understood [8][9].

Far less is known about the function and architecture of the **M-bands**. A closer investigation of electron micrographs from longitudinal sections of myofibrils revealed that these approximately 100 nm wide, dark traverse bands in the middle of the A-bands [10] are composed of a distinct symmetric pattern of up to 8 pairs of separable lines with one central single line (denoted according to [10] as M-lines M1 – M9/M9'; larger numbers correspond to a larger distance from the central M1 line). Their appearance differs in different types of muscle fibers and stages of muscular development [10][11]. Five M-lines (M1, M4/M4', M6/M6') are most pronounced and frequently observed at an apparently regular inter-line distance of approximately 22 nm. Thereby the M4/M4' lines appear constantly strong in all types of muscle while the intensities of the central M1 line and the M6/M6' line varies according to the investigated fiber type [10].

These M-lines are likely to be the top-projection of the so-called M-bridges, thin lines of density with a central thickening that appear to interconnect the stems of the thick filaments in myofibrilar cross-sections at the M-band levels [12] [13] **(Figure 1-3 B)**. The thickening within the M-bridges represents an apparent structure designated as 'M-filament' running parallel to the thick filaments. This M-filament is according to a first structural model proposed by [14], cross-linked by secondary cross-bridges, resulting in M-lines of weaker density (for example M3 in **Figure 1-3 A)**.
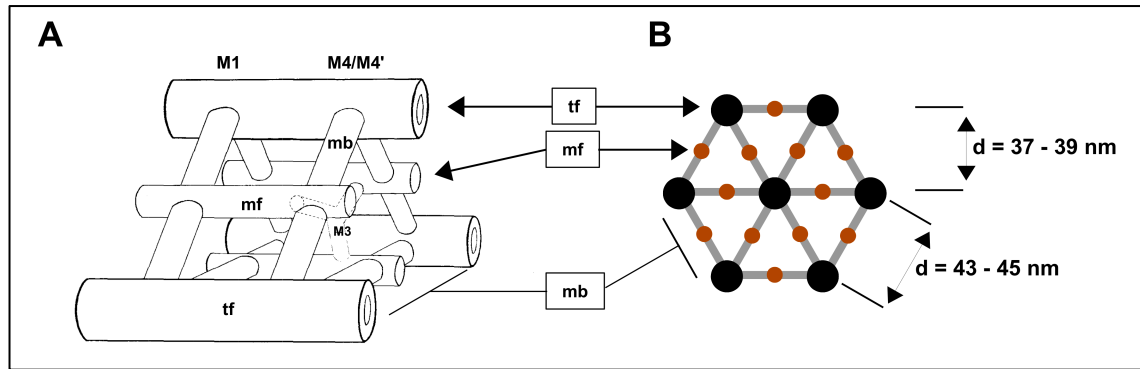
Figure 1-3 Appearance of the M-bands.

(A) Three-dimensional model of an M-band fragment adapted from [14]. The model consists of thick filaments (tf) interconnected by M-bridges (mb) appearing as pronounced lines M1, M4/M4' and M6/M6' (not shown) with a regular line distance of 22 nm. The centers of the M-bridges are interconnected by structures designated as M-filaments (mf), which run in parallel to the thick filaments. Weaker M-lines (e.g. M3) were assigned to secondary M-bridges that interconnect the M-filament structures between the M-bridges **(B)** The 3-dimensional model is based on the **appearance of the M-band lattice in electron micrographs of myofibrillar cross-sections** in which the equally spaced thick filaments (black dots) are interconnected by thin lines (grey) that show a thickening in their centers (brown dots).

The positioning of M-bridges and the M-filament as a lattice between the thick filaments suggests that they play an important **functional role** in the **maintenance of the constant axial distance** between the thick filaments. By this, they might ensure an appropriate spacing between the thick filaments [13] and so ensure appropriate spacing of the thick and the thin filament lattices prior to and during sarcomere contraction. A **second** proposed functional role is the **lateral alignment** of the thick filaments within the sarcomeres. Being bipolar structures, thick filaments do not necessarily produce equal forces on both half sarcomeres. Force production may run out of balance for different reasons. For example shortage of ATP after extensive contractions may result in the inability of a fraction of myosin heads to dissociate from the thin filament and therefore locally 'block' filament sliding. In addition, being dynamic polymeric structures, the thin filaments within the same sarcomere differ in their length [15][16] and therefore the number of possible cross-bridging sites may vary locally. Despite these possible sources of force-induced disorder, the A-bands have been shown to be able to move as a complete band (dependent on sarcomere tension) within the sarcomeres and possibly reach the Z-discs on one side [17]. These experiments suggest a strong and possibly elastic intra-A-band cross-linker that may be provided by

components of the M-bridges and/or the M-filaments.

Finally the M-bands, lying in the center of the force-generating thick filaments have been proposed to represent a regulatory node, which senses mechanical load and translates it into regulatory responses (reviewed in [18]). This is emphasized by the fact that a small number of myopathies, for example a subclass of the limb-girdle muscular dystrophies (types 2A and 2J) have been linked to proteins found within the M-bands [19][20]. In these types of myopathies, the sarcomeres seem to develop normally and signs of muscle failure become apparent only in the time of adolescence and in a certain set of muscles.

1.2.1 Components of the sarcomeric cytoskeleton associate in the M-band

The exact composition of the M-bands is not known to date and due to their differing appearance it can be safely assumed that their composition varies. In vertebrates, apart from myosin, three families of proteins of the sarcomeric cytoskeleton have been identified. **Titin** and proteins of the **myomesin** family have been located in and assigned to M-band specific sub-structures. The recently described **obscurin** and its small homolog **obscurin-like 1 (OBSL1)** occupy several myofibrillar sites but have both been found at the M-bands [21][22].

1.2.1.1 Titin: the third filament

Passively stretched skinned myofibrils rapidly return to their original length. Therefore, the presence of a component of the sarcomeres that donates its elastic properties to the sarcomeres has already been proposed at the time when the sliding filament model was postulated [23]. However, this behavior could not be assigned to a specific sarcomeric structure until an additional density was observed to appear in between the I-bands and A-bands of frog muscle, after these had been separated by extreme stretching of myofibrils [24]. The observed density, with a diameter of less than 30 Å, was by far thinner than the thin filaments and therefore likely to be composed of a single filamentous protein rather than a polymer. The identity of this 'filament' remained unknown until it could be purified as the insoluble fraction of bovine heart myofibril preparations and was named connectin, as it seemed to connect A- and I-bands [25].

Connectin was shortly thereafter identified by electrophoresis of muscle preparations from different species as two separately running large molecular weight proteins and in reference to its size named titin [26]. First coding sequences of the titin encoding gene *Ttn* identified a regularity of two types of repeats of ca. 100 residues on the protein [27] which fitted well with the previously observed appearance of the protein as 'string of pearls' with a periodicity of ca. 40Å [28]. The identity of these 'pearl-like' repeats as modules of the immunoglobulin-like (Ig-like) folding class was later confirmed by structure determination of the domain M5, an Ig-like domain of the I-set and the domain A71, a fibronectin type 3 (Fn3) module by NMR spectroscopy [29][30]. Today the complete sequence of the titin gene is known. It consists of 363 exons [31] of which approximately one third are subject to alternative splicing [32]. The size of the gene products is in a range of ca. 700 kDa (Novex-3 titin, expressed in heart [31]) to 3.7 MDa (found in soleus muscle [33]) and the expression of all exons would result in a product of 38,138 residues or 4.2 MDa. This makes titin the largest single polypeptide known to date. Analysis of the primary sequence resulted in the prediction that ca. 80% of the molecule is folded as repetitive Ig-like modules of the I-set and Fn3 class (current assignment Pfam/SMART: 152 Ig-like domains; 132 Fn3 domains). The remaining ca. 20% of the protein consists of non-modular sequences and a single catalytic kinase domain located close to the C-terminus.

1.2.1.1.1 Titin connects Z-disks to the M-bands

Titin has a maximum length of 1.25 μm (measured for skeletal muscle titin [34]) and spans more than half of the sarcomere from the Z-disc [35] to the M-band [36].
The **Z-disk region** of titin is located at its N-terminal region and composed of 10 Ig-like domains (termed Z1-Z10) and a number of repetitive sequences with a length of ca. 45 amino acids and a partial α-helical fold [37] named Z-repeats. The number of Z-repeats is variable due to alternative splicing and differs in different types of skeletal and heart muscle fibers [38][39]. Z-disk titin interacts directly with muscle α-actinin (α-actinin 2) [40] [41], the main component of the Z-discs. This interaction is mediated by the C-terminal non-metal binding EF-hand domain of α-actinin and Z-repeat number 5 [42]. Evidence for additional α-Actinin binding sites in the Z-repeat region exists (our lab, not published and [37]). In addition a direct interaction between titin domains Z10-I1 and thin filaments has been found [43]. Finally a third (and maybe most remarkable) Z-disc complex has been reported to link two molecules of titin at the periphery of the Z-

disks. The two N-terminal Ig-like domains of titin, Z1 and Z2 interact with the small protein telethonin (also named T-Cap) [44]. To interact, titin crosses the Z-discs [45] and forms an anti-parallel 2:1 complex. Additional evidence for the formation of larger complexes of both proteins by self-association of the unfolded C-terminus of telethonin exists [46]. The crystal structure of the trimeric 2:1 complex has been determined in our lab **(Figure 1-4)**. The structure revealed that the N-terminal half of telethonin is sandwiched in a palindromic way between the two antiparallel Z1Z2 Ig-like tandems by the formation of a large intermolecular ß-sheet [47]. Subsequent analysis of its resistance against externally applied forces revealed that disruption of this complex requires a force of ca. 700 pN [48] if the force is applied from the oppositely lying C-termini of the Z2 domains, which is the assumed pulling direction *in vivo*. This is the highest stability of a protein-protein complex ever measured [48] and emphasizes the role of titin for mechanical linkage of the sarcomeres. In addition to the interactions described here which are of importance for Z-disc anchoring, a number of proteins that interact with the Z-disc titin have been found and these have been assigned to a variety of regulatory functions [49].
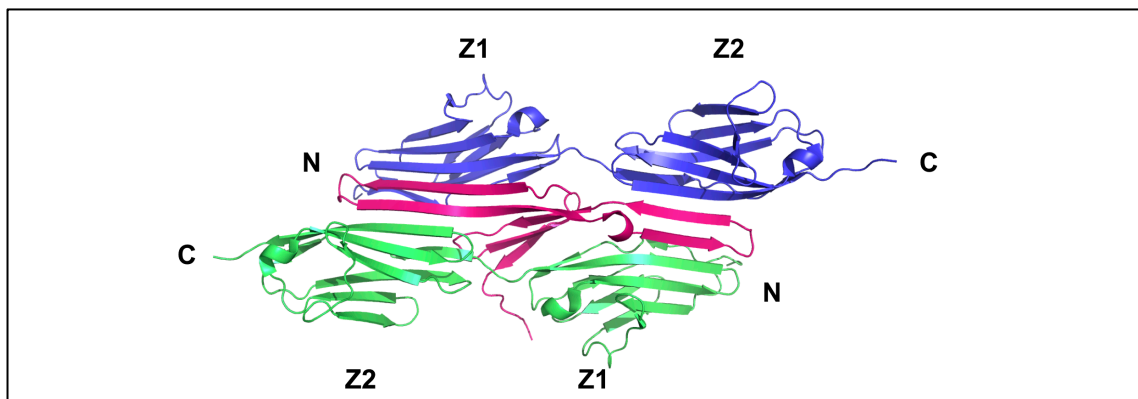
Figure 1-4 Crystal structure of the trimeric titin – telethonin complex.
Two Z1Z2 Ig-like tandems (green and blue) are connected to each other by the N-terminal half of telethonin (pink). Pulling from their C-termini revealed an enormous mechanostability, presumably based on the large intermolecular ß-sheet.

The **M-band** region of titin comprises the C-terminal approximately 240 kDa of the protein [50]. This part is encoded by 6 *Ttn* exons (called Mex1-6). This region consists of a single a N-terminally located Ser-/Thr-kinase domain (TK) and ten predicted Ig-like domains, separated by 6-7 interspersed regions (Is1-7) of up to 500 amino acids (Is2 which locates between domains M3 and M4) in length and have a unpredicted fold.

TK is likely to be activated by force-induced removal of a C-terminal inhibitory loop [51][52][53]. The main part of M-band titin, comprising all Ig-like domains and Is1-6 encoded by the exons Mex1-4 and 6 are constitutively expressed in all titin isoforms that contain the M-band region. Mex5, coding for Is7, the domains M9 and M10 linking region of ca. 100 residues length is subject to muscle type specific alternative splicing [50]. Electron miscroscopic mapping using monoclonal antibodies has shown, that the kinase domain and the C-terminal segment of M-band titin (domain M9) both locate at the periphery of the M-bands. It has therefore been suggested that this part of titin spans the whole M-band and therefore overlaps with itself [54]. However, a site of self-association that would directly link two titins from both half sarcomeres as known at the periphery of the Z-discs [47] has not yet been described.

A small number of interaction partners of the M-band region of titin are known **(Figure 1-5)**: Is7 provides one of the two known titin binding sites for calpain-3 (also called P94) [55], a skeletal muscle specific, calcium regulated Cys-protease that acts as a key regulator of sarcomeric protein turnover (reviewed in [56] and [57]). Is2 has been shown to bind FHL-2/DRAL and is proposed to act as a platform for metabolic enzymes in the M-band [58]. Three cytoskeletal proteins with an Ig-like rich architecture bind M-band titin in its middle region and at the very C-terminus. The M-band integral protein myomesin-1 **(section 1.2.1.2)** assembles with domain M4 of titin via three predicted Fn3 domains (domains 4-6) located in its center. Complex formation has been proposed to be negatively regulated by a phosphorylation event in the linker region between myomesin domains 4 and 5 [59]. The very C-terminal domain M10 interacts with the N-terminal Ig-like domains of the homologous proteins obscurin and OBSL1. Both also bind myomesin-1 in the linker region between myomesin domains 4 and 5 [60]. In this way, titin and obscurin/OBSL1 share the same site of assembly on myomesin-1. The consequences are not yet clear. However, the disruption of the respective titin-obscurin/OBSL1 and obscurin/OBSL1-myomesin complexes by over-expression of single interacting domains in rat cardiomyocytes only affected the M-band localization of obscurin while no effect on myomesin or titin was observed [60].
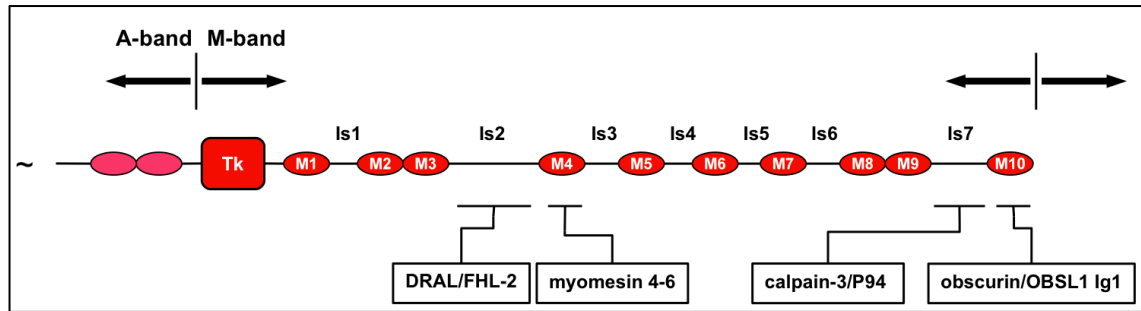
Figure 1-5 Schematic representation of M-band titin.

The region spans the whole M-band and consists of 10 Ig-like and one Ser-/Thr-kinase domain. The binding site of the 5 known binding partners are shown at the bottom. Note that the length of Is2 is not proportioally. It is approximately five times longer than e.g. Is7.

Taken together the findings on the location of titin termini within the sarcomeres revealed a surprising property of this third filament: As it crosses both the Z-discs and the M-bands, titin may not only be seen as a connector within the sarcomeres but also as a connector between them. These interconnected titins are likely to create a filament system spanning along whole myofibrils.

1.2.1.1.2 The central portion of titin acts as sarcomeric ruler and spring

The orientation of titin along the two halves of the sarcomeres and its anchoring at the borders suggest that it could play a role in sarcomere elasticity and could also act as a molecular template for the arrangement of the force generating components. Indeed, the radiation induced disruption of titin and nebulin (a Z-disc anchored thin filament binding protein) in skinned myofibers has been shown to be accompanied by a loss of resistance towards passive tension and by the intra-sarcomeric misalignment of the thick-filaments after calcium-induced contraction [61]. Two regions on titin have been found to be responsible for elasticity with respect to filament alignment:

Elastic properties are mainly achieved by spring like force-dependent length changes in the titin **I-band** region. This region consists of three different sequence types: Ig-like domains of the I-set arranged in tandem, the N2 region and the PEVK region. The PEVK region is named after its high content in proline (P), glutamate (E), valine (V) and lysine (K). These are not equally distributed but clustered into sequence region patches of strongly differing pI [62]. Length changes were proposed to result from changes in tilting angles between Ig-like domains [63] and electrostatic folding as well as unfolding of differently charged stretches in the PEVK region [62]. In addition reversible unfolding of I-band Ig-like domains were proposed to be involved in the extension and possible intra-domain disulfide bond formation under oxidizing and reducing conditions to modulate this process [64]. However Ig unfolding is today regarded as a process of secondary importance for the elasticity because of the slow process of domain refolding but it may play a role in the recovery of overly stretched sarcomeres. Tension of the molecule beyond its resting length is assumed to be first compensated by straightening of the Ig rich region followed by unwinding of the PEVK segments [65][66]. Alternative splicing of the extendable I-band part of titin is the main origin of differences in the protein's length and a source of muscle/fiber type specific adaptation. For example skeletal muscle isoforms can contain up to 53 additional Ig-like segments when compared to shorter cardiac isoforms [32]. Conversely the length of the elastic PEVK region tends to be much shorter in the rather stiff cardiac myocytes (minimal 163 residues) than in more flexible skeletal muscle fibres (up to >2200 residues).

The titin part spanning the **A-bands** in both half sarcomeres comprises a length of ca.

0.8 µm and a calculated weight of ca. 2 MDa. It is less variable in its composition than the flexible I-band spanning region and contains according to fold predictions domains of the Fn3 class as the only titin segment, while only Ig-like domains of larger classes are predicted in the other parts. A hallmark of the A-band region is its very regular composition of repetitive modules designated as '**super-repeats**'. Two types of repeats exist: The first type of repeat is found in the inner part of the A-bands denoted as C-zone (zone of 'constant packing') [10]. Here 11 copies of a repeat comprising 11 predicted domains in the order 'Ig–Fn3–Fn3–Ig–Fn3–Fn3–Fn3–Ig–Fn3–Fn3–Fn3' exist in a continuous series. The estimated repeat length of 44 nm (11 domains of ca. 4 nm length) is in good agreement with a regular pattern of 11 traverse stripes that can be found in the C-zone **(**visible as faint stripes in **Figure 1-2 A)** [4]. These have a periodicity of 43 nm and Bennett *et al.* [67] demonstrated that the seven or eight outer stripes contain the 130 kDa protein MyBP-C (or Protein C).
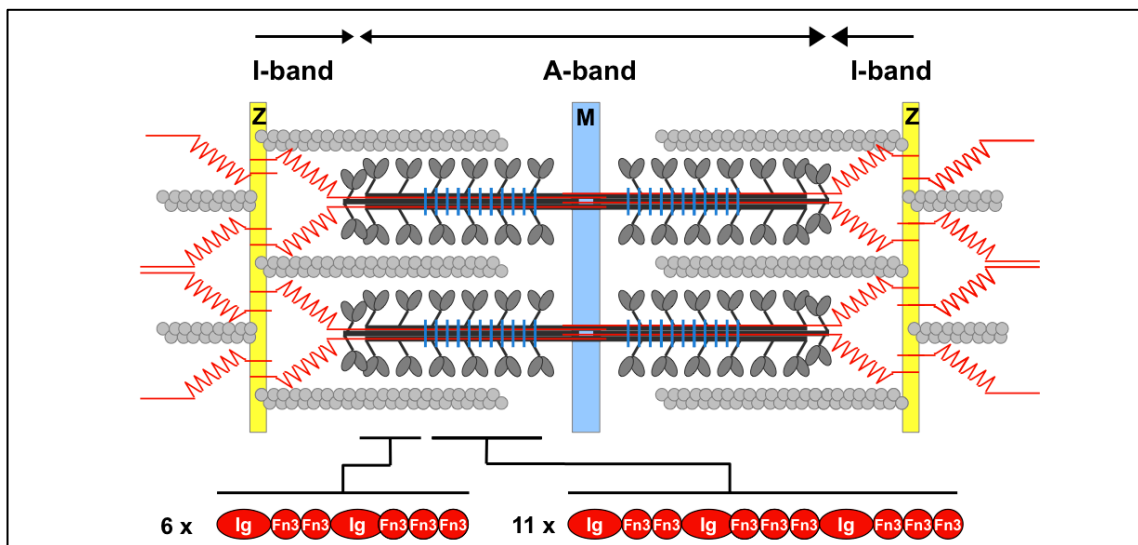
Figure 1-6 Titin in the sarcomeres.

Titin (red line) spans the half sarcomere between Z-disks and M-bands which it both crosses. The I-band region (jagged) is the main source of sarcomeric elasticity and site of protein variability due to alternative splicing. The blue lines represent a pattern of 11 traverse stripes, which cross over the sarcomeres within the C-zone of the A-band. This pattern reflects the pattern of A-band titin 11-domain repeats.

MyBP-C interacts with the C-terminal tail of myosin [68] and the N-terminal Ig-like domain of each 11-domain super-repeat [69]. Connected to myosin through MyBP-C, the 11 domain repeats could in this way act as a template to define the size of the thick filament region [4]. The N-terminal region of A-band titin spans the so-called D-zone of

the A-band and comprises six copies of a second, smaller repeat 'Ig–Fn3–Fn3–Ig–Fn3–Fn3–Fn3'. The role of these is not yet understood. Their predicted length (ca. 27 nm) is nearly twice the length of the axial spacing between myosin-heads (14.3 nm). Therefore, a role in the observed pairing of the myosin heads groups in the D-Zone (close to the tip of the thick filaments) has been proposed [4], but experimental evidence for this is still missing. In addition to spatial reasoning based on its structure, complete or partial knockout experiments of titin confirm its role in the development and organization of the myofibrillar structure: Complete knockouts of the titin gene results in a failure of myofbril assembly [70] in which, although expressed at normal levels, myosin remains diffusely distributed. Similar observations were made in a knockout experiment where only the C-terminal M-band region of titin was disrupted in cardiomyocytes. Although phenotypically normal when heterozygously deleted, a homozygous knockout failed to assemble a striated sarcomeric pattern, as well as the distant Z-disc structures [71].

1.2.1.2 Members of the myomesin family: thick filament cross-linkers

The myomesin protein family consists of three homologous proteins located at the levels of the M-bands in vertebrate striated muscles: Myomesin-1 (185 kDa) [72], M-protein (myomesin-2, 165 kDa) [73] and the recently identified myomesin-3 (162 kDa) [74]. They are encoded by separate genes found in all vertebrate families, but are absent in invertebrates and share, within the same species primary structure identity levels below 50% [74]. Even so, all three share a **common domain architecture (Figure 1-7 A)**: a unique N-terminal domain followed by two domains of the Ig-like class, a central portion composed of five domains of the Fn3 Ig-like subclass and a C-terminal region composed of five domains of an unpredicted class of Ig-like domains**.** In addition, myomesin-1 contains a unique insertion sequence called EH-segment (embryonic heart) between domains 6 and 7 that is only present during developmental stages, but is later spliced out in fully differentiated muscle [75]. This EH-segment was shown to act as a non-folded, entropic spring based on CD-spectroscopy and AFM experiments. This awards the protein elastic properties in a manner similar to that observed for the PEVK region of I-band titin [76].

Myomesin proteins have been suggested to be the major comppnent of the M-bridges and thus crosslink the thick filaments. Their expression patterns match the appearance

of the M-lines in different types of muscle fibers.

With a mass of 185 kDa **myomesin-1** [72] is the largest member of the family and the only one that is constitutively expressed in all types of striated muscle fibers [75][77]. It appears to be the main constituent of the ubiquitously present **M4/M4' lines** of the M-band [54] and was therefore proposed as a major cross-linking protein of the thick filaments. The association of myomesin-1 with five different proteins has been described **(Figure 1-7 A)**. The N-terminal non-Ig-like domain binds to the rod-domain of myosin present in the M-band [54]. Three Fn3-class domains in its central portion (My4-6) were found to bind domain M4 of M-band titin in a solid phase binding assay. As previously mentioned, phosphorylation of a serine on the short linking sequence between domains My4 and My5 has been proposed to act as a negative regulator for this assembly [59]. The titin binding-site overlaps with the binding sites for the third predicted Ig-like domain of obscurin and OBSL1 binding in the My4 and My5 linking sequence [60]. The last binding partner is the muscle type creatine kinase (MM-CK) a protein involved in muscular ATP-regeneration, which binds myomesin-1 Fn3 domains My7-8 in a pH-dependent manner [78]. Hence, myomesin-1 does link major constituents of the M-band. Although its length apparently allows the direct crossing between two thick filaments (center to center distance = 43-45 nm), a single binding site for myosin cannot explain physically thick filament cross-linking.

The discovery of a C-terminal site of homo-dimerization mediated by the Ig-like domain My13 [79] may shed new light on the cross-linking function of myomesin-1 within the M-band. Indeed, the assembly of two myomesin-1 proteins could allow the binding of myosin on the two tips of such a myomesin-filament. The crystal structure of the assembly of two My12-13 domain tandems [80] showed that two My13 domains assemble in a symmetric antiparallel way by the formation of a symmetric intermolecular ß-sheet. Surprisingly, this crystal structure also revealed the presence of a short α-helical linker between the My12 and My13 **(Figure 1-7 C)**. Reversible unfolding of this and further predicted helical linkers on the filament were proposed to account for the compensation of physical stress, a hypothesis which was later confirmed *in vitro* in single molecule AFM experiments and molecular dynamics simulations [81].
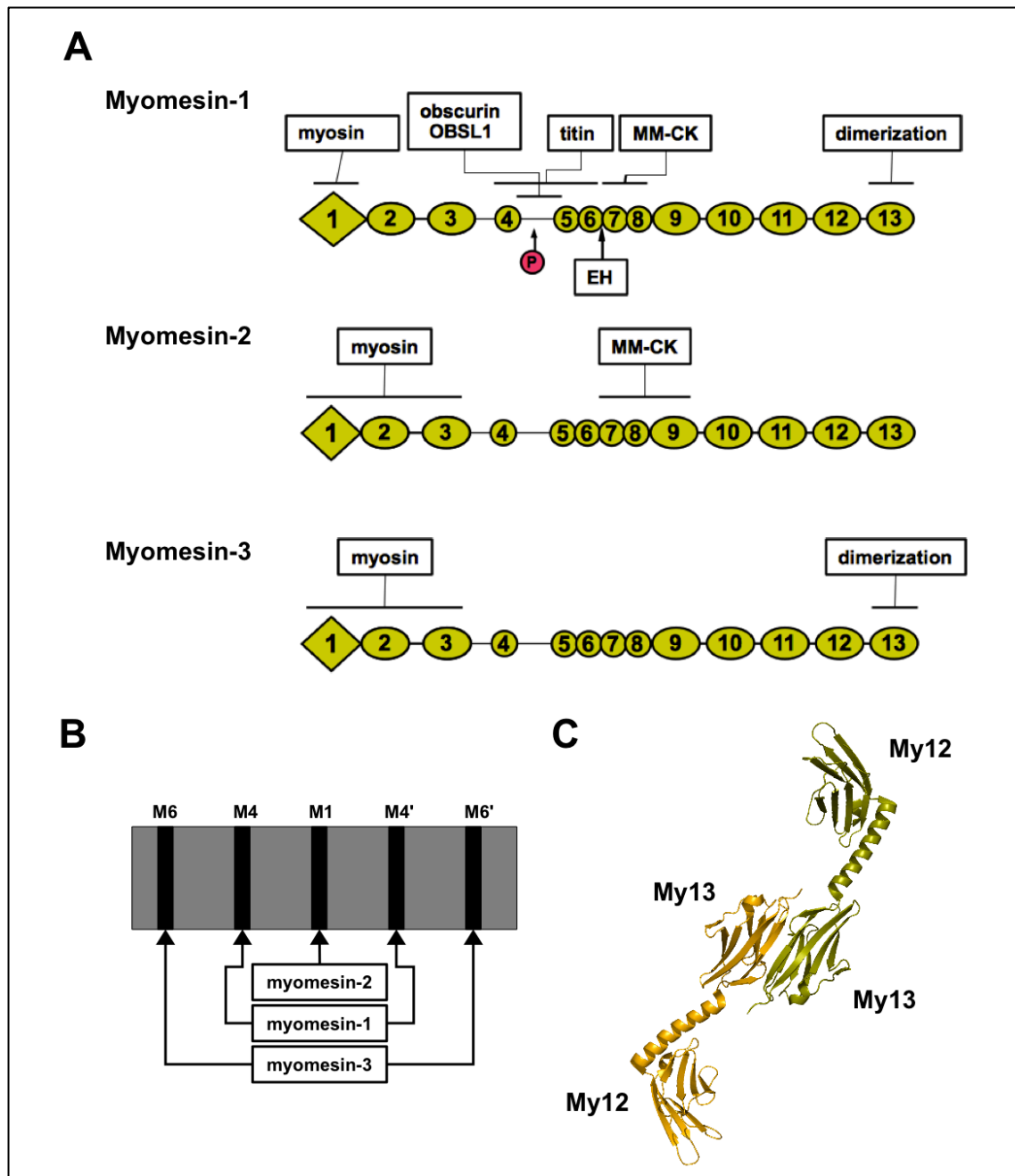
Figure 1-7 Properties of proteins of the myomesin family.

(A) The three known myomesins share the same domain architecture comprising an N-terminal myosin-binding domain (rhomb), seven predicted domains of the Ig-like folding class (ellipsoids) at the N- and C-terminal side and a middle segment composed of Fn3 domains (circles). The locations of know interaction sites are labeled by bars on top, the known phosphorylation site in the My4-5 linker and the location of the alternatively spliced EH segment in myomesin 1 are shown at the bottom. **(B)** Assignment of the three family members to the three most prominent pairs of M-lines as proposed by [74]. **(C)** Crystal structure of the C-terminal self-assembly of myomesin-1 mediated by domain 13.

Myomesin-2 and **myomesin-3** are less well characterized. Both proteins are smaller than myomesin-1 (165 and 162 kDa) and show a fiber type specific expression pattern, which matches the occurence of different M-lines (in mouse). For example slow fibres of skeletal muscle show a four M-line pattern (M4/M4' and M6'M6') and express myomesin-1 and myomesin-3 while the fastest fibres display a 3 line pattern (M1 and M4/M4') and express only myomesin-1 and 2. The conclusion is therefore that **myomesin-2** is the constituent of the **single central M1-line** [54] and **myomesin-3** the constituent of the outer **M6/M6' lines** [74] **(Figure 1-7 B)**. Although significantly differing in length and primary sequence, the N-terminal domains of both homologs are responsible for the targeting to the thick filaments [82][74]. A C-terminal self-assembly of myomesin-3 has been demonstrated [74] while myomesin-2 lacks this property [79]. There is therefore still no plausible explanation for thick filament cross-linking of myomesin-2.

1.2.1.3   Obscurin and OBSL1

1.2.1.3.1 Obscurin and OBSL1: Novel sarcomeric filaments

The third class of cytoskeletal proteins present in the vertebrate M-bands is composed of the two known members of the obscurin family: obscurin and obscurin-like 1 (OBSL1). **Obscurin** has only recently been discovered [21] as a protein mainly expressed in heart and skeletal muscle and was named for the fact that it could not be purified from muscular extracts. This is probably due to its large size and association with muscular membrane systems. The complete gene structure of obscurin revealed 119 exons coding for a protein of up to 800 kDa in size [83]. Domain predictions show a modular architecture resembling that of titin: The N-terminal part of the protein contains a long serial array of 65 predicted Ig-like domains of which two belong to the Fn3 sub-class. In contrast to titin, this Ig-array does not contain predictable linker sequences of significant length or display obvious domain segmentation. The C-terminal region differs from the poly-Ig region in its composition: Next to a small number of Ig-like domains it contains a variety of domains involved in signaling processes. The poly Ig-array is preceded by a calmodulin binding IQ-motif [21], six Ig-like domains and a SH3-RhoGEF-PH domain triplet, which is involved the activation of the small GTPases RhoA and RhoQ [84][85]. Downstream of this GEF-region, differential splicing creates two structurally different C-termini giving rise to the two

known main isoforms obscurin A and B: In **obscurin A**, the GEF region is followed by a ca. 500 amino-acid long C-terminal sequence with no predictable domain structure. Two stretches within this sequence have been assigned to link to the SR-integral protein small ankyrin-1 (sAnk1) and to create a direct link between the sarcomeres and muscular membrane systems [79][86]. In **obscurin B**, this region is replaced due to an exon-skipping event [21] by a segment containing two predicted Ig-like domains, one Fn3 and two kinase domains of the myosin light chain kinase (MLCK) class. The catalytic activities and substrates of these domains are yet to be determined.

OBSL1 is a shorter (130-230 kDa) filamentous protein with strong homology to obscurin's N-terminal poly-Ig segment, but with a broader tissue specific expression pattern [83]. Three alternatively spliced mRNAs have been described, in which splicing occurs at the 3'-end. This gives rise to proteins containing 9-20 predicted Ig-like and a single Fn3 domain. Unlike obscurin with whose N-terminal segment it shares primary sequence identity levels of approximately 30%, OBSL1 completely lacks domains with signaling functions. OBSL1 is expressed from a separate gene that was proposed to have evolved by gene duplication from the invertebrate obscurin homolog UNC89. This duplicated gene then split into separate genes coding for OBSL1 and the tandem kinase domain containing SPEG (striated muscle preferentially expressed gene [21]). Its broad expression pattern, which includes expression in non-muscular tissues, suggests OBSL1 acts as a general cytoskeletal adaptor protein [84], a hypothesis that is supported by the discovery that some OBSL1 variants are involved in the development of the non-muscular developmental disease 3M-syndrome [85].

1.2.1.3.2 Obscurin and OBSL1 assemble with titin and myomesin-1 in the M-bands

Obscurin and OBSL1 have been shown to locate at different regions of the myofibrillar structure. Antibodies directed against Ig-domains in the center of the obscurin N-terminal Ig-like segment or against its GEF domain stained primarily at the level of the M-bands of fully differentiated rat myofibrils, while other antibodies stained simultaneously at the level of the Z-discs and M-bands during early sarcomere development [86][87]. In contrast, antibodies directed against the non-modular C-terminal region of the obscurin A isoform stained adult rat myofibrils simultaneously at the levels of Z-discs and M-bands [88]. There is further evidence, that obscurin might relocate between the Z-discs, A/I junction and I-bands depending on muscular stress

[89]. The myofibrillar location of OBSL1 differs from that of obscurin. Both proteins share the location at M-band levels, but location of OBSL1 at the Z-disc level could not be observed. Instead antibodies directed against the N-terminal OBSL1 domain bind around the nuclei of rat cardiomyocytes [22], an observation that could reflect OBSL1's proposed function as a general cytoskeletal adaptor protein.

Given the homology between these two proteins, the way by which they are discriminated and targeted to specific locations within the sarcomere has been investigated. Several studies have identified binding sites for both proteins on titin. Z-disc and I-band targeting of obscurin seems to be mediated by the Ig-like domain tandem Ig58-59 located at the C-terminal end of obscurin's poly-Ig segment. This tandem has been reported to bind to the two titin domains Z9-Z10 at the Z-disc periphery [21], as well as to a single I-band internal Ig-like domain which is exclusively expressed in the short heart specific Novex-3 titin isoform [87]. The M-band location of obscurin and OBSL1 is mediated by the binding of the very N-terminal Ig-like domain 1 (Ig1) to the C-terminal titin domain M10. The second M-band targeting interaction is the direct binding of Ig3 of both proteins to a linker region connecting myomesin-1 domains My4 and My5 [60].

The parameters determining the location of obscurin at the Z-disc, I-band or M-band levels, as well as the location of OBSL1 at the M-band or perinuclear region are yet to be determined. Both known M-band binding epitopes of obscurin are present in either of the known isoforms A and B. The removal of one of the binding sites on obscurin by alternative splicing or steric inhibition of binding due to structural rearrangements in the protein may define the positioning. Yet our knowledge on the structure and possible further isoforms is too limited to explain alternate locations. The same is true for OBSL1. Its binding sites for titin and myomesin-1 are the only known points for sarcomeric attachment. A potential nuclear binding partner has still to be determined.

1.2.2 Alterations of M-band proteins are involved in the development of hereditary diseases

The importance of the M-band is highlighted by the finding that alterations of the two M-band proteins titin and calpain-3 are related to the development of different rare hereditary muscle diseases. Three groups of M-band related diseases have been described:

Several diseases of the tibial muscular dystrophy (**TMD**) and limb-girdle muscular dystrophy (**LGMD**) group are phenotypically accompanied by a slowly progressing weakening of muscles mainly of the lower extremities. Other skeletal and cardiac muscles do not seem to be affected. The time of onset varies between the first and sixth decade of life. Speed of disease progression and extent of weakening vary between patients.

Two types of LGMD have been correlated to alterations in calpain-3 and M-band titin:

LGMD2A (2 designates the autosomal recessive inheritance) is caused by mutations on the calpain-3 gene [19], a regulatory protease bound to M-band titin through interaction with the Is7 segment [90]. Autosomal recessive **LGMD2J** and dominant **TMD** are associated with nucleotide substitutions or insertions on the last two exons, Mex5 and Mex6, of M-band titin, resulting in amino-acid substitution or truncation of the encoded regions, the calpain-3 binding Is7 segment or the obscurin/OBSL1 binding domain M10 (known mutations are summarized in [91]). The mechanism of LGMD and TMD development is not clear. However, a secondary loss of calpain-3 from the M-bands upon alterations in the M10 domain *in vivo* [92], and the inability of two M10 disease variants to bind obscurin/OBSL1 *in vitro* [60] points towards an involvement of the binding partners.

Hereditary myopathy with early respiratory failure (**HMERF**) is an autosomal dominantly inherited myopathy characterized by a late onset (average age: 35 years) and a weakening of the upper and lower extremities, as well as respiratory muscles, but lack of a cardiac phenotype [93]. An involvement of a single Arg to Trp residue exchange on the kinase domain of titin was proposed to be linked to the development of this disease. Although no modulation of kinase activity could be observed, the mutation was found to interfere with the binding of the domain to NBR1, an adaptor protein that connects ubiquitin-ligases of the MURF family to titin [94].

The third group of disease phenotype designated as early onset myopathy with fatal

cardiomyopathy (**EOMFC**) is characterized by an early onset (5-12 years) of skeletal muscle weakening and dilated cardiomyopathy with early fatal consequences [95]. The titin genes of affected patients carried two different homozygous nucleotide insertions that result in a frameshift and thereby create early stop codons at the end of the exons Mex1 (within the coding region for domain M5) or Mex3 (within the coding region for M7-M8 connecting Is6).

1.2.3   A current model of the vertebrate M-band

How are titin, myomesin-1 and obscurin/OBSL1 assembled to form 'an' M-band? This question was, with respect to titin and myomesin-1/-2 first approached by Obermann and colleagues [54]. They used monoclonal antibodies against four epitopes on titin, six epitopes on myomesin-1 and six epitopes on myomesin-2 to determine distances of epitope-bound antibodies to the M-band central M1-line by electron microscopy. While all antibodies against myomesin-2 bound close to the M1 line, the assumed location of the protein, only antibodies against N-terminal myomesin-1 epitopes were found at the levels of the M4/M4' lines, whereas C-terminal epitopes showed shorter distances to the center. The labeling of titin epitopes revealed the close proximity of N-terminal M-band (titin kinase domain) and C-terminal M-band domains (domains M8 and M9). Based on these data, they proposed a model in which M-band titin spans the complete M-band in an anti-parallel way and myomesin-1 is bound through its N-terminus to the thick filaments at the M4/M4' levels to then protrude with its C-terminal part towards the M1 line. Titin could therefore possibly be the 'M-filament', which runs in parallel to the thick filaments. Although based on a very low resolution method (an antibody has approximately the dimensions of 1/3 of a whole myomesin protein) and limited available data (labeling of only one epitope in the central portion of M-band titin) this model was later indirectly confirmed by the discoveries of the C-terminal homo-dimerization of myomesin-1 [79] and the binding of the titin M4-domain to the center of myomesin-1 [59]. Thus myomesin-1 appears to crosslink neighboring thick filaments and to anchor titin in the M-band. Later the model was further constrained by the discovery of simultaneous binding between two N-terminal domains of obscurin/OBSL1 and the titin C-terminal domains and the center of myomesin-1 respectively. Both homologs seem to provide an additional titin-myomesin-1 link but *in vivo* complex disruption experiments have shown that this assembly is not essential for the correct localization of titin or myomesin-1 [60]. The positioning of the titin and

myomesin-1 binding sites at the N-terminal or third Ig-like domains , respectively, suggests that the very C-terminal M10 domain of titin, of which the position in the M-band have not been determined, are directed back towards the center of the M-band. A graphical representation of this constrained model is given in **Figure 1-8**.
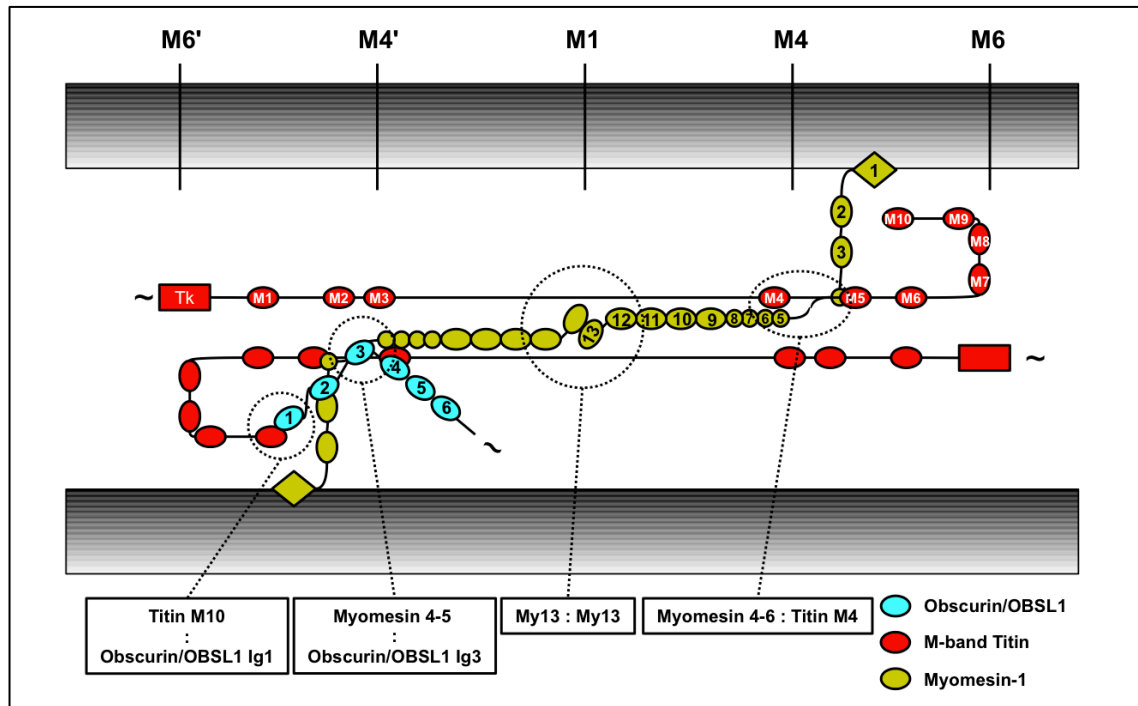
Figure 1-8 Schematic model of the vertebrate M-bands.

The model incorporates results from epitope-localization experiments [59], described assembly sites [59][60] and spatial constrains. Only the parts of titin and obscurin/OBSL1, which are related to M-band localization are shown. For a better visualization, only one molecule of obscurin/OBSL1 is shown. The locations of the M-lines M1-M6/M6' are designated on top and inter-filament assembly sites are designated by a dashed circle. The non-structural proteins calpain-3, FHL-2 and MM-CK are not shown.

1.3 Aim of the work

Previous work has shown, that the M-bands are structures in which the thick filaments are cross-linked by a lattice of regular appearance. Efforts to understand their structure and function have led to the discovery that the filamentous proteins titin, myomesin-1 and obscurin/OBSL1 locate in and interact with each other on several sites in the M-bands. The characterization of these, potentially lattice forming protein-protein interactions, has to date been limited to the qualitative confirmation of complex formation by biochemical methods and low-resolution co-localization studies with the exception of the homo-dimeric myomesin-1 tail-to-tail interaction.

To gain a better understanding of the M-bands, the formation of two known interactions of titin in the M-bands should be characterized *in vitro* by biophysical methods and X-ray crystallography in this thesis.

As a first target the complexes of titin and obscurin/OBSL1 were chosen. These only recently described complexes [60] are of special interest. They represent the currently only known assembly between head and tail domains of sarcomeric filaments and the impairment of their formation is potentially involved in the development of hereditary muscle diseases.

Second, the assembly of titin and myomesin-1 was chosen. In contrast to other mapped binding sites of sarcomere filaments, several domains have been shown to be necessary for assembly and potentially post-translational modifications are involved in its regulation [59]. This indicates that a potentially novel type of interaction between Ig-like domains may be present.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals, kits and consumables

All chemicals used in this work were of highest available purity and were purchased from Roth (antibiotics, buffers, organics and salts for *E. coli* culture, electrophoresis and protein purification except TCEP and Agarose) and Sigma-Aldrich/Fluka (chemicals for crystallization). Additionally used chemicals, kits and consumables are listed in **Table 2-1**.

Table 2-1 Chemicals, kits and consumables

Chemicals and solutions	Supplier
Agarose	Serva
DNA standard 1 kB ladder	New England Biolabs
Ni-NTA agarose	Qiagen
Rotimark 10-150 and Rotimark Standard	Roth
PageRuler unstained protein ladder 10-200 kDa	Fermentas
Silicone / Paraffin oil	Hampton research
Spectinomycin	Sigma-Aldrich
TBE-buffer (10x)	Roth
TCEP	Soltec Ventures
Kits	
NucleoSpin Plasmid	Macherey & Nagel
Pre-crystallization test	Hampton research
Wizard SV Gel and PCR cleanup kit	Promega
Consumables	
Centrifuge filters 0.6 ml (0.2/0.45 μM pore size)	Millipore
Centrifugal ultrafiltration devices (6 ml/ 20 ml; 3 kDa, 5 kDa, 10 kDa MWCO)	Sartorius, Millipore
Crystallization plates 24 well with screwed lids	Nextal / Qiagen
Dialysis membranes from regenerated cellulose (3.5 kDa, 8 kDa MWCO)	Roth
Limbro 24 well crystallization plates	Crystalgen / Molecular dimensions
Plastic tubes (50, 15, 2, 1.5 ml)	Saarstedt
Plastic petri dishes	Saarstedt
PCR-tubes (0.2 ml)	Millipore

Siliconized glass cover slides	Hampton research
Syringe filters (0.2/0.45 µM pore size) filtropur	Sartorius
Under oil vapor batch plates	Jena Bioscience / Douglas Instruments

2.1.2   Culture media

All media were mixed from components in distilled water and autoclaved (121°C at 2 bar for 20 min) prior to use.

Table 2-2 Culture media compositions

	1/l			1/l	
LB (Lennox formulation)			**TB**		
Tryptone	10	g	Tryptone	12	g
Yeast extract	5	g	Yeast extract	24	g
NaCl	5	g	K$_2$HPO$_4$	12.5	g
			KH$_2$PO$_4$	2.3	g
			Glycerol	4	ml
LB agar					
Trypton	10	g			
Yeast extract	5	g			
NaCl	5	g			
Agar	15	g			

2.1.3   Enzymes, primers, vectors and expression strains

Oligonucleotides were purchased from Eurofins MWG Operon in HPSF purified quality and lyophilized form. They were diluted in double distilled water to a final concentration of 10 pmol/µl and frozen until needed. Phusion™ polymerase, *DpnI* restriction endonuclease and dNTP stock solutions were purchased from New England Biolabs, Taq 2x Mastermix from Fermentas.

Human-Rhinovirus-14 3C-protease (commercially available as Precision™ protease, Novagen) was expressed using an in-house expression-vector (pET24D-3C; Arie Geerlof, unpublished) in *E. coli* BL21(DE3) and purified by IMAC standard protocols. The 6x-His-tagged protease was stored in aliquots at -80°C in 20 mM Tris pH 8.0, 300 mM NaCl, 20% v/v glycerol.

Plasmid backbones for the cloning of expression vectors were provided by the EMBL-Hamburg vector collection (pETM-14) or custom made by RF-free cloning using existing vector backbones (pETM-14SMT3, pCDF-3C, pCDF-3C-SMT3).

Template DNA for the amplification of gene fragments was either purchased as cDNA from imaGenes-Bio (Berlin, Germany), as codon-optimized DNA from MrGene (Regensburg, Germany) or received as a kind gift from Peter van der Ven (Fürst lab, University of Bonn, Germany).

The expression strain *E. coli* BL21 (DE3) star pRARE2 used in this work is a commercially available strain, optimized for increased mRNA stability (RnaseE knockout, Invitrogen) that has been transformed with the pRARE2 plasmid (Novagen) to allow the expression of proteins from genes with rare codons. *E. coli* DH5α cells were used for cloning and plasmid propagation. Cells were made chemically competent using an in-house protocol based on the Inoue method [96].

Table 2-3 DNA-templates used for creation of expression vectors

Gene	Clone	Supplier
M-band Titin	DKFZp451A172Q	imaGenes, Berlin
OBSL1	IRALp962B2418Q	imaGenes, Berlin
Obscurin	Obs1-3	MrGene, Regensburg
Myomesin-1	1-7 (pET23A)	Peter van der Ven

2.1.4   Buffers for protein purification and characterization

Table 2-4 Protein purification buffers

IMAC buffers

	Lysis		Wash		Elution	
Tris HCl, pH 8.0	25	mM	25	mM	25	mM
NaCl	300	mM	1	M	300	mM
Imidazole HCl, pH 8.0	-		10	mM	250	mM
TCEP, pH 7.0	0.5	mM	0.5	mM	0.5	mM
DnaseI	5	µg/ml				
Hen egg white lysozyme	10	µg/ml				

3C cleavage buffer   **SEC buffer**

Tris, pH 8.0	25	mM		20/10	mM
NaCl	300	mM		150	mM
1-thioglycerol	6	mM	TCEP pH 7.0	1	mM
EDTA	1	mM			

PBS for ITC			**Lämmli SDS-PAGE running buffer (1x)**		
K_2HPO_4	1.8	mM	Tris base	15.15	g
NaH_2PO_4	8.1	mM	Glycine	72	g
KCl	2.7	mM	SDS	50	g
NaCl	137	mM			
TCEP	1	mM			

→ adjusted to pH 7.2 with NaOH

2.1.5   Machines and columns

Table 2-5 Machines and columns

Machine/column	Manufacturer
Aekta purifier	Amersham / GE healthcare
Aekta FPLC	Amersham / GE healthcare
Superdex 75 16/60 pg	GE healthcare
Superdex 200 16/60 pg	GE healthcare
Superdex 75 HR 10/300	GE healthcare
Superdex 200 HR 10/300	GE healthcare
Nanodrop	PEQlab
Tristar MiniDawn	Wyatt Technology

2.1.6 Software and databases

Unless otherwise specified, software, server based tools and databases were of the latest version published at the time of use.

Table 2-6 Softwares and databases

Software	Version		Reference
APBS	1.3		[97]
ARP/wARP	7.1	server based (EMBL-Hamburg)	[98]
Auto-rickshaw		server based (EMBL-Hamburg)	[99]
BALBES		server based (University of York)	[100]
COOT	0.6.1		[101]
CCP4-suite	6.1.13		[102]
Dali		server based (University of Helsinki)	[103]
DaliLite		server based (EBI Hinxton)	[104]
NCBI BLAST (blastn, blastp, blastX)		server based (NCBI)	[105]
PDB2PQR			[106]
Phenix-suite	1.6-486 1.7.1-743		[107]
PROPKA			[108]
PISA		server based (EBI Hinxton)	[109]
Pymol	1.4.1	DeLano Scientific LLC; open source version	
SMART/Pfam		server based (EMBL-Heidelberg, Sanger institute, Hinxton)	[110]
XDS	2009/10	MPI Göttingen	[111]
Databases			
NCBI Genbank		www.ncbi.nlm.nih.gov/genbank/	[112]
RCSB-PDB		www.pdb.org	[113]
UniProt		www.uniprot.org	[114]

2.2 Methods

2.2.1   Cloning of expression constructs and site directed mutagenesis

Vectors for the T7 promotor driven prokaryotic expression of recombinant proteins with a cleavable N-terminal 6x-histidine or 6x-histidine-SMT3 tag [115] were cloned between the 3C protease cleavage site and the BamHI restriction site of pETM-14, pETM-14SMT3 (EMBL-Hamburg expression vector backbones, unpublished), pCDF-3C or pCDF-SMT3 (custom made, unpublished) using the restriction free cloning (RF) method [116]. For this, the gene of interest was amplified from cDNA using primers that contained at the 5'-half sequences, which are complementary to the 3C-site (forward primer: 5'-CTGGAAGTTCTGTTCCAGGGGCCC-3') and the multiple cloning site (reverse primer 5'-CGACGGAGCTCGAATTCGGATCCTTA-3'), respectively, of the destination vectors and in the 3'-half, sequences of 20-26 bp length complementary to the gene of interest. The resulting PCR-product was then column purified with the Wizard SV Gel and PCR cleanup kit (Promega) and used as a "mega-primer" in the final cloning reaction with the destination vector as a template. The methylated original template vector was then removed by enzymatic digestion with 20U *DpnI* followed by incubation at 37°C for 1 h. 10 μl of the digested PCR reaction were transformed into *E. coli* DH5α cells. Colonies were screened for gene insertion by colony PCR with T7 promotor (5'-TAATACGACTCACTATAGGG-3') and T7 terminator primers (5'-CTAGTTATTGCTCAGCGGT-3') followed by estimation of the amplified DNA fragment size by agarose gel electrophoresis. Plasmid DNA of single insertion-positive clones was purified from 10 ml cultures (LB supplemented with 50 μg/ml kanamycin or spectinomycin) with the NucleoSpin plasmid mini Kit (Macherey & Nagel) and the correctness of the expression cassette was verified by di-deoxy sequencing ([117] Eurofins MWG). The composition of reaction mixtures and thermo-protocols is summarized in **Table 2-7**.

Table 2-7 Reaction mixtures and thermo protocols for RF-cloning

Mega-primer amplification		Cloning reaction
	Reaction mixtures	
0.5 µl	Phusion polymerase (2U/ µl)	0.5 µl
1 µl	dNTPs (10mM each)	1 µl
10 µl	5x buffer HF	10 µl
2.5 µl	fw-primer (10 pmol/µl)	
2.5 µl	rv-primer (10 pmol/µl)	
	mega-primer (>100 ng/µl)	10 µl
0.5 µl	template plasmid (50-100 ng/µl)	0.5 µl
33 µl	H_2O	28 µl

			Thermo protocol			
	98°C	30 s	initial denaturation		98°C	30 s
	98°C	10 s	denaturation		98°C	10 s
30 x	60°C	20 s	primer annealing	20 x	55°C	20
	72°C	30 s/kbp	primer extension		72°C	3 min
	72°C	10 min	final extension		72°C	10 min
	4°C		storage		4°C	
			DpnI digestion		37°C	1 h

Point-mutations were introduced by partial RF re-cloning into expression constructs. A mega-primer containing the desired mutation was amplified by PCR using a gene internal mutant primer (length >35 bp, Tm >75 °C, non-complementary bases lying in the center of the primer) in fw or rv direction and T7 promotor or T7 terminator primer as gene-external primer. The purified mega-primer was then used for re-cloning the mutated gene-version over the existing construct. Compositions of PCR reaction mixtures and thermo-protocols were identical to the ones used for initial cloning (**Table 2-7**). The presence of the desired mutations was then verified by sequencing. Constructs cloned in this work and their mutants are listed in **Table 2-8**.

Table 2-8 Constructs for protein expression in *E. coli*

Protein	Uniprot sequence ID	Construct	Residue range	N-terminal tag	Vector backbone
OBSL1	O75147				
		Ig 1	1 – 105	His-3C	pCDF-3C
Obscurin A	Q5VST9				
		Ig 1	1 - 103	His-3C	pCDF-3C
Titin	Q8WZ42				
		M4	33294 - 33395	His-3C	pETM-14
		M10	34253 - 34350	His-3C	pETM-14
		M10-V22P		His-3C	pETM-14
		M10-T24P		His-3C	pETM-14
		M10-A26P		His-3C	pETM-14
		M10-D61R		His-3C	pETM-14
		M10-R98E		His-3C	pETM-14
		M10-H55P		His-3C	pETM-14
		M10-I56N		His-3C	pETM-14
		M10-L65P		His-3C	pETM-14
Myomesin-1	P52179-2				
		My5	632 - 735	His-3C	pETM-14
		My4L	503 - 635	His-3C	pETM-14
		My4-5	503 - 735	His-3C	pETM-14
		My4-7	503 - 946	His-SMT3-3C	pETM-14SMT3

2.2.2   Protein expression in *E. coli*

Proteins were recombinantly produced by T7 promotor driven [118], *lac*-repressor controlled [119] expression in *E. coli* BL21star (DE3) pRARE2 (Novagen). Large-scale expression was carried out in TB-medium supplemented with 100 μg/ml kanamycin (expression from pET vectors) or 50 μg/ml spectinomycin (expression from pCDF vectors) and 34 μg/ml chloramphenicol. Cultures were inoculated with 0.5% (v/v) of an overnight pre-culture of freshly transformed cells and grown in volumes of 1l in 2l baffled Erlenmeyer-flasks under continuous shaking (200 rpm) at 37°C until an OD_{600nm} of 2.0 – 2.5 was reached. Then the protein expression was induced by addition of IPTG to a final concentration of 0.5 mM (single domain constructs), 0.2 mM (multi-domain constructs) and expressed at 20 °C / 220 rpm for 16 – 20 h. Cells were harvested by centrifugation at 4000 x g for 15 min. and the resulting cell pellets were stored frozen at -20°C.

2.2.3   Protein purification

Two methods were used for protein purification:

(1) Immobilized metal-affinity chromatography (**IMAC**) [120] utilizes the propensity of deprotonated His side-chains to complex divalent metal cations. Proteins that are recombinantly tagged with an array of His (a 'polyhistidine tag') can be separated from the majority of non-tagged proteins by adsorbance on a solid phase exposing divalent metal cations (e.g. Ni^{2+} or Co^{2+} immobilized with NTA on agarose beads). The bound protein can be eluted competitively by application of high concentrations of imidazole, protonation of the His-tag by pH-change, or by detachment of the bound cations from the solid phase using chelators.

(2) Size exclusion chromatography (**SEC**) [121] is a method that separates molecules according to their hydrodynamic volume (their "size"). The method is based on the differing probability of molecules to traverse a porous bead shaped column material like cross-linked dextran. While large molecules enter the bead pores with low probability, smaller molecules can easily enter and need to pass a larger volume before eluting from the column.

2.2.3.1   IMAC affinity purification of poly-histidine tagged proteins

Frozen cell pellets were resuspended in 10 ml/g of pellet of pre-cooled IMAC lysis-buffer and lysed by sonication on ice. Insoluble lysate fractions were removed by centrifugation (30 000 x g at 4°C for 1h). The filtered supernatant was then applied twice at room temperature to a gravity flow IMAC column (5 ml bed volume / 100 ml lysate), washed with 10 cv IMAC wash buffer and column-bound proteins were eluted from the column with 2 cv of IMAC elution buffer. The initial protein yield was estimated by measurement of the OD_{280nm}. For proteolytic removal of the N-terminal 6xHis / 6xHis-SMT3 tag, recombinant HRV14-3C protease was added in a 1:100 (w/w) ratio to the IMAC eluate and the mixture was dialyzed over night at 4°C against 2l 3C-cleavage buffer. Cleaved tags and residual un-cleaved protein were removed by reapplication of the dialysate to a 5 ml IMAC column. Due to the cleavage of tags with HRV14-3C protease (recognition sequence: LEVLFQ↓GP), all purified proteins contained in addition to their native sequence at their N-terminus a Gly-Pro dipeptide originating from the 3C-site.

2.2.3.2 Purification by size-exclusion chromatography

Final purification and buffer exchange of the purified proteins was done by size-exclusion chromatography in SEC buffer which contained 10 mM Tris-HCl pH 8.0 if the protein should be directly subjected to crystallization trials or with 20 mM Tris-HCl pH 8.0 if the protein was needed for a different purpose. Proteins were, if necessary divided into portions of $\leqq$ 50 mg, concentrated to a volume of 2 ml using centrifugal ultrafiltration devices, injected on a pre-equilibrated Superdex 75 16/60 pg (constructs titin M4 and M10; OBSL1 Ig1; myomesin-1 My5 and My4-5;) or a Superdex 200 16/60 pg column (myomesin-1 My4-7) and eluted from the column at a constant flow rate of 1.2 ml/min. Protein containing fractions (as determined by OD$_{280nm}$) were checked for purity by SDS-PAGE and the purest fractions were pooled, concentrated, flash-frozen in liquid nitrogen and stored at -80°C.

2.2.4 Protein and protein complex characterization

2.2.4.1 Estimation of protein concentrations

Concentrations of protein solutions were estimated by the measurement of light absorption at a wavelength of 280 nm on a NanoDrop ND-1000 UV/Vis spectrophotometer and application of the Lambert-Beer law [122] [123]:

Equation 1: Lambert-Beer law

$$c = \frac{A}{\varepsilon \bullet d}$$

where c is the protein concentration [mol l^{-1}], A the absorption [no dimensions], d the path length of light [cm] and ε the molar extinction coefficient [L mol^{-1} cm^{-1}] of the absorbing material. In case of protein absorption at 280 nm wavelength, ε can be calculated by summation of ε of absorbing residues (Trp, Tyr, disulfides) multiplied by their number per protein molecule:

Equation 2: Molar extinction coefficient of proteins

$$\varepsilon = n_{Trp} \bullet 5690 + n_{Tyr} \bullet 1280 + n_{disulfide} \bullet 120$$

As all proteins characterized in this work were of human cytoplasmic origin, no disulfides were expected and their absorption was ignored.

2.2.4.2 Estimation of protein weight and purity by SDS-PAGE

SDS - polyacrylamide gel electrophoresis (SDS-PAGE, [124]) allows the estimation of a protein's molecular weight and purity virtually independently of their fold and amino acid composition.

The proteins are denatured by heat in the presence of a reducing agent and sodium dodecyl sulfate (SDS), a detergent with a strongly anionic head group that binds the protein main chain (approx. 1 SDS molecule per 2 residues). When bound to SDS, the denatured protein's negative charge is directly proportional to its molecular weight. The proteins are then run along an electric field through an alkaline gel made of polyacrylamide, where their electrophoretic mobility varies according to their denatured size and charge. The molecular weight is then estimated by comparison to molecular weight standards run on the same gel.

The original Lämmli SDS-PAGE system (12 – 15% SDS-PAA separation gels pH 8.8, Tris -Glycine SDS running buffer pH 8.0) was used in this study. Gels were run at constant 200 V for 40 - 60 min and then stained with Coomassie G250 or R250 according to standard protocols.

2.2.4.3 Estimation of secondary structure content by circular dichroism

Circular dichroism spectroscopy is a method that allows the estimation of a macromolecule's secondary structure content by the measurement of differences in the absorption of left and right handed circularly polarized light.

The method was used for the assessment of the secondary structure content of mutated versions of titin M10. For this the proteins were dialyzed for 4-6 h against 20 mM sodium phosphate buffer pH 7.5 and then diluted in the same buffer to final concentrations of 0.11 – 0.25 mg/ml. Spectra between 190 and 260 nm wavelength were recorded in 0.5 nm steps on a Jasco J-810 spectropolarimeter at 20 °C in a quartz cuvette of 1 mm path length. Raw data were converted to molar ellipticity (mdeg M^{-1}

cm^{-1}) and the secondary structure content was analyzed by DICHROWEB [125].

2.2.4.4   Complex characterization by ITC

Isothermal titration micro-calorimetry (ITC) is a method used for the determination of thermodynamic reaction parameters in solution. For this, two separated reaction partners are titrated into each other and the energy that is required to keep the sample cell at a constant temperature during the reaction is measured in comparison to a reference cell filled with an identical solvent.

The method allows the direct measurement of reaction stoichiometry (n), enthalpy (ΔH) and association / dissociation constant (K_a / K_d). From these parameters the calculation of a reaction's Gibbs' free energy (ΔG) and entropy (ΔS) is possible by application of:

Equation 3: Gibb's free energy

$$\Delta G = -RT \ln K_D = \Delta H - T\Delta S$$

where R is the ideal gas constant (8.314472 J mol^{-1} K^{-1}) and T the absolute temperature (K). The association constant K_a describes the relation between bound and unbound reactants, in case of a reaction with two reactants A and B:

Equation 4: Association constant

$$K_D = \frac{[AB]}{[A][B]} = \frac{1}{K_D}$$

In this work a VP-ITC calorimeter (Microcal / GE Healthcare) with a cell volume of 1.42 ml and an injection syringe volume of 273 µl was used to determine the thermodynamic parameters of protein – protein interactions.

Measured raw data were integrated and reaction parameters were calculated by non-linear curve fitting with the Origin^{TM} software.

Measurements of protein – protein association: Before the experiment both complex partners were dialyzed over night at 4°C against PBS pH 7.2 with 1 mM Na-TCEP and protein concentrations were adjusted by dilution with the original buffer. Binding energies were measured by titration of a least ten-fold concentrated (300 – 700 µM) ligand protein in steps of 10 µl into its binding partner (30 – 70 µM) at a constant temperature of 25°C.   All experiments were repeated in triplicate to allow the

calculation of standard deviations. Dilution effects caused by the injection of the ligand protein were probed by injection of the ligand into the original buffer. In case significant effects were observed, the dilution heat was subtracted from the original measurements.

2.2.4.5   X-ray crystallography

X-ray crystallography was used as a tool to determine the three dimensional structure of the proteins and protein complexes presented. Although other techniques exist for protein structure determination, these techniques suffer from certain disadvantages like protein size limitation (NMR), limited resolution (cryo EM) or strong effects on sample integrity (electron crystallography). X-ray crystallography in contrast allows the determination of a crystallized macromolecule's structure of, in principle, unlimited size and to very high resolution with only minor destructive interference with the sample itself. Major drawbacks of this well developed method lie in the necessity of potentially time consuming crystallization and the fact that the molecule's structure is determined in a crystalline state. This makes the confirmation of observed structural features by complementary methods mandatory.

A detailed description of the theory and practice behind X-ray crystallography is beyond the scope of this thesis. Therefore only a limited description of diffraction theory and phasing methods, which are relevant to this work will be given. The reader is referred for more detailed explanations to this literature [126], [127] and [128] from which parts of the following explanations are taken.

2.2.4.5.1 Principles

X-ray crystallography uses the elastic scattering of monochromatic X-rays by electrons from the electron cloud of molecules to determine their density within a crystal and from this to derive a three-dimensional model of the crystal's atomic composition.

A crystal is composed of a large number of identical building blocks, the unit cells which by translation in three dimensions fill the complete crystal volume.

When X-rays are scattered by electrons in a molecule, interference occurs between the scattered waves. Scattered waves are only in phase and therefore interfere constructively when certain geometrical conditions are fulfilled which are described by **Bragg's law [129]**:

Equation 5: Bragg's law

$$2d_{hkl}\sin\theta = n\lambda$$

where d is the distance of two parallel planes between which the diffraction occurs, θ the incoming and outgoing angle of the electromagnetic wave relative to the plane normal, n an integer number and λ the wavelength of the electromagnetic wave.

Each unique measured reflection with the reciprocal indices *hkl* represents the sum of electromagnetic waves, which have been scattered by the electrons of all atoms in a crystal lattice. The result of the wave summation is a single wave, the **structure factor** **F**(*hkl*), which can be described as a complex vector that can be calculated by:

Equation 6: Structure factor equation

$$F(hkl) = \sum_{i=1}^{N} f_j(h,k,l) \exp(2\pi i(hx_j + ky_j + lz_j))$$

where F(*hkl*) is the structure factor of the reflection *hkl*, f_j the normalized element specific atomic scattering factor of the atom *j*, N the number of ordered scattering atoms within the unit cell, i=$\sqrt{-1}$ (the imaginary number) and *hx_j, ky_j* and *lz_j* the fractional coordinates of the atom *j*.

Instead of as a function of atomic scattering factors, the structure factor equation can be expressed as a function of the local electron density $\rho(x,y,z)$:

Equation 7: Structure factor equation in dependence of electron density

$$F(hkl) = \int_h \int_k \int_l \rho(x,y,z) \exp(2\pi i(hx + ky + lz)\, dx\, dy\, dz$$

As the structure factor (and so diffraction itself) is the Fourier transform of electron density, the electron density in a unit cell can be calculated from the Fourier transform of the structure factor equation, which leads to the electron density equation

Equation 8: Electron density equation

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F(hkl)| \exp(-2\pi i(hx + ky + lz) + i\alpha(hkl))$$

from which the electron density ρ at each point x,y,z can be calculated if the cell volume

V, the structure factor amplitudes F(hkl) and the structure factor phases α are known.

In a diffraction experiment only the unit cell volume and reflection intensities $I(hkl)$ can be measured. The structure factor amplitude $F(hkl)$ is calculated from the measured reflection intensities $I(hkl)$, the radiation source dependent combined geometry and polarization factor LP and the absorption correction factor A by:

Equation 9: Intensity – amplitude conversion

$$|F(hkl)| = \frac{\sqrt{I(hkl)}}{LP \bullet A}$$

The phases φ of the recorded reflections get lost during the experiment. This leads to a loss of the more important part of information on the electron density in a crystal and is designated as the crystallographic **phase problem**.

2.2.4.5.2 Phase determination

Different methods have been developed for phase determination, either by pure mathematical approaches (direct methods; molecular replacement, difference fourier synthesis), by exploitation of a modified experimental setup which allows the calculation of the reflection phases from changes in F(hkl) (isomorphous replacement methods – SIR/MIR/RIP, anomalous dispersion methods - SAD/MAD and their combinations SIRAS/MIRAS) or by combination of both (MR-SAD, phased molecular replacement). From these methods, **molecular replacement** (MR), single / multiple wavelength anomalous dispersion (**SAD/MAD**) and combinations of SAD and molecular replacement were used in this work for the determination of initial phases.

In **molecular replacement** a similar, non-isomorphous structure (the search model) is placed into the unit cell and used to calculate the initial phases. The positioning of the model involves two consecutive steps in which it is iteratively rotated along three axes (rotation function) and three axes of translation (translation function) through the unit cell space followed by the search for additional model positions within the asymmetric unit. The quality of the resulting model is then estimated by calculation of global parameters like the model's **R-factor** which indicates the agreement between the observed structure factor amplitudes F_{obs} and the structure factors F_{calc} calculated from the model and corrected by a scale factor k.

Equation 10: Model R-factor calculation

$$R = \frac{\sum_{hkl} \left\| F_{obs}(hkl) \right| - k \left| F_{calc}(hkl) \right\|}{\sum_{hkl} \left| F_{obs}(hkl) \right|}$$

or the scaling insensitive standard linear **correlation coefficient C**:

Equation 11: Correlation coefficient

$$C = \frac{\sum_{hkl} \left(\left| F_{obs} \right|^2 - \overline{\left| F_{obs} \right|^2} \right)\left(\left| F_{calc} \right|^2 - \overline{\left| F_{obs} \right|^2} \right)}{\left[\sum_{hkl} \left(\left| F_{obs} \right|^2 - \overline{\left| F_{obs} \right|^2} \right)\left(\left| F_{calc} \right|^2 - \overline{\left| F_{obs} \right|^2} \right) \right]^{1/2}}$$

Low R-values and high correlation coefficients indicate a good solution.

In addition geometric plausibility of the packing of the positioned model within the unit cell, e.g. by measurement of contacts between placed molecules are used as indicators of the correctness of the found solution.

Anomalous dispersion methods: The diffraction resulting from elastic scattering of photons is centro-symmetric. This means that Bragg reflections, which are related to each other by inversion through the origin have the same structure factor amplitude but opposite phase angles. These reflections are called Friedel's pairs and their relationship is designated as Friedel's law [130]**.**

Equation 12: Friedel's law

$$\left| F(hkl) \right| = \left| F(\overline{hkl}) \right|$$

$$\alpha(hkl) = -\alpha(hkl)$$

If they have an energy, which is higher than the element dependent **X-ray absorption edge** (**Figure 2-1)**, photons are not only elastically scattered by the atom, but some are absorbed by the electron shell. This absorption excites electrons from inner orbitals and the loss of energy to the original level results in photon reemission (**X-ray fluorescence**). These reemitted photons are retarded behind the elastically scattered photons and influence the scattering of all atoms by addition of a real dispersive component f' and an imaginary anomalous component f'' to each atom's scattering factor f^0 (**Equation 13**).
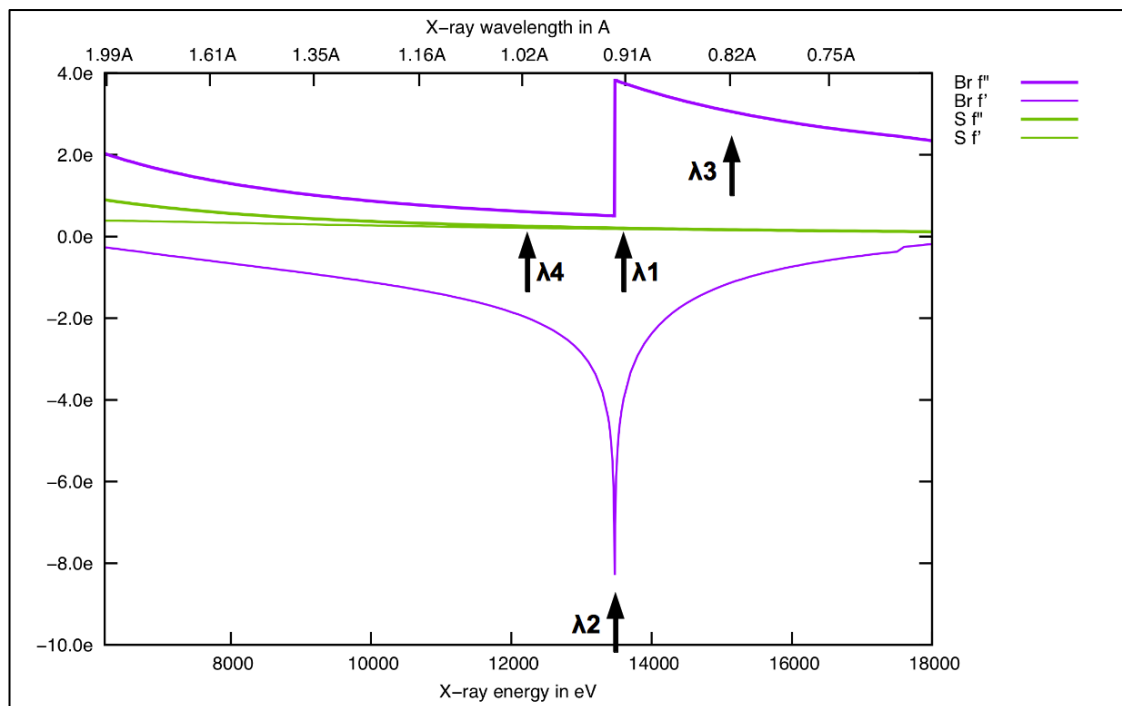
Figure 2-1 Anomalous contributions of sulfur and bromine.

A plot of theoretical values for f' and f" of sulfur and bromine shows, that sulfur has a weak anomalous signal of <1 electron (e) only at long X-ray wavelengths while the absorption edges of heavier elements like bromine can be readily reached at photon energies which are suitable for X-ray diffraction experiments. Black arrows indicate the wavelengths at which datasets for a typical bromine MAD experiment would be collected: λ1 = peak (also for SAD), λ2 = inflection point, λ3 = high energy remote, λ4=lowenergy remote.(Plot generated by http://skuld.bmsc.washington.edu/scatter/)

Equation 13: Cromer Mann formula with anomalous contributions

$$f_{(S,\lambda)} = f_{(s)}^{0} + f_{(\lambda)}^{'} + i \bullet f_{(\lambda)}^{''}$$

They therefore cause a phase shift of waves which are scattered by non-emitting atoms. This results in a break of the phase- and amplitude relationship of the resulting reflection pairs. The influence of anomalous scattering of a single atom on a structure factor F_P can be visualized by an Argand diagram **(Figure 2-2)**.
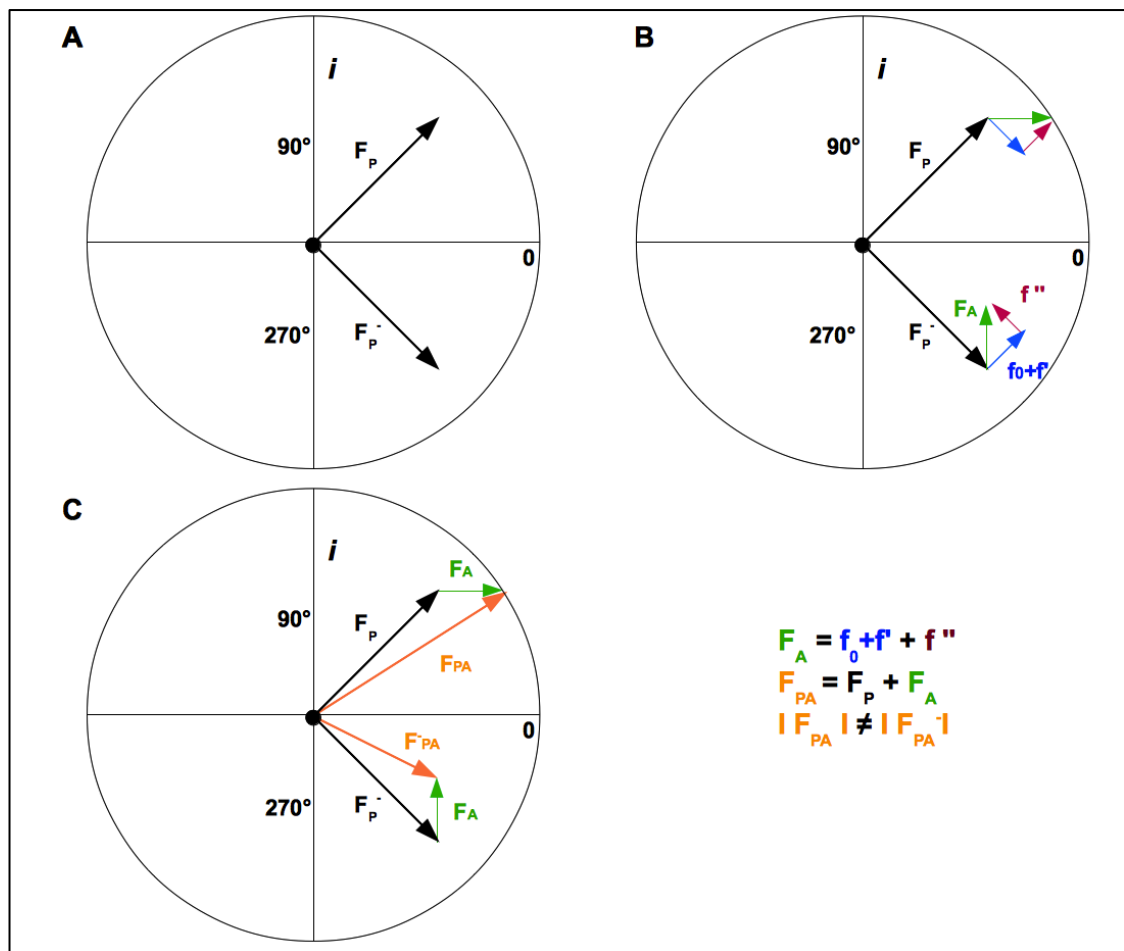
Figure 2-2 Argand_diagram.

(A) According to Friedel's law, centrosymmetrically related reflections have the same amplitude but opposite phase angles. **(B)** If an anomalously scattering atom contributes to the diffraction, its scattering component f_0+f' is mirrored by the real axis so that Friedel's law stays true. Addition of the component f'' which is mirrored by the imaginary axis (always positive!) to f_0+f' creates the anomalous contribution F_A. **(C)** Upon addition of F_A to F_P and F_P^- the resulting vectors F_P and F_{PA}^- loose their relation in magnitude and phase. (Diagram modified from Rupp, 2010)

Amplitude changes can be measured in a diffraction experiment. Therefore the phases of the single reflections $\mathbf{F}^+$ and $\mathbf{F}^-$ can be calculated from the measured amplitude difference $\Delta F = F^+ - F^-$. They can be calculated either from a single anomalous dataset recorded at a wavelength where the anomalously scattering element has large f'' (SAD-phasing) or from several datasets collected at maximum f'' (peak data), minimum f' (inflection data) and optionally high and low energy remote wavelengths (**Figure 2-1**, 2, 3, 4 wavelength MAD phasing [131] [132]). For further details on how phases are practically calculated from anomalous data, the reader may be referred to recent reviews on SAD-phasing [133] and MAD-phasing [134]. Both methods require the presence of a

well-ordered anomalous scatterer in the crystal combined with data collection at corresponding X-ray energy (energies) to achieve measurable large ΔF. Recent reports have shown that phases can be obtained from anomalous scattering of sulfur atoms, which are naturally present in proteins (reviewed in [135]), but the method is still very limited due to technical and physical restraints. Therefore the incorporation of heavier elements showing strong anomalous properties is necessary. This can be achieved by different methods. From these, quick soaking of bromide ions into an existing crystal and co-crystallization with bromide containing salts was used in this study to determine experimental phases by the two wavelength-MAD and the SAD method.

2.2.4.5.3 Estimation of diffraction data quality

The measurement of diffraction intensities is accompanied by the introduction of a variety of experimental errors [136] e.g. by fluctuations of the X-ray beam intensity, variation of the exposed crystal volume or the use of large area CCD-detectors, which measure X-rays indirectly after signal conversion to visible light. These errors make raw intensity data practically unusable for the direct determination of a crystal structure. Intensities of related reflections need to be put on the same scale and the remaining error needs to be estimated. This is done partly during the integration and mainly during the scaling step of diffraction data.

To assess the strength and quality of the obtained scaled diffraction data a number of global descriptors have been introduced and will be used here. The include the signal to noise ratio **[$I/\sigma(I)$]** of the measured intensities and the **multiplicity N**. In addition the following R values provide evidence on the accuracy if the measured intensities.

The symmetry merging R-value R_{merge} [137] is calculated by averaging the multiple intensities measured for the same Miller index ($\langle I(hkl)\rangle$) and its symmetry related mates:

Equation 14: R_{merge}

$$R_{merge} = \frac{\sum\limits_{hkl}\sum\limits_{i}|I_i(hkl) - \langle I(hkl)\rangle|}{\sum\limits_{hkl}\sum\limits_{i}I_i(hkl)}$$

It is dependent on data redundancy and becomes worse the more often a reflection has been measured although in absence of significant radiation damage, the precision of the intensity measurement should increase. Therefore, quality indicators have been

introduced to describe the precision of an individual reflection (R_{meas} [138]) and the averaged reflections ($R_{p.i.m.}$ [139]) independently of how often they have been measured [140].

Equation 15: R_{meas}

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_{i} I_i(hkl)}$$

Equation 16: $R_{p.i.m.}$

$$R_{p.i.m.} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_{i} I_i(hkl)}$$

2.2.4.5.4 Protein crystallization

2.2.4.5.4.1 Principles

In order to determine the three-dimensional structure of a protein by X-ray crystallography, large (>10-50 μm) single protein crystals are required. These can be grown from concentrated solutions of pure protein by different methods all of which aim to bring the protein to a level of supersaturation by controlled addition of precipitating chemicals and/or changes in protein concentration (

Figure 2-3). In this work the vapor diffusion method was used exclusively for crystallization. In this method the protein solution is mixed with a precipitant solution and the mixed drop is stored in an airtight container either hanging on the container lid ("hanging drop vapor diffusion") or sitting on an elevated pedestal ("sitting drop vapor diffusion") over a large volume of the precipitant solution ("mother liquor"). The difference in precipitant concentration between crystallization drop and mother liquor leads to exchange of solvent by evaporation until an equilibrium of vapor pressure is reached. The resulting increase of protein and precipitant concentration in the drop leads eventually to the formation of protein crystals. In addition to traditional sitting / hanging drop vapor diffusion, the vapor batch method [141] was used for optimization trials. In this method, the sitting crystallization drop is covered by a partially water permeable oil

which slows down evaporation. The drops are stored in the absence of a mother liquor reservoir in an unsealed container. This leads to slow continuous evaporation of the solvent beyond an equilibrium point at which usually diffusion stops in traditional vapor diffusion experiments.
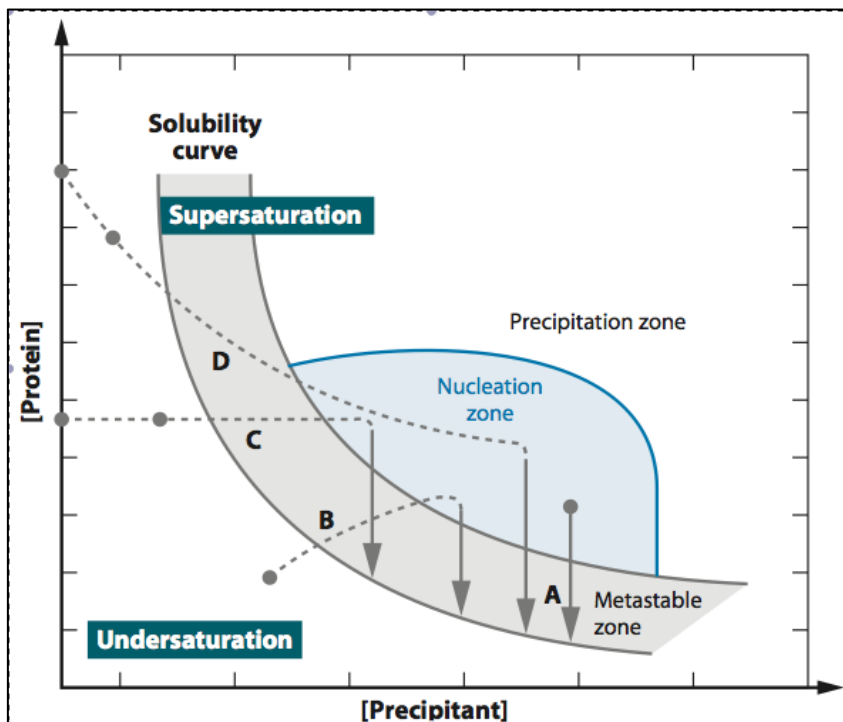

Figure 2-3 Crystallization phase diagram.

Proteins crystallize from nuclei in a phase of supersaturation which can be reached by different methods like Microbatch (A), Vapor diffusion (B), Microdialysis (C) or free interface diffusion (D). Note, that after the nucleation zone is reached (dashed lines), the concentration of protein in solution decreases (solid arrows) which allows the passage to the supersaturated phase. Image taken from [142]

2.2.4.5.4.2   Initial screening, crystal reproduction and optimization

Initial screening for crystallization was carried out at the high throughput crystallization facility at EMBL Hamburg [143]. Before submission of the sample an appropriate protein concentration was estimated with the pre-crystallization test (Hampton research). Proteins were initially screened against 192 – 480 different crystallization solutions using the sitting drop vapor diffusion method with drop sizes of 600-800 nl and at

constant 20°C.

Once one or several initial crystallization hits were identified, manual trials of crystal reproduction and optimization were carried out by variation of different parameters such as crystallization setup (hanging/sitting drop vapor diffusion; under oil vapor batch crystallization), crystallization temperature (20°C/4°C), drop volume and shape and modifications of the crystallization condition (pH, precipitant concentration, additives). Detailed descriptions on the crystallization procedures of individual proteins are given in the results section.

2.2.4.5.5 Data collection, structure solution and refinement

Diffraction data were collected at several synchrotron radiation beamlines at EMBL/DESY (Hamburg, Germany) and ESRF (Grenoble, France) **(Table 2-9)**. Crystals were harvested from the crystallization drop, cryo-protected in an appropriate solution (mother liquor mixed with 20-25% glycerol, 20-30% ethylene glycol or PEG200), mounted on a nylon- or Mitegen-loop and flash-frozen either directly in the cryo-stream (diffraction experiment at EMBL/DESY) or liquid nitrogen (diffraction experiment at ESRF).

Table 2-9 Synchrotron beamlines used for diffraction experiments.

Synchrotron	beamline	wavelength	detector
ESRF	ID14-1	0.934 Å	ADSC q210
	ID23-2	0.873 Å	mar mosaic 225
	ID23-1	tunable	ADSC q315r
EMBL/DESY	X12	tunable	mar mosaic 225
	MX2	tunable	Rayonix 225 HE

All datasets were collected at a temperature of 100 K using the single axis rotation method. After the collection of initial diffraction images, the crystal was indexed with MOSFLM [144] and BEST [145] was used for calculation of an appropriate collection strategy. Data were integrated with XDS [111]. Laue and space group of the crystal were determined with POINTLESS [102] and the integrated data were scaled with XSCALE [111] or SCALA [146] .

Initial phases for the structure solution of different proteins were either determined experimentally by multiple/single wavelength anomalous dispersion (MAD/SAD)

methods [147] or by molecular replacement (MR) using PHASER (CCP4 or Phenix module [148]), MOLREP [149] or BALBES [100]. Details of the applied phasing procedures will be given in the following sections. Starting from the initial phase information, a model was built with ARP/wARP [98] and then manually completed with COOT [101]. Structure refinement was done in phenix.refine [107]. Groups for the refinement of TLS parameters were determined using the TLSMD web server [150]. Molprobity [151] and phenix.model_vs_data [152] were used for model validation.

2.2.4.5.6 Structure analysis and visualization

Secondary structural elements were assigned with the STRIDE program [153]. Structural alignments were calculated with DaliLite [104]. Electron density maps were calculated with the phenix.maps tool. Model images were created with Pymol (open source version 1.4.1).

For the calculation of surface electrostatic potentials, protonation states of the models were estimated with PROPKA [108] for a pH of 7.0 and the resulting files were converted to the PQR format with PDB2PQR [106]. Eletrostratic potentials were then calculated with APBS [97] under utilization of the AMBER forcefield for a temperature of 37°C and an ionic strength of 150 mM NaCl. Surface potentials were then visualized with the APBS Pymol plugin.

3 Results and Discussion

3.1 Characterization of the human titin – OBSL1 / obscurin head – to – tail complex

3.1.1   Results

3.1.1.1   Titin M10 and obscurin/OBSL1 Ig1 form a binary complex

The three complex partners titin domain M10, OBSL1 Ig1 and obscurin Ig1 could be over-expressed as separate proteins in large amounts (>50 mg per liter of culture) in *E. coli* and purified by IMAC and SEC to a high purity. Complex formation between titin M10 and OBSL1 Ig1 / Obs Ig1 was confirmed by analytical size exclusion chromatography. The mixture of titin M10 with equal stoichiometric amounts of OBSL1/Obs Ig1 just prior to the injection on an analytical column led to a significant decrease of approximately 1 ml in the observed elution volume when compared to the volumes observed for single proteins **(Figure 3-1)**. This shows that the proteins associate spontaneously and form stable binary complexes.
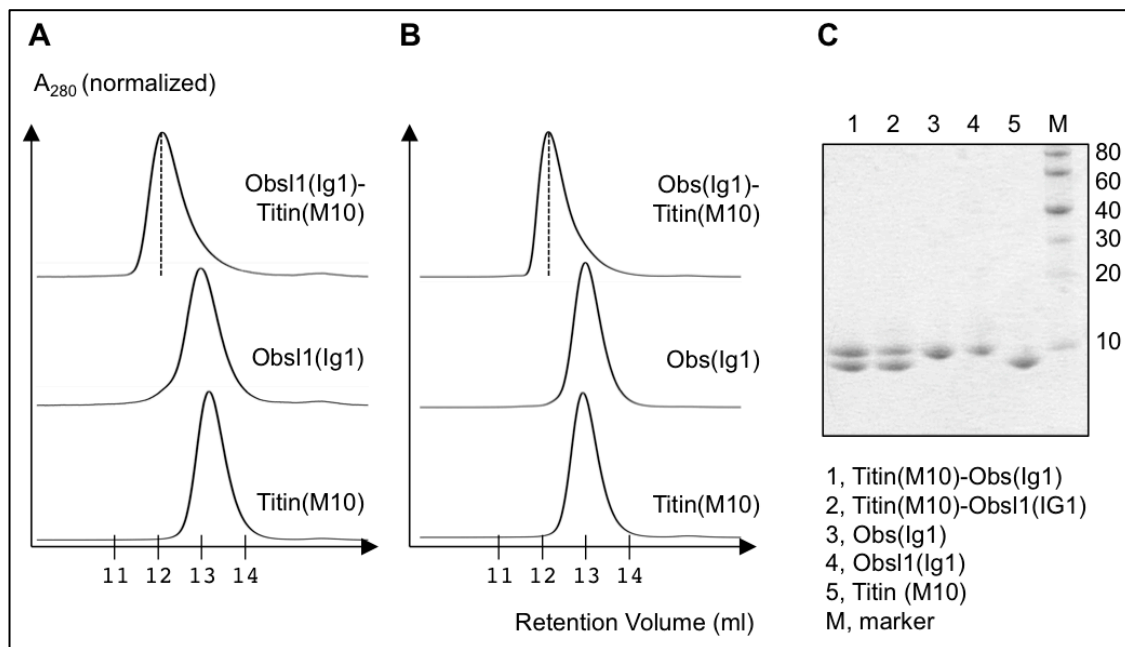

Figure 3-1 Confirmation of complex formation between titin M10 and OBSL1/Obs Ig1.

 (A)/(B) Elution profiles of complexes between titin M10 and OBSL1 / obscurin Ig1 and separate domains injected on an analytical size exclusion chromatography column (Superdex 75 HR 10/300). **(C)** Peak analysis. Fractions collected from complex peaks contain both complex partners; 4-12% SDS-PAGE; Coomassie stained.

3.1.1.2 Crystallization

The OBSL1 Ig1 – titin M10 and obscurin Ig1 – titin M10 complexes, that had been formed by mixture of the complex partners in a 1:1 molar ratio, were submitted for crystallization trials at a concentration of 40 mg/ml. Crystals of the obscurin Ig1 – titin M10 complex could not be grown. A condition with fine crystalline precipitate was identified in bicine pH 9.0, 1.6 M $(NH_4)_2SO_4$ but could not be optimized to yield crystals of a size suitable for diffraction experiments.

Crystals of the OBSL1 Ig1 – titin M10 complex grew within 1-7 days in a variety of conditions at pH 6.0 – 9.0 and in different precipitants like 20 – 30 % (w/v) PEG1000 / 2000MME / 5000MME / 3350 / 4000, 1.6 – 2.4 M $(NH_4)_2SO_4$ / Li_2SO_4 or 25 – 50% v/v MPD. All crystals were of a similar pyramidal shape, but the ones that had grown in PEG or MPD and/or at low pH tended to be crooked and showed visible defects such as an irregular surface and inclusions. Crystals grown in $(NH_4)_2SO_4$ at pH 8.0-9.0 were in contrast sharply edged and grew up to a size of more than 1 mm **(Figure 3-2 C)**. The crystals used for the final diffraction experiments had been grown by hanging drop vapor diffusion in 0.1 M imidazole pH 8.0, 2M $(NH_4)_2SO_4$ and had a size of ca. 0.5 mm in all dimensions.
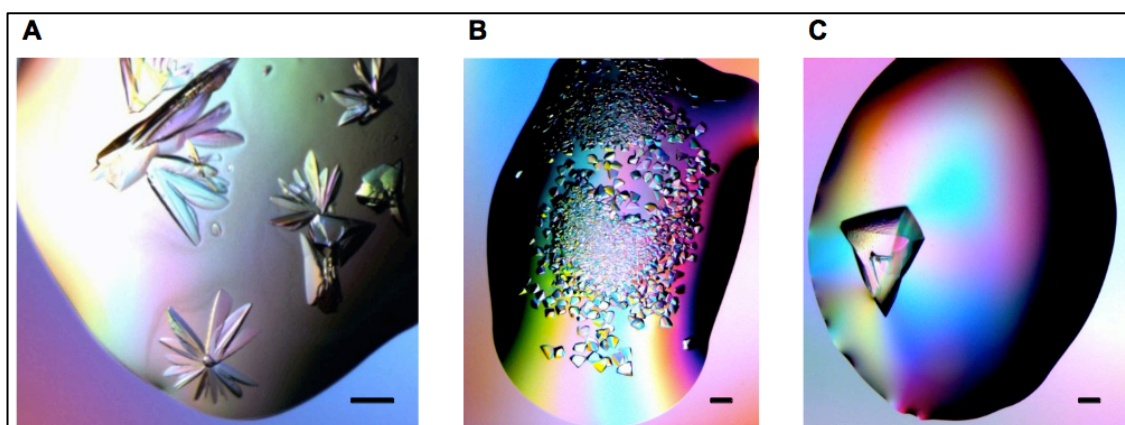

Figure 3-2 Crystallization of the titin M10 – OBSL1 Ig1 complex.

(A) crooked crystals grown in MES pH 6.5, 20% PEG2000 MME; **(B)** small crystals grown in MES pH 6.0, 2M $(NH_4)_2SO_4$; **(C)** large crystals grown in imidazole pH 8.0, $(NH_4)_2SO_4$; The black bars indicate a length of 0.2 mm;

Diffraction studies on different crystal forms confirmed that all crystals belonged to the same space group with similar unit cell parameters but differed in maximum attainable resolution and crystal mosaicity. They were therefore considered as identical crystal forms.

3.1.1.3   Data collection and structure solution

Native data were collected to a maximum resolution of 1.4 Å on the DORIS beamline X12 from crystals that had been cryo-protected in mother liquor with 20% (v/v) glycerol. Due to the strong intensity of low resolution reflections the data were collected in two separate high- and low-resolution data sets, which were independently integrated with XDS and then scaled together with XSCALE. Trials to solve the complex structure by molecular replacement with different available structures of I-set Ig-like domains of titin were not successful. The crystals were therefore soaked for ca. 1 min. into a drop containing the cryo-protectant solution (0.1M imidazole pH 8.0, 2M $(NH_4)_2SO_4$, 20% glycerol) supplemented with 0.5 M LiBr. The anomalous signal at the wavelengths of the peak and inflection point of the bromide absorption edge was then exploited to solve the structure of the complex in the spacegroup $P3_1$, using the two wavelengths MAD method. Two datasets at the Br-peak and inflection point wavelengths were collected at beamline X12, integrated with XDS and submitted to the Auto-Rickshaw structure solution pipeline. A nearly complete structural model of the complex was obtained from the Auto-Rickshaw / ARP/wARP server. It was then used to solve the structure from the native high-resolution dataset by molecular replacement using PHASER. The model was manually completed in COOT and individual coordinates and anisotropic/isotropic B-factors were refined with phenix.refine. Diffraction data, phasing and refinement statistics are given in **Table 3-1**.

Table 3-1 Data collection, phasing and refinement statistics of the titin M10 – OBSL1 Ig1 complex.

	Native	Br-derivative	
		Inflection	Peak
Data collection and processing:			
Beamline		X12	
Detector		Mar mosaic 225	
Wavelength [Å]	0.99184	0.91706	0.91740
Resolution range [Å]	1.40 (1.44 – 1.40)	1.90 (1.95 - 1.90)	1.69 (1.73 – 1.69)
Space group		$P3_1$	
Unit cell dimensions [Å]	a/b = 60.9, c = 42.1	a/b = 61.2, c = 42.8	
Mosaicity (°)	0.26	0.21	0.22
No. of measured reflections	113955	153326	106892
No. of unique reflections	34179	27160	39569
Redundancy	3.3	5.8	2.7
$I/\sigma(I)$	20.6 (3.3)	41.9 (11.6)	27.9 (5.5)
Completeness [%]	99.5 (99.6)	98.0 (82.7)*	98.6 (96.3)*
R_{merge} [%]	2.7 (33.3)	2.6 (13.4)*	4.1 (28.1)*
R_{meas} [%]	3.4 (38.5)	3.0 (14.0)*	2.9 (22.5)*
Overall B-factor from Wilson plot [Å^2]	20.0		

(* : calculated from separate I^+ / I^-)

numbers in parentheses correspond to the highest resolution shell

Phasing			
Figure of merit (SHELXE)		0.709	
Pseudo free CC (SHELXE)		72.39	
Number of sites (SHELXD)		15	

Refinement statistics:	
No. of reflections (free)	34177 (1711)
R_{work} / R_{free} [%]	16.4 / 19.5
No. of atoms: protein / ions / water	1585 / 5 / 266
Average B-factor protein / ions / water [Å^2]	26.4 / 68.5 /40.8
R.m.s. deviations	
Bond lengths [Å]	0.006
Bond angles [°]	1.1
Ramachandran plot	
Favored [%]	95.04
Additionally allowed [%]	4.96

3.1.1.4 Overall structure

The crystallographic model contains one copy of each protein monomer, one sulfate ion and 266 water molecules in the asymmetric unit. The two domains are oriented in an anti-parallel way in the asymmetric unit. Assignment of their secondary structure using the DSSP algorithm [154] confirms the primary structure based predictions that both OBSL1 Ig1 and titin M10 are folded as Ig-like domains of the I-set fold class [155]. The ß-strand A, which is continuous in C-set Ig-like domains is divided into two sub-strands A and A' because of the presence of a Pro in its center (Pro 12 in M10; 20-ProArgPro-22 in OBSL1 Ig1) and crosses over between the two ß-sheets ABED and A'GFCC' [155] **(Figure 3-3, A and B)**. A C" strand as found in domains of the V-set is not present. In contrast to canonical domains of the I-set, the strand D is fragmented of OBSL1 Ig1 by the presence of Pro 63 into two short sub-strands D (residues 60-61) and D' (residues 64-65). Both strand fragments contribute to the formation of the same ß-sheet ABEDD' by hydrogen-bonding with strand E. M10 and OBSL1 Ig1 share only 22% sequence identity (21 residues) with the highest identity level in the strand B - C connecting loop, strand C and C' (consensus: GEPXPXVXWXXGG). Despite this rather low similarity level, the two domains are highly similar at the structural level and can be superimposed with an r.m.s. deviation of 1.2 Å over the backbone Cα atoms of 95 aligned residues **(Fig 3-3 C)**.
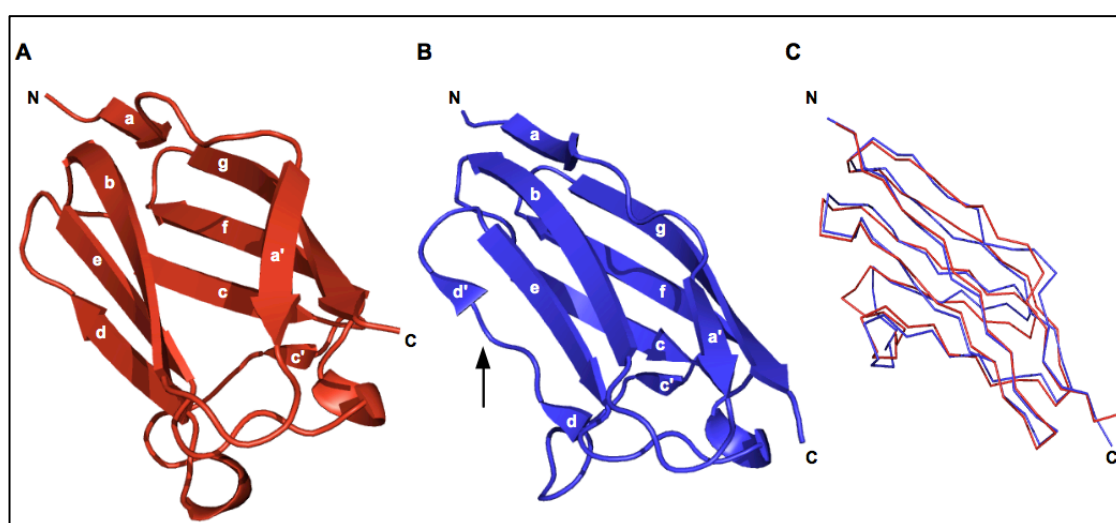

Figure 3-3 Structures of titin M10 and OBSL1 Ig1.

(A)/(B) Ribbon representation of of titin M10 (A) and OBSL1 Ig1 (B). The position of Pro63 in OBSL1 Ig1 dividing the D-strand in two sub-strands d and d' is marked with an arrow. (C) C-α trace of the superposed domains. Red: titin M10; blue: OBSL1 Ig1;

3.1.1.5   Complex formation by intermolecular ß-sheet formation

A single significant complex interface could be identified with PISA [109]. This interface is found between the two Ig-like domains of the same asymmetric unit and creates an anti-parallel V-shaped titin M10 and OBSL1 Ig1 complex. An average surface of 670 $Å^2$ is covered on each domain by the complex partner. The core of the interface consists of a number of hydrogen bridges which form between main-chain atoms of the N-terminal portions of the parallel oriented strands B in M10 (residues 22 - 26) and G (residues 92 – 96) in OBSL1 Ig1 resulting in the formation of an asymmetric intermolecular ß-sheet DEB – G(A')FCC' **(Figure 3-4)**.

Two additional hydrogen-bridges exist between the hydroxyl group of OBSL1 Tyr 96 and main-chain atoms of M10 Ser 16. Two salt-bridges M10 Asp 61 – Ig1 Arg 21 and M10 Arg 98 – Ig1 Glu 94 that are located at opposite domain ends complete the interface. Side- and main-chain interactions as well as complexation parameters as calculated from the model by PISA are ssummearized in **Table 3-2**. Surface potential calculations show that the portions of the M10/OBSL1 strands involved in the formation of the intermolecular sheet create the only electrostatically neutral surface patches on both domains. Peaks of opposing potentials are present at the site where polar interactions are formed between M10 Asp 61 and OBSL1 Arg 21 while the calculated potentials at the sites forming the second salt bridge are rather low **(Figure 3-4 C)**

Figure 3-4 Structure of the titin-OBSL1 complex.

(A) Ribbon representation of the complex in top and longitudinal view. Titin M10 and OBSL1 are colored in red and blue respectively. Interacting parts of the parallel strands B and G in M10 and Ig1 are colored in yellow. The resulting intermolecular ß-sheet DEB-G(A')FCC' in orange. Side-chains of the residues involved in formation of the sheet proximal (M10 Arg98 – OBSL1 Glu94) and distal (M10 Asp61 – OBSL1 Arg21) salt bridges are shown as sticks **(B)** Detailed view on the interfacing backbone residues M10 V22-A26 and Ig1 Ala92-Tyr96. The dashed lines indicate backbone atom mediated hydrogen bonds. The surrounding 2Fo-Fc electron density map is contoured at 2σ (0.75 e⁻/$Å^2$). **(C)** Electrostatic surface potentials calculated for pH 7.0 and 150 mM NaCl. Interfacing strands marked by ellipsoids show a nearly neutral local potential. The location of residues involved in polar interactions is marked with arrows. (Figure made with Pymol and APBStools. Coloring scale in C from -5 to +5 kT).
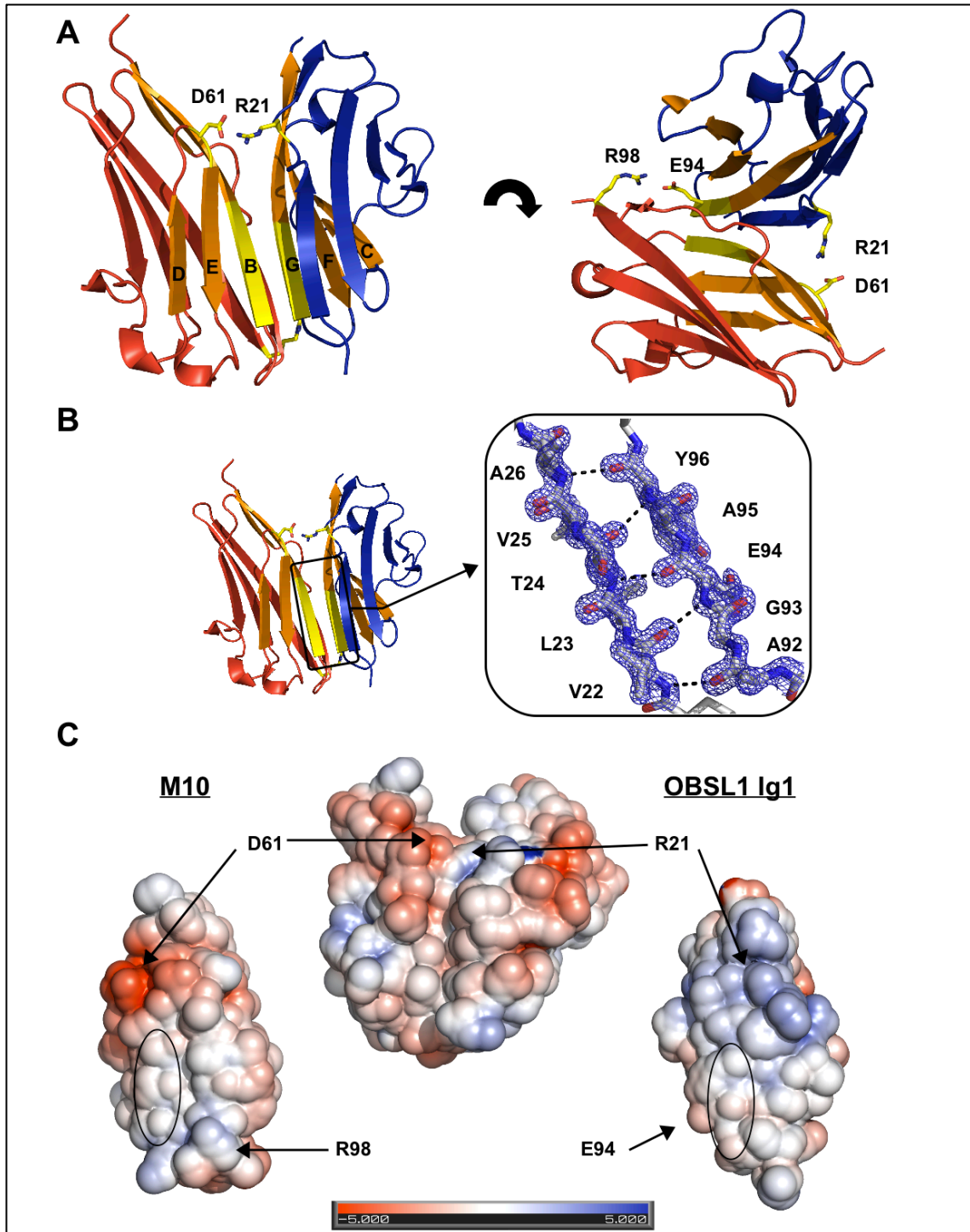
Table 3-2 Complexation parameters.

	Titin M10	OBSL1 Ig1		
Global parameters (calc. by PISA)				
Interfacing residues (no.)	23	21		
Interfacing atoms (no.)	74	69		
Buried solvent accessible area ($Å^2$)	649.9	687.9		
Gain of solvation energy upon complexation (kJ/mol)	-19.7	-8.8	**Total:**	-28.5

Main chain H-bridges					**Distance ($Å$)**
	Val 22	N	O	Ala 92	3.09
		O	N	Glu 94	2.90
	Thr 24	N	O		2.93
		O	N	Tyr 96	2.87
	Ala 26	N	O		3.18

Main / side chain H-bridges					
	Ser 16	O	OH	Tyr 96	2.95
		N	OH		3.24

Salt bridges					
	Asp 61	OD1	NH1	Arg 21	3.01
		OD2	NH1		3.74
		OD1	NH2		3.70
		OD2	NH2		3.93
	Arg 98	NH2	OE1	Glu 94	3.56
		NH1	OE2		3.15
		NH2	OE2		3.32

3.1.1.6 Characterization of the interface and disease related mutations

3.1.1.6.1 Homology between OBSL1 and obscurin

The structure of the complex between M10 and obscurin domain Ig1 could not be solved. However the close evolutionary relationship of obscurin and OBSL1 [22] suggests that the assembly of both head domains with the tail domain of titin follows the same principles. If so, their primary sequence should be similar in regions that are important for the association. Indeed, a Needleman-Wunsch global sequence alignment (parameters: match +2, gap opening: -2, gap extension -1) [156] between OBSL1 and obscurin Ig1 shows that the two domains are identical in 46 of 107 aligned residues (43% identity). While low identity levels are observed in regions contributing to the formation of the (non M10 binding) sheet **ABED**, the identity level is much higher in the regions, which contribute to the formation of the Ig sheet **C'CFG**. More specifically, within the strand F and the M10-binding part of strand G (OBSL1 residues 82-98), only 3 out of 17 residues are not strictly conserved in obscurin **(Figure 3-5)**. Among the OBSL1 residues, which contribute to the intermolecular sheet formation, only Tyr 96 is not identical in obscurin but replaced by a Phe. Hydrogen bonding with Ser 16 on M10 through a side-chain hydroxyl group is therefore impossible in obscurin.

Both positions found in OBSL1 to form polar contacts with M10 are conserved in obscurin. Glu 94 in OBSL1 is conserved in obscurin as Glu 92, while OBSL1 Arg 21 is replaced by Lys 19 in obscurin. The formation of salt-bridges between obscurin Ig1 and M10 is therefore possible in a similar way as in the structurally characterized complex.
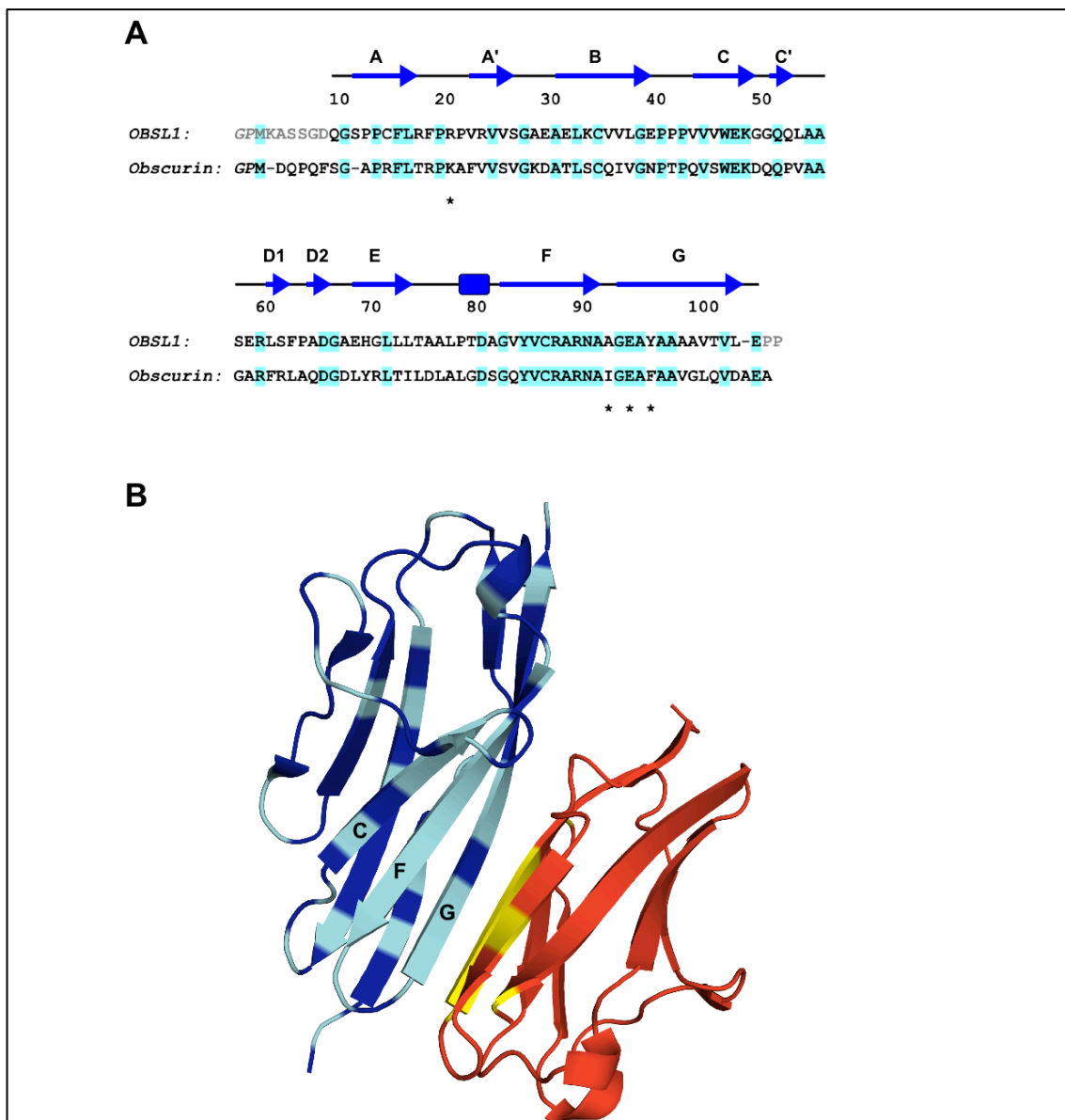
Figure 3-5 Homology between OBSL1 and Obscurin Ig1.

(A) Sequence alignment of human OBSL1 and obscurin Ig1. Identical residues are colored in cyan. The positions of secondary structure elements are shown on top. Residues contributing to the assembly are marked with a star. **(B)** Mapping of identical residues on the structural model: The positions of identical residues are shown in cyan. They are mostly located in regions that contribute to the formation of the complex forming sheet C'CFGA' with the highest levels identity in close proximity to the complex interface. Interfacing residues on M10 are colored in yellow.

3.1.1.6.2 Calorimetric characterization

The OBSL1/obscurin Ig1 – M10 complexes are hetero-dimeric and all partners could be purified as separate proteins. This provided the unique opportunity to characterize the thermodynamic aspects of Ig-Ig complex formation by strand addition and to validate the observed interface by isothermal titration calorimetry (ITC).

The measurements showed that both OBSL1 and obscurin Ig1 bind M10 with an affinity of about 1 μM at standard conditions (PBS pH 7.2, 1 mM TCEP, 25 °C). Titration of the complex partners led to a cooling of the sample cell. This indicates that an increase in the system entropy is driving the association reaction rather than a decrease in system enthalpy. The increase of entropy for the obscurin-M10 association is more than twice the one of OBSL1-M10 and the same is true for reaction enthalpy ΔH. The calculated resulting free enthalpy ΔG of the both associations lies with approximately -34 kJ/mol close to the free enthalpy of complexation calculated based on the crystallographic model (-28.5 kJ/mol).

To assess the importance of polar interactions and backbone hydrogen bonding for the complex formation, M10 variants carrying point mutations were produced. Their correct folding was confirmed by CD-spectroscopy. Charge inversion led to no significant effect when the interface proximal salt bridge was modified (variant M10 R98E) but to a ca. 7 or ca. 30 fold decrease in the binding affinity while the free enthalpy remained the same when the formation of the interface distal bridge was suppressed (variant M10 D61R). Additional M10 variants in which single residues which contribute to the formation of the intermolecular ß-sheet were changed to a Pro residue were also tested for binding toobscurin/OBSL1. Such mutations (variants M10 V22P and T24P) are expected to prevent the peptide amide hydrogen bonding and indeed completely abolished the ability to associate with both complex partners. The measured thermodynamic parameters are listed in **Table 3-3**.

3.1.1.6.3 Characterization of disease related mutations

Next to several mutations that lead to the absence of a M10 domain on the titin protein due to introduction of stop codons 5' of the M10 coding exon Mex6, four missense-mutations on human M10 are known. These are associated with the development of autosomal dominant tibial muscular dystrophy (TMD; OMIM #600334) and recessive

limb girdle muscular dystrophy type 2J (LGMD2J; OMIM # 608807) **(Figure 3-6)**. The more severe childhood onset phenotype LGMD2J originates from an 11 bp exchange [157], which alters both length and composition of the M10 C-strand (35-GluValThrTrpSer-39 → 34-ValLysGluLys-38). More specifically it results in a shortening of the strand by one residue and at the same time the replacement of the core Trp by charged residues. An aromatic residue in this position is found in virtually all Ig-like domains as essential part of the domain's hydrophobic core [158]. It can therefore be safely assumed that the LGMD2J M10 variant will not fold correctly due to disturbance of hydrophobic core formation. The effect of the remaining three M10 variants is less obvious from the primary and tertiary structure. They originate from single nucleotide mutations and cause the exchange of single residues on protein. The two mutations found in Belgian [159] and Italian [160] patients with late onset TMD result in the exchange of Ile 56 to Asn and His 55 to Pro, respectively.
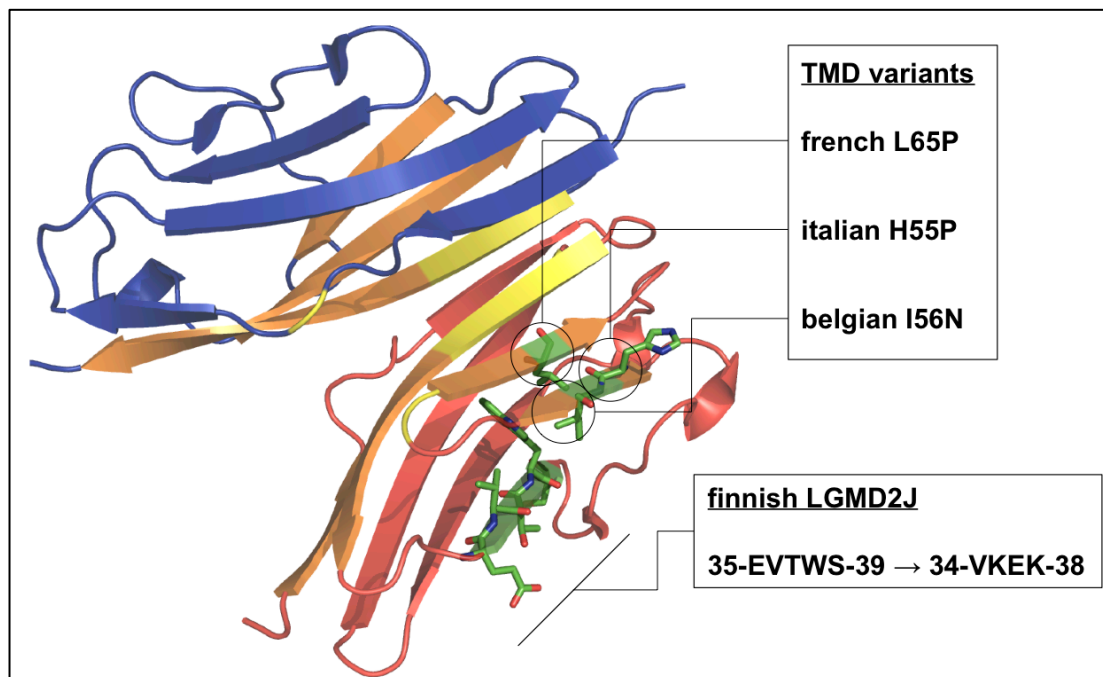

Figure 3-6 Location of residues found to be altered in patients with TMD and LGMD2J.

 (Same representation and color code as in figure 3.4)

Both residues are part of M10 strand D, which contributes to the formation of the OBSL1/obscurin interacting sheet ABEG but is located on the side opposite of the complex interface. The third TMD mutation, which was found in a French family [157]

leads to an exchange of Leu 65 with Pro. The mutated site is located in the center of M10 strand E and therewith in the center of interacting sheet ABED. None of the known single residue exchanges contributes directly to the described complex interface but two of them cause a change to Pro and a local or global effect the domain fold cannot be excluded. The three variants were therefore purified from recombinant source and their folding and ability to bind OBSL1/obscurin Ig1 were investigated by CD spectroscopy and ITC, respectively. The two D-strand variants, H55P and I56N of M10 could be expressed with high yields and purified with SEC profiles comparable to the ones observed for native M10. In contrast, variant L65P could only be expressed in smaller amounts and eluted at a lower volume from a Superdex 75 16/60 pg column compared to the native protein **(Figure 3-7)**.
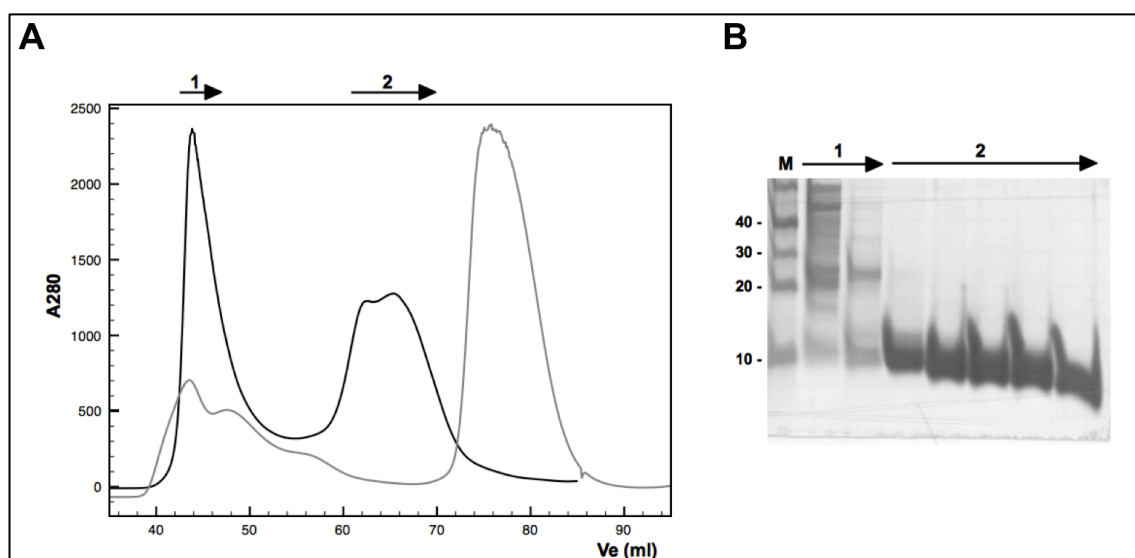

Figure 3-7 Purification of M10 L65P.

 (A) A comparison of the elution profiles of M10 L65P (black profile) and native M10 (grey profile) shows that the disease variant elutes approx. 10 ml earlier from a preparative S75 column. **(B)** Confirmation of L65P purity by SDS-PAGE. Numbered arrows correspond to the analyzed fractions. M = marker (Rotimark 10-150). Numbers on the left side correspond to the marker band mass in kDa.

CD spectroscopy confirmed a virtually identical ß-sheet content for D-strand variant I56N when compared to native M10 (native: 43%; I56N 41%) while, probably due to local disturbance of the strand conformation, it is slightly reduced for the Pro mutant H55P (39%). Both variants can therefore be considered as correctly folding Ig-like domains. In contrast the CD-spectrum of the E-strand variant L65P indicates significant fold changes. A ß-sheet characteristic peak of positive ellipticity at 200 nm wavelength

as observed with native M10, turns negative for L65P, indicating the strong presence of randomly coiled parts in the protein. Thus, this mutation interferes significantly with the folding of M10 **(Figure 3-8)**.
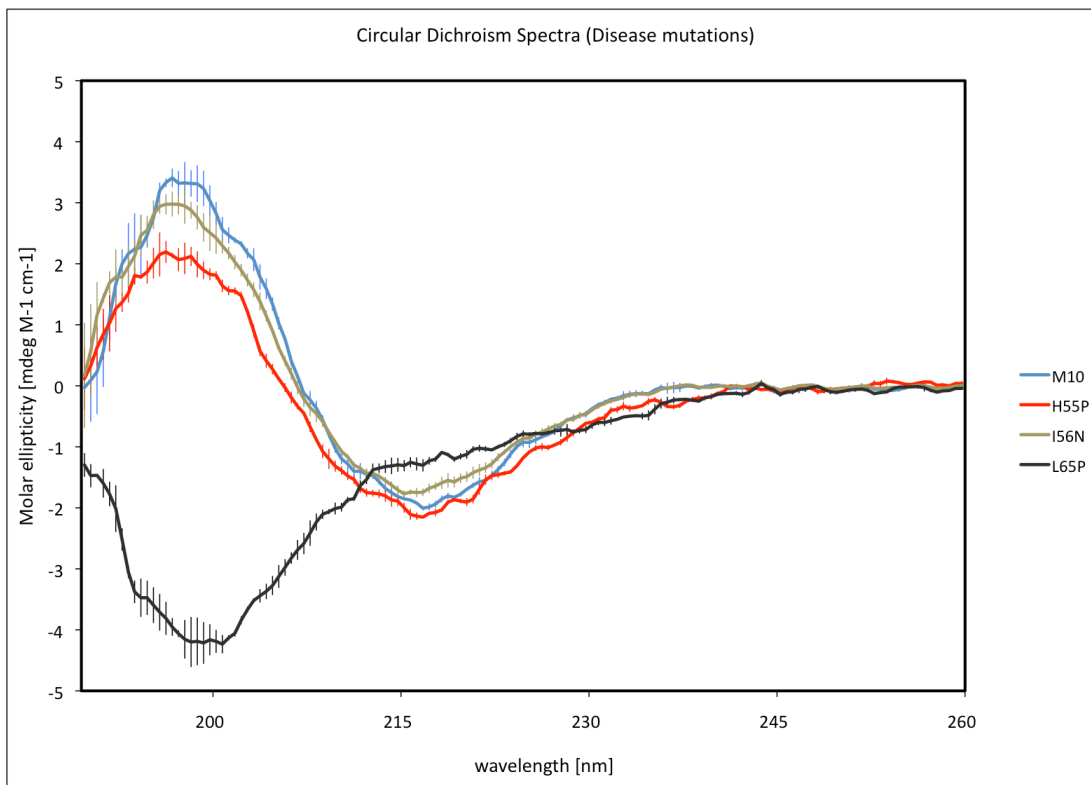

Figure 3-8 Circular dichroism spectra of titin M10 disease variants.

Variants I56N and H55P show spectra which are similar to native M10 while the spectrum for variant L65P shows drastic changes in the domain fold. The vertical bars indicate experimental errors measured from three independent measurements

In agreement with the correct folding only minor changes in the binding affinities for OBSL1/obscurin could be measured for the folded I56N and H55P variants. Binding of the mis-folded L65P variant was in contrast not detectable **(Table 3-3)**
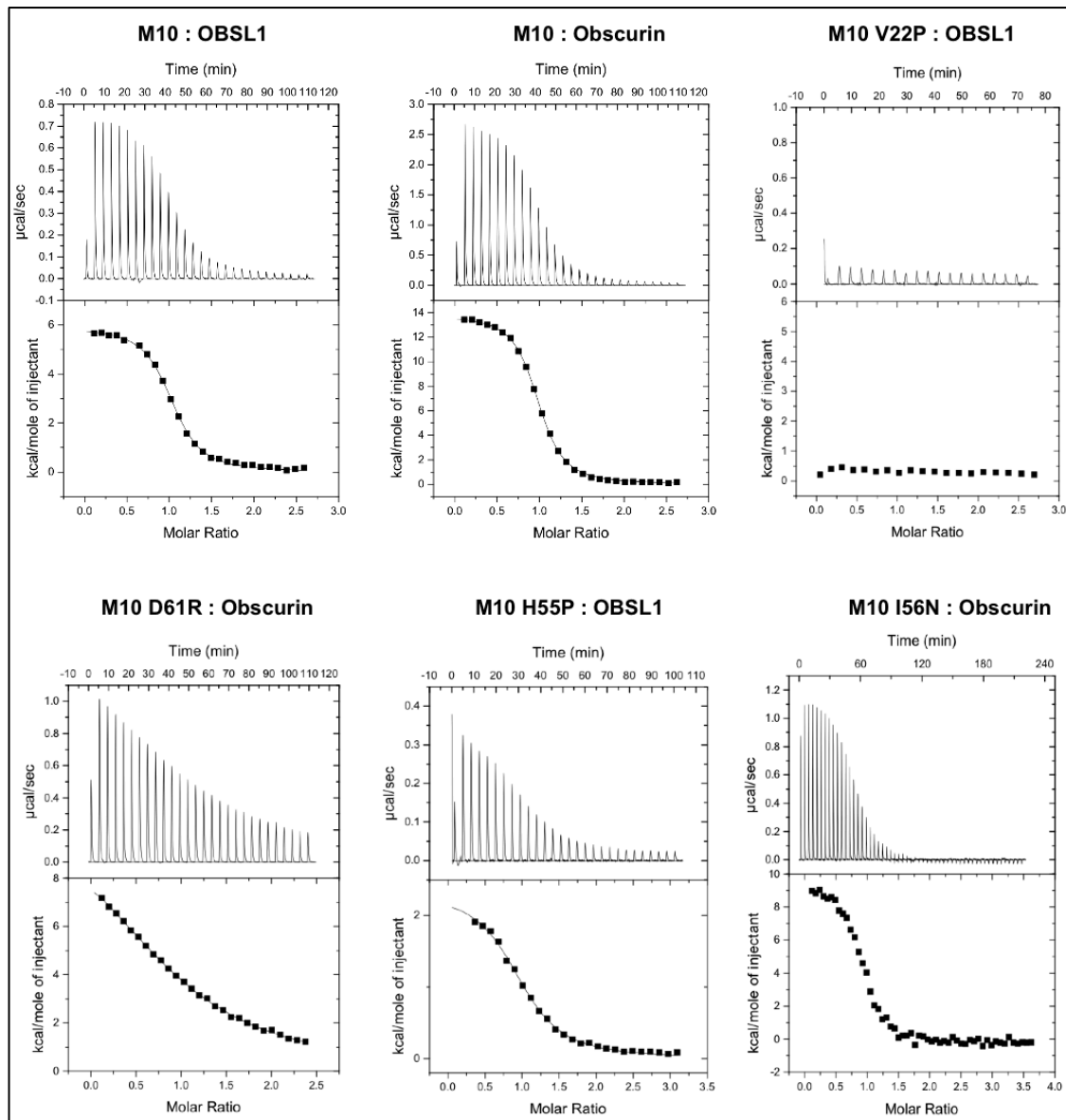
Figure 3-9 Calorimetric characterization of the M10 – OBSL1/Obscurin association.

Raw and integrated data as calculated by the Microcal Origin[TM] software are shown for representative measurement of association reactions between OBSL1/Obscurin (Ig1) and wt M10 functional and disease related variants, respectively. Note the difference in scale of the Y-axis. Although consistent between replicate measurements the enthalpy taken up differs significantly between OBSL1 and obscurin and between different M10 variants (See also **Table 3-3**).

Table 3-3 Thermodynamic parameters determined by ITC.

M10	Binding partner	N	ΔH (kJ/mol)	$-T\Delta S$ (kJ/mol)	ΔG (kJ/mol)	K_D (µM)
native	OBSL1	1.01±0.04	25±0.5	-59.4±0.3	-34.4±0.2	0.94±0.08
	obscurin	0.96±0.01	59.6±1.7	-93.6±1.6	-33.9±0.1	1.16±0.08
Interface mutants						
R98E	OBSL1	0.96±0.05	11.5±0.5	-45.3±0.9	-34.4±0.2	0.92±0.07
	obscurin	1.19±0.05	26.7±1.4	-63.4±1.3	-33-7±0.2	1.22±0.07
D61R	OBSL1	0.95±0.06	12.6±2.0	-42.0±1.6	-29.4±0.6	28.5±4.8
	obscurin	1.01±0.06	54.1±4.8	-80.0±4.4	-25.9±0.4	7.2±1.5
V22P	OBSL1	ND	ND	ND	ND	ND
	obscurin	ND	ND	ND	ND	ND
T24P	OBSL1	ND	ND	ND	ND	ND
	obscurin	ND	ND	ND	ND	ND
TMD disease mutants						
H55P	OBSL1	0.98±0.02	9.6±0.2	-41.5±0.1	-32.0±0.2	2.5±0.2
	obscurin	0.88±0.02	26.9±1.3	-58.9±1.1	-32.0±0.3	2.4±0.3
I56N	OBSL1	0.98±0.01	11.1±0.7	-44.1±0.5	-33.0±0.6	1.5±0.2
	obscurin	0.97±0.03	42.2±1.2	-75.3±1.1	-33.1±0.1	1.6±0.1
L65P	OBSL1	ND	ND	ND	ND	ND
	obscurin	ND	ND	ND	ND	ND

Calculated errors are standard deviations from three independent measurements; ND = not detectable

3.1.2 Discussion

3.1.2.1 A novel variation of a known assembly mode

The structural basis for the terminal domain assembly of the filamentous sarcomeric proteins myomesin-1 [80] and filamin-C [161] as well as their homologs filamin-A [162] and the *D. discoideum* filamin homolog ABP120 [163] has been extensively studied in the past (reviewed in [164]). A striking similarity between all these structurally characterized Ig-like complexes is the formation of homo-dimers through ß-strand addition in a symmetric anti-parallel way. In all cases, one outer strand of each ß-sheet is added to its oppositely oriented homologous strand on the interacting domain to form two continuous intermolecular ß-sheets.

A variation of this binding mode was observed for the N-terminal dimerization of titin [47][46]. Here, the N-terminal half of the non-Ig-like protein telethonin adds one ß-strand to each of the G-strands of the two N-terminal titin Ig-like domains Z1 and Z2 and so forms a bridge between two anti-parallel Z1Z2 tandems. Although, in contrast to the above-mentioned homo-dimeric Ig-assemblies, only one sheet on each Ig-like domain is contributing to the assembly in this case, it is very resistant to externally applied forces [48].

The complex between titin and obscurin-like 1 is not only the first hetero-dimeric but also the first assembly of N- and C-termini of sarcomeric filaments for which high resolution structural data are available. Despite the structural similarity to homo-dimeric Ig-complexes with respect to its domain composition, it revealed some **previously unobserved features**:

(1) As also described for the homo-dimeric filamin and myomesin-1 complexes, the M10 and OBSL1 domains assemble in an anti-parallel way. However, in this case the assembly is mediated only by the N-terminal portions of the parallel oriented strands B (M10) and G (OBSL1) that originate from different ß-sheets of both domains. Due to this asymmetry, backbone hydrogen bonding between additional strands is impossible **(Fig. 3-10 block schemes)**. With an average of 670 $\mathring{A}^2$ of buried domain surface the interface between both domains is of a similar size as the myomesin C-terminal interface (670 $\mathring{A}^2$) but significantly smaller than C-terminal filamin interfaces (984 – 2275 $\mathring{A}^2$) [161] [162] [163].

(2) As a result of this asymmetry, the two Ig-like domains do not form a compact

protein block but their assembly creates a V-shaped complex in which the domain termini pointing to the C-terminus of OBSL1 and the N-terminus of titin are distant from each other.
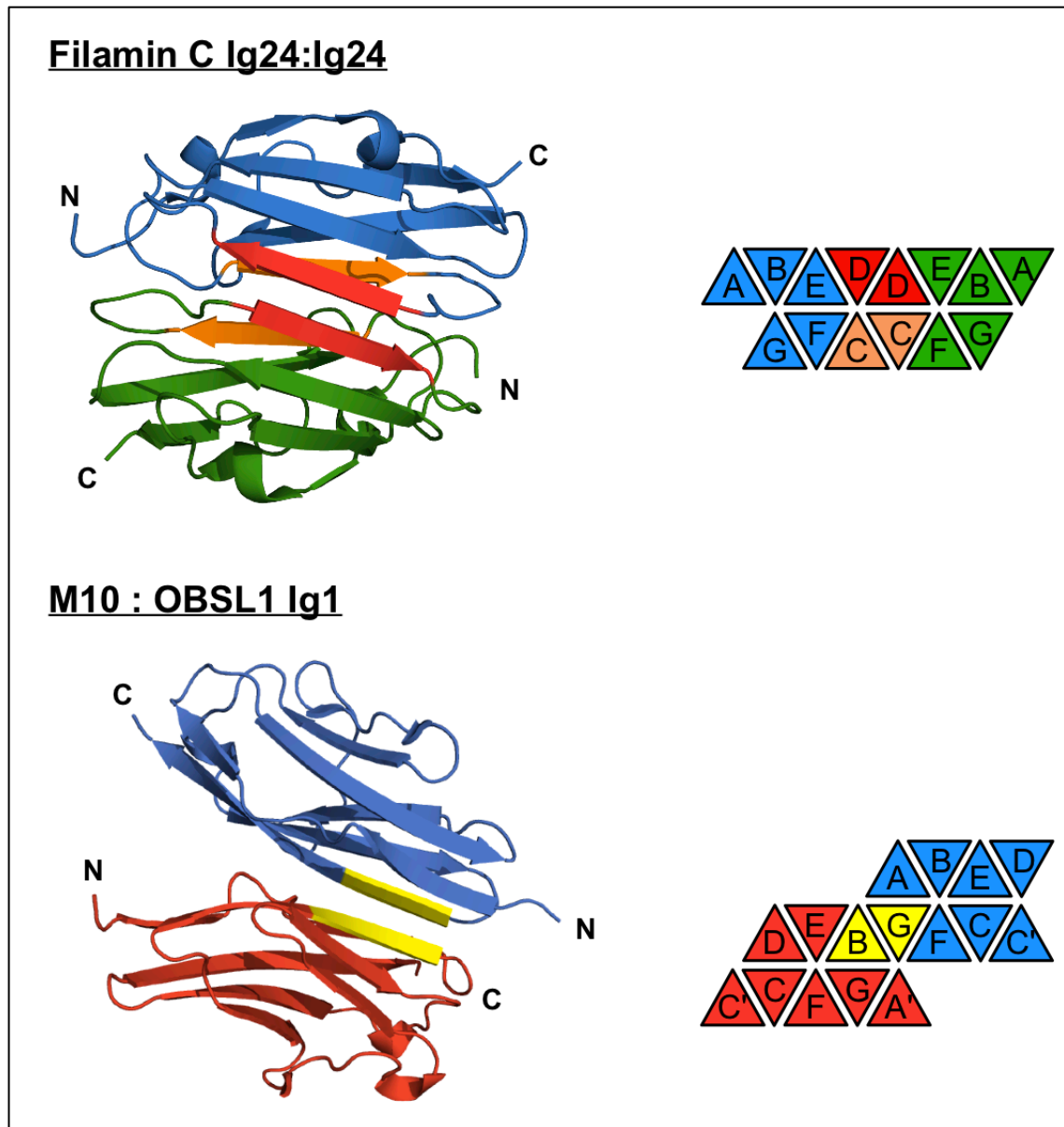

Figure 3-10 Comparison between Ig-like homo- and hetero-dimerization.

Homo-dimerizing Ig-like domains like the C-terminal filamin-C complex shown here interact through association of homologous strands. In contrast, in the titin-OBSL1 complex described here, non-homologous strands from opposite sheets of the domains associate creating a unique asymmetric arrangement. Ribbon diagrams of both complexes are shown on the left side. In filamin-C Ig24, the interacting strands are shown in red and orange, in M10 – OBSL1 in yellow. At the right side a schematic representation of the strand arrangement in both complexes is shown. Color code as in the ribbon diagrams.

Despite these differences, the affinities of the M10 – Ig1 complexes (K_D = 1 µM measured by ITC in PBS) are similar to the ones measured for the dimeric filamin-C Ig24 complex (K_D = 2.5 µM, measured by sedimentation equilibrium in PBS [161]). As the M10-OBSL1 complex could be assembled *in vitro*, its association could be thermodynamically characterized by ITC. The experiments with native domains and variants carrying mutations in the dimer interface suggest that the assembly is driven by the spontaneous displacement of solvent molecules from the surface of the complex partners, which leads to an increase of the system entropy. In contrast, electrostatic interactions play a subordinate role for complexation but they may modulate the association e.g. by facilitating the correct positioning of both partners or increasing the speed of association.

This independence from polar contacts distinguishes the M10-OBSL1/obscurin complex from the homo-dimeric myomesin-1 C-terminal assembly that could be completely suppressed by the disruption of a single salt bridge (confirmed by analytical size exclusion chromatography [80]). The fact that dimerization of titin C- and OBSL1/obscurin N-termini mainly depends on the formation of an ordered H-bonding network may have specific implications on their function for the structural integrity of the sarcomeres. It may indeed make this assembly less sensitive to environmental factors such as the cytosolic pH and ionic strength, which can drastically change in skeletal muscle.

3.1.2.2   A safeguard for titin kinase function?

What consequences could this unique type of terminal arrangement have *in vivo*?
Our knowledge about the structures and functions of the very C-terminal parts of M-band titin, and the N-terminal parts of OBSL1 and obscurin is limited. Primary structure-based domain predictions show that the whole OBSL1 and the N-terminal two thirds of obscurin consist of a tight array of domains of the Ig-like class with only a single inter-domain spacing region of about 25 residues in length located in between domains two and three of both proteins. In muscle tissues, both obscurin and OBSL1 are mainly found at the level of the M-bands. Based on localization studies by electron microscopy, it has been postulated, that OBSL1 localizes to the inner part of the myofibrils, while the much bigger obscurin binds the sarcomeres at the myofibrilar surface (personal communication, Matthias Gautel, Kings College, London) where it

aligns the sarcoplasmic reticulum with the sarcomeres by C-terminal binding to small ankyrin 1 [86][165] and takes effect on the regulation of small Rho GTPases [85][84]. A complete knockout of obscurin in mice was not accompanied by significant changes in sarcomeric structures. However, the sarcomplamic reticulum was altered [165]. The disruption of the obscurin/OBSL1 complex by over-expression of isolated obscurin/OBSL1 Ig1 or M10 or OBSL1 Ig3 (which forms a complex with myomesin domains 4-5), respectively, in neonatal rat cardiomyocytes did only lead to a diffusion of endogeneous obscurin and OBSL1 from the M-band. M-band localization of the partners titin and myomesin-1 were in contrast not affected [60]. A structural role of the complex for the physical linkage of the major M-band components is therefore unlikely. So is the binary titin – obscurin/OBSL1 complex exclusively providing an anchor for obscurin and OBSL1?

One hallmark of titin is its kinase domain located at the border between the sarcomeric A- and M-bands [166]. Based on its crystal structure, this domain has been proposed to require mechanical removal of a C-terminal inhibitory loop for the binding of ATP and activation [51], and indeed further evidence exists that the application of weak forces promotes the binding of ATP. Exposure to bigger pulling forces of 30-50 pN in contrast leads to its complete unfolding [52].

In a recent report a second structure of the M10-OBSL1 Ig1 complex has been published [167]. The published structural model is virtually identical to the one presented here. In addition, the authors examined the mechanic stability of the titin – OBSL1/obscurin complexes by AFM and showed that the complexes are mechanically rather labile. The forces which are required to disrupt single complexes are with ca. 30 pN more than one order of magnitude lower [167] than the extreme forces which are required to disrupt the titin Z-disc anchoring titin – telethonin complex [48]. From the comparably small size of the interface and the tilted arrangement of complex partners this is not a surprising finding and also excludes the hypothesis that the complex could mechanically anchor titin in the M-band. This measured complex disruptive force is, however, at the lower end of the force range required to unfold (and permanently destroy) the kinase domain of titin being located opposite end of M-band titin. If there is no second site for titin M-band fixation and the obscurin/OBSL1 are fixed in the M-band (e.g. by their second binding site to myomesin-1), the complex could though act as safeguard for the integrity of the kinase domain through a force-sensitive 'hold and release' mechanism **(Figure 3-11)**. A labile binding of M10 to 'M-band immobilized'

obscurin or OBSL1 termini could trigger the mechanical activation of Tk at normal moderate forces. As soon as excessively strong forces occur, e.g. from excentric muscle contractions, strong passive muscle tension or muscle shearing, the dissociation of the M10-OBSL1/obscurin assembly could then release titin's C-terminus, relax and inactivate the kinase domain. Spontaneous re-assembly of the head-to-tail complex could then again facilitate kinase activation.
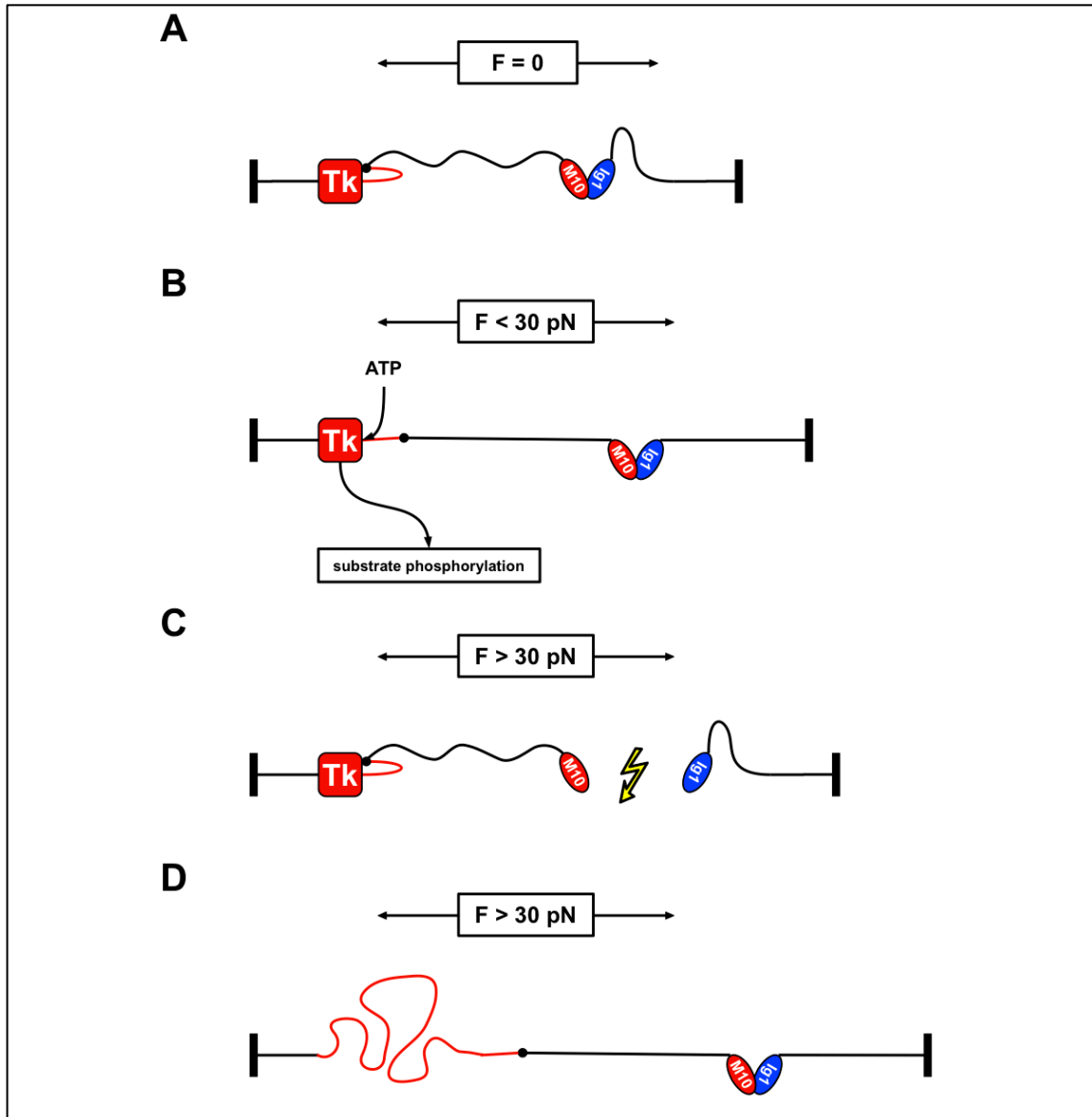

Figure 3-11 Hypothetical function of the M10/Ig1 complexes as a mechanical safeguard for Tk function.

(A) In the absence of force, Tk (red square) is inhibited by a C-terminal loop (red) which is connected to the titin – OBSL1/obscurin complex through M-band titin (M1-M9; black rolling line). The site of mechanical obscurin/OBSL1 fixation is unknown but may be provided by its interaction with myomesin-1. **(B)** The inhibitory loop is

pulled off the kinase as soon as a moderate force is applied. This facilitates ATP binding and substrate phosphorylation. **(C)** Application of stronger forces release the C-terminal M10 domain from OBSL1/obscurin and thus inactivates Tk. The binary complex could spontaneously re-form as soon as the external force decreases to a lower level. **(D)** If the M10-obscurin/OBSL1 complex was mechanically more stable, forces beyond the strength required for its activation could act on Tk. This would lead to irreversible unfolding and so, permanent destruction of the domain.

The direct interaction of titin and myomesin-1 represents the only additionally described titin M-band anchoring site. However, as this interaction could not be confirmed (this study **section 3-2** and by Y2H [60]) it is unlikely to occur in the sarcomere. Direct mechanical coupling of the titin kinase and M10 domains of the same molecule seems therefore possible.

3.1.2.3   Two hypothetical models for the origin of M10 linked myopathies

The structure of the OBSL1 – titin M10 complex presented here and additional data on their folding and binding to OBSL1 and obscurin allowed me to draw conclusions on the effect of two out of four known missense mutations leading to hereditary myopathies on complex formation. These M10 variants, which are associated with the childhood onset LGMD2J and severe cases of TMD in a single French family, have a strong effect on the domain fold and do therefore inhibit complex formation, although the affected residues are not directly located in the interface. A causal connection between the resultant inability to locate obscurin and its homolog to the M-band (which can be concluded from competition experiments in [60]) and disease development cannot be excluded. Especially since obscurin (but not OBSL1) with its several signaling domains of which the functions are poorly understood, and with its connection to the SR, certainly needs an exact positioning within the sarcomeres.

However a causal relation between disease and obscurin/OBSL1 becomes less clear if the two other M10 variants H55P/I56N are considered. Although causing a late onset (40-60 years) TMD and comparably mild symptoms, all known mutations result in the same disease phenotype, which differs from the strong phenotypes only in symptom strength and time point of occurrence. The altered residues in M10 variants of the mild phenotypes are located on the opposite side of the complex interface and result in only minor fold alterations. At the same time the M10 variants retain their ability to bind their complex partners with only a moderate (1.5-2.5 fold) decrease in binding affinities.

How then can the occurrence of the inherited diseases be associated with all mutations? Two hypothetical disease mechanisms will be discussed here. These might be true also in combination with each other.

(1) LGMD2J and TMD variants of M10 modulate sarcomeric ubiquitination through modulation of titin kinase activity.

Although its function in terms of substrate phosphorylation remains to date nebulous, strong evidence exists for a functional connection between the titin kinase domain and the muscular ubiquitination system. The open (active) form of the domain binds the two zinc finger proteins Nbr-1 and p62 in which direct interaction with TK is mediated by Nbr-1. P62 itself forms complexes with the muscle specific E3 ubiquitin ligases MURF1-3 [94]. This signaling complex of titin, Nbr-1, p62 and MURFs dissociates in non-contracting cardiomyocytes and, at least for MURF-2 translocation into the nucleus after dissociation has been shown [94] (reviewed in [168]). A connection of kinase activation and ubiquitination-mediated protein turnover combined with the 'hold and release' mechanism (previous section) would therefore directly link pathogenic mutations of M10 with changes in sarcomeric protein turnover, and, at the same time, explain differences in symptom onset and strength.

A total lack of M-band fixation as it occurs with misfolded M10 variants would lead to a permanently inactive kinase domain, similar to the proposed state when the complex dissociates due to the exposition to strong forces **(Figure 3-11 C)**. The MURF linking complex would never form and protein turnover would be permanently altered and lead to an early onset of symptoms. The effect would be more subtle with M10 variants that only slightly decrease the binding affinity to OBSL1/obscurin. The affinity of these has been determined with isolated protein domains, so it can be safely assumed that the 'real' affinities for the complexes differ due to linkage to the rest of the proteins. However, already a decrease in binding affinity from a K_D of 0.94 μM (M10:OBSL1) to a K_D of 2.5 μM (M10 H55P:OBSL1) would increase the proportion of free ligands to complexed ones from 1:1064 to 1:630 thus nearly double the number of present unbound titin C-termini and lead to increased inhibition of kinase domains and non-forming Nbr-1-p62-MURF complexes. The dissociation of the TK-NBR-p62-MURF complex is supposed to be only the beginning of a signaling cascade with MURF-2 transport into the nucleus as slow intermediate step. Therefore, the only small calculated change in the proportions of titin-OBSL1/obscurin complexes and free partners can, if taking the presence of thousands of titin molecules in a myofibril into account, result in

an attenuation of downstream effects over time. For example a critical concentration of MURF2 ubiquitin ligase in the nuclei may be reached after a large number of muscle contractions and then result in a noticeable effect. Apparent effects would become significant only after a long time and in heavily used muscle. This is the case in dominant TMD.

(2) Modulation of calpain-3 activity

A second line of evidence for the origin of M10 variant pathogenicity can be derived from the facts that TMD and LGMD2J affect only certain skeletal muscles, while heart muscle, although continuously in use, remains unaffected, and that TMD mutations do only poorly interfere with obscurin and OBSL1 complex formation. Thus, it might be possible that a second, currently unknown interaction partner which binds M10 at the opposite side of the OBSL1/obscurin interface, is responsible for TMD development. This partner should be present only in skeletal muscle tissues and possibly even only in a fraction of these.

Indeed a strong candidate for this exists with calpain-3, a large skeletal muscle specific cysteine protease [169]. The protein is not expressed in heart muscle and mutations on the *CAPN3* gene have been identified as the origin of autosomal recessive limb-girdle muscular dystrophy 2A [19]. The function and activation mechanism of the protease are still under intense debate (for reviews see e.g. [56][57][170]) but it seems to be clear that auto-proteolytic activation plays a role and the binding to one of the two identified binding sites on the titin molecule [90][171] inhibits autolysis. A loss of control of calpain-3 activation is supposed to be one cause of LGMD2A. One of the described calpain-3 binding sites on titin is located directly N-terminal of the M10 domain, in a region that consists of ca. 100 amino acids and spans domains M9 and M10. This region, denoted as Is7, is the only part of M-band titin that is subject to muscle tissue and fiber-type specific alternative splicing [87]. A connection of M10 folding or mutation-induced local structural changes and calpain-3 activation could therefore be the source of muscle malfunction in TMD and LGMD2J.

Indeed a (mal-) functioning connection between the two proteins and their disease phenotypes was recently demonstrated in a TMD/LGMD2J mouse model [172]. Calpain-3 knockout mice (which show a muscular dystrophy phenotype in a homozygous but no phenotype in heterozygous form [173]) were crossed with a LGMD2J model mouse with a 'damaged' M10 domain (showing significant signs of muscle disease similar to human LGMD2J). The crossed calpain$^{+/-}$ LGMD2J$^{+/-}$ mice lost

their dystrophic phenotype. A reduced expression of calpain-3 can thus compensate for the misfolding of M10.

This is clear evidence for a connection between the malfunction of both proteins and the resulting diseases. It may have different implications, and a direct or indirect physical linkage of both in the M-bands may be the simplest. Yet it provides an easily testable hypothesis, await experimental analysis.

3.2 Characterization of the direct assembly between titin and myomesin-1

3.2.1 Results

3.2.1.1 Purification of the complex partners

In order to characterize the interaction between human titin and myomesin-1 in the M-band, constructs of the described [54][59] interacting protein fragments titin M4 and myomesin 4-7 (containing the mapped minimal binding site My4-6) were cloned from human cDNA and recombinantly expressed in *E. coli*. Both proteins could be expressed as His (M4) or His-SMT3 tagged (My4-7) protein in a soluble form and purified by IMAC and subsequent SEC. M4 eluted from a size exclusion chromatography column as a single symmetric peak. The monomeric nature of the protein was confirmed by SLS (**Figure 3-12 B**; SLS data not shown). In contrast, the SEC profile of My4-7 appeared asymmetric **(Figure 3-12 A)**. Analytical size exclusion chromatography allowed the partial separation of the heterogeneous components into two peaks. This indicates that the purified My4-7 protein contained at least two distinct proteins or protein assemblies of different hydrodynamic radii **(Figure 3-13 A)**.
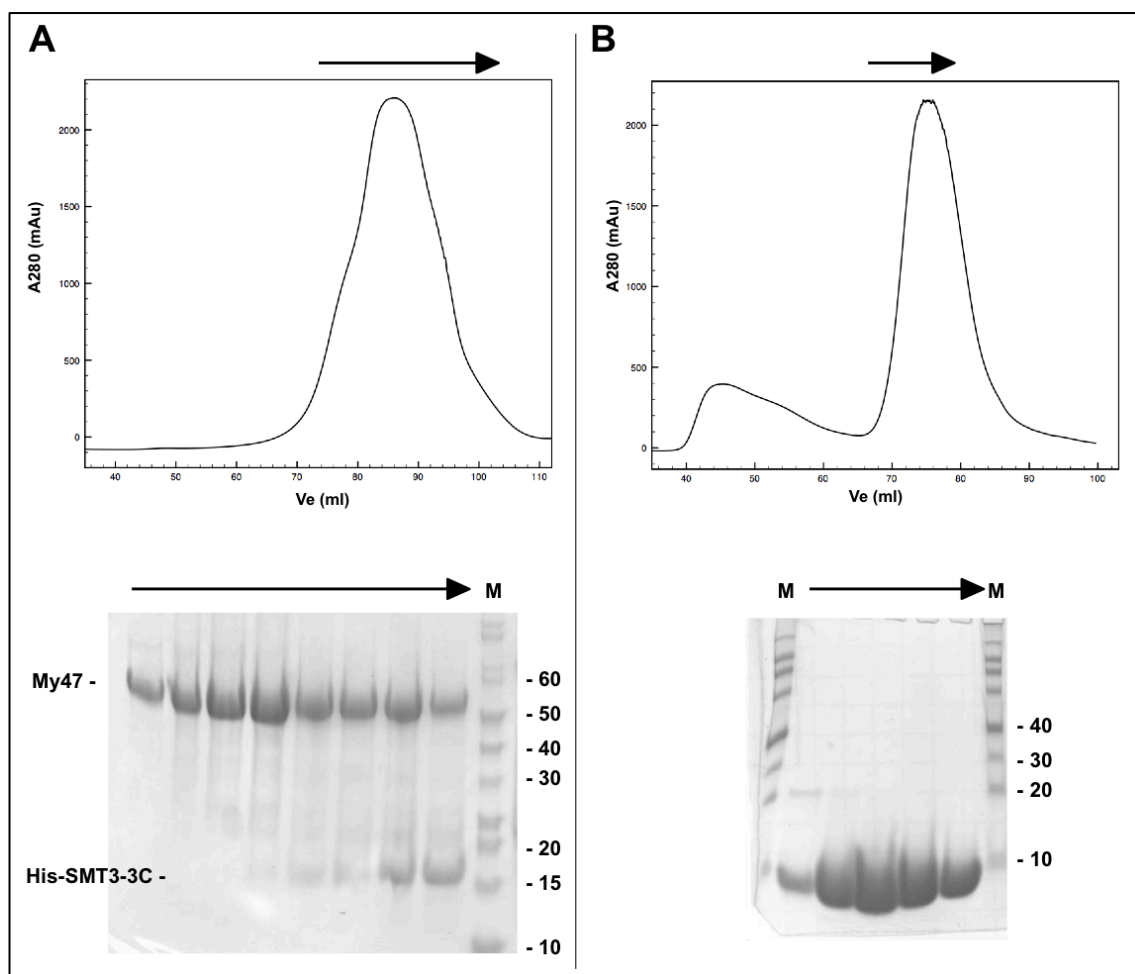
Figure 3-12 Purification of My4-7 and M4.

(A) Top panel: Elution profile of My4-7 from a Superdex 200 16/60 column. The protein eluted in an asymmetric peak. Lower panel: SDS-PAGE analysis. The analyzed range is marked with an arrow in the top panel. The impurity of approximately 15 kDa weight found in of high elution volume fractions is the cleaved His-SMT3-3C tag. Note that this tag does with a molar extinction coefficient of 1490 only insignificantly contribute to the absorption at 280 nm wavelength, when compared to cleaved My4-7 (ε = 77350). Its presence can thus be excluded as source of the observed peak-asymmetry. **(B)** Top panel: Elution profile of M4 from a Superdex 75 16/60 pg column. Lower panel: Peak content analysis by SDS-PAGE. M = molecular weight marker (My4-7: PageRuler; M4: Rotimark); marker band sizes in kDa.

3.2.1.2 *In vitro* titin does not spontaneously form a complex with myomesin-1

Titin M4 and myomesin-1 My4-7 were assayed for their ability to interact using analytical size exclusion chromatography and ITC. For analytical size exclusion chromatography M4 was mixed with equimolar amounts of a mixture of both My4-7 forms (approximately 80% of the late eluting and 20% of the early eluting fraction) in purification buffer (20 mM Tris pH 8.0, 150 mM NaCl, 1 mM TCEP). The mixture was then incubated for 1 h at 4 °C and injected on an analytical Superdex 200 10/300 column. A significant shift in the peak elution volumes compared to separately injected reference samples could not be observed **(Figure 3-13 A)**. Subsequent SDS-PAGE analysis confirmed that the My47 and M4 did not co-elute **(Figure 3-13 B)**. In agreement with this result, significant binding could not be measured by ITC in the following condition: PBS pH 7.2 with 1 mM TCEP at 25 °C or 30 °C (**Figure 3-13 C**). A direct interaction of titin M4 with the domains 4-7 of myomesin-1 *in vitro* could therefore not be confirmed.

Figure 3-13 Analysis of the myomesin-1 - titin interaction.

(A) Left panel: A mixture of equimolar amounts of My47 and M4 which was incubated for 1 h at 4 °C elutes as three molecular species from a Superdex 200 10/300 column in 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM TCEP. Right panel: Elution profiles of separately injected reference complex partners show that My4-7 elutes in two distinct peaks (blue) and M4 as a single one (red). The peak elution volumes do not significantly differ from the ones observed with mixed proteins. **(B)** Analysis of SEC fractions by SDS-PAGE. M4 does not co-elute with My4-7 (15% SDS-PAGE gel; marker sizes on the left in kDa). The black arrow corresponds to the analyzed range of eluate indicated by the same in (A) **(C)** ITC measurement of 0.88 mM titin M4 titrated into 0.08 mM My4-7 at 25 °C (left) and 30 °C (right).
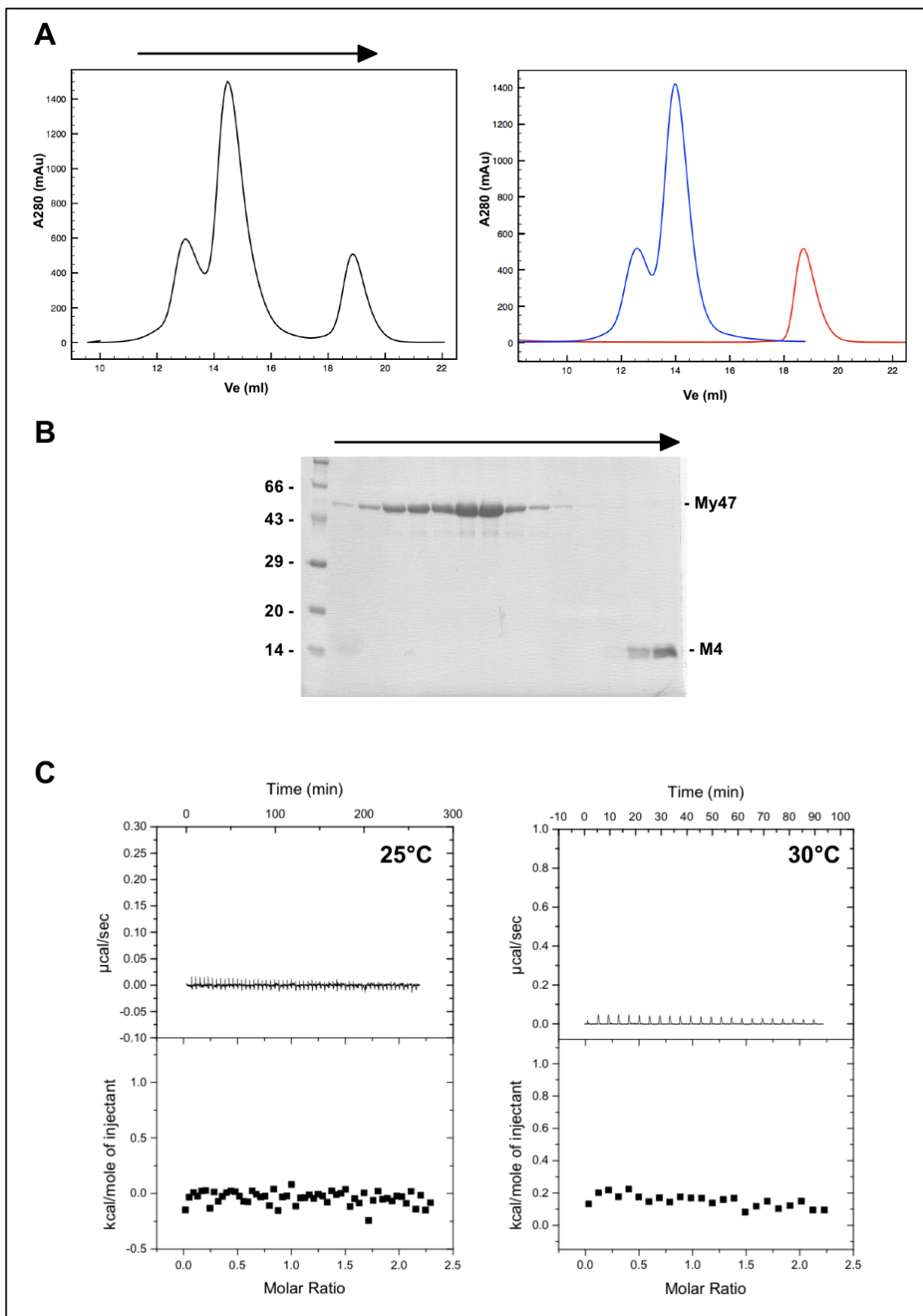
3.2.2   Discussion

The formation of a complex between titin domain M4 and myomesin-1 domains My4-6 has been described in a single study [59]. In this study, recombinant human M-band titin fragments bound to a nitrocellulose membrane were incubated with different recombinant fragments of human myomesin-1. Of these My1-8, My4-8 and My4-6 did bind the membrane via titin M4, while shorter constructs (My4-5, My5-6) did not. The direct binding between both seemed to be in agreement with previous data, in which myomesin-1 co-purified with titin from muscle tissue [36] and electron microscopy data, in which neighboring domains of the interaction partners co-localized at similar positions in the M-bands [54]. However, since this original report, neither a functional role, apart from 'M-band binding' has been proposed nor an attempt was made to repeat the binding experiment. During the time of my PhD thesis it was reported, that the assembly of both could not be confirmed by either Y2H or by pull-down experiments [60]. The data presented here confirm this finding.

In this work for the first time a quantitative characterization of both binding partners and their assembly by analytical size exclusion chromatograph and ITC was attempted. The binding partners could be expressed and purified from a recombinant source. A direct assembly could not be found with either method. What could explain the different findings? Three conceivable explanations will be discussed here:

(1) Differences in M4 construct sizes: In the present work a construct of the titin M4 domain comprising 101 residues (11.5 kDa) was used to probe the interaction with myomesin-1. The correct folding as an Ig-like domain was later confirmed by determination of its crystal structure **(section 3.4)**. It was assumed to be similar to the one originally found to bind myomesin-1. Indeed, a graphical representation (Fig. 1 of [59]) showed that the M4 construct covered a range of approximately 100 amino acids located between the domain linking sequences Is2 and Is3. However, the exact residue range of the construct was not given, and the SDS-PAGE analysis of the purified recombinant protein revealed an apparent molecular weight close to 20 kDa (Fig 4b of [59]). The almost two-fold greater molecular mass cannot be explained by the presence of the tags used for affinity purification (hexa-histidine tag) and epitope detection (EEF-tag), or an unusual migration behavior on an SDS-PAGE gel **(Figure 3-13 B)**. It can therefore not be ruled out, that the recombinant protein found to bind myomesin-1 indeed contained large parts of Is2 or Is3 and that these latter are required for the

binding.

(2) Protein oxidation or misfolding: The crystal structure of titin M4 **(section 3.4)** shows that the domain contains three cysteine residues of which one (Cys 22) is in an exposed position on the protein surface and can form a disulfide bond as observed in one M4 crystal form. The exact sequence of the myomesin-1 construct originally found to bind M4 is not known. The primary sequence of the predicted myomesin-1 domains 4-6 (residues 503-851) reveals the presence of ten Cys residues. For both proteins, having been purified and analyzed under reducing conditions, the formation of inter-molecular disulfide bridges between exposed Cys residues or unspecific binding between mis-folded proteins can be ruled out in the ITC and SEC experiments presented here. This is however not the case for the original assay. In it, the recombinant binding partners had also been purified under reducing conditions (in the presence of 1 mM DTT), but the binding assay involved the fixation of the M4 domain on a blotting membrane followed by a subsequent air-drying step. The protein may have altered its structure and in a dried state, the protein environment may have changed to an oxidizing one. It is therefore possible that, in the blot-based assay, myomesin-1 bound titin M4 in a non-specific way by non-specific disulfide formation or a different mechanism.

(3) Dependence of the assembly on features only present with a certain quaternary structure: The closer examination of both binding partners revealed surprising properties: Both have the ability to form dimers *in vitro*. Titin M4 dimerization could only be observed in one out of two crystal forms and but never detected in solution. Because the titin M4 was only purified as a monomeric species, the dependence of the assembly on dimeric M4 cannot be ruled out **(section 3.4)**.

In the case of myomesin-1, dimerization was found to be mediated by the domain My5 located in the center of the described titin interacting construct **(section 3.3)**. The myomesin-1 protein used to assay the interaction with titin, contained mono- and oligomeric protein forms in an estimated 4:1 ratio. The dependence of the assembly on dimeric myomesin-1 would therefore have led to at least the binding of a sub-stoichiometric fraction of M4. In contrast, the dependence of the assembly on dimeric M4 be ruled out.

In conclusion, the formation of a titin – myomesin-1 complex in the M-bands is unlikely, but cannot be entirely ruled out. Currently several *in vitro* findings ([60] and this work) cast doubts on the complex's existence and stand in opposition to the single *in vitro* experiment reporting it [59].

3.3 Characterization of the central assembly site of myomesin-1

3.3.1 Results

3.3.1.1 Myomesin-1 dimerization is mediated by domain 5

During the work on the putative complex of myomesin-1 and titin M4 **(section 3.2.1.2)**, it became apparent that the recombinant myomesin-1 construct My4-7 was not purified in a homogeneous form. The protein showed irregularities in peak symmetry when eluted from a SEC column. A two-domain construct, My4-5, eluted as two partially separated peaks from a size exclusion column **(Figure 3-14 A)**. This observation was reproducible (although the relation of peak heights changed between different protein preparations) and led to the hypothesis that it could be the result of a second, previously unreported site of oligomerization on myomesin-1. To map the domain responsible for the self-association two single domain constructs were produced: My4L (residues 503-635, including the linker region between domains 4 and 5) which eluted as a single symmetric peak from an analytical size exclusion chromatography column **(Figure 3-14 B)**. In contrast, the isolated domain My5 (myomesin-1 residues 632-735) eluted as three peaks from a SEC column **(Figure 3-15 A/B)**. Thus My5 is sufficient for the observed myomesin-1 self assembly.

To determine the oligomeric state of the different myomesin-1 forms, their molecular weights were determined by static-light scattering [174]. The measurements confirmed that the two predominant peaks of the constructs My4-7, My4-5 and My5 contained monomeric and dimeric protein forms (**Table 3-4**). The oligomeric state of the protein eluting in the third peak, only observed with the My5 single domain construct, could not be determined with acceptable accuracy. This suggests, that it could possibly contain a mixture of larger assemblies or protein aggregates.
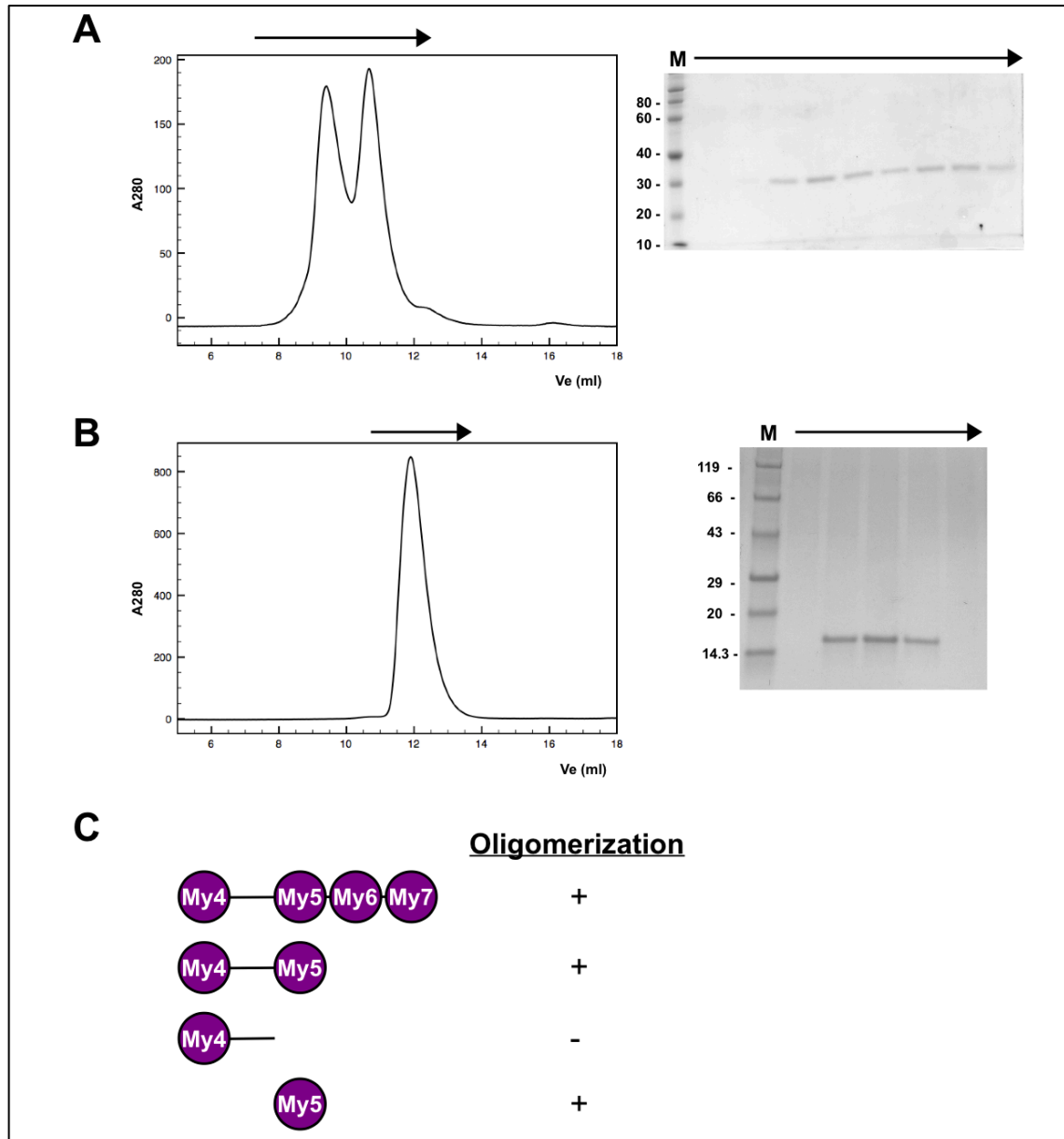
Figure 3-14 Analysis of My4L and My45.

(A) My4-5 elutes from a Superdex 75 10/300 column reproducibly in two partially separable peaks. **(B)** The truncated My4L construct elutes as a single peak from the same column. Right panel in A/B: Peak analysis by SDS-PAGE (12%). The black arrows indicate the ranges of analyzed fractions. M: Marker (My4-5: Rotimark 10-150; My4L: Rotimark Standard). Numbers on the left correspond to the marker band masses in kDa. **(C)** Schematic representation of the recombinant constructs used for the mapping of the myomesin-1 oligomerization site.

The presence of mono- and dimeric forms is supported by the comparison of the measured elution volumes of domain My5 with the ones measured for monomeric Ig-like domains (M10/obscurin Ig1), the complex of both and the titin Z1Z2 Ig-like domain tandem [175]. The smaller form of My5 eluted at a volume of 13.0 ml from a

Superdex 75 10/300 column and thereby at nearly the same volume as monomeric Ig-like domains (e.g. M10: 13.1 ml). The dimeric My5 form eluted 1 ml earlier (11.2 ml) than the laterally aligned M10-obscurin hetero-dimer (12.2 ml) but with a similar volume as the titin Z1Z2 Ig-like domain tandem, which adopts an extended conformation (11.5 ml) [175] **(Fig. 3-15 C)**. It can therefore be concluded that the dimeric form of My5 adopts an extended, tandem-like domain conformation rather than a compact one **(Fig. 3-15 D)**. This indicates that the mechanism that leads to the dimerization of My5 is different from the ones previously described for titin/OBSL1/obscurin and the myomesin-1 and filamin Ig-like domains, which assemble in all cases side-by-side.

Table 3-4 Molecular weights of recombinant myomesin-1 proteins as determined by SLS

Construct	Peak 1 (kDa)	Peak 2 (kDa)	Peak 3 (kDa)	Calc. monomer weight (kDa)
My4-7	61.9 (36%)	131.8 (21%)	-	48.4
My4-5	25.3 (21%)	48.3 (13%)	-	25.5
My5	11.9 (27%)	25.9 (19%)	>35	11.4
My4L	13.1 (13%)	-	-	14.5

Weights in kDa as calculated by the Astra software.

Numbers in parentheses correspond to the standard deviations of the calculated weights.

To assess the stability of the different My5 forms, samples were stored at 4 °C for 24 h, then concentrated and re-injected on an analytical SEC column. Monomeric and dimeric forms preserved their oligomeric state. The larger oligomers or aggregates dissociated partially into monomers **(Figure 3-15 C)**. This indicates the presence of at least two different types of My5 oligomerization. The first type is responsible for My5 dimerization. Its stability indicates a slow monomer-dimer equilibrium. The second type accounts for the formation of larger oligomers. Fast dissociation into monomers occurs. Thus it appears to follow a faster equilibrium. However as the stoichiometry of the larger oligomers is not known and dissociation releases mainly monomers, it is possible that the oligomer peak contained a mixture of different oligmomeric My5 forms of which only a fraction with a certain composition dissociates rapidly, while a different fraction dissociates slowly.
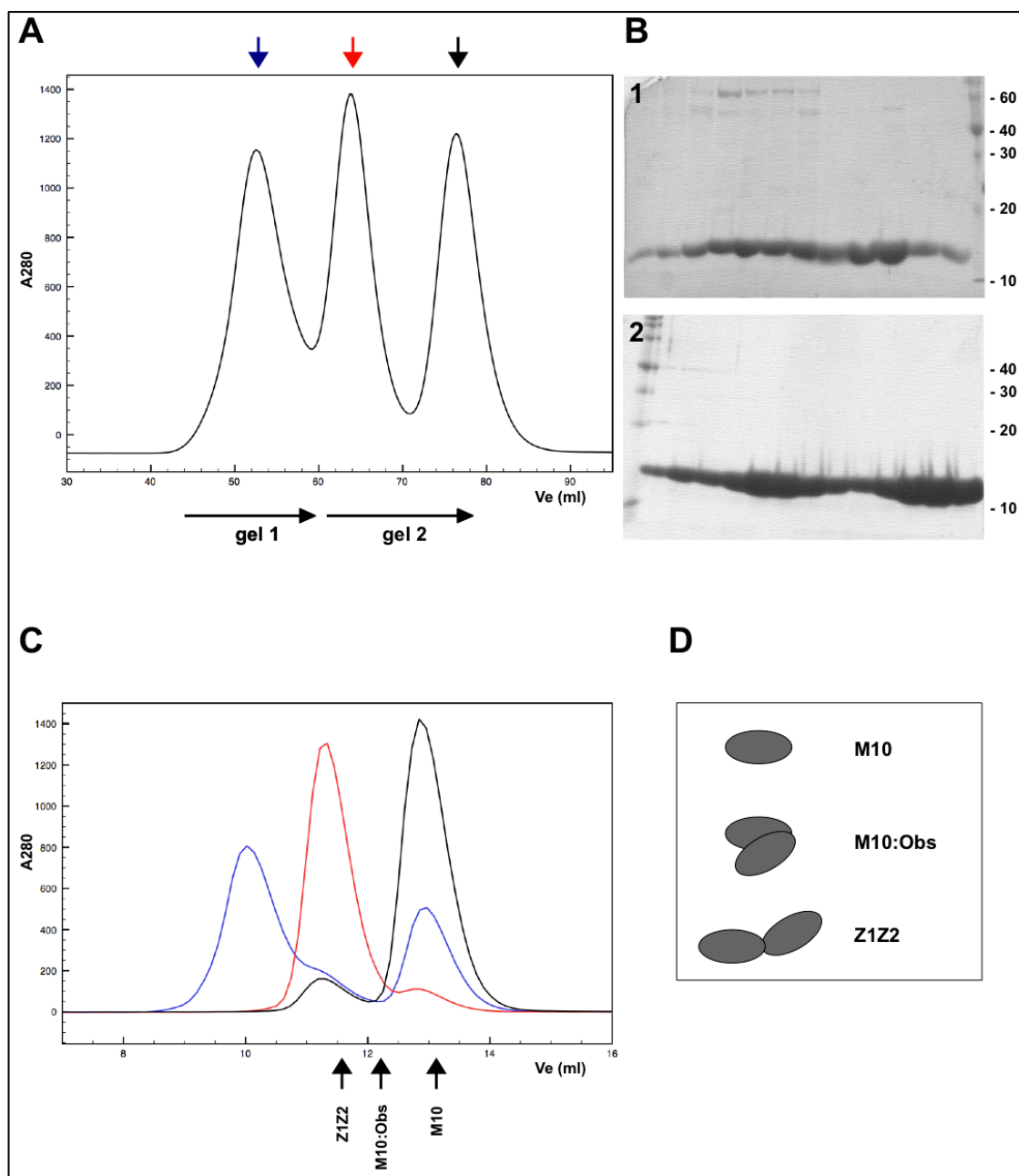
Figure 3-15 Purification and stability of My5.

(A) My5 elutes in three distinct peaks from a Superdex 75 16/60 pg column. The arrows denote positions of fractions from which samples were taken to assay complex stability. **(B)** All peaks contain only the recombinant protein except for the lowest elution volume one, which contains minor amounts of high molecular weight contaminants. 15% SDS-PAGE gel; numbers correspond to the protein standard weight in kDa **(C)** Analytical size exclusion chromatography of samples stored for >24h. Mono- and dimeric My5 (red and black, respectively) retain their state while the larger form (blue) partially dissociates into the monomeric form. Bottom arrows denote peak elution volumes of reference proteins. **(D)** Schematic representation of the shape of reference proteins based on their crystal structures.

3.3.1.2   Crystallization of the myomesin-1 My5 domain

To elucidate the assembly mechanism in the central portion of myomesin-1, crystallization trials for the three dimerizing constructs My4-7, M4-5 and My5, were carried out. Crystals could be obtained from all constructs after initial screening using the sitting drop vapor diffusion method. However, the crystals grown from monomeric or dimeric forms of the multi-domain constructs My4-5 (20 mg/ml) and My4-7 (32 mg/ml) had a hair- or needle like morphology. The crystals of My4-5 could be reproduced by the hanging drop vapor diffusion method, and My4-7 crystals could be grown under oil in the conditions obtained from the initial screening **(Fig. 3-16 D/E)**. However, optimization trials by variation of protein or precipitant concentrations, additive screening or changes in the crystallization setup (crystallization method, drop sizes) failed to improve crystal quality and size. Due to their fragility, single crystals could not be harvested for a diffraction experiment and attempts to use them for content determination by mass-spectrometry or SDS-PAGE resulted in rapid crystal decomposition.

In contrast, crystals suitable for diffraction experiments could be obtained from the minimal dimerizing construct My5. Purified monomeric and dimeric forms of myomesin-1 domain My5 were submitted for high throughput screening at concentrations of 13 mg/ml (dimeric form) and 26 mg/ml (monomeric form). Initial crystals grew within 1-2 weeks after set-up as crystal clusters from both oligomeric forms in different unbuffered solutions containing 2 - 2.4 M $(NH_4)_2SO_4$ and 0.2 M Na_3-citrate, Na_2-tartrate, Na_2-malonate or Mg-formate.

From these, only crystals grown in Na_3-citrate could be reproduced. Manual crystallization using the hanging drop vapor diffusion and vapor batch crystallization methods (under 30% paraffin oil, 70% silicone oil) with larger precipitant and protein drop volumes of 1-4 μl, decreased the crystal nucleation rate and resulted in the growth of single crystals.
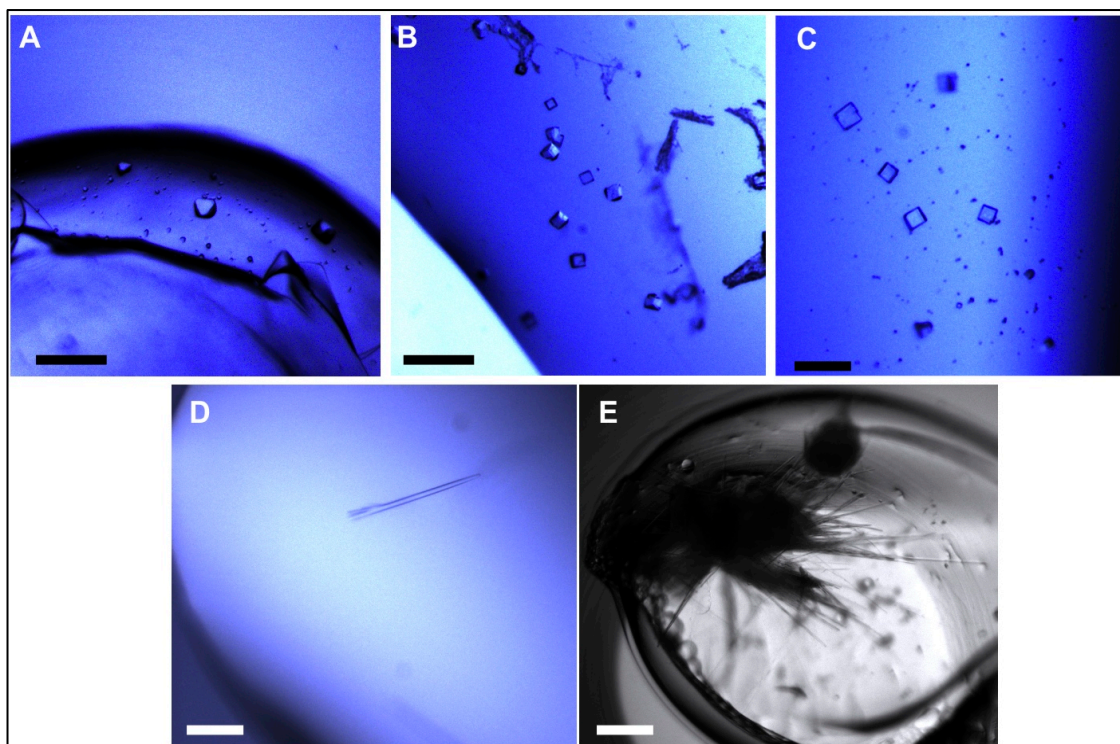
Figure 3-16 Crystallization of myomesin-1 proteins.

Rhombic crystals of dimeric My5 grown under oil **(A)** or by vapor diffusion **(B)**. **(C)** Cubic crystals grown from monomeric My5 by vapor diffusion. **(D)** Crystals of My4-5 the dimer (25 mg/ml) grown after ca. one month from HEPES pH 7.0, 0.8 M $(NH_4)_2SO_4$. **(E)** Crystals of My4-7 monomer (35 mg/ml) grown under oil after ca. 2 weeks from MES pH 6.0, 7.5% (w/v) PEG 3350. The bars indicate a length of 0.2 mm.

The initial condition was optimized by variations of salt concentrations. Finally, 4 µl of protein solution (10-30 mg/ml) mixed with an equal volume of 1.9 M $(NH_4)_2SO_4$, 0.3 M Na_3-citrate as precipitant yielded crystals with a size of approximately 50 µm in all dimensions for both monomeric and dimeric forms. Although crystals from both oligomeric forms grew under the same conditions, their morphology differed. Crystals from the monomeric form had a cubic shape while the dimeric form crystallized in a rhombic shape **(Figure 3-16 A, B, C)**. The crystals grown by the hanging-drop method grew within three to seven days. Under oil crystals grew within two weeks to their final size.

The content of both crystal forms was analyzed by MALDI-TOF. The measured molecular weights 11.412 kDa (monomer crystals) and 11.409 kDa (dimer crystals) are in good agreement with the calculated protein weight of 11.390 kDa and confirm that both forms contained the complete My5 protein.

3.3.1.3   Data collection, structure solution and refinement

For diffraction experiments the crystals of monomeric and dimeric My5 from both types of crystallization setups were cryo-protected in mother liquor supplemented with 20% (v/v) glycerol, mounted onto a litho-loop and flash frozen in liquid nitrogen. Diffraction data were collected on the microfocus beamline ID23-2 at ESRF. Despite their similar size, crystals displayed very different diffraction properties. Crystals of the monomeric My5 form did not diffract at all. The same was the case for most of the crystals grown from dimeric My5. Only a single out of ca. 30 tested crystals showed diffraction and a complete dataset to a maximum resolution of 2.68 Å could be collected. The data were integrated with XDS and scaled with Scala. The structure was solved in the rhombohedral space group $R3_2$ by molecular replacement using BALBES and the Fn3 module A71 of human titin [176] as a search model.

An initial model comprising 188 of 210 residues in two protein chains was automatically built by ARP/wARP. The protein chains and solvent molecules were manually rebuilt in COOT. Individual isotropic B-factors, TLS parameters and atomic positions were refined with phenix.refine. Despite the limited number of unique reflections due to the high symmetry of the $R3_2$ space group and the maximum measured resolution of 2.68 Å, the model could be refined to a good free R-value of 24.8.

The final crystallographic model contains two My5 domains (myomesin-1 residues 632-735) together with eight sulfate ions and 23 water molecules in each asymmetric unit. The model is complete for both domains except for six N-terminal (Gly 631 – Ile 636) and four C-terminal residues (Val 732 – Lys 735) on molecule A as well as four N-terminal (Gly 631 – Glu 634), four C-terminal (Val 732 – Lys 735) and three loop forming residues (Arg 664 - His 666) on molecule B for which no electron density could be observed. In addition traceable electron density was missing for the side-chains of eight residues (molecule A: Val 637; molecule B: Lys 659, Gln 663, Glu 681, Arg 685, Lys 693, Gln 716, Glu 722). Diffraction data, model- and refinement statistics are summarized in **Table 3-5**.

Table 3-5 Data collection, refinement and model statistics of My5

Data collection and processing:	
Beamline	ID23-2
Detector	Mar mosaic 225
Wavelength [Å]	0.8726
Resolution [Å]	48.9 - 2.68 (2.83 – 2.68)
Space group	$R3_2$
Unit cell parameters [Å]	a/b = 97.82, c = 111.69
Mosaicity (°)	0.57
No. of measured reflections	29473 (4148)
No. of unique reflections	5931 (859)
Redundancy	5.9 (4.8)
Mean $I/\sigma(I)$	11.0 (3.1)
Completeness [%]	99.9 (99.9)
R_{merge} [%]	8.9 (43.5)
$R_{p.i.m.}$ [%]	4.8 (24.8)
R_{meas} [%]	10.9 (55.9)
Overall B-factor from Wilson plot [Å^2]	56.3

numbers in parentheses correspond to the highest resolution shell

Refinement and model statistics:	
No. of reflections (free)	5931 (582)
R_{work} / R_{free} [%]	20.6/ 24.8
No. of atoms: protein / ions / water	1424/40/23
Average B-factor Protein / ions / water [Å^2]	32.8/90.3/26.7
R.m.s. deviations	
Bond lengths [Å]	0.015
Bond angles [°]	1.266
Ramachandran plot	
Favored [%]	98.4
Additionally allowed [%]	1.6

3.3.1.4   Crystal structure of the My5 dimer

The structure reveals how two My5 molecules in the same asymmetric unit interact with each other to form a dimer with two domains of a Fn3-like fold that are reconstituted from fragments of both molecules, a type of interaction termed three dimensional domain swapping [177].

The two isolated molecules show an all beta fold with six strands organized in two ß-sheets. Compared to the fold of fibronectin type 3 domains, which consists of seven anti-parallel ß-strands organized in two ß-sheets ABE and C'CFG [178], a partially open topology is apparent **(Figure 3-17).**
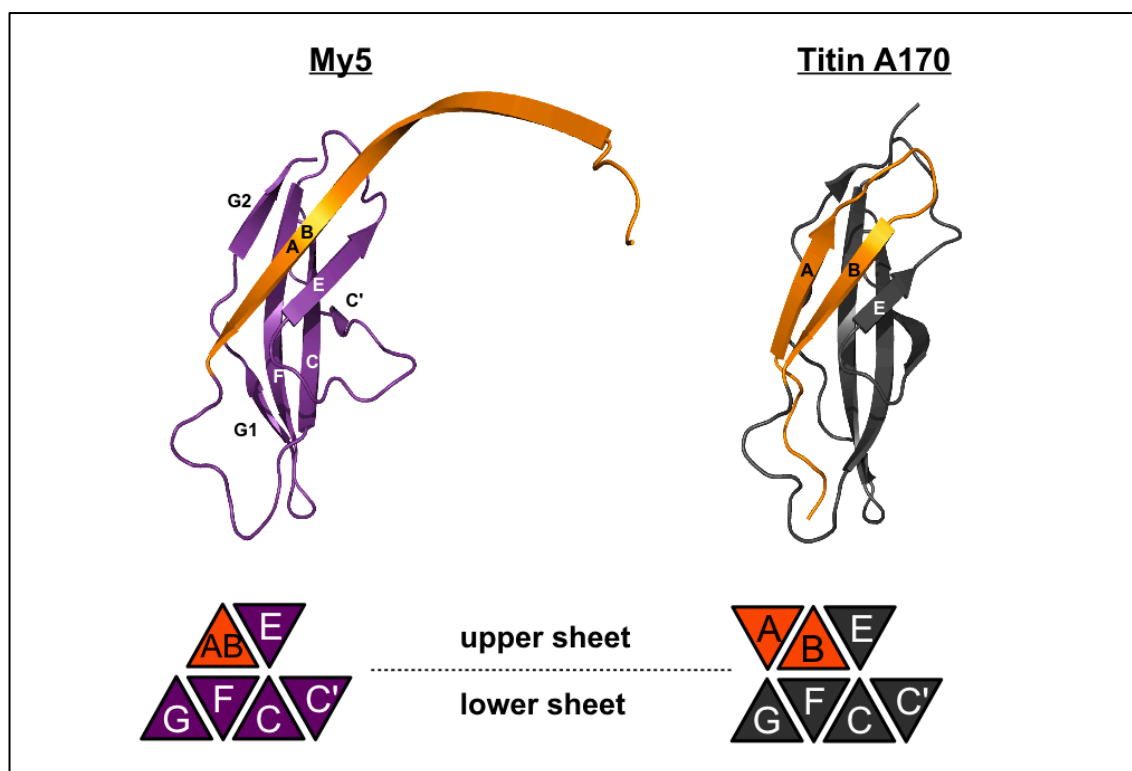
Figure 3-17 Open topology of an isolated My5 domain.

Left: In its dimeric form My5 possesses an open fold in which the N-terminal segment forms a long continuous AB ß-strand (orange). On the right side as an example for a canonical Fn3 fold, the A170 domain of titin [179] is shown. Here the N-terminal segment is divided into two anti-parallel strands A and B that form with strand E the 'closed' upper Fn3-sheet. At the bottom, a schematic view on the strand topology of both domains is shown.

One sheet possesses the common strand topology of the 'lower' sheet of a fibronectin type 3 domain. Here four ß-strands C', C, F and G are assembled in an anti-parallel way.

The sheet bordering C'-strand, which is located in the center of the 18 residue long sheet crossing segment connecting strands C and E, and is assembled in an anti-parallel way with strand C, is with a length of only two residues (Gln 684 – Arg 685) particularly short.

The C-terminal G-strand located at the opposite sheet border is fragmented into two sub-strands G1 and G2. This is due to the presence of Pro 723 at the end of sub-strand G1, which causes a twist in the backbone conformation of the following three residues (Ser 724-Ala 726). Such Pro-induced fragmentation of the G-strand has previously been found for example in all structurally characterized titin Fn3 modules: A71 [176], A77/A78 [180] and A170 [179].

The second sheet consists of only two ß-strands in positions corresponding to the anti-parallel strands B and E in Fn3 domains. A third N-terminal A strand that assembles anti-parallel with the B-strand is not present. The 'upper' Fn3 sheet is thus incomplete in the isolated protomers **(Figure 3-17)**. Instead of folding back as an A-strand, the corresponding N-terminal segments are swapped between both molecules in the same asymmetric unit. Residues usually forming A- or B-strands respectively and the AB-connecting loop, form an 18 residues long continuous ß-strand (Thr 642 - Lys 659) **(Figure 3-17)** here called AB-strand. This AB-strand assembles in an anti-parallel way with the homologous AB-strand of the opposite molecule. The N-terminal strand swapping thus complements on both domains the Fn3 like domain topology and thereby creates a large, nearly symmetric intermolecular E1-AB1-AB2-E2 'upper' ß-sheet.

The resulting swapped My5 dimer resembles the shape of a saddle with approximate dimensions of 58 x 40 x 18 Å **(Figure 3-18)** in which both globular domains enclose an angle of circa 105°. Due to the anti-parallel arrangement of the protomers in the complex, both C-termini are directed towards the dimer center while the N-termini point towards the opposite tips of the dimer. The interface consist of a total of 31 hydrogen bonds of which 20 are formed between main-chain atoms and covers with an average of 1920 $Å^2$ nearly one third (26.8 %) of each molecule's surface.
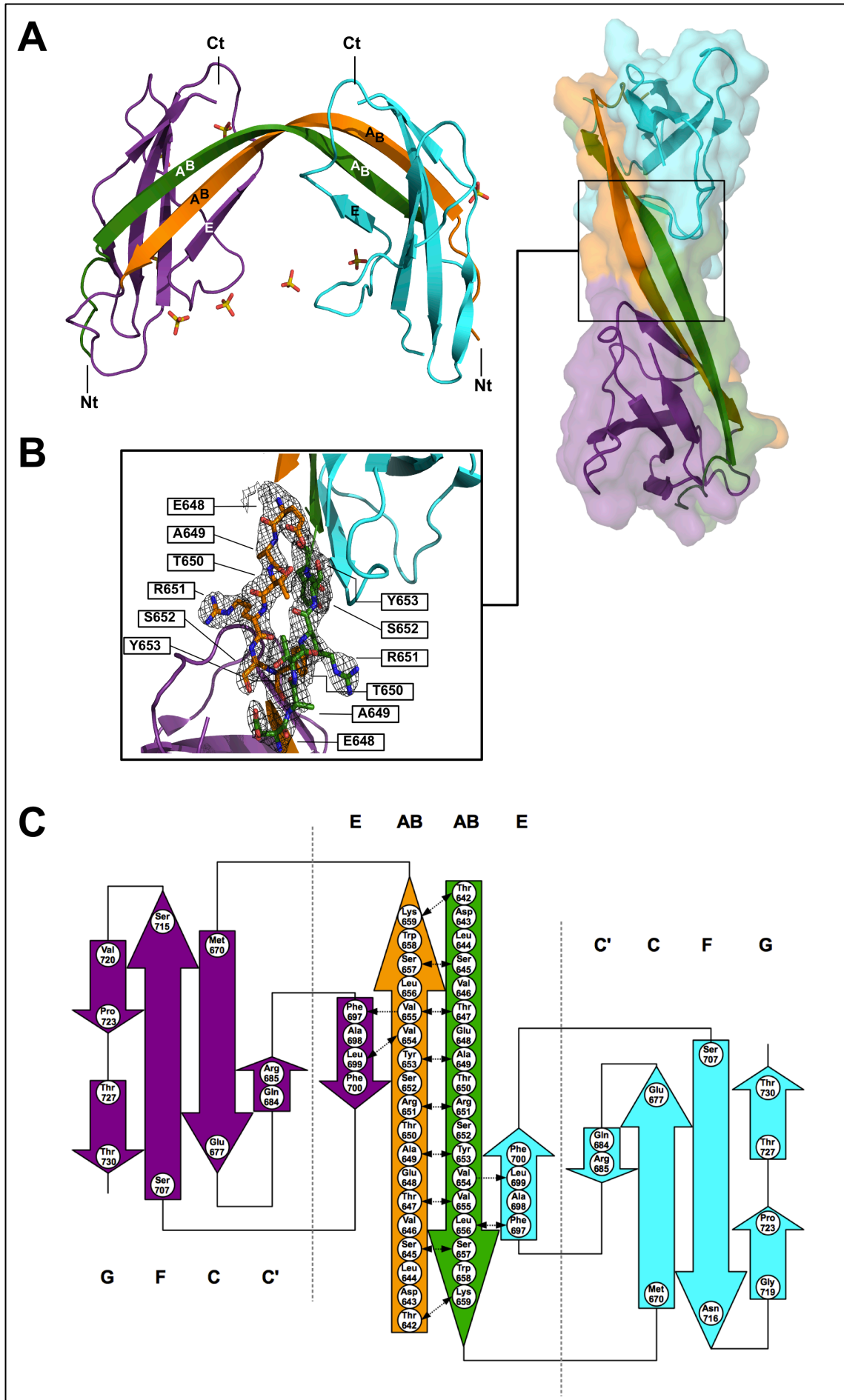
Figure 3-18 Crystal structure of the My5 dimer

(A) Dimerization by strand swapping. Left panel: Cartoon representation of the My5 dimer in side view. Both molecules (A: purple; B: cyan) interact by the anti-parallel assembly of their N-terminal AB-strands (orange and green). Sulfate ions are shown as sticks. Nt: N-terminus; Ct: C-terminus. Right panel: Top view on the dimer. The dimer surface is shown transparently. **(B) Confirmation of strand swapping.** An Fo-Fc omit map (2σ) of the dimer bridge region (Glu 648 – Tyr 653) shows the direct connectivity between both monomers. Figure created with Pymol. **(C) Schematic representation of the strand topology.** The dimerization creates a quasi-symmetrical E-AB-AB-E sheet. Coloring as in (B). H-bonds between backbone atoms in the interacting sheet are marked with arrows (N-H -> O). The three ß-sheets of the dimer are separated by dashed vertical lines.

3.3.2 Discussion

The discovery of a second possible site of self-interaction on myomesin-1 was unexpected. A significant amount of work in the past has been spent on the detection and characterization of homo- and heteromeric assemblies of the protein. These have led to the discoveries of its N-terminal binding site for myosin [59][82][181], central binding sites for titin [59], obscurin/OBSL1 [60] and MM-CK [78] and its C-terminal site of self-assembly [79][80]. The self-assembly of the Fn3 domain My5 described here has however not yet been found. The explanation for this may lie in the methodology. In all cases, ligands were initially found by yeast-two-hybrid (Y2H) screening, and the binding of the identified partners was then confirmed by either pull-down assays, e.g. of GST- and MBP-tagged proteins, or by blot-overlay assays. The detection of interactions using Y2H screening requires spontaneous association of the partners after the transport of bait and prey proteins into the yeast nucleus, an event that is very unlikely given the type of interaction and stability of the My5 dimers.

3.3.2.1 Dimerization by three dimensional domain swapping

The observed location and mode of interaction distinguish the assembly of My5 from the previously described direct assemblies of Ig-like domains in sarcomeric proteins. In all cases, these proteins interact by the spontaneous formation of intermolecular ß-sheets between readily folded terminal Ig-like domains (this work **section 3.1**; [164][182]). My5 in contrast is located in the center of the myomesin-1 protein and dimerizes by the swapping of an N-terminal ß-strand.

This type of interaction, known as three-dimensional domain swapping has been described for more than 40 proteins of different fold [183][184]. More specifically, the dimerization mediated by the swapping of a single or multiple ß-strands has been found in a small number of structurally characterized proteins of the Ig-like superfamily (**Table 3-6**). In all cases, the interaction was observed on isolated Ig-like domains and the propensity for strand swapping could be traced back to factors that either thermodynamically destabilize the monomeric form (introduction of conformational strain by the shortening of the 'hinge loop' in CD2 [185] and Llama VHH [186]), or originate from recombinant protein expression (tagging of CD2 with dimeric GST [187]). In one case, the oligomer formation occurred only during the crystallization

process (CD47 [188]).

Table 3-6 Ig-like proteins found to interact by strand-swapping

Protein	Exp. host	Fold class	Oligomer fraction present in solution	Swapped strands	Postulated reason for swapping	Reference
CD2 (Ig1)	*E. coli*	V-set	15 %	complete	Expressed as a GST tagged protein	[187]
CD2 (Ig1)	*E. coli*	V-set	Up to 90 %	complete	2 aa deletion in hinge loop	[185]
Llama VHH	*Llama*	V-set	ND	GG'	Hinge loop truncation	[186]
CD47	*E. coli*	V-set	0 %	G	Swapping occurred during crystallization	[188]
NCAM2 (Ig1)	*E. coli*	I-set	20% (estimated from the SEC profile)	A	Formation of a secondary hydrophobic interface	[189]
CTLA-4	*E. coli*	V-set	ND	GG'	Engineered disulfide bridge	[190]
						ND: not determined

These factors are unlikely to represent the source of dimerization of My5. The proteins were tagged with either a short 6xHis-tag (My5, M4-5) or a monomeric 6xHis-SMT3 tag (My4-7). The influence of these tags on the folding process is unlikely. Both, the isolated My5 domain and multi-domain constructs were purified as a mixture of monomers and dimers. Although slightly differing between different protein preparations, mono- and dimeric forms of My5 (if the larger, potentially mis-folded form is ignored) and My4-5 appeared in an approximate 1:1 ratio (see e.g. **Figure 3-14 A** and **Figure 3-15 A**) and both forms were stable in solution. This indicates, that none of these forms is thermodynamically unstable. Moreover the lack of higher oligomers in multi-domain constructs implies that the strand swapping propensity is a unique feature of the My5 domain.

It appears therefore likely that the ability of My5 to swap its N-terminal strand is associated with its folding pathway. As My5 is apparently the only out one of five Fn3 domains of myomesin-1 (and the currently only known Fn3 domain at all) that forms swapped dimers, its pathway should be distinct from the folding pathway of other Fn3 domains.
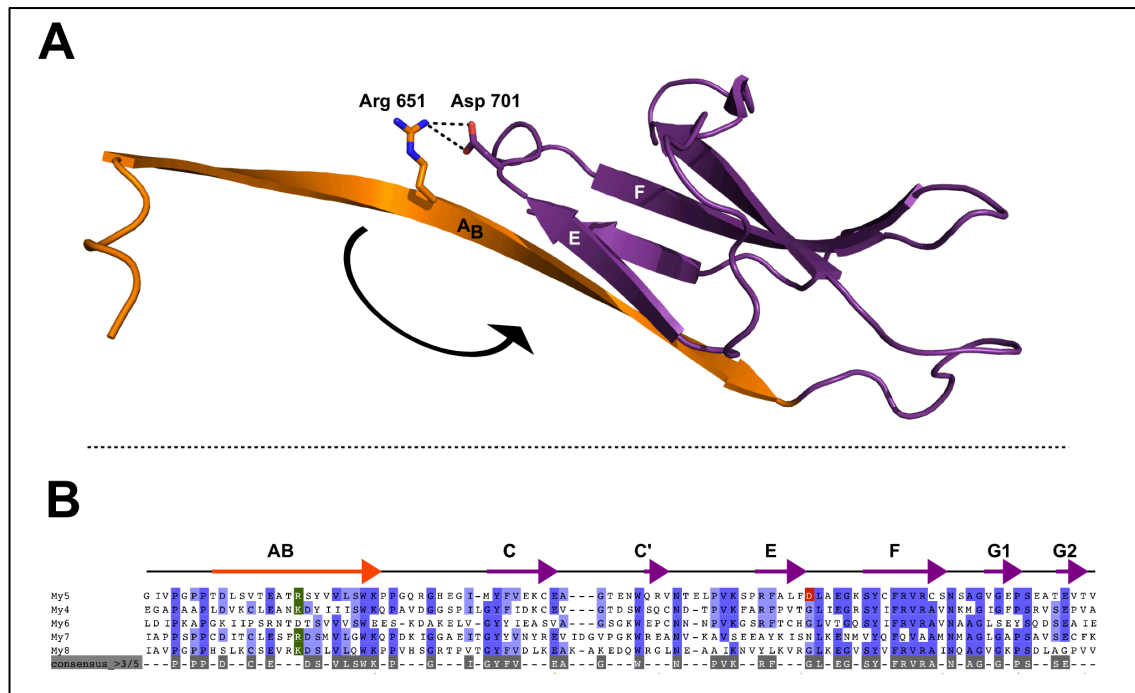
Figure 3-19 Formation of an intramolecular salt bridge in the My5 dimer.

(A) Ribbon representation of an isolated molecule of the My5 dimer showing the salt-bridge that connects the swapped AB-strand (orange) and the E-F connecting loop segment. The black arrow indicates the direction into which the N-terminal part of the AB-strand would need to turn to fold into a distinct A-strand. **(B)** Multiple sequence alignment of My5 with the four other Fn3 domains of human myomesin-1. Identical positions are shown in blue (bright: 40% -> dark: 100% identical). The positions of the salt bridge forming residues are shown in green (Arg 651) and red (Asp 701).

Being one of the nine 'superfolds' [191] which are found in a large number of evolutionarily unrelated proteins and possess a large variety of different functions [192], the folding process of Ig-like domains has been extensively studied. Based on a number of *in vitro* and *in silico* studies, it was suggested that all domains adopting the Ig-like family fold via a conserved 'nucleation - condensation' mechanism. The folding is initiated by the formation of native-like side-chain interactions between a very small number of hydrophobic residues of the domain core strands B, C, E and F. These interactions create a 'critical folding nucleus' that induces the formation of a metastable **transition state** in which additional parts of the hydrophobic domain core and the core strands reach their final positions. At this point, the parts forming the peripheral strands (strands A and G in the case of Fn3 domains) are still disordered. In a last step, these latter finally condense on the already defined parts of the transition state to fold into secondary structure elements and complete the Ig fold [193][194][195].

Hypothetically, the temporary or permanent formation of intramolecular interactions between readily defined and still disordered parts of the transition state, might delay this final condensation step of the folding process. Such an event would create an extended 'time frame' in which the disordered parts could fold into the acceptor site of a second 'open' molecule and so lead to the formation of strand swapped dimers.

A closer analysis of the My5 structure shows, that My5 indeed contains one site, which could put such temporal contstraint on the conformation of the domain's N-terminal part. This site consists of an intramolecular salt-bridge between the guanidinium group of Arg 651 in the center of the swapped AB-strand and the carboxyl group of Asp 701, which is located in the loop segment connecting strands E and F **(Figure 3-19 A)**. A sequence alignment of all myomesin-1 Fn3 domains **(Figure 3-19 B)** shows that a residue with a positively charged side-chain is present in an equivalent position to Arg 651 also in My4 (Lys 523), My7 (Arg 851) and My8 (Lys 956) while the position of Asp 701 is in three out of five domains (My4, My6 and My8) occupied by a Gly and in one domain (My7) by an Asn. The salt-bridge can thus only form in My5 of myomesin-1.

This or maybe other still unknown factors may be the origin of the unique interaction of My5 described here. In the light of the certainly too limited data which could be presented here, it is too early to draw any conclusions about the validity of the interface and the possibility that it represents a false positive result, eventually only originating from the usage of *E. coli* as expression host cannot be entirely dismissed.

However, the data presented here indicate that My5 has an unusually high propensity for domain swapping and indirect evidence for a second site of oligomerization on myomesin-1 exists. This evidence will be discussed in the next section.

3.3.2.2   Biological implications: A hypothetical model for the incorporation of tetrameric myomesin-1 in the M-bands

The M-bands represent structures that cross-link the thick filaments in the centers of the sarcomeric A-bands. Although our knowledge of their architecture and composition is certainly still limited, myomesin-1 has been proposed to act primarily as a thick filament cross-linking protein [79][181]. This assignment is based on the finding that the protein is expressed in all types of muscle fibers, in contrast to its homologs myomesin-2 and myomesin-3 [74]. Its vital importance for the sarcomeres is highlighted by the finding that sarcomeres do not form if its expression is knocked

down [60]. The cross-linking function is facilitated by its ability to bind myosin-tails through its N-terminal domain [59] and to dimerize through its C-terminal domain [79][80]. These data together with localization studies of several myomesin-1 domains [54] have been integrated into structural models describing how the protein is incorporated in the M-bands: Lange *et al.* [79] proposed a model according to which a myomesin-1 filament consisting of two tail-to-tail assembled monomers would attach to the stems of neighboring thick filaments and span between the M4 and M4'-lines. The binding site for titin M4 (domains My4-6) would create a hinge ([79] Fig. 5) between the dimeric C-terminal protein parts (My7-13), which run in parallel, and the N-terminal part (My1-4), which are oriented vertically to the thick filaments. The results presented here (section 3.2) and in [60] could not confirm a direct titin – myomesin-1 interaction. This finding has been integrated into the most recent model of the M-band [18] where the myomesin-1 dimer simply diagonally spans the distance between two thick filaments. However, under the assumption that myomesin-1 indeed represents the major physical myosin cross-link, and that a C-terminally linked dimer has a length of approximately 100 nm (24 Ig-like domains with a length of ca. 4 nm), this model displays also some weaknesses. It notably fails to explain how a dimeric myomesin-1 filament could restrain the maximum distance between the thick filaments by a simple diagonal spanning with a single attachment point on each side. X-ray diffraction experiments have shown that, even if strong lattice swelling is induced by bathing skinned muscle fibers in hypo-osmotic solutions, the axial inter-filament distances increase by only ca. 15% when compared to native fibers (measurements from muscles of different animals reviewed in [196]). A single, dimeric myomesin-1 filament could though allow for a much greater extension of the lattice, in an extreme case, such as the aforementioned swelling experiments, even up to the whole filament length. Therefore either an (unknown) additional component of the M-bands would be necessary to further restrain the axial thick filament distance, or myomesin-1 itself needs to be structurally subdivided into 'sub-filaments' of a shorter length.

The potential second site of self-assembly on myomesin-1 that we described here for the first time supports the second hypothesis. Based on the current knowledge on the self-assembly properties of myomesin-1 and the appearance of the M-bands in electron micrographs, we propose a new model of myomesin-1 organization within the M-bands. As My5 dimerizes through a strand swap, the orientation of the domain termini at opposite domain ends is preserved compared to a monomeric Fn3/Ig-like domain. This

implies a formation of parallel myomesin-1 dimers.

By C-terminal, anti-parallel association [80] of two My5 linked parallel dimers, a H-shaped tetrameric myomesin-1 complex can be created which is structurally sub-divided into two segments **(Figure 3-20)**:
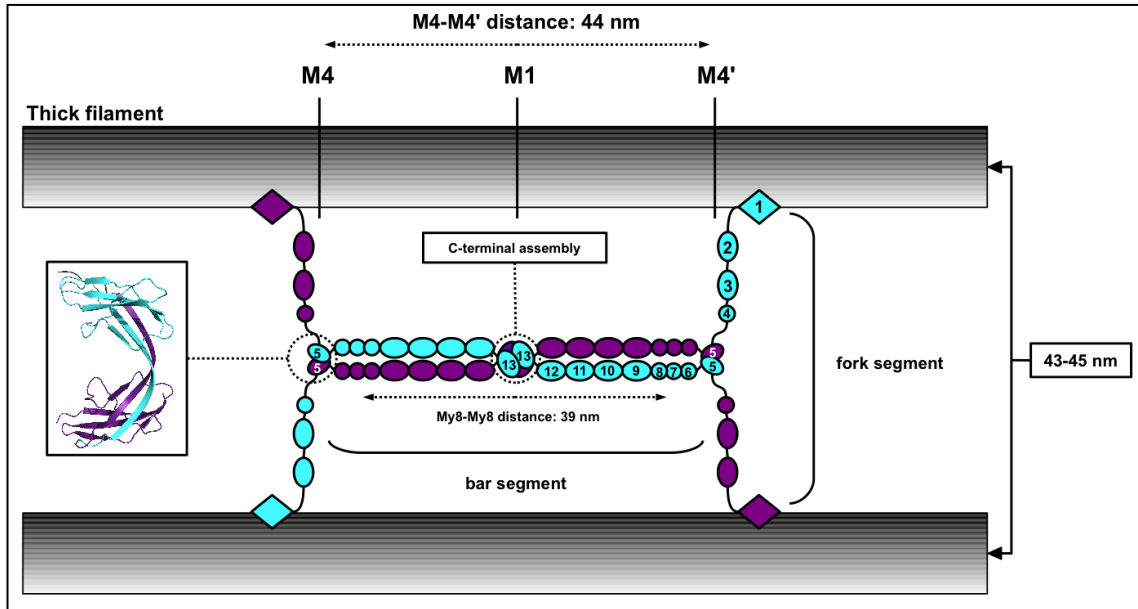
Figure 3-20 Proposed assembly of tetrameric myomesin-1.

The discovery of a second central site of self-assembly on myomesin-1 allows the construction of a myomesin-1 tetramer. One C-terminally connected dimer (cyan) assembles with a second one (purple) through two My5 oligomerization sites. From the fixation in three points (My5-dimer: My13-dimer: My5-dimer), the C-terminal parts of each myomesin-1 molecule form a central tetrameric **'bar segment'** while the N-terminal parts fixed in only one point can create two **'fork segments'** which interconnect neighboring thick filaments. The estimated locations of the sub-segments 'fork' and 'bar' are in good agreement with the appearance of the M-bands in electron micrographs where they appear as M4/M4' lines (fork) and M-filament (bar).

The central '**bar**' of the tetramer is composed of four protein segments located C-terminally of My5. Each C-terminal segment consists of an array of eight Fn3 and Ig-like domains (domains My6-13), which are only separated by short helical linkers between domains My9 and My13 [80]. As it is restrained by two My5 assemblies at its ends and two My13 assemblies at its center, the bar segment is likely to form a conformationally rather uniform unit, that upon exposure to weak forces can extend longitudinally by unfolding of the linker helices [81]. The region located N-terminally of My5, although accounting for more than one third of the protein mass, consists of only three domains with a predicted fold (2 x Ig-like, 1x Fn3) and the N-terminal

myosin-binding site of unknown structure. The folded parts are interspersed by linker regions of 30-40 residues in length. Within the myomesin-1 tetramer this part is only fixed in a single point (My5). Therefore the N-termini of each protein in the tetramer can probably extend in opposite directions and form two **'fork segments'** on either end of the bar segments which directly cross-link thick filaments.
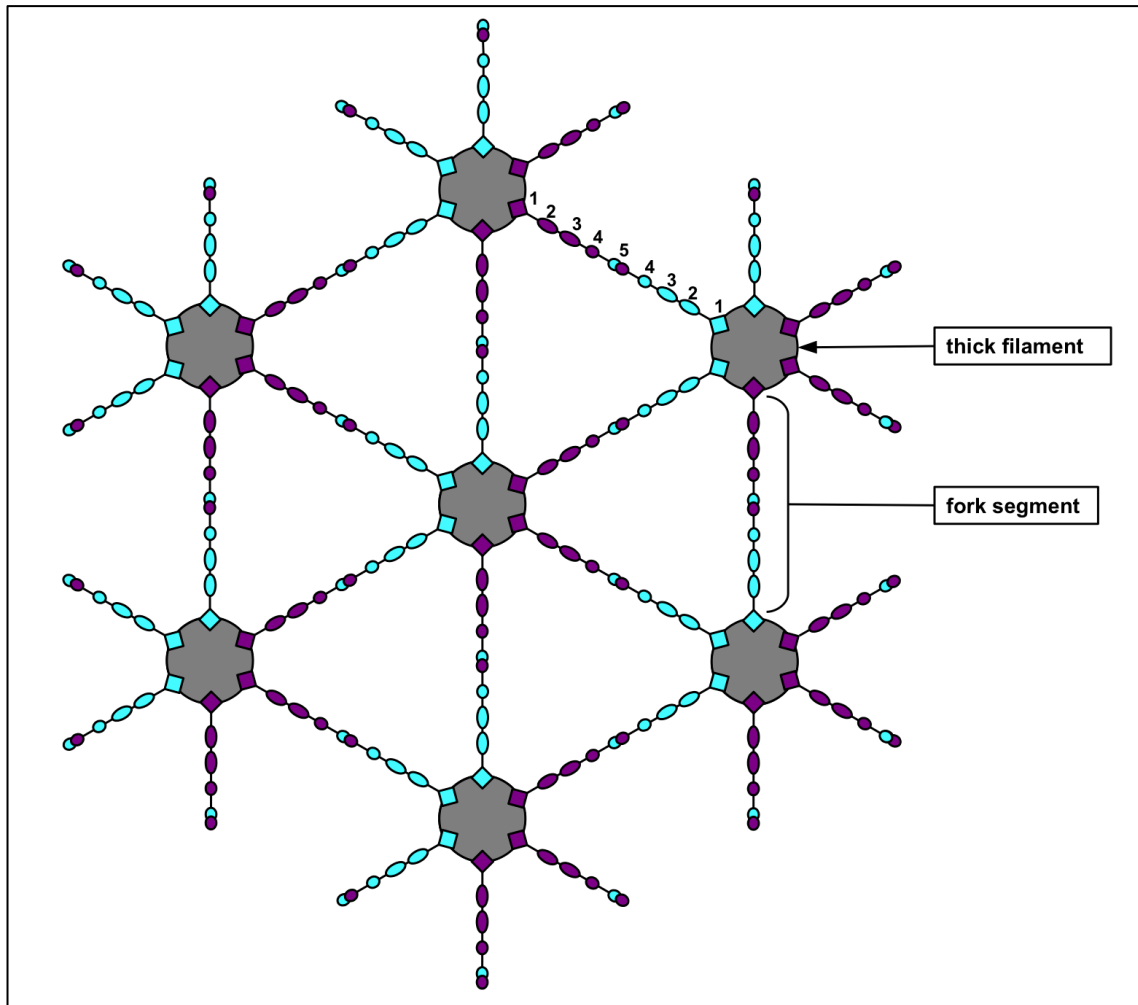
Figure 3-21 Cross sectional view of the proposed model.

The formation of a fork segment by My5-mediated dimerization of myomesin-1 allows the direct interconnection of neighboring thick filaments. In electron micrographs of myofibrillar cross-sections such direct interconnections are observable as linear density features at the levels of the M4 and M4' lines.

Thus in the proposed model, the direct connection between the thick filaments is mediated only by the My5 interconnected N-terminal parts of myomesin-1, while the C-terminally connected bar segment forms a bridge between two fork segments on both sides of the M-band center.

The proposed model is in good agreement with the electron microscopic appearance of the M-bands [54]: The **fork segment** in the model spanning between the thick filaments could correspond to the straight filament connecting density, the so called M-bridges, that can be observed in electron micrographs of myofibrillar cross-sections at the M4-line levels **(Figure 3-21).** As it runs parallel to the thick filaments in the model, the **bar segment** can, in cross-sections, only be seen as a thickening in the center of the connecting lines [1]. In a side view **(as in Figure 3-20)** both segments would be visible as densities that vertically connect (**fork**) and density that runs parallel to the thick filaments (**bar**). Such density has indeed been described from longitudinal sections as filaments connecting M4/M4' lines (vertical density) and M-filament (parallel density) [14]. Thus, if the postulated model is correct, the M4/M4'-line cross-bridges and the parallel M-filament described as distinct structures [14] represent different sub-structures of the same tetrameric myomesin-1 complex.

The distances of different myomesin-1 epitopes from the M-band center are in good agreement with the model: N-terminal myomesin-1 epitopes assigned to the fork segment have been found to show only moderate variation of distances to the M1-line (My1: 27.1 nm; My4-5 linker 19.9 nm) while epitopes assigned to the bar segment protrude towards the M1-line (My8: 13.2 nm; My11: 3.1 nm) [54].

Further support for the model comes from work on the structure of the C-terminal region of myomesin-1 (S. Chatziefthimiou, EMBL-Hamburg; unpublished data). A segment comprising domains My8-13 has in its My13-mediated dimeric form a length of 39.5 nm **(Figure 3-20, arrows)**. This is in good agreement with the described distance of 44 nm between the M4 and M4' lines [14] that would be, according to our model, bridged by the C-terminal **bar segment** (My6-13).

3.4 Crystal structure of titin M4

3.4.1   Results

3.4.1.1   Crystallization

Small crystals of titin M4 initially grew in drops from high throughput crystallization trials of different myomesin-1 constructs mixed in a 1:1 molar ratio with titin M4 at a protein concentration of 10 mg/ml. These crystals had grown after several days in sodium citrate pH 4.0 and 0.5 M $(NH_4)_2SO_4$ or MES pH 6.5 and 1 M NaCl respectively. Both forms did not diffract beyond 5 Å resolution. After it had turned out that myomesin-1 and titin do not form a complex and that the crystals contained only the M4 domain, the crystallization conditions were optimized manually using the hanging drop vapor diffusion method and different concentrations of pure, monomeric M4. The optimization trials led to crystals of a larger size and better diffraction quality in two conditions **(Figure 3-22)**:

(1) Rhombic, ca. 0.3x0.15x0.15 mm:

> 10 mg/ml Titin M4; 0.1 M Na_3-Citrate pH 4.0, 0.2 M $MgSO_4$, 0.3 M $(NH_4)_2SO_4$

(2) Hexagonal rods, ca. 0.5x0.25x0.25 mM:

> 15-20 mg/ml Titin M4; 0.1 M MES pH 6.5, 2 M RbBr

Both crystal forms grew to their final size within 1-2 days.
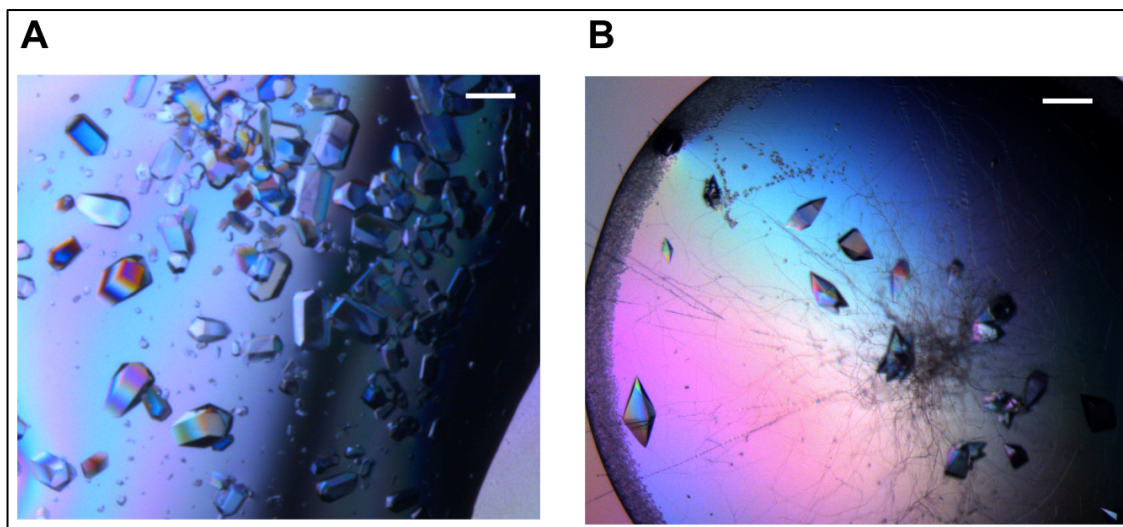
Figure 3-22 Crystals of titin M4.

M4 crystallized in a trigonal **(A)** and a tetragonal **(B)** form. The white bars indicate a length of 0.2 mm.

3.4.1.2   Data collection, structure solution and model refinement

The structure of the M4 domain was solved from hexagonal rod shaped crystals in the trigonal space group $P3_221$ by the SAD method using the anomalous signal of Br-ions. The data were collected at the Br-peak wavelength on beamline X12, integrated with XDS and scaled with Scala. Sub-structure determination with SHELXD [197] found 21 heavy atoms and phases calculated by SHELXE [198] after density modification were of good enough quality to allow the building of a nearly complete model (96/103 residues) with ARP/wARP [98]. The model contained one M4 domain per asymmetric unit. It was manually completed with Coot and refined with phenix.refine to a resolution of 1.9 Å. Despite the good agreement between the protein component of the model and the electron density map, the model could not be refined to R-factors below $R_{work}/R_{free} =$ 0.26/0.28 because of the following two reasons: **(1)** In the crystal, the protein molecules are arranged to form the wall of a hexagonal 'tube', with a 70 Å large solvent channel (overall solvent content > 70%) in its center. This channel contains a large amount of non-interpretable electron density which could originate from partially ordered solvent, but also from disordered protein chains present within the channel. **(2)** The problem was exacerbated by the presence of highly concentrated heavy atoms (2M Rb^+ and Br^+ ions)

in the crystal. Although a number of heavy atoms located close to the protein could be clearly modeled into the electron density according to their appearance in an anomalous difference map, a large number of anomalous difference peaks remained, presumably representing partially occupied or high B-factor ions.

To facilitate model refinement, the RbBr (which appeared to be crucial for the growth of large and well diffracting crystals) was exchanged to 2 M NaCl by soaking the crystals four times for ca. ten minutes in a solution containing 0.1 M MES pH 6.5, 2M NaCl directly before flash freezing in the same solution supplemented with 20% v/v glycerol. The absence of Rb^+ and Br^- ions was then confirmed by the measurement of X-ray fluorescence around the corresponding elemental absorption edges. Although the efficient removal of RbBr could be thus confirmed, a residual anomalous signal was still present. A broader X-ray fluorescence scan revealed the presence of zinc in the crystal **(Figure 3-23)**. To locate the zinc ion, additional data were collected below and above its elemental absorption edge.
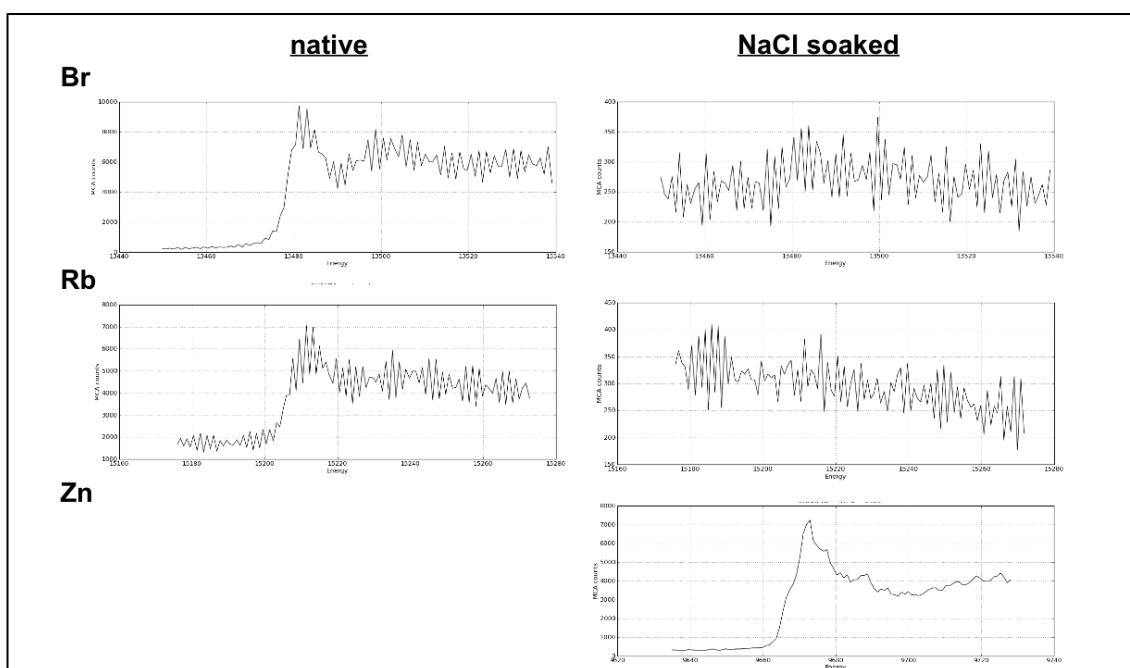
Figure 3-23 Confirmation of efficient ion exchange in trigonal M4 crystals.

X-ray fluorescence of Br^- and Rb^+ -ions measured around the peak wavelengths of each element is no longer detectable after back-soaking with NaCl, while Zn-fluorescence is still strong. Measurements were carried out at the ESRF beamline ID23-1. X-axis: photon energy [eV]; Y-axis: photon count []

Atomic positions, individual B-factors and TLS-parameters of the model could subsequently be refined against these 'native' data to a maximum resolution of 1.6 Å and lower R-values were obtained although a significant amount of non-interpretable electron density remained in the solvent channel. Surprisingly, this higher resolution native data revealed the presence of a second mainly disordered protein chain in the solvent channel, as indicated by the appearance of electron density with ß-strand like connectivity.

Crystals of the rhombic crystal form were cryo-protected in mother liquor supplemented with 20% (w/v) glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected to a maximum resolution of 2.0 Å on beamline ID14-1. The structure was solved in the tetragonal space group $P4_12_12$ by molecular replacement with PHASER using the final model from the trigonal form as search model. Atomic coordinates, isotropic B-factors and TLS-parameters were refined with phenix.refine. Data collection, phasing and refinement statistics for both crystal forms are summarized in **Table 3-7**. Model statistics for both crystal forms are summarized in **Table 3-8**.
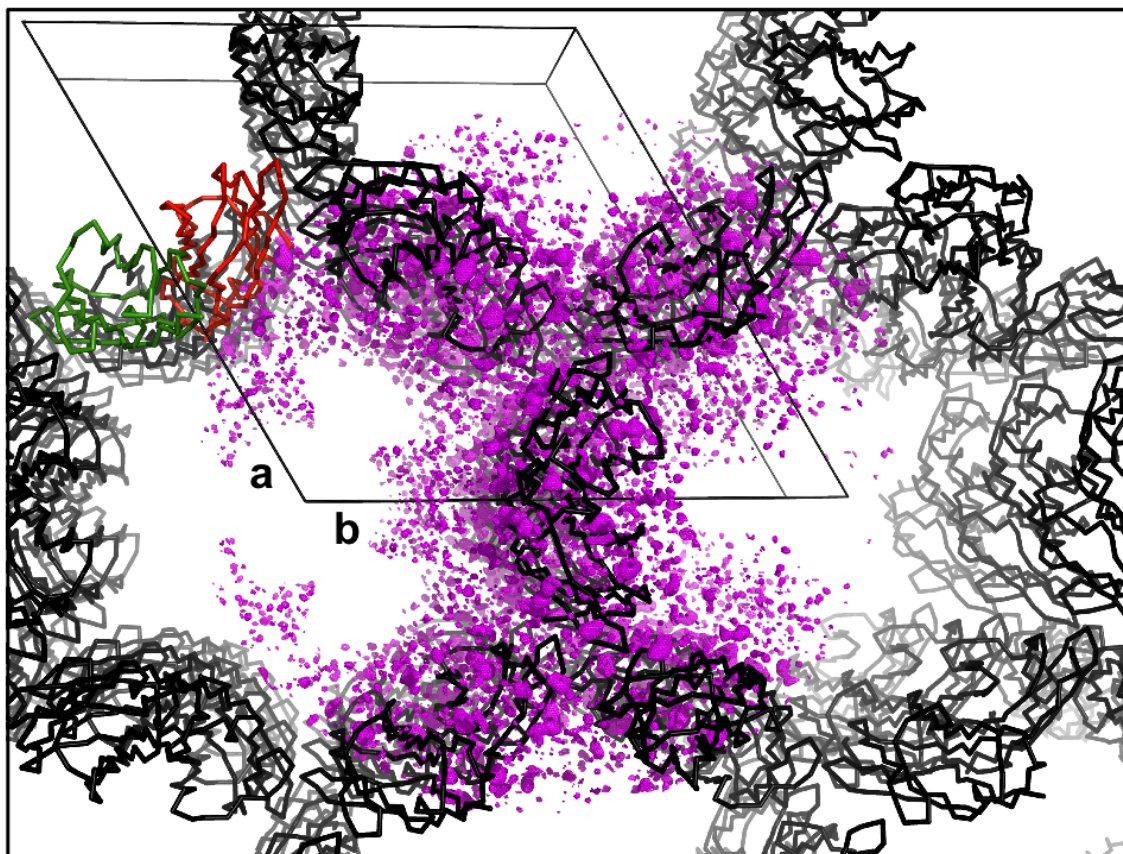
Figure 3-24 Crystal packing and anomalous scattering of the trigonal crystal form.

A view along the unit-cell c-axis shows that M4 packs in the trigonal crystal form as a tubular structure with >70% solvent content. The resulting large solvent channel of approximately 70Å diameter contains a large amount of non-interpretable electron density. An anomalous difference map calculated from data collected at the Br-peak wavelength (magenta, σ level = 2.5) indicates that the channel is filled with anomalously scattering ions in only partially occupied positions or displaying high B-factors. For clarity only the C-alpha traces of the protein are shown (black). Two individual molecules are shown in green and red.

Table 3-7 Data collection and phasing statistics for titin M4.

	Tetragonal form			Trigonal form	
	native	Br-peak energy	Zn-peak energy	Zn-low energy	NaCl back-soaked
Data collection and processing					
Beamline	ID14-1	X12	MX2		ID23-1
Detector	ADSC Q210	Mar mosaic 225	Rayonix 225 HE		ADSC Q315r
Wavelength [Å]	0.934	0.917	1.258	1.311	1.286
Resolution range [Å]	48.6-2.0 (2.1-2.0)	42.4-1.9 (2.0-1.9)	19.5-1.95 (2.06-1.95)	19.5-2.1 (2.21-2.1)	42.4-1.6 (1.69-1.6)
Space group	$P4_12_12$		$P3_22_1$		
Unit cell dimensions [Å]	a/b = 63.5, c = 151.1	a/b = 84.8, c = 49.0	a/b = 84.4, c = 49.9		a/b = 84.9, c = 49.9
Mosaicity (°)	0.19	0.18	0.48		0.27
No. of measured reflections	166745 (22646)	98011 (14301)	314399 (41612)	256994 (34975)	307132 (20252)
No. of unique reflections	21839 (3050)	16297 (2343)	15171 (2198)	12184 (1751)	27114 (3942)
Redundancy	7.6 (7.4)	6.0 (6.1)	20.7 (18.9)	21.1 (20.0)	11.3 (5.1)
Mean $I/\sigma(I)$	21.3 (4.8)	24.2 (4.0)	36.3 (6.0)	40.9 (8.0)	24.8 (3.1)
Completeness [%]	99.8 (98.9)	99.9 (100)	99.9 (100)	99.9 (100)	99.6 (99.9)
R_{merge} [%]	8.2 (49.3)	4.9 (43.5)*	5.7 (55.5)	5.5 (43.4)*	5.1 (45.8)
R_{meas} [%]	8.8 (52.9)	5.9 (52.6)*	6.0 (58.6)	5.7 (45.7)*	5.3 (50.5)
$R_{p.i.m.}$ [%]	3.1 (18.4)	3.3 (29.3)*	1.8 (18.6)	1.7 (14.1)*	1.5 (20.3)
Overall B-factor from Wilson plot	21.5	24.4	28.2	33.7	24.4
Phasing					
Figure of merit (SHELXE)		0.698			
Pseudo free correlation coefficient (SHELXE)		73.74			
Number of sites (SHELXD)		21			

numbers in brackets correspond to the highest resolution shell
(*: calculated from separate I^+/I^-)

Table 3-8: Refinement and model statistics for M4.

	Tetragonal form	Trigonal form
Refinement:		
Resolution	48.6 – 2.0	27.8 – 1.60
No. of reflections (free)	21658 (1100)	27076 (1741)
R_{work} / R_{free} [%]	17.5 / 21.4	21.2 / 22.2
No. of atoms: protein / ions / solvent	2310 / 35 / 291	838 / 4 / 185
Model:		
Average B-factor: protein / ligand / solvent [Å^2]	22.4 / 39.4 / 33.0	32.5 / 28.2/ 46.8
R.m.s. deviations		
Bond lengths [Å]	0.006	0.006
Bond angles [°]	1.038	1.115
Ramachandran plot		
Favored [%]	99.64	99.04
Additionally allowed [%]	0.36	0.96

3.4.1.3   Overall structure

The crystallographic models of M4 contain three (tetragonal form; **Figure 3-26 A**) or one (trigonal form; **Figure 3-26 B**) copies of the domain (titin Q8WZ42 residues 33294-33395; here renumbered 1-103 including an N-terminal Gly residue originating from the 3C-cleavage site) in each asymmetric unit. The models cover the complete sequence except for a small number of residues located at the N- and C-termini of the crystallized protein **(Table 3-9)**.

Table 3-9: Completeness of M4-models

		sequence coverage
tetragonal	**A**	1-101
	B	1-100
	C	4-100
trigonal	**A**	4-99

The four molecules are structurally highly similar to each other (maximum r.m.s. deviation between Cα-atoms: 0.9 Å). Significant structural differences are present in the flexible N-termini and the conformations of surface exposed side-chains.

Both models show that M4 folds with a Greek-key like ß-sandwich topology that is characteristic for Ig-like domains [192]. Eight strands form two separate open ß-sheets ABED and CFGA' **(Figure 3-26 B)** in which all strands except strands A' and G on the second sheet assemble in an anti-parallel way. The division of the N-terminal A-strand into two separate strands, A and A', which contribute to the formation of opposite ß-

sheets and the lack of an outer C'' strand identify the domain to be a member of the I-set of the Ig-like folding family [155].
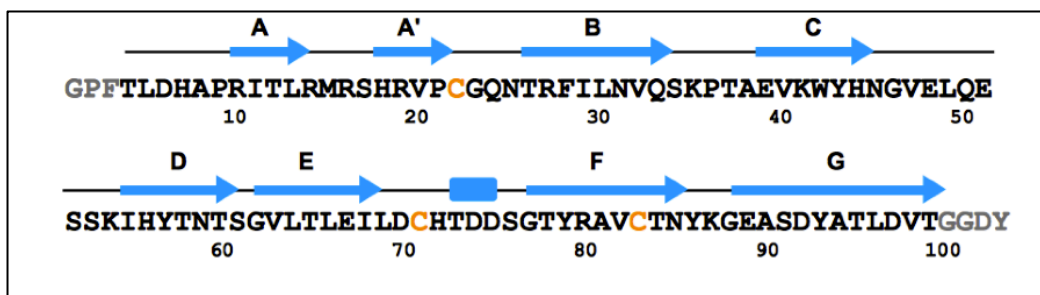
Figure 3-25 Schematic representation M4 secondary structure.

The domain folds into an eight-stranded Ig-like domain (blue arrows A-G). A small helix is present in the E-F connecting loop. The positions of Cys residues are marked in yellow. Residues not modeled in the trigonal form are shown in grey.

In addition to the single fully ordered chain, weak electron density of a second protein molecule has been found in the trigonal form. This electron density appears to originate from two partially ordered anti-parallel ß-strands, which are connected through a single long side-chain (modeled as Arg) pointing towards Asp 97 in strand G of the ordered molecule **(Figure 3-26 C)**. Although in absence of clear side-chain density the identity of this second molecule could not be determined, it is likely that it represents a second, mainly disordered M4 molecule, which is connected to the ordered one by only a single salt bridge. The discovery of this molecule within the solvent channel provides, in addition to the large amount of ordered solvent (185 water molecules on 108 modeled amino acids), a possible explanation for the only mediocre R-values that could be reached during refinement. Although Rb^+ and Br^- ions had been efficiently removed (and surprisingly data to a higher resolution could be measured) after the back-soaking procedure, the model could not be refined to R-values below 20 as is usually expected with this resolution and data quality.

The M4 domain contains three Cys residues. Cys 83 is part of strand F and has its thiol group directed to the domain core. The second conserved Cys, usually found on a ß-strand of the opposite ß-sheet in various I-band titin Ig-like modules and pointing inward [199], is not present in M4. The condition for reversible intramolecular disulfide formation with a fold stabilizing effect under oxidizing conditions [200] is therefore not given. Instead, the two remaining Cys are located at unusual positions at the C-terminal domain tip. Cys 22 is part of the strands A'-B connecting loop and its thiol group is

exposed at the domain surface, while Cys 71 is located in the strands E-F connecting loop with its thiol group directed towards the domain core.
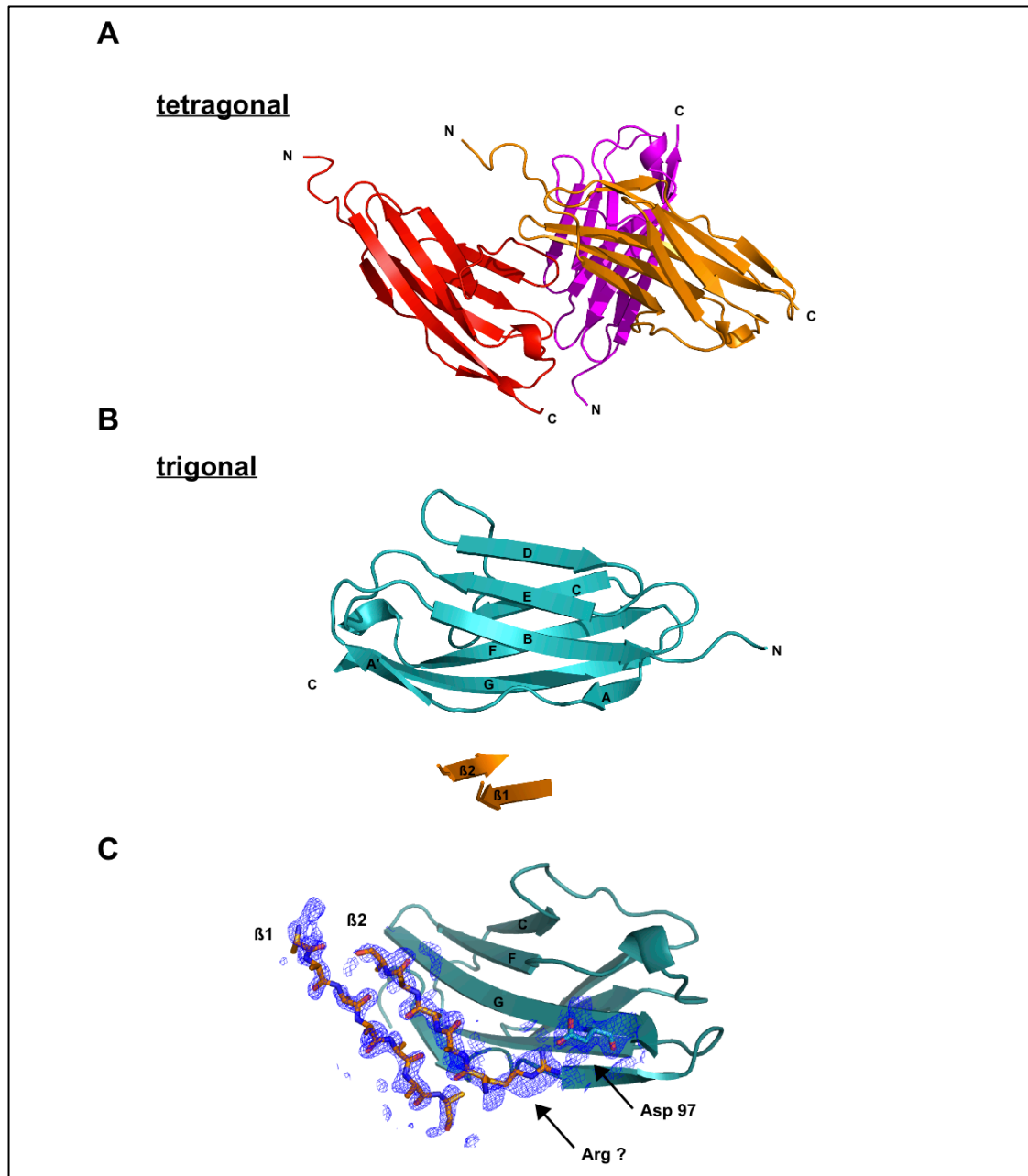
Figure 3-26 Crystal structures of the M4 domain.

Ribbon representations of the crystallographic models. The model from the tetragonal crystal form **(A)** contains three domains in each asymmetric unit. The trigonal form **(B)** contains a single ordered chain with additional 12 additional modeled residues of unknown origin (orange strands ß1 and ß2). **(C)** 2Fo-Fc map of the refined model at a sigma level of 1.0 (blue) indicates the presence of a second protein chain (orange). As no clear side-chain density could be observed for 11 residues they were modeled as Ala. A single long side chain (modeled as an Arg) seems to connect to Asp 97 of the ordered chain. The model is related to the model in B by a rotation of 180° around the horizontal axis.

3.4.1.4   Dimerization of M4

The M4 domain was purified as a monomeric protein **(Figure 3-12)** and crystallized as a monomer in the tetragonal form. In the trigonal crystal, however, it formed a dimer between two molecules related to each other by a crystallographic two-fold axis. The dimerization is mediated by a small symmetric interface. It consists of two sites of specific interaction on the loops connecting the A'-B and E-F strands which are both directed towards the C-terminal domain end.

The **first** interaction involves a disulfide bridge, formed between the surface exposed thiol groups of Cys 22 on the tip of the loop connecting the A'-B strands of both molecules. Apart from the repositioning of the Cys 22 side-chain, the disulfide formation does not induce any conformational change in the A'-B connecting loop when compared to the domains in the tetragonal crystal form. The covalent bond has formed between the two molecules despite the fact that the protein has been crystallized in a reducing environment (at a starting concentration of 0.5 mM TCEP in the crystallization drop) and within a short time. The disulfide bridge connects the two molecules at their C-terminal side to form a tip-to-tip dimer in which the molecules are twisted against each other by 180° along their N-C axis **(Figure 3-28 A/B)**.

The **second** attachment site is created by the twisted arrangement. A zinc ion is coordinated with a distorted tetrahedral binding geometry by the carboxyl group of Asp 70 and the imidazole group of His 72 on the tip of the strands E-F connecting loop of each domain. The distances observed between the Zn-coordinating atoms of both residues are in good agreement with documented ideal values [201] for zinc coordination and clearly show that Asp 70 is a monodentate Zn-ligand **(Table 3-10)**.

Table 3-10 Zn distances in the M4 coordination site

Residue	Atom	Zn distance	Ideal value *
Asp 70	OD1	2.11 Å	1.99 Å
	OD2	2.94 Å	< 2.5 Å**
His 72	NE2	2.19 Å	2.03 Å
Cys 22	SG	5.30 Å	2.31 Å

* ideal distances calculated by [201] from structures with a resolution < 1.25 Å

** largest accepted distance in [201] for consideration as bidentate coordination for Asp or Glu

The presence of Zn ions in the crystal was initially found by an X-ray fluorescence scan and the location of one ion in the domain interface was confirmed by calculation of anomalous difference maps from data collected at Zn-peak and Zn-low energy wavelengths **(Figure 3-27)**.
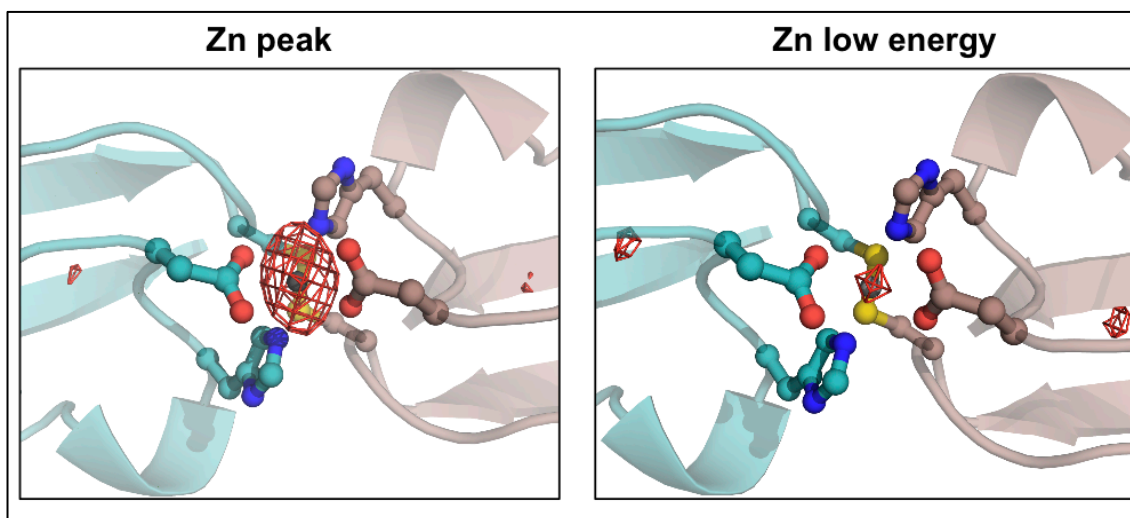
Figure 3-27 Zinc coordination in the M4 dimer.

The coordination of a zinc ion by Asp 70 and His 72 was confirmed by the calculation of anomalous difference maps from data collected at Zn-peak (1.258 Å; theoretical f''= 3.7 e⁻) and Zn-low energy wavelengths (1.311 Å; theoretical f''=0.51 e⁻). The maps, here shown at 5σ (red mesh), display a strong anomalous signal at the peak wavelength. The signal is nearly lost by an X-ray energy shift below the absorption peak. The weak difference peaks visible on both sides of the interface represent the positions of the sulfur atoms of Cys 71.

As it has not been added to the crystallization solution, the **source of the bound zinc** in the crystal is not entirely clear. It is unlikely that the metal has been co-purified from the recombinant source as the binding site requires a M4 dimer to be formed and the protein has been dialyzed against a cleavage buffer containing 1mM EDTA during the proteolytic cleavage of the His-tag. One possibility is that the metal has been added with the RbBr salt used for crystallization. According to the manufacturer's description, it contained up to 0.1% (w/w) of Zn-salts. The usage of 2 M RbBr for crystallization could therefore have led to a final concentration of up to 2 mM Zn^{2+} ions in the crystallization drop.
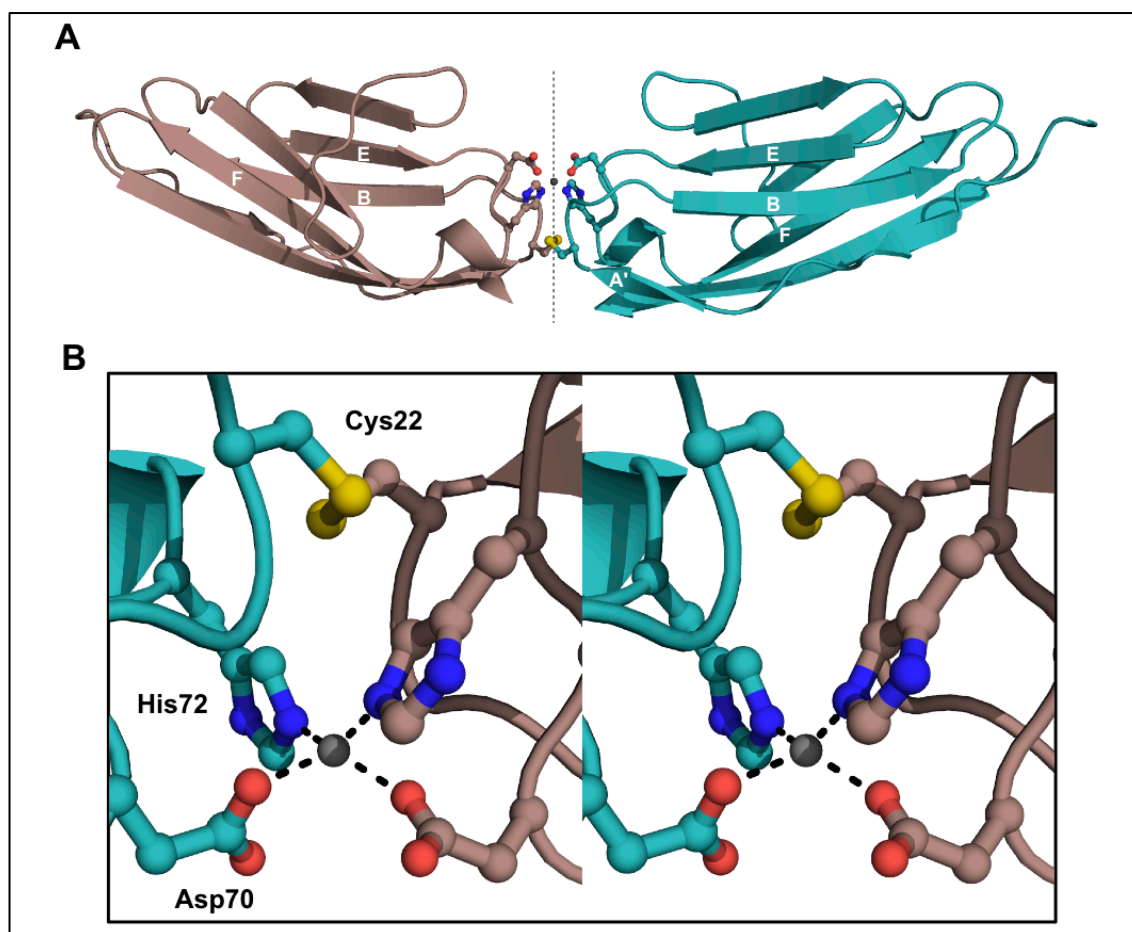
Figure 3-28 Dimerization of titin M4.

(A) Structural overview of the M4-dimer identified in the trigonal crystal form. Both monomers are related to each other by a crystallographic two-fold axis (dashed line). The interaction is mediated by the side chains of only three residues on the E-F strand and A'-B strand connecting loops and a single Zn^{2+} ion. **(B)** Detailed stereo view on the interface. Side chains of interfacing residues are shown with ball and sticks. Note that the coordinated Zn^{2+} ion and the disulfide bond lie on the same twofold axis.

3.4.1.5   The position of Cys22 is unique within titin Ig-like domains and conserved among vertebrates

The main interaction between the two M4 molecules in the dimer, whose existence is reported for the first time in this work, is mediated by the formation of a disulfide bridge between the side chains of Cys 22. As the dimeric form has only been found in one out of two available M4 crystal structures, and the protein has been purified as a monomer, it seems likely that the dimer is an artifact of the crystallization process. However, the tetragonal crystal form in which M4 is present as a monomer, was grown

at a non-physiological pH of 4.0, and the protein was purified in the presence of 1mM TCEP. Both factors are very likely to prevent disulfide bond formation. The inverse assumption that, if dimer formation had been observed *in vitro* in solution, this would indicate that M4 could dimerize *in vivo* is not valid. With Cys 22 being the only surface exposed Cys on M4, it constitutes the only candidate for potential formation of a disulfide bond under oxidative conditions.

In the absence of experimental data showing whether M4 can dimerize *in situ* in the M-bands, the existence of a dimer observed in the trigonal crystal form cannot be undoubtedly confirmed. However, a comparative analysis can provide at least qualitative evidences on whether the observed interaction is likely to play a functional role or simply represents an experimental artifact.

In a recently published large scale bioinformatic study, the location of cysteines and their clustering in proteins within available structures and homology models of orthologous proteins were examined for their correlation with this residue's conservation [202]. The authors found that Cys has the tendency to be either a highly or a poorly conserved residue. The conservation is strongly correlated with its position in the protein and its proximity to other cysteines. Cys with a presumed functional role, e.g. spatially clustered Cys involved in the formation of fold-stabilizing disulfide bridges or part of metal coordination sites, were found to be the most conserved residues of all. Single Cys that are buried inside the protein showed conservation levels, which are similar to most other amino acids. In contrast, isolated, surface exposed Cys were found to represent the most degenerated positions among all residues. This may, as the authors speculate, be due to the reactivity of the exposed thiol group that may lead to undesired effects for the protein under oxidative conditions, which create a comparably high position-specific selective pressure.

Thus, if M4 Cys 22 possesses a specific functional role, two conditions which are testable by simple comparisons should be fulfilled:

(1) Residues required for the interface formation should be a unique feature of the M4 domain.

(2) Cys22 should have been more conserved during protein evolution than comparable, surface exposed cysteines.

(1) Uniqueness of the interface:

A structure- and sequence-based multiple alignment of M4 and the 17 structurally

characterized Ig-like domains of titin (mouse: I65-I70: [63]; human: I1: [200]; I27: [203]; Z1Z2: [47]; A168-A169: [204]; M5: [29]; M10 [182]; M1: unpublished PDB 2BK8; A164-A165: unpublished PDB 3LCY) was generated. The comparison shows that the potentially zinc-coordinating residues (Glu, Asp, His, Gln, Asn) can be found in homologous positions of the M4 Zn-binding site in nearly all Ig-like domains of titin. In contrast, a Cys in an A'-B-strand connecting loop has not been observed previously **(Figure 3-29)**. The inspection of the coordinates revealed that apart from M4 Cys 22, none of the side-chains of Cys residues on those domains are exposed at the protein surface. A structural superposition with Dali [103] shows, that the fold of M4 is highly similar to all other structurally characterized titin domains. Despite primary structure identity levels ranging from only 22% (M4 : I70) to 35% (M4 : I1), the M4 model can be superimposed with r.m.s. deviations between Cα atoms of 0.9 Å (M4 : A168) to 1.7 Å (M4 : A164). A sufficiently accurate mapping of the positions of A'-B strands connecting loops from the primary structure of predicted titin Ig-like domains was therefore possible. The mapping was performed by alignment of the M4 sequence against all 152 Ig-like domains of human titin (Uniprot Q8WZ42) as predicted by SMART/Pfam [110] with blastp [105]. Cysteines in regions that are likely to fold into an A'-B connecting loop could not be detected in any predicted titin Ig-like domain. The uniqueness of the interface within human titin is thus demonstrated.
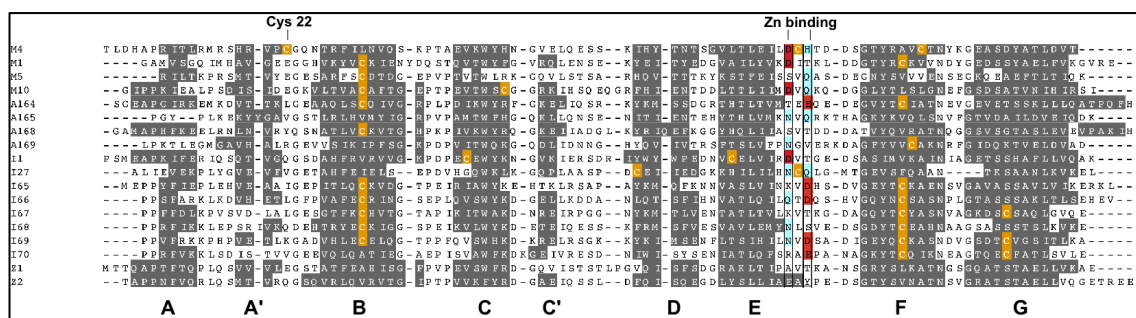
Figure 3-29 Structure based sequence alignment of titin domains of known structure.

A sequence alignment between M4 and 17 titin Ig-like domains of known structure reveals that a Cys in the A'-B strands connecting loop is present only in M4. ß-strands are marked in grey and labeled in the bottom row. Cysteines are highlighted in orange and potentially zinc-coordinating residues in homologous positions to the M4 binding site are highlighted in cyan (His/Gln) or red (Asp/Glu). Figure prepared with Jalview [205].

(2) Cys22 conservation

On M-band titin, the M4 domain is separated from the adjacent domains M3 and M5 by the presumably non-folded Is2 (approx. 500 aa) and Is3 (approx 100 aa) linkers which both do not contain any cysteines. Thus Cys 22 on M4 is within more than 70 kDa of M-band titin the only Cys with a surface exposed side-chain. Clustering with a Cys from an adjacent part of titin or burial from the solvent by an adjacent folded domain is therefore unlikely.

Consequently Cys 22 belongs to the group of isolated, exposed cysteines and should, according to [202] be poorly conserved if it doesn't play any functional role. On the other hand, if the observed formation of a disulfide has relevance *in vivo,* Cys 22 should be comparably well conserved.

A statistical evaluation of how well surface exposed Cys positions in sarcomeric proteins are preserved is for two reasons currently not possible. First, a reasonable number of structures where such positions have been described are not available. Secondly, titin is exclusively expressed in vertebrates. Although invertebrate titin homologs as twitchin (*C. elegans*) and projectin (*D. melanogaster*) and molluscan mini-titins are known [206], these are far shorter than vertebrate titin and homologous regions from which the M-band part of vertebrate titin has evolved cannot be reliably determined.

However, it is possible to gain qualitative insights on whether Cys 22 on M4 is well conserved. The crystal structure of the My5 dimer described in this work showed that myomesin-1 contains a single Cys with a surface exposed side chain in position 709 **(section 3.3)**. This residue on strand F of My5 is currently the only known Cys with a surface exposed thiol group amongst M-band proteins of known structure. It should therefore be exposed to similar oxidative stress as M4 Cys 22 and so be subject of similar selective pressure.

The M4 and My5 homologous sequences were identified by a search with the tblastn tool [105] in the NCBI genome and genomic shotgun library sequence databases. A dataset of homologous sequences from 36 species of all vertebrate classes was compiled and a multiple sequence alignment was created with MUSCLE [207].

Table 3-11 Cysteine conservation in M4 and My5

	M4			My5		
Overall identity	47/102			52/102*		
Cys conservation	Cys 22 (exposed)	Cys 71 (buried)	Cys 83 (buried)	Cys 676 (buried)	Cys 709 (exposed)	Cys 714 (buried)
	31/36	36/36	36/36	30/32*	10/36	35/36
additional degenerated Cys positions	5			2		

* calculated only from available sequences

The alignments **(Figure 3-30)** show that the overall conservation levels of both domains are similar. In My5 52 out of 102 and in M4 47 out of 102 aligned residues were found to be identical in all aligned sequences of My5 and M4, respectively. Non-exposed cysteines (M4 Cys 71 and 83; My5 Cys 676 and 714) are virtually completely conserved among all species. M4 Cys 22 is present in 31 out of 36 aligned sequences (89% conservation) and is absent in only two out of 32 sequences (cat and dog) from terrestrial vertebrates (reptiles, birds and mammals). In contrast, My5 Cys 709 is only present in ten out of 36 sequences (28% conservation). Six out of these ten Cys 709 containing My5 sequences originate from primates and the remaining four from rodents and birds. Furthermore, the alignments show that a number of additional Cys sites exist in different species on both domains. Sites with (based on the present structural models) presumed surface exposed and buried side-chains exist, but in all cases these are only poorly conserved. For example in position 27 on the B-strand of M4, an exposed Cys is present in four mammalian species of different families, while in all other, this position is occupied by a Arg, Lys or His.

Thus, the position of M4 Cys 22 is nearly as well conserved as buried cysteines while the position of My5 Cys 709 and other Cys are more degenerated.

Figure 3-30 Conservation of M4 and My5.

An alignment of sequences of M4 and My5 from 36 vertebrate species shows that the positions of all three cysteines (yellow) of human M4 have been well conserved throughout titin evolution. In contrast, in My5 only the two buried Cys show high conservation levels while the surface exposed Cys 709 is conserved in only one third of all sequences. The M4 zinc-binding site (Asp 70 / His 72; red and pink) is conserved in all but one of the M4 sequences (*Oryzias latipes*). In one third of the sequences His 72 is substituted by also potentially Zn-coordinating Gln or Asn. Figure prepared with Jalview [205]. Conservation levels: dark blue – highly conserved; white - not conserved;

Titin M4

Myomesin-1 My5

3.4.2   Discussion

The solution of the M4 crystal structure revealed the previously unreported ability of the titin M4 domain to form dimeric assemblies. The tip-to-tip interaction is mediated by only six residues, three on each molecule, that form a disulfide bridge and a symmetric zinc coordination site between equivalent residues of the two monomers. A search in the protein interface database SCOPPI [208] did not reveal a single structure in which both a metal coordination site as well as a disulfide bond contribute to the formation of an Ig-like domain dimer. Purely disulfide or metal mediated oligomerization though has been described in a number of cases. For example, the tenth Ig-like domain of human filamin A was found to form a tip-to-tip dimer which is linked by a disulfide between homologous Cys on the E-F strands connecting loop in a crystal structure [209]. The authors argue that the observed bond is a crystallographic artifact, as dimerization has not been observed under reducing conditions (in 5 mM BME) in solution.

In the light of the presented data, the possibility that the observed M4 dimer is non-relevant *in vivo* cannot be excluded. However, as shown, the M4 binding site is comparably well conserved, and intra- as well as inter-molecular disulfide bond formation, although classically seen as being unlikely in the reducing environment of the eukaryotic cytoplasm, has been shown to occur at least temporarily and to modulate the tertiary and quaternary structure of proteins in the form of so-called thiol-based redox switches [200][210][211][212]. In the following section, existing evidences indicating whether M4 could indeed dimerize *in vivo* will be discussed.

3.4.2.1   The M4 assembly may represent a temporary stable anchor of titin in the M-bands

Due to the fact that titin spans half of the sarcomere, the protein has been proposed to act as a third filament that provides striated musculature elasticity and resistance against passive tension [24]. While elasticity is believed to be mainly provided by extensible parts of the protein located in the I-band region (reviewed in [65] and [66]), the protein's anchoring on both ends of the half-sarcomeres is only partially understood. In the sarcomere bordering Z-discs, titin is anchored by direct interactions with alpha actinin 2 [40][42] and by the indirect association with itself through the trimeric assembly of two titin N-termini with the protein telethonin [45][47]. The question as to

how its C-terminus is physically anchored or whether it is anchored at all remains to date unanswered. Early electron microscopic work on natively purified titin showed that a fraction of the protein existed in stable oligomeric forms (here named titin IIa). In these oligomers two or more molecules appeared to be cross-linked at their C-terminus by a globular 'head-domain' that contained small amounts of myomesin-1 and moyomesin-2. This head could be removed only by protease treatment [36]. Stable interactions at the C-terminus can thus be inferred.

However, the two described interactions of titin with obscurin/OBSL1 and myomesin-1 are unlikely to provide this kind of stability. A direct interaction with myomesin-1 is unlikely to exist (this work section 3.2 and [60]), and the assembly of the M10 domain with the N-terminal domains of obscurin and OBSL1 has been shown to provide only resistance against moderate forces [167].

The covalent complex between two M4 domains described here could provide the molecular basis for how, at least temporarily or under certain conditions, a physically strong titin anchoring in the M-band could be achieved. If the M4 dimer represents a true assembly *in vivo*, three conditions must be fulfilled for its formation: **(1)** Two M4 domains need to be located in close proximity of each other in the M-bands. **(2)** The formation of a disulfide bond must be possible. **(3)** Zinc ions must be available.

(1) Proximity: The location of M4 in the M-bands is not exactly known. In the single study that attempted to characterize the location of distinct parts of M-band titin, N- and C-terminal epitopes (domains Tk, M8 and M9) were found at the level of the M-band borders (approx. 50 nm from the M1 line). The location of only a single central epitope, the Is4 region between domains M5 and M6, was determined and found approximately in the middle between the M4/M4' and M6/M6' lines (M4/M4'-lines: 22 nm; M6/M6'-lines: 44 nm; Is4 32 nm distance to M1 line) [54]. Under the assumption that M-band titin, as proposed, spans the M-bands and so overlaps in an anti-parallel manner with M-band titin from the neighboring half-sarcomere, M4 domains are therefore likely to be located in the region between the central M1 and the M4/M4' lines **(Figure 3-31).** Due to the presence of long flanking regions of unknown folding the exact location of M4 cannot be accurately estimated. However, it appears possible that the domains of titin molecules from both halves of the sarcomeres are located in close proximity to each other, either permanently or temporarily depending on the state of contraction.

(2) Disulfide bonding: Although further data on the susceptibility of M4 to oxidation are not presented here, the fast crystallization as a covalently linked dimer in a weakly

reducing environment (in one day and in presence of 0.5 mM TCEP) demonstrated that (quasi)-spontaneous dimerization is possible. The factors required for the assembly need to be further evaluated. As already argued, the Cys 22 side-chain is surface exposed and as M4 is encompassed by the Cys-free and presumably unfolded, Is2 and Is3 regions, its thiol group therefore likely represents a truly free-standing reactive group.

In the strongly reducing cytosolic environment of eukaryotic cells, disulfide bridge formation in and between proteins is classically seen as unlikely. However, under certain conditions the environment can change. The sarcomeres are sites of large metabolic turnover where most of the energy required for contraction is produced by the rapid oxidation of fatty acids.
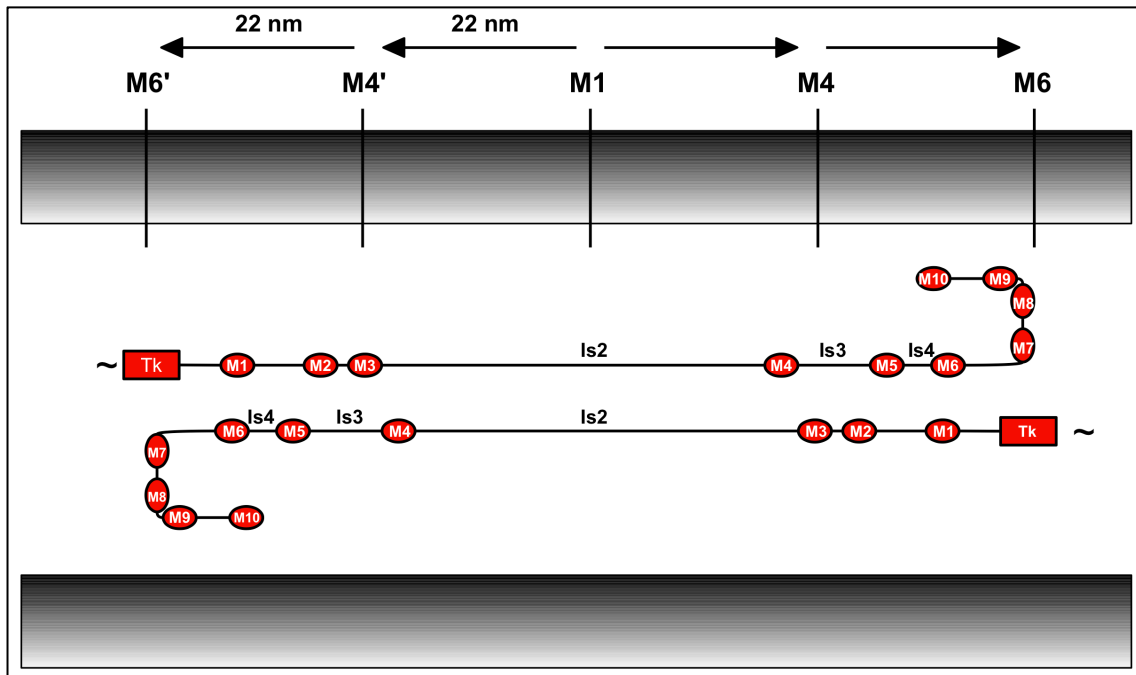
Figure 3-31 Proposed incorporation of M-band titin into the M-bands.

Based on the distances of four titin epitopes to the central M1-line (Tk: 51.8 nm; Is4: 31.9 nm; M8: 47.8 nm; M9: 51.8 nm), [54] have proposed a model in which the C-termini of titin from both half sarcomeres overlap with each other. The position of the M4 domain has not been determined. Because it is separated from the adjacent domains by the long Ig-like domain linking sequences, Is2 and Is3, featuring unpredicted structures, its location in the M-bands cannot be accurately calculated, but it is likely to be between the M1 and M4/M4' lines.

These oxidative reactions are accompanied by the formation of reactive oxygen species such as H_2O_2 and the O_2^- radical, and their accumulation during extensive muscular

work can lead to a temporary loss of the reducing environment in the cytoplasm [213]. This loss of 'reducing power' can then lead to oxidation of proteins with consequences for their activity. For example, for muscular calcium and potassium channels, reversible disulfide formation due to changes in the oxidative environment has been described as a mechanism regulating their activity [211][212]. Moreover, intramolecular disulfide bond formation has been proposed to stabilize the fold of Ig-like domains in the protein's I-band segment and so to regulate the protein's force resistance and elasticity properties in an oxidation state dependent manner [200][214]. Thus, under certain conditions, disulfide bond formation appears to occur on muscular proteins.

A factor that could increase the susceptibility of M4 for oxidation could be the close proximity between the zinc coordination site and the reactive Cys side chain. Being composed of His and Asp, the M4 zinc binding site is certainly less sensitive to oxidation than the more frequently found Cys-based binding sites. Thus, as soon as two M4 domains come close to each other in an appropriate orientation, a complete zinc binding site can be formed, which is independent of the cell's redox state. The distance between the zinc ion and the thiol group is with approximately 5 Å too large for a direct contribution of Cys 22 to zinc coordination in the reduced state. However, the proximity of Zn^{2+} can potentially lower the pKa of the thiol group and so alleviate covalent SS-bonding.

(3) Availability of zinc: Being an intermolecular coordination site in which both molecules are moving in the course of sarcomere contraction, the zinc binding site of M4 is unlikely to be permanently present. The mechanical resistance of such a site has to my knowledge never been examined, but it appears unlikely that once formed, in the absence of the disulfide bridge it could persist under substantial mechanical stress. Therefore, non protein-bound cytosolic zinc must be available at the time when M4 dimerizes.

In the strongly reducing intracellular environment of a myofibril, approximately 70% of intracellular zinc ions are compartmentalized in the mitochondria and nuclei, while the remaining 30% are bound to cytosolic metalloproteins [215]. The concentration of free zinc ions in the myofibrilar cytoplasm is therefore very low. In rat cardiomyocytes for example, their concentration has been determined to be only approximately 0.23 nM in a resting state [216]. Under oxidizing conditions, however, a large amount of protein-bound zinc can be released due to the disturbance of Cys-based metal binding sites on cytoplasmic proteins and lead to an up to a 30-fold increase of unbound ions [216].

Thus, under oxidizing conditions zinc ions are available in comparably large amounts, while under normal, reducing conditions they are hardly available in their free form.

In conclusion the dimerization of titin through oxidation and metal coordination appears theoretically possible under extreme conditions, which could be reached during heavy muscular work. Certainly the limited data provided here do not allow the validity of the M4 complex to be evaluated. Further data on how extreme conditions are required for the assembly to occur *in vitro* and *in vivo* are needed to test whether the observed dimerization is biologically relevant.

In the absence of additional data, the functional role of such a C-terminal titin interaction remains speculative. Being a covalent bond, the disulfide bridge between two domains can certainly resist forces far greater than the ones required to unfold most of titin's Ig-like domains and to disrupt the very resistant Z-disc anchoring titin-telethonin complex [48]. Therefore, a mechanical role of the complex under appears conceivable. For example, the temporary formation of the complex could prevent titin's C-terminal region from being permanently extracted from the M-bands during heavy muscle stress. As the M4 domain is located in the center of M-band titin, parts of the protein located C-terminally of the assembly site (Is3-M10) could upon formation of such an M4-based titin anchor be uncoupled from destructive force related effects as irreversible domain unfolding and thus their function e.g. as anchoring sites for obscurin, OBSL1 and calpain-3 could be preserved even under heavy load.

4   Conclusions and Future Perspectives

The present work had the aim to gain insight into the architecture of the M-bands by investigating direct interactions between the proteins titin on the one hand and obscurin/OBSL1 and myomesin-1, respectively, on the other hand.

The crystal structure of the titin-OBSL1 head-to-tail complex revealed an in sarcomeric proteins previously unobserved type of Ig-ike : Ig-like domain interaction and additional calorimetric data imply that the M10 – obscurin complex forms in a similar way. With respect to the data obtained it is currently unclear, whether there exists a direct connection between the impairment of the formation of these complexes and the development of hereditary muscle diseases. A number of hypothetical explanations exist confirming (e.g. modulation of Tk activity) or arguing against (e.g. the proposed involvement of additional M10 binding partners) an involvement of obscurin/OBSL1 M-band binding and disease development.

An interaction of the titin M4 domain with the central segment of myomesin-1 could not be confirmed. While the exact consequences of this finding are not entirely clear (a functional role for the interaction has neither been proposed nor examined), it shows that titin is possibly anchored to the M-band integral myomesin-1 only indirectly by association with OBSL1/obscurin.

However, the inability of both proteins to bind each other can be seen as a lucky coincidence. The characterization of the putative binding partners has led to the discovery, that both can, at least *in vitro* form homo-dimeric assemblies. It is questionable whether the disulfide mediated interaction of titin M4 and the strand-swapped dimer of myomesin-1 My5 could have been detected by methods such as yeast-two-hybrid screening, co-immunoprecipitation or dot-blot analysis that have previously been used to identify the interactions of M-band proteins.

Both assemblies have the potential to change our view of the architecture of the M-band lattice. Certainly as a next step, **validation** experiments for both assemblies *in vivo* need to be performed.

The model of a tetrameric myomesin-1 and its incorporation in the M-bands (**Figure 3-20** and **Figure 3-21**) proposed here are in good agreement with microscopic data. Thus, if My5 dimerization indeed takes place *in vivo,* and the model is correct, the disruption of the complex e.g. by over-expression of isolated My5 in neonatal rat cardiomyocytes as performed in [60] to validate the M10-obscurin interaction, should

cause detectable effects on the assembly of the M-bands, if not of the entire sarcomeres. Furthermore, it would be interesting to investigate whether a similar kind of self-assembly occurs in the centers of the myomesin-1 homologs myomesin-2 and 3 which are expressed in a muscle fiber-type specific way. Both have a similar domain-architecture as myomesin-1 and bind myosin. However, at least myomesin-2 is unable to form binary complexes at its C-terminus [74]. Additional cross-links on the myomesin proteins could thus provide the basis for a myosin cross-linking lattice with fiber-type specific properties.

The validation of the M4 complex will certainly prove more difficult. As it can be safely assumed to require rather 'harsh' conditions to form, simple competition experiments will not be sufficient. A possible method of validation would be the genetic replacement of M4 in a mouse with an M4 domain lacking the key residues that are required for complex formation, or expressing an additional isolated M4 domain. Similar mouse models, e.g. carrying the LGMD2J variant of the M10 domain, have been created in the past [172] and proven as a valuable tool to investigate the effects of regional alterations of titin.

Further assemblies in the M-bands remain to be characterized. With respect to the current work, two are of special interest:

As already discussed, the protease **calpain-3**, which binds titin N-terminally of the M10 domain appears to be a 'key-player' in the maintenance of sarcomere integrity, and a direct connection to the pathology of M10 variants seems possible. Calpain-3 itself forms (in contrast to other members of the calpain family) homo-dimeric assemblies [217], a fact that could not only have an impact on the protein's regulatory mechanism but, could also lead to indirect physical interconnection of two or more titin proteins in muscle cells where this protein is expressed. Preliminary results of our group indicate that large parts of calpain-3 can be expressed and purified from a recombinant source. Interaction studies with M-band titin fragments are currently ongoing.

The second candidate is the complex between the third Ig-like domain of **OBSL1/obscurin** and the linker between **myomesin-1** domains My4 and My5 [60]. This complex has not been direct subject of this work but the finding of the My5 mediated myomesin-1 assembly site addresses the question on what impact the oligomeric state of myomesin-1 could have on binding to obscurin and OBSL1. In the only published study on this interaction [60], it was solely investigated by pull-down assays and Y2H screening, however it remained unclear which form of myomesin-1

bound to obscurin/OBSL1.

5   Literature

1.   Squire JM, Al-Khayat HA, Knupp C, Luther PK (2005) Molecular architecture in muscle contractile assemblies. Adv. Protein Chem 71: 17-87. doi:10.1016/S0065-3233(04)71002-5

2.   Howard L (n.d.) Human skeletal muscle tissue 1 - TEM.. Available: http://commons.wikimedia.org/wiki/File:Human_skeletal_muscle_tissue_1_-_TEM.jpg.

3.   Ebashi S, Endo M, Otsuki I (1969) Control of muscle contraction. Q. Rev. Biophys 2: 351-384.

4.   Trinick J (1996) Titin as a scaffold and spring. Cytoskeleton. Curr. Biol 6: 258-260.

5.   Huxley AF, Niedergerke R (1954) Structural changes in muscle during contraction; interference microscopy of living muscle fibres. Nature 173: 971-973.

6.   Huxley H, Hanson J (1954) Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. Nature 173: 973-976.

7.   Reedy MK, Holmes KC, Tregear RT (1965) Induced changes in orientation of the cross-bridges of glycerinated insect flight muscle. Nature 207: 1276-1280.

8.   Luther PK (2009) The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. J. Muscle Res. Cell. Motil 30: 171-185. doi:10.1007/s10974-009-9189-6

9.   Gautel M (2011) The sarcomeric cytoskeleton: who picks up the strain? Curr. Opin. Cell Biol 23: 39-46. doi:10.1016/j.ceb.2010.12.001

10.   Sjöström M, Squire JM (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. doi:16/S0022-2836(77)80045-4

11.   Thornell LE, Carlsson E, Kugelberg E, Grove BK (1987) Myofibrillar M-band structure and composition of physiologically defined rat motor units. Am. J. Physiol 253: C456-468.

12.   Knappeis GG, Carlsen F (1968) The ultrastructure of the M line in skeletal muscle. J. Cell Biol 38: 202-211.

13.   Luther PK, Munro PMG, Squire JM (1981) Three-dimensional structure of the vertebrate muscle A-band: III. M-region structure and myosin filament symmetry. Journal of Molecular Biology 151: 703-730. doi:16/0022-2836(81)90430-7

14.   Luther P, Squire J (1978) Three-dimensional structure of the vertebrate muscle M-region. J. Mol. Biol 125: 313-324.

15.   Robinson TF, Winegrad S (1979) The measurement and dynamic implications of thin filament lengths in heart muscle. J. Physiol. (Lond.) 286: 607-619.

16.   Traeger L, Goldstein MA (1983) Thin filaments are not of uniform length in rat skeletal muscle. J. Cell Biol 96: 100-103.

17.   Horowits R, Podolsky RJ (1987) The positional stability of thick filaments in activated skeletal muscle depends on sarcomere length: evidence for the role of titin filaments. J. Cell Biol 105: 2217-2223.

18.   Gautel M (2011) The sarcomeric cytoskeleton: who picks up the strain? Curr. Opin. Cell Biol 23: 39-46. doi:10.1016/j.ceb.2010.12.001

19.   Richard I, Broux O, Allamand V, Fougerousse F, Chiannilkulchai N, et al. (1995) Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. Cell 81: 27-40.

20.   Pénisson-Besnier I, Hackman P, Suominen T, Sarparanta J, Huovinen S, et al. (2010) Myopathies caused by homozygous titin mutations: limb-girdle muscular dystrophy 2J and variations of phenotype. J. Neurol. Neurosurg. Psychiatr 81: 1200-1202. doi:10.1136/jnnp.2009.178434

21.   Young P, Ehler E, Gautel M (2001) Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly. J. Cell Biol 154: 123-136.

22.   Geisler SB, Robinson D, Hauringa M, Raeker MO, Borisov AB, et al. (2007) Obscurin-like 1, OBSL1, is a novel cytoskeletal protein related to obscurin. Genomics 89: 521-531. doi:10.1016/j.ygeno.2006.12.004

23.   Natori R (1954) The property and contraction process of isolated myofibrils. Jikeikai Med. J. 1: 119 f..

24.   Sjöstrand F (1962) The connections between A- and I-band filaments in striated frog muscle1. Journal of Ultrastructure Research 7: 225-246.

25.   Maruyama K, Kimura S, Kuroda M, Handa S (1977) Connectin, an elastic protein of muscle. Its abundance in cardiac myofibrils. J. Biochem 82: 347-350.

26.   Wang K, McClure J, Tu A (1979) Titin: major myofibrillar components of striated muscle. Proc. Natl. Acad. Sci. U.S.A 76: 3698-3702.

27.   Labeit S, Barlow DP, Gautel M, Gibson T, Holt J, et al. (1990) A regular pattern of two types of 100-residue motif in the sequence of titin. Nature 345: 273-276. doi:10.1038/345273a0

28.   Trinick J, Knight P, Whiting A (1984) Purification and properties of native titin. J. Mol. Biol 180: 331-356.

29.   Pfuhl M, 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.

30.   Goll CM, Pastore A, Nilges M (1998) The three-dimensional structure of a type I module from titin: a prototype of intracellular fibronectin type III domains. Structure 6: 1291-1302.

31.   Bang ML, Centner T, Fornoff F, Geach AJ, Gotthardt M, 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. Circ. Res 89: 1065-1072.

32.   Freiburg A, Trombitas K, Hell W, Cazorla O, Fougerousse F, et al. (2000) Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. Circ. Res 86: 1114-1121.

33.   Maruyama K (1997) Connectin/titin, giant elastic protein of muscle. FASEB J 11: 341-345.

34.   Suzuki J, Kimura S, Maruyama K (1994) Electron microscopic filament lengths of connection and its fragments. J. Biochem 116: 406-410.

35.   Fürst DO, Osborn M, Nave R, 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. J. Cell Biol 106: 1563-1572.

36.   Nave R, Fürst DO, 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? J. Cell Biol 109: 2177-2187.

37.   Atkinson RA, Joseph C, Kelly G, Muskett FW, Frenkiel TA, et al. (2001) Ca2+-independent binding of an EF-hand domain to a novel motif in the alpha-actinin-titin complex. Nat. Struct. Biol 8: 853-857. doi:10.1038/nsb1001-853

38.   Gautel M (1996) The super-repeats of titin/connectin and their interactions:

Glimpses at sarcomeric assembly. Advances in Biophysics 33: 27-37. doi:16/0065-227X(96)81661-3

39. Sorimachi H, Freiburg A, Kolmerer B, Ishiura S, Stier G, et al. (1997) Tissue-specific expression and alpha-actinin binding properties of the Z-disc titin: implications for the nature of vertebrate Z-discs. J. Mol. Biol 270: 688-695. doi:10.1006/jmbi.1997.1145

40. Ohtsuka H, Yajima H, Maruyama K, Kimura S (1997) Binding of the N-terminal 63 kDa portion of connectin/titin to alpha-actinin as revealed by the yeast two-hybrid system. FEBS Lett 401: 65-67.

41. Sorimachi H, Freiburg A, Kolmerer B, Ishiura S, Stier G, et al. (1997) Tissue-specific expression and alpha-actinin binding properties of the Z-disc titin: implications for the nature of vertebrate Z-discs. J. Mol. Biol 270: 688-695. doi:10.1006/jmbi.1997.1145

42. Ohtsuka H, Yajima H, Maruyama K, Kimura S (1997) The N-terminal Z repeat 5 of connectin/titin binds to the C-terminal region of alpha-actinin. Biochem. Biophys. Res. Commun 235: 1-3. doi:10.1006/bbrc.1997.6534

43. Linke WA, Ivemeyer M, Labeit S, Hinssen H, Rüegg JC, et al. (1997) Actin-titin interaction in cardiac myofibrils: probing a physiological role. Biophys. J 73: 905-919. doi:10.1016/S0006-3495(97)78123-2

44. Mues A, van der Ven PF, Young P, Fürst DO, Gautel M (1998) Two immunoglobulin-like domains of the Z-disc portion of titin interact in a conformation-dependent way with telethonin. FEBS Lett 428: 111-114.

45. Gregorio CC, Trombitás K, Centner T, Kolmerer B, Stier G, 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. J. Cell Biol 143: 1013-1027.

46. Pinotsis N, Petoukhov M, Lange S, Svergun D, Zou P, et al. (2006) Evidence for a dimeric assembly of two titin/telethonin complexes induced by the telethonin C-terminus. J. Struct. Biol 155: 239-250. doi:10.1016/j.jsb.2006.03.028

47. Zou P, Pinotsis N, Lange S, Song Y-H, Popov A, et al. (2006) Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature 439: 229-233. doi:10.1038/nature04343

48. Bertz M, Wilmanns M, Rief M (2009) The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. Proc. Natl. Acad. Sci. U.S.A 106: 13307-133310. doi:10.1073/pnas.0902312106

49. Kontrogianni-Konstantopoulos A, Ackermann MA, Bowman AL, Yap SV, Bloch RJ (2009) Muscle giants: molecular scaffolds in sarcomerogenesis. Physiol. Rev 89: 1217-1267. doi:10.1152/physrev.00017.2009

50. Kolmerer B, Olivieri N, Witt CC, Herrmann B, Labeit S (1996) Genomic Organization of M Line Titin and its Tissue-specific Expression in Two Distinct Isoforms. J. Mol. Biol 256: 556-563.

51. Mayans O, van der Ven PFM, Wilm M, Mues A, Young P, et al. (1998) Structural basis for activation of the titin kinase domain during myofibrillogenesis. Nature 395: 863-869. doi:10.1038/27603

52. Puchner EM, Alexandrovich A, Kho AL, Hensen U, Schäfer LV, et al. (2008) Mechanoenzymatics of titin kinase. Proc. Natl. Acad. Sci. U.S.A 105: 13385-13390. doi:10.1073/pnas.0805034105

53. Gräter F, Shen J, Jiang H, Gautel M, Grubmüller H (2005) Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. Biophys. J 88: 790-804. doi:10.1529/biophysj.104.052423

54. Obermann WM, Gautel M, Steiner F, van der Ven PF, Weber K, et al. (1996) The structure of the sarcomeric M band: localization of defined domains of myomesin,

M-protein, and the 250-kD carboxy-terminal region of titin by immunoelectron microscopy. J. Cell Biol 134: 1441-1453.

55.   Kinbara K, Sorimachi H, Ishiura S, Suzuki K (1997) Muscle-specific calpain, p94, interacts with the extreme C-terminal region of connectin, a unique region flanked by two immunoglobulin C2 motifs. Arch. Biochem. Biophys 342: 99-107. doi:10.1006/abbi.1997.0108

56.   Beckmann JS, Spencer M (2008) Calpain 3, the "gatekeeper" of proper sarcomere assembly, turnover and maintenance. Neuromuscul. Disord 18: 913-921. doi:10.1016/j.nmd.2008.08.005

57.   Duguez S, Bartoli M, Richard I (2006) Calpain 3: a key regulator of the sarcomere? FEBS J 273: 3427-3436. doi:10.1111/j.1742-4658.2006.05351.x

58.   Lange S, Auerbach D, McLoughlin P, Perriard E, Schäfer BW, et al. (2002) Subcellular targeting of metabolic enzymes to titin in heart muscle may be mediated by DRAL/FHL-2. J. Cell. Sci 115: 4925-4936.

59.   Obermann WM, Gautel M, Weber K, Fürst DO (1997) Molecular structure of the sarcomeric M band: mapping of titin and myosin binding domains in myomesin and the identification of a potential regulatory phosphorylation site in myomesin. EMBO J 16: 211-220. doi:10.1093/emboj/16.2.211

60.   Fukuzawa A, Lange S, Holt M, Vihola A, Carmignac V, et al. (2008) Interactions with titin and myomesin target obscurin and obscurin-like 1 to the M-band: implications for hereditary myopathies. J. Cell. Sci 121: 1841-1851. doi:10.1242/jcs.028019

61.   Horowits R, Kempner ES, Bisher ME, Podolsky RJ (1986) A physiological role for titin and nebulin in skeletal muscle. Nature 323: 160-164. doi:10.1038/323160a0

62.   Greaser M (2001) Identification of new repeating motifs in titin. Proteins 43: 145-149.

63.   Lee EH, Hsin J, von Castelmur E, Mayans O, Schulten K (2010) Tertiary and Secondary Structure Elasticity of a Six-Ig Titin Chain. Biophys J 98: 1085-1095. doi:10.1016/j.bpj.2009.12.4192

64.   Gao M, Wilmanns M, Schulten K (2002) Steered molecular dynamics studies of titin I1 domain unfolding. Biophys. J 83: 3435-3445. doi:10.1016/S0006-3495(02)75343-5

65.   Linke WA (2000) Stretching molecular springs: elasticity of titin filaments in vertebrate striated muscle. Histol. Histopathol 15: 799-811.

66.   Linke WA, Ivemeyer M, Olivieri N, Kolmerer B, Rüegg JC, et al. (1996) Towards a molecular understanding of the elasticity of titin. J. Mol. Biol 261: 62-71.

67.   Bennett P, Craig R, Starr R, Offer G (1986) The ultrastructural location of C-protein, X-protein and H-protein in rabbit muscle. J. Muscle Res. Cell. Motil 7: 550-567.

68.   Okagaki T, Weber FE, Fischman DA, Vaughan KT, Mikawa T, et al. (1993) The major myosin-binding domain of skeletal muscle MyBP-C (C protein) resides in the COOH-terminal, immunoglobulin C2 motif. J. Cell Biol 123: 619-626.

69.   Freiburg A, 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.

70.   van der Ven PF, Bartsch JW, Gautel M, Jockusch H, Fürst DO (2000) A functional knock-out of titin results in defective myofibril assembly. J. Cell. Sci 113 (Pt 8): 1405-1414.

71.   van der Ven PF, Fürst DO (1997) Assembly of titin, myomesin and M-protein

into the sarcomeric M band in differentiating human skeletal muscle cells in vitro. Cell Struct. Funct 22: 163-171.

72. Grove BK, Kurer V, Lehner C, Doetschman TC, Perriard JC, et al. (1984) A new 185,000-dalton skeletal muscle protein detected by monoclonal antibodies. J. Cell Biol 98: 518-524.

73. Masaki T, Takaiti O, Ebashi S (1968) "M-substance", a new protein constituting the M-line of myofibrils. J. Biochem 64: 909-910.

74. Schoenauer R, Lange S, Hirschy A, Ehler E, Perriard J-C, et al. (2008) Myomesin 3, a novel structural component of the M-band in striated muscle. J. Mol. Biol 376: 338-351. doi:10.1016/j.jmb.2007.11.048

75. Agarkova I, Auerbach D, Ehler E, Perriard JC (2000) A novel marker for vertebrate embryonic heart, the EH-myomesin isoform. J. Biol. Chem 275: 10256-10264.

76. Schoenauer R, Bertoncini P, Machaidze G, Aebi U, Perriard J-C, et al. (2005) Myomesin is a molecular spring with adaptable elasticity. J. Mol. Biol 349: 367-379. doi:10.1016/j.jmb.2005.03.055

77. Agarkova I, Schoenauer R, Ehler E, Carlsson L, Carlsson E, et al. (2004) The molecular composition of the sarcomeric M-band correlates with muscle fiber type. Eur. J. Cell Biol 83: 193-204.

78. Hornemann T, Kempa S, Himmel M, Hayess K, Fürst DO, et al. (2003) Muscle-type creatine kinase interacts with central domains of the M-band proteins myomesin and M-protein. J. Mol. Biol 332: 877-887.

79. Lange S, Himmel M, Auerbach D, Agarkova I, Hayess K, et al. (2005) Dimerisation of myomesin: implications for the structure of the sarcomeric M-band. J. Mol. Biol 345: 289-298. doi:10.1016/j.jmb.2004.10.040

80. Pinotsis N, Lange S, Perriard J-C, Svergun DI, Wilmanns M (2008) Molecular basis of the C-terminal tail-to-tail assembly of the sarcomeric filament protein myomesin. EMBO J 27: 253-264. doi:10.1038/sj.emboj.7601944

81. Berkemeier F, Bertz M, Xiao S, Pinotsis N, Wilmanns M, et al. (2011) Fast-folding {alpha}-helices as reversible strain absorbers in the muscle protein myomesin. Proc Natl Acad Sci U S A. Available: http://www.ncbi.nlm.nih.gov/pubmed/21825161. Accessed 12 Aug 2011.

82. Auerbach D, Bantle S, Keller S, Hinderling V, Leu M, et al. (1999) Different domains of the M-band protein myomesin are involved in myosin binding and M-band targeting. Mol. Biol. Cell 10: 1297-1308.

83. Fukuzawa A, Idowu S, Gautel M (2005) Complete human gene structure of obscurin: implications for isoform generation by differential splicing. J. Muscle Res. Cell. Motil 26: 427-434. doi:10.1007/s10974-005-9025-6

84. Ford-Speelman DL, Roche JA, Bowman AL, Bloch RJ (2009) The rho-guanine nucleotide exchange factor domain of obscurin activates rhoA signaling in skeletal muscle. Mol. Biol. Cell 20: 3905-3917. doi:10.1091/mbc.E08-10-1029

85. Coisy-Quivy M, Touzet O, Bourret A, Hipskind RA, Mercier J, et al. (2009) TC10 controls human myofibril organization and is activated by the sarcomeric RhoGEF obscurin. J. Cell. Sci 122: 947-956. doi:10.1242/jcs.040121

86. Kontrogianni-Konstantopoulos A, Jones EM, Van Rossum DB, Bloch RJ (2003) Obscurin is a ligand for small ankyrin 1 in skeletal muscle. Mol. Biol. Cell 14: 1138-1148. doi:10.1091/mbc.E02-07-0411

87. Bang ML, Centner T, Fornoff F, Geach AJ, Gotthardt M, 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. Circ. Res 89: 1065-1072.

88. Kontrogianni-Konstantopoulos A, Bloch RJ (2003) The hydrophilic domain of small ankyrin-1 interacts with the two N-terminal immunoglobulin domains of titin. J. Biol. Chem 278: 3985-3991. doi:10.1074/jbc.M209012200

89. Bowman AL, Kontrogianni-Konstantopoulos A, Hirsch SS, Geisler SB, Gonzalez-Serratos H, et al. (2007) Different obscurin isoforms localize to distinct sites at sarcomeres. FEBS Lett 581: 1549-1554. doi:10.1016/j.febslet.2007.03.011

90. Sorimachi H, Kinbara K, Kimura S, Takahashi M, Ishiura S, et al. (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.

91. Hackman P, Marchand S, Sarparanta J, Vihola A, Pénisson-Besnier I, et al. (2008) Truncating mutations in C-terminal titin may cause more severe tibial muscular dystrophy (TMD). Neuromuscul. Disord 18: 922-928. doi:10.1016/j.nmd.2008.07.010

92. Haravuori H, Vihola A, Straub V, Auranen M, Richard I, et al. (2001) Secondary calpain3 deficiency in 2q-linked muscular dystrophy: titin is the candidate gene. Neurology 56: 869-877.

93. Edström L, Thornell LE, Albo J, Landin S, Samuelsson M (1990) Myopathy with respiratory failure and typical myofibrillar lesions. J. Neurol. Sci 96: 211-228.

94. Lange S, Xiang F, Yakovenko A, Vihola A, Hackman P, et al. (2005) The kinase domain of titin controls muscle gene expression and protein turnover. Science 308: 1599-1603. doi:10.1126/science.1110463

95. Carmignac V, Salih MAM, Quijano-Roy S, Marchand S, Al Rayess MM, et al. (2007) C-terminal titin deletions cause a novel early-onset myopathy with fatal cardiomyopathy. Ann. Neurol 61: 340-351. doi:10.1002/ana.21089

96. Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of Escherichia coli with plasmids. Gene 96: 23-28.

97. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. Proceedings of the National Academy of Sciences 98: 10037 -10041. doi:10.1073/pnas.181342398

98. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol 6: 458-463. doi:10.1038/8263

99. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA (2005) Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. Acta Crystallogr. D Biol. Crystallogr 61: 449-457. doi:10.1107/S0907444905001307

100. Long F, Vagin AA, Young P, Murshudov GN (2008) BALBES: a molecular-replacement pipeline. Acta Crystallogr. D Biol. Crystallogr 64: 125-132. doi:10.1107/S0907444907050172

101. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr 60: 2126-2132. doi:10.1107/S0907444904019158

102. The CCP4 suite: programs for protein crystallography (1994) Acta Crystallogr. D Biol. Crystallogr 50: 760-763. doi:10.1107/S0907444994003112

103. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W545-549. doi:10.1093/nar/gkq366

104. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16: 566-567.

105. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local

alignment search tool. J. Mol. Biol 215: 403-410. doi:10.1006/jmbi.1990.9999

106. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 32: W665-W667. doi:10.1093/nar/gkh381

107. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221. doi:10.1107/S0907444909052925

108. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins: Structure, Function, and Bioinformatics 61: 704-721. doi:10.1002/prot.20660

109. Krissinel E, Henrick K (2005) Detection of Protein Assemblies in Crystals. In: R. Berthold M, Glen RC, Diederichs K, Kohlbacher O, Fischer I, editors. Computational Life Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg, Vol. 3695. pp. 163-174. Available: http://www.springerlink.com/content/7wgf5v6kxc35p5a6/. Accessed 4 Sep 2011.

110. Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic Acids Res 37: D229-232. doi:10.1093/nar/gkn808

111. Kabsch W (2010) XDS. Acta Crystallogr. D Biol. Crystallogr 66: 125-132. doi:10.1107/S0907444909047337

112. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2011) GenBank. Nucleic Acids Res. 39: D32-37. doi:10.1093/nar/gkq1079

113. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242.

114. Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, et al. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics 10: 136. doi:10.1186/1471-2105-10-136

115. Weeks SD, Drinker M, Loll PJ (2007) Ligation independent cloning vectors for expression of SUMO fusions. Protein Expr. Purif 53: 40-50. doi:10.1016/j.pep.2006.12.006

116. Chen GJ, Qiu N, Karrer C, Caspers P, Page MG (2000) Restriction site-free insertion of PCR products directionally into vectors. BioTechniques 28: 498-500, 504-505.

117. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A 74: 5463-5467.

118. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol 189: 113-130.

119. Dubendorff JW, Studier FW (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. J. Mol. Biol 219: 45-59.

120. Porath J, Carlsson J, Olsson I, Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 258: 598-599.

121. PORATH J, FLODIN P (1959) Gel filtration: a method for desalting and group separation. Nature 183: 1657-1659.

122. Beer A (1852) Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. Annalen der Physik und Chemie. 86: 78–88.

123. Pouger P (1729) Essai d'optique, Sur la gradation de la lumière. Sur la gradation de la lumière.: S. 164ff.

124. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

125. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein

secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res 32: W668-673. doi:10.1093/nar/gkh371

126. Rupp B (2007) Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. 1st ed. Taylor & Francis Ltd. p.

127. Giacovazzo C, Catti M, Monaco HL, Artioli G, Viterbo D, et al. (2011) Fundamentals of Crystallography. 3rd ed. Oxford University Press. p.

128. Merritt EA (n.d.) Anomalous scattering.. Available: http://skuld.bmsc.washington.edu/scatter/.

129. Bragg WL (1913) The Diffraction of Short Electromagnetic Waves by a Crystal. Proceedings of the Cambridge Philosophical Society: 43-57.

130. Friedel G (n.d.) Sur les symétries cristallines que peut révéler la diffraction des rayons. C.R. Acad. Sci. Paris 157: 1533 - 1536.

131. Hendrickson WA (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254: 51-58.

132. Phillips JC, Hodgson KO (1980) The use of anomalous scattering effects to phase diffraction patterns from macromolecules. Acta Cryst A 36: 856-864. doi:10.1107/S0567739480001817

133. Dauter Z, Dauter M, Dodson E (2002) Jolly SAD. Acta Crystallogr D Biol Crystallogr 58: 494-506. doi:10.1107/S090744490200118X

134. Ogata CM (1998) MAD phasing grows up. Nat. Struct. Biol 5 Suppl: 638-640. doi:10.1038/1330

135. Ramagopal UA, Dauter M, Dauter Z (2003) Phasing on anomalous signal of sulfurs: what is the limit? Acta Crystallogr. D Biol. Crystallogr 59: 1020-1027.

136. Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D Biol. Crystallogr 67: 282-292. doi:10.1107/S090744491003982X

137. Blundell TL, Johnson LN (n.d.) Protein Crystallography. New York: Academic Press. p.

138. Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat. Struct. Biol 4: 269-275.

139. Weiss MS, Hilgenfeld R (1997) On the use of the merging R factor as a quality indicator for X-ray data. J Appl Crystallogr 30: 203-205. doi:10.1107/S0021889897003907

140. Weiss MS (2001) Global indicators of X-ray data quality. J Appl Crystallogr 34: 130-135. doi:10.1107/S0021889800018227

141. Baldock P, Mills V, Stewart PS (1996) A comparison of microbatch and vapour diffusion for initial screening of crystallization conditions. Journal of Crystal Growth 168: 170-174. doi:16/0022-0248(96)00350-8

142. Li L, Ismagilov RF (2010) Protein crystallization using microfluidic technologies based on valves, droplets, and SlipChip. Annu Rev Biophys 39: 139-158. doi:10.1146/annurev.biophys.050708.133630

143. Mueller-Dieckmann J (2006) The open-access high-throughput crystallization facility at EMBL Hamburg. Acta Crystallogr. D Biol. Crystallogr 62: 1446-1452. doi:10.1107/S0907444906038121

144. Leslie AGW (2006) The integration of macromolecular diffraction data. Acta Crystallogr. D Biol. Crystallogr 62: 48-57. doi:10.1107/S0907444905039107

145. Bourenkov GP, Popov AN (2010) Optimization of data collection taking radiation damage into account. Acta Crystallogr. D Biol. Crystallogr 66: 409-419. doi:10.1107/S0907444909054961

146. Evans P (2006) Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr 62: 72-82. doi:10.1107/S0907444905036693

147. Guss JM, Merritt EA, Phizackerley RP, Hedman B, Murata M, et al. (1988) Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic "blue" copper protein from cucumbers. Science 241: 806-811.

148. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658-674. doi:10.1107/S0021889807021206

149. Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr 66: 22-25. doi:10.1107/S0907444909042589

150. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62: 439-450. doi:10.1107/S0907444906005270

151. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr 66: 12-21. doi:10.1107/S0907444909042073

152. Afonine PV, Grosse-Kunstleve RW, Chen VB, Headd JJ, Moriarty NW, et al. (2010) phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. J Appl Crystallogr 43: 669-676. doi:10.1107/S0021889810015608

153. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res. 32: W500-502. doi:10.1093/nar/gkh429

154. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-2637. doi:10.1002/bip.360221211

155. Harpaz Y, 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. doi:10.1006/jmbi.1994.1312

156. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol 48: 443-453.

157. Hackman P, Vihola A, Haravuori H, Marchand S, Sarparanta J, et al. (2002) Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. Am. J. Hum. Genet 71: 492-500. doi:10.1086/342380

158. Halaby DM, Poupon A, Mornon J (1999) The immunoglobulin fold family: sequence analysis and 3D structure comparisons. Protein Eng 12: 563-571.

159. Van den Bergh PYK, Bouquiaux O, Verellen C, Marchand S, Richard I, et al. (2003) Tibial muscular dystrophy in a Belgian family. Ann. Neurol 54: 248-251. doi:10.1002/ana.10647

160. Pollazzon M, Suominen T, Penttilä S, Malandrini A, Carluccio MA, et al. (2010) The first Italian family with tibial muscular dystrophy caused by a novel titin mutation. J. Neurol 257: 575-579. doi:10.1007/s00415-009-5372-3

161. Pudas R, Kiema T-R, Butler PJG, Stewart M, Ylänne J (2005) Structural basis for vertebrate filamin dimerization. Structure 13: 111-119. doi:10.1016/j.str.2004.10.014

162. Seo M-D, Seok S-H, Im H, Kwon A-R, Lee SJ, et al. (2009) Crystal structure of the dimerization domain of human filamin A. Proteins 75: 258-263. doi:10.1002/prot.22336

163. McCoy AJ, Fucini P, Noegel AA, Stewart M (1999) Structural basis for

dimerization of the Dictyostelium gelation factor (ABP120) rod. Nat Struct Mol Biol 6: 836-841. doi:10.1038/12296

164. Pinotsis N, Abrusci P, Djinović-Carugo K, Wilmanns M (2009) Terminal assembly of sarcomeric filaments by intermolecular beta-sheet formation. Trends Biochem. Sci 34: 33-39. doi:10.1016/j.tibs.2008.09.009

165. Lange S, Ouyang K, Meyer G, Cui L, Cheng H, et al. (2009) Obscurin determines the architecture of the longitudinal sarcoplasmic reticulum. J. Cell. Sci 122: 2640-2650. doi:10.1242/jcs.046193

166. Labeit S, Gautel M, Lakey A, Trinick J (1992) Towards a molecular understanding of titin. EMBO J 11: 1711-1716.

167. Pernigo S, Fukuzawa A, Bertz M, Holt M, Rief M, et al. (2010) Structural insight into M-band assembly and mechanics from the titin-obscurin-like-1 complex. Proc. Natl. Acad. Sci. U.S.A 107: 2908-2913. doi:10.1073/pnas.0913736107

168. Gautel M (2011) Cytoskeletal protein kinases: titin and its relations in mechanosensing. Pflugers Arch. Available: http://www.ncbi.nlm.nih.gov/pubmed/21416260. Accessed 23 Mar 2011.

169. Sorimachi H, Imajoh-Ohmi S, Emori Y, Kawasaki H, Ohno S, et al. (1989) Molecular cloning of a novel mammalian calcium-dependent protease distinct from both m- and mu-types. Specific expression of the mRNA in skeletal muscle. J. Biol. Chem 264: 20106-20111.

170. Kramerova I, Beckmann JS, Spencer MJ (2007) Molecular and cellular basis of calpainopathy (limb girdle muscular dystrophy type 2A). Biochim. Biophys. Acta 1772: 128-144. doi:10.1016/j.bbadis.2006.07.002

171. Ono Y, Torii F, Ojima K, Doi N, Yoshioka K, et al. (2006) Suppressed disassembly of autolyzing p94/CAPN3 by N2A connectin/titin in a genetic reporter system. J. Biol. Chem 281: 18519-18531. doi:10.1074/jbc.M601029200

172. Charton K, Danièle N, Vihola A, Roudaut C, Gicquel E, et al. (2010) Removal of the calpain 3 protease reverses the myopathology in a mouse model for titinopathies. Hum. Mol. Genet 19: 4608-4624. doi:10.1093/hmg/ddq388

173. Richard I, Roudaut C, Marchand S, Baghdiguian S, Herasse M, et al. (2000) Loss of calpain 3 proteolytic activity leads to muscular dystrophy and to apoptosis-associated IkappaBalpha/nuclear factor kappaB pathway perturbation in mice. J. Cell Biol. 151: 1583-1590.

174. Nettleship JE, Brown J, Groves MR, Geerlof A (2008) Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. Methods Mol. Biol. 426: 299-318. doi:10.1007/978-1-60327-058-8_19

175. Marino M, Zou P, Svergun D, Garcia P, Edlich C, et al. (2006) The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. Structure 14: 1437-1447. doi:10.1016/j.str.2006.07.009

176. Goll CM, Pastore A, Nilges M (1998) The three-dimensional structure of a type I module from titin: a prototype of intracellular fibronectin type III domains. Structure 6: 1291-1302.

177. Bennett MJ, Schlunegger MP, Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. Protein Sci 4: 2455-2468. doi:10.1002/pro.5560041202

178. Dickinson CD, Veerapandian B, Dai XP, Hamlin RC, Xuong NH, et al. (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin. J. Mol. Biol 236: 1079-1092.

179. Mrosek M, Labeit D, Witt S, Heerklotz H, von Castelmur E, et al. (2007) Molecular determinants for the recruitment of the ubiquitin-ligase MuRF-1 onto

M-line titin. FASEB J. 21: 1383-1392. doi:10.1096/fj.06-7644com

180. Bucher RM, Svergun DI, Muhle-Goll C, Mayans O (2010) The structure of the FnIII Tandem A77-A78 points to a periodically conserved architecture in the myosin-binding region of titin. J. Mol. Biol. 401: 843-853. doi:10.1016/j.jmb.2010.06.011

181. Obermann WM, Plessmann U, Weber K, Fürst DO (1995) Purification and biochemical characterization of myomesin, a myosin-binding and titin-binding protein, from bovine skeletal muscle. Eur. J. Biochem 233: 110-115.

182. Sauer F, Vahokoski J, Song Y-H, Wilmanns M (2010) Molecular basis of the head-to-tail assembly of giant muscle proteins obscurin-like 1 and titin. EMBO Rep 11: 534-540. doi:10.1038/embor.2010.65

183. Liu Y, Eisenberg D (2002) 3D domain swapping: as domains continue to swap. Protein Sci. 11: 1285-1299. doi:10.1110/ps.0201402

184. Shameer K, Shingate PN, Manjunath SCP, Karthika M, Pugalenthi G, et al. (2011) 3DSwap: curated knowledgebase of proteins involved in 3D domain swapping. Database 2011: bar042-bar042. doi:10.1093/database/bar042

185. Murray AJ, Head JG, Barker JJ, Brady RL (1998) Engineering an intertwined form of CD2 for stability and assembly. Nat Struct Mol Biol 5: 778-782. doi:10.1038/1816

186. Spinelli S, Desmyter A, Frenken L, Verrips T, Tegoni M, et al. (2004) Domain swapping of a llama VHH domain builds a crystal-wide beta-sheet structure. FEBS Lett. 564: 35-40. doi:10.1016/S0014-5793(04)00304-7

187. Murray AJ, Lewis SJ, Barclay AN, Brady RL (1995) One sequence, two folds: a metastable structure of CD2. Proc. Natl. Acad. Sci. U.S.A. 92: 7337-7341.

188. Hatherley D, Graham SC, Turner J, Harlos K, Stuart DI, et al. (2008) Paired Receptor Specificity Explained by Structures of Signal Regulatory Proteins Alone and Complexed with CD47. Molecular Cell 31: 266-277. doi:10.1016/j.molcel.2008.05.026

189. Rasmussen KK, Kulahin N, Kristensen O, Poulsen J-CN, Sigurskjold BW, et al. (2008) Crystal structure of the Ig1 domain of the neural cell adhesion molecule NCAM2 displays domain swapping. J. Mol. Biol. 382: 1113-1120. doi:10.1016/j.jmb.2008.07.084

190. Sonnen AF-P, Yu C, Evans EJ, Stuart DI, Davis SJ, et al. (2010) Domain metastability: a molecular basis for immunoglobulin deposition? J. Mol. Biol 399: 207-213. doi:10.1016/j.jmb.2010.04.011

191. Orengo CA, Jones DT, Thornton JM (1994) Protein superfamilies and domain superfolds. Nature 372: 631-634. doi:10.1038/372631a0

192. Williams AF, Barclay AN (1988) The immunoglobulin superfamily--domains for cell surface recognition. Annu. Rev. Immunol 6: 381-405. doi:10.1146/annurev.iy.06.040188.002121

193. Lappalainen I, Hurley MG, Clarke J (2008) Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. J. Mol. Biol. 375: 547-559. doi:10.1016/j.jmb.2007.09.088

194. Clarke J, Cota E, Fowler SB, Hamill SJ (1999) Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. Structure 7: 1145-1153.

195. Paci E, Clarke J, Steward A, Vendruscolo M, Karplus M (2003) Self-consistent determination of the transition state for protein folding: Application to a fibronectin type III domain. Proceedings of the National Academy of Sciences 100: 394 -399. doi:10.1073/pnas.232704999

196. Millman BM (1998) The filament lattice of striated muscle. Physiol. Rev 78: 359-

391.

197. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr 58: 1772-1779.

198. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr., A, Found. Crystallogr 64: 112-122. doi:10.1107/S0108767307043930

199. von Castelmur E, Marino M, Svergun DI, Kreplak L, Ucurum-Fotiadis Z, et al. (2008) A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. Proc. Natl. Acad. Sci. U.S.A 105: 1186-1191. doi:10.1073/pnas.0707163105

200. Mayans O, Wuerges J, Canela S, Gautel M, 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.

201. Harding MM (2006) Small revisions to predicted distances around metal sites in proteins. Acta Crystallogr. D Biol. Crystallogr 62: 678-682. doi:10.1107/S0907444906014594

202. Marino SM, Gladyshev VN (2010) Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. J. Mol. Biol. 404: 902-916. doi:10.1016/j.jmb.2010.09.027

203. Stacklies W, Vega MC, Wilmanns M, Gräter F (2009) Mechanical network in titin immunoglobulin from force distribution analysis. PLoS Comput. Biol. 5: e1000306. doi:10.1371/journal.pcbi.1000306

204. Müller S, Lange S, Gautel M, Wilmanns M (2007) Rigid conformation of an immunoglobulin domain tandem repeat in the A-band of the elastic muscle protein titin. J. Mol. Biol 371: 469-480. doi:10.1016/j.jmb.2007.05.055

205. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191. doi:10.1093/bioinformatics/btp033

206. Bullard B, Linke WA, Leonard K (2002) Varieties of elastic protein in invertebrate muscles. J. Muscle Res. Cell. Motil. 23: 435-447.

207. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792-1797. doi:10.1093/nar/gkh340

208. Winter C, Henschel A, Kim WK, Schroeder M (2006) SCOPPI: a structural classification of protein-protein interfaces. Nucleic Acids Res. 34: D310-314. doi:10.1093/nar/gkj099

209. Page RC, Clark JG, Misra S (2011) Structure of filamin A immunoglobulin-like repeat 10 from Homo sapiens. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 67: 871-876. doi:10.1107/S1744309111024249

210. Brandes N, Schmitt S, Jakob U (2009) Thiol-based redox switches in eukaryotic proteins. Antioxid. Redox Signal 11: 997-1014. doi:10.1089/ARS.2008.2285

211. Wang G, Strang C, Pfaffinger PJ, Covarrubias M (2007) Zn2+-dependent redox switch in the intracellular T1-T1 interface of a Kv channel. J. Biol. Chem. 282: 13637-13647. doi:10.1074/jbc.M609182200

212. Eu JP, Sun J, Xu L, Stamler JS, Meissner G (2000) The skeletal muscle calcium release channel: coupled O2 sensor and NO signaling functions. Cell 102: 499-509.

213. Winder WW (1998) Intramuscular mechanisms regulating fatty acid oxidation during exercise. Adv. Exp. Med. Biol. 441: 239-248.

214. Leake MC, Grützner A, Krüger M, Linke WA (2006) Mechanical properties of cardiac titin's N2B-region by single-molecule atomic force spectroscopy. J. Struct. Biol. 155: 263-272. doi:10.1016/j.jsb.2006.02.017

215. Rossowska MJ, Dinh C, Gottschalk SB, Yazdani M, Sutton FS 3rd, et al. (1990)

Interaction between caffeine intake and heart zinc concentrations in the rat. Br. J. Nutr. 64: 561-567.

216. Turan B, Fliss H, Désilets M (1997) Oxidants increase intracellular free Zn2+ concentration in rabbit ventricular myocytes. Am. J. Physiol. 272: H2095-2106.

217. Ravulapalli R, Diaz BG, Campbell RL, Davies PL (2005) Homodimerization of calpain 3 penta-EF-hand domain. Biochem. J 388: 585-591. doi:10.1042/BJ20041821

Appendix

Publications

Molecular basis of the head-to-tail assembly of giant muscle proteins obscurin-like 1 and titin.

Florian Sauer, Juha Vahokoski, Young-Hwa Song and Matthias Wilmanns.

EMBO reports July 2010; 11(7):534-40. doi:10.1038/embor.2010.65

Presentations and participations

(1) Participation in 9th International EMBL PhD Student Symposium "Patterns in Biology – Organisation of Life in Space in Time", held in Heidelberg, Germany, October 25 - 27, 2007

(2) Poster presentation at the 38th European Muscle Conference, held in Lille, France, September 12 – 16, 2009

Title: **Structural characterization of filament assemblies in the sarcomeric M-band**

<u>Florian Sauer</u> and Matthias Wilmanns

Abstract published in: Journal of Muscle Research and Cell Motility; December 2009; 30 (7-8);

(3) Invited talk at the second Marie-Curie 'MUZIC' meeting, held at the Karolinska Institute, Stockholm, Sweden, June 17-18, 2010

Molecular basis for the head-to-tail assembly of the muscle proteins obscurin-like 1 and titin.

(4) Practical course in **high-throughput expression screening by CoFi-blotting** at with Evitraproteoma AB, Stockholm, Sweden, June 14-17, 2010

(5) Poster presentation at the EMBO-conference: **"The Molecular and Cellular Mechanisms, Regulating Skeletal Muscle, Development and Regeneration"**, held in Wiesbaden, Germany, May 10 – 15, 2011

Title: **Structural determinants of Calpain-3 homodimerisation.**

Florian Sauer, <u>Marie-Cécile Pelissier</u> and Matthias Wilmanns

Contributions

The here presented work was entirely performed by myself with the following exceptions:

(1) Parts of the work on expression and purification of titin M10 variants and folding analysis by CD-spectroscopy (section 3.1) were carried out by Juha Vahokoski, EMBL-Hamburg (now: University of Oulu).

(2) The construct My4L (section 3.3) was purified by Marie-Cécile Pelissier, EMBL-Hamburg. Its size was determined by SLS by Viktor Deineko, EMBL-Hamburg.

(3) A number of diffraction datasets collected at ESRF-beamlines were collected for me by: Matthew Groves, Marie-Cécile Pelissier, Vivian Pogenberg and Juha Vahokoski. Diffraction data from the PETRA3 MX2 beamline was collected by Gleb Bourenkov.

Acknowledgements

Curriculum Vitae

Personal data:

Name	:	Florian Sauer
Nationality	:	German
Date of birth	:	13.08.1982
Place of birth	:	Würzburg, Germany

Education:

10. 2007 – present	:	**PhD student** at EMBL-Hamburg. Supervisor: Dr. Matthias Wilmanns. Doktorvater: Prof. Dr. Hermann Schindelin
10. 2001 – 10. 2007	:	**Diploma Student** of biology at the Julius-Maximilians Universität Würzburg. Major subject: Biochemistry Minor subjects: (1) Bioinformatics (2) Behavioral Physiology and Sociobiology Diploma thesis under supervision of Prof. Dr. Hermann Schindelin. Title: Strukturelle Charakterisierung von Gephyrin
09. 1992 – 06. 2001	:	**High School**, Johann-Schöner Gymnasium Karlstadt
09. 1988 – 07. 1992	:	**Elementary School**, Grund- und Volksschule Thüngen

__Erklärung__

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben.

Diese Dissertation hat weder in gleicher, noch in ähnlicher Form in einem anderen Prüfungsverfahren vorgelegen.

Ich habe außer den mit dem Zulassungsgesuch urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht.

Würzburg, den

Florian Sauer