Pro gradu –thesis

EFFECT OF NEBULIN VARIANTS ON NEBULIN-ACTIN INTERACTION

Svetlana Sofieva

November 2019

Folkhälsan research center

University of Helsinki
Genetics
Specialization in Human Genetics

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Faculty of biological and environmental sciences	Laitos – Institution– Department Department of Biosciences	
Tekijä – Författare – Author Svetlana Sofieva		
Työn nimi – Arbetets titel – Title Effect of nebulin variants on nebulin-actin interaction		
Oppiaine – Läroämne – Subject Human genetics		
Työn laji – Arbetets art – Level Master's thesis	Aika – Datum – Month and year 11/2019	Sivumäärä – Sidoantal – Number of pages 61 pages + 15 pages in Appendix

Tiivistelmä – Referat – Abstract

Nemaline myopathy (NM) is a rare congenital disorder, the most common of congenital myopathies. It affects primarily the skeletal muscles and it is recognised by nemaline bodies in muscle tissue samples and muscle weakness. Mutation of eleven genes are known to lead to NM and the most frequent disease-causing variants are either recessive *NEB* variants or dominant *ACTA1* variants. Variants in *NEB* are thought to be well tolerated and only 7% of them are hypothesized to be pathogenic. Over 200 pathogenic *NEB*-variants have been identified in Helsinki and the majority occurred in patients as a combination of two different variants. The missense variants were speculated to have a modifying effect on pathogenicity by affecting nebulin-actin or nebulin-tropomyosin interactions.

Nebulin is a gigantic protein coded by *NEB* and is one of the largest proteins in vertebrates. It is located in the thin filament of the skeletal muscle sarcomere. Enclosed by terminal regions, nebulin has an extensive repetitive modular region that covers over 90% of the protein. The repetitive zone comprises of 26 modules called super repeats (SR). SRs consist of seven simple repeats. There are seven conserved SDXXYK actin-binding sites at each super repeat, one per simple repeat, and one conserved WLKGIGW tropomyosin-binding site. Due to its enormous size and highly repetitive sequence, nebulin is one of the least studied proteins *in vivo, in vitro* or *in silico*.

In the NM patient database used for this study, there are 70 families with verified pathogenic mutations and in 30 of them, there were additional missense variants in *NEB*. These missense variants can be pathogenic modifying factors or have no impact on the phenotype. Seven missense variants were selected to study the effect of these mutations on actin-binding capacity compared to wild-type nebulin using the SR panel constructed previously by Laitila and Lehtonen. Also, due to the differences in actin-binding capacity of SRs compared to each other, one of the aims was to determine whether corresponding mutations in different SRs would have a similar or different effect on actin-binding capacity. For this aim, one missense mutation in the strongly actin-binding SR 1, and one in the weakly actin-binding SR 7 were selected from the NM database, and corresponding variants were created. Also, an in-frame deletion in SR7 found in the ExAC database and the corresponding mutation in SR1 were constructed for this study.

The actin-binding strength was determined using actin co-sedimentation assay and actin affinity assay. The results for co-sedimentation assay indicate that missense variants can have an effect on nebulin-actin interactions and, therefore, can be a possible cause for NM. The corresponding mutations had no correlation in their effect on actin-binding strength, just the opposite. S1-m-2 decreased actin-binding strength of SR1 and S7-m-2 had no effect on SR7. Likewise, S7-m-1 and S7-del-1 decreased actin-binding strength of SR7 and corresponding mutations had no effect on SR1. The selected missense mutations found in NM patients in SRs 2 and 4 decreased actin-binding strength, if located at the actin-binding sites and in SR 10 increased the actin-binding strength, if located at the actin-binding site. The change in actin binding strength was defined as significant if the P-value was below 0.005. The more accurate affinity assay was performed as a trial only for S16 and S16-m-1, a variant at a tropomyosin-binding site close to an actin-binding site. It indicated a difference in actin-binding affinity missed by the actin co-sedimentation assay. The results are preliminary, but show big promise and should be optimized and implemented in the future missense mutation affinity studies.

In an attempt to understand if the effect missense mutations have on nebulin-actin interaction is based on the change in nebulin structure, the 3D-structure of each produced fusion protein was predicted *in silico*. Considering that the variants were produced as GST-fusion proteins, the position and effect of GST in them is also a point of interest. In order to predict the structure of these large proteins, a combined approach was implemented using I-TASSER (Iterative Threading ASSEmbly Refinement) software. The software uses *ab initio* modeling, threading methods and atomic-level structure refinement to build an accurate 3D-model of a protein from sequence.

According to the predicted 3D models of the fusion proteins, the GST-part of the proteins folds into a globular structure and acts as a core around which the nebulin fragments fold. The GST does not bind to actin and is positioned on the inside, which indicates minimal effect on nebulin-actin interaction, but may be a reason for an alternative nebulin fragment folding. The accuracy of the default set of programs in software does not give the definitive answer of the possible effect missense mutations can have on structural changes. However, I-TASSER approach for 3D-modeling is promising with further software optimization and can possibly serve as an effective bioinformatic tool in the future.

Avainsanat – Nyckelord – Keywords
Nemaline myopathy, nebulin, *NEB*, actin-binding, functional assay, co-sedimentation, protein affinity, protein structure prediction, 3D-model, I-TASSER

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Katarina Pelin, Vilma-Lotta Lehtokari, Johanna Lehtonen

Säilytyspaikka – Förvaringställe – Where deposited
Department of Biosciences and Viikki Campus Library

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Bio- ja ympäristötieteiden tiedekunta	Laitos – Institution– Department Biotieteiden laitos	
Tekijä – Författare – Author Svetlana Sofieva		
Työn nimi – Arbetets titel – Title Nebuliinivarianttien vaikutus nebuliini-aktiini-vuorovaikutukseen		
Oppiaine – Läroämne – Subject Ihmisgenetiikka		
Työn laji – Arbetets art – Level Pro Gradu-tutkielma	Aika – Datum – Month and year 11/2019	Sivumäärä – Sidoantal – Number of pages 61 sivu + 15 liitesivua

Tiivistelmä – Referat – Abstract

Nemaliinimyopatia (NM) on harvinainen synnynnäinen sairaus, mutta yleisin syntyperäisistä lihastaudeista. Se vaikuttaa ensisijaisesti luustolihaksiin ja se tunnistetaan nemaliinikappaleista lihaskoepaloissa ja lihasheikkoudesta. Tällä hetkellä on tunnistettu mutaatioita 11 geenissä, jotka johtavat NM:aan ja yleisimmät sairautta aiheuttavat variantit ovat joko resessiiviset *NEB*-variantit tai dominantit *ACTA1*-variantit. *NEB*-varianttien uskotaan olevan hyvin siedettyjä ja vain 7% niistä oletetaan olevan patogeenisiä. Helsingissä on tunnistettu yli 200 patogeenistä *NEB*-varianttia ja suurin osa niistä esiintyi kahden eri variantin yhdistelmänä. Missense-variantteilla on ajateltu olevan modifioiva vaikutus patogeenisyyteen vaikuttamalla nebuliini-aktiini- ja nebuliini-tropomyosiini-vuorovaikutuksiin.

Nebuliini on jättikokoinen proteiini, jonka koodaa *NEB*-geeni, ja se on yksi suurimmista selkärankaisten proteiineista. Se sijaitsee luustolihaksen sarkomeerissä, ohuessa säikeessä. Suurin osa proteiinista, yli 90%, on toistuvaa modulaarista aluetta, jonka päissä sijaitsevat terminaaliset alueet. Toistoalue koostuu 26 moduulista, joita kutsutaan supertoistoiksi (super repeat, SR). SR:t rakentuvat seitsemästä yksinkertaisesta toistosta. Jokaisessa supertoistossa on seitsemän konservoitunutta SDXXYK-aktiinisitoutumiskohtaa, yksi per yksinkertainen toisto, ja yksi konservoitunut WLKGIGW-tropomyosiinisitoutumiskohtaa. Valtavan kokonsa ja hyvin toistuvan sekvenssinsä vuoksi nebuliini on yksi vähiten tutkituista proteiineista *in vivo*, *in vitro* tai *in silico*.

Tässä tutkimuksessa käytettiin NM-potilastietokantaa, jossa 70:ssä perheessä on vahvistettu olevan patogeeninen mutaatio ja niistä 30:ssä on lisäksi löydetty missense-variantteja *NEB*:ssa. Nämä missense-variantit voivat olla patogeenisyyttä muuntelevia tekijöitä tai voi olla, ettei niillä ole vaikutusta fenotyyppiin. Seitsemän missense-varianttia oli valittu ja mutaatioiden vaikutukset nebuliinin aktiinisitomiskykyyn verrattuna villityyppi-nebuliiniin oli tutkittu käyttäen aiemmin Laitilan ja Lehtosen kehittämää SR-paneelia. Lisäksi, toistojaksojen aktiinisitoutumiskyky vaihtelee toisiinsa nähden, joten yhtenä tavoitteista oli selvittää, onko vastaavilla mutaatioilla eri SR:ssa samanlainen vai erilainen vaikutus aktiinin sitomiskykyyn. Tähän tarkoitukseen oli valittu yksi missense-mutaatio vahvasti sitovasta SR 1:stä ja yksi heikosti sitovasta SR7:stä NM-potilastietokannasta, ja rakennettiin vastaavat variantit molemmissa SR:ssa. Lisäksi, perustuen ExAC-tietokannasta löydettyyn in-frame deleetioon SR 7:ssä, tutkimusta varten oli rakennettu SR 7-variantti ja sitä vastaava SR 1-variantti.

Aktiinisitoutumisen vahvuus määriteltiin käyttäen aktiinin ko-sedimentaatioanalyysia ja aktiiniaffiniteettianalyysia. Kosedimentaaatio-analyysin tulokset viittaavat siihen, että missense-varianteilla voi olla vaikutus nebuliini-aktiini-vuorovaikutukseen ja siten voi mahdollisesti aiheuttaa NM:aa. Vastaavien mutaatioiden vaikutukset aktiinin sitomisvahvuuteen eivät korreloineet, vaan olivat päinvastaisia. S1-m-2 vähensi SR1:n aktiinin sitomisvahvuutta ja S7-m-2:lla ei ollut siihen vaikutusta. Samalla tavalla S7-m-1 ja S7-del-1 laskivat SR7:n aktiinin sitomisvahvuutta ja vastaavilla mutaatioilla ei ollut vaikutusta. Potilastietokannasta valitut missense-mutaatiot SR2:ssa ja SR4:ssa laskivat aktiinin sitomisvahvuutta, mikäli ne sijaitsivat aktiinin sitoutumiskohdissa ja SR10:ssä aktiinin sitoutumiskohdissa olevat mutaatiot lisäsivät aktiinin sitomisvahvuutta. Aktiinin sitoutumisvahvuuden muutos määriteltiin merkitseväksi, jos sen P-arvo oli alle 0.005. Tarkempi affiniteettianalyysi suoritettiin kokeiluna vain S16:lle ja S16-m-1:lle, tropomyosiinin sitoutumiskohdassa ja lähellä aktiinin sitoutumiskohtaa sijaitsevalle variantille SR16:ssa. Affiniteettianalyysin tulosten mukaan näiden kahden nebuliinifragmentin aktiiniaffiniteetissä on eroa, jota ei havaittu aktiinin kosedimentaatioanalyysissa. Alustavat tulokset näyttävät lupaavilta ja affiniteetinmääritysmenetelmä tulisi optimoida ja käyttää tulevissa missense-mutaatioiden affiniteettitutkimuksissa.

Selvittääksemme johtuuko missense-mutaatioiden vaikutus nebuliini-aktiini-vuorovaikutukseen muutoksesta nebuliinin rakenteessa, jokaisen tuotetun GST-nebuliini-fuusioproteiinin 3D-rakenne oli ennustettu *in silico*. GST:n sijainti ja vaikutus fuusioproteiinin rakenteeseen olivat myös mielenkiinnon kohteena. Mallien ennustamiseen käytettiin I-TASSER (Iterative Threading ASSEmbly Refinement)-ohjelmistoa, jossa yhdistyvät kolme proteiinien mallintamisen lähestymistapaa: *ab initio*-mallinnus, laskostumisen tunnistusmenetelmät ja atomitason rakenteen tarkennus. Ohjelmisto rakentaa tarkan 3D-mallin aminohapposekvenssistä.

Ennustettujen 3D-mallien mukaan, fuusioproteiinien GST-osa laskostuu pallomaiseksi rakenteeksi ja toimii ytimenä, jonka ympärille nebuliinifragmentit laskostuvat. GST ei sido aktiinia ja sijaitsee rakenteen keskustassa, mikä viittaa siihen, että sen vaikutus nebuliini-aktiini-vuorovaikutukseen on hyvin pieni tai olematon. On kuitenkin mahdollista, että GST:n läsnäolo johtaa vaihtoehtoiseen nebuliinifragmenttien laskostumiseen. Ohjelmiston tarkkuus ei oletusohjelmilla anna selkeää vastausta mahdollisista missense-mutaatioiden vaikutuksista proteiinirakenteeseen. Kuitenkin, I-TASSER-lähestymistapa nebuliinivarianttien 3D-mallintamiseen näyttää lupaavalta ohjelmiston edelleenoptimisaatiolla ja se voi mahdollisesti toimia tehokkaana bioinformaattisena työkaluna tulevaisuudessa.

Avainsanat – Nyckelord – Keywords
Nemaliinimyopatia, nebuliini, *NEB*, sitoutuminen aktiiniin, funktionaalinen analyysi, kosedimentaatio, proteiiniaffiniteetti, proteiinirakenteen ennustaminen, 3D-malli, I-TASSER

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Katarina Pelin, Vilma-Lotta Lehtokari, Johanna Lehtonen

Säilytyspaikka – Förvaringställe – Where deposited
Biotieteiden laitos ja Viikin kampuskirjasto

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

Abbreviations

3D	3-dimensional
A	adenine
ab initio	in bioinformatics, a biological feature prediction method based solely on computational modeling without references to extrinsic sources
ADP, ATP	adenosine diphosphate, adenosine triphosphate
Arg	amino acid, arginine, R
C	cytosine
Ca^{2+}	calcium ion
Ca^{2+}-ATPase	transport protein that transfers Ca^{2+} through the membrane against electrochemical gradient using ATP; in sarcoplasmic reticulum also known as SERCA
Cys	amino acid, cysteine, C
C-score	confidence score
G	guanine
Glu	amino acid, glutamic acid, E
Gln	amino acid, glutamine, Q
His	amino acid, histidine, H
in silico	studies performed on a computer
in vitro	studies performed on a component of an organism, isolated from usual biological surroundings
in vivo	studies performed on living organisms or cells as whole, living entity
kDa	kilodalton, 1000 daltons, 1000 atomic mass units
LASP-1, LASP-2	LIM and SH3 domain protein 1, LIM and SH3 domain protein 2
LINE	Long Interspersed Nuclear Element

Lys	amino acid, lysine, K
L-score	local accuracy score
Met	amino acid, methionine, M
µg	microgram
µl	microliter
µM	micromoles per liter
N-RAP	Nebulin-related-anchoring protein
Ser	amino acid, serine, S
SNV	single nucleotide variation
stem cell	cell with the capacity of self-renewal and differentiation into various cell types
T	thymine
Thr	amino acid, threonine, T
Tyr	amino acid, tyrosine, Y

Literature review

1. Muscle tissues

Muscles are part of the musculoskeletal system and their main function is to create voluntary and involuntary movement. Thus, the ability to contract is a general trait of all muscle tissues. There are three types of muscle tissue: visceral or smooth, cardiac or semi-striated and skeletal or striated (see **Figure 1**). Approximately 50% of average human body weight is composed of muscle tissue. Majority of muscles, roughly 80%, are formed of skeletal muscle tissue, which totals circa 40% of body weight. The muscles are formed of highly differentiated myocytes that can be repaired and replaced adapting to body growth, physical training, and trauma. The regeneration of muscle tissue happens via muscle stem cells called satellite cells located in sublaminal niches adjacent to the muscle fibers. Satellite cells are formed during embryonic development and form approximately 5% of skeletal muscle cell mass in adults. (Alberts, Johnson et al. 2008, Qaisar, Bhaskaran et al. 2016, Belizário, Fontes-Oliveira et al. 2016)

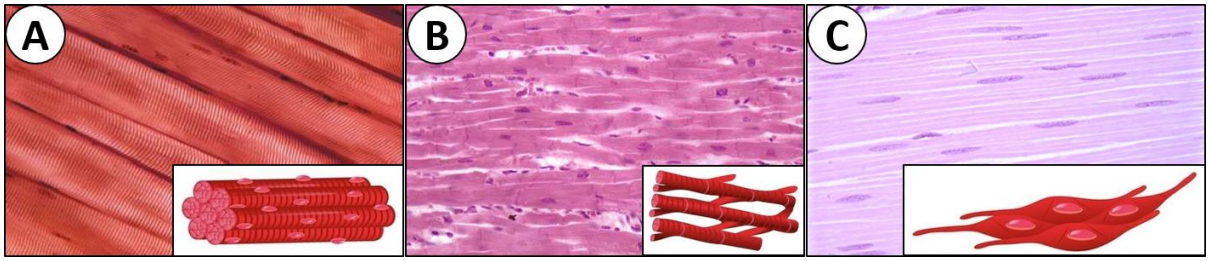


Figure 1. **Muscle tissue types. (A) Skeletal muscle. (B) Cardiac muscle. (C) Smooth muscle.** Modified from HistologyOLM by Stephen Gallik, Ph. D. © 2009 (http://histologyolm.stevegallik.org/node/143) and online resource Science Learning Hub © 2007-2018 The University of Waikato: muscle dissection (https://www.sciencelearn.org.nz/resources/1932-muscle-dissection).

1.1. Skeletal muscle

Skeletal muscles form the majority of muscles found in the human body. It is the only type of muscle tissue that can be consciously controlled. Under a microscope, skeletal muscle appears to look striated. It is generally attached to two bones through tendons, tough bands of connective tissue made of collagen fibers. The movement occurs due to shortening of the muscle during contraction, which pulls on tendons and leads to bone movement. (Silverthorn 2007)

Skeletal muscles are highly arranged and have a hierarchical structure. Muscle cells, also called muscle fibers, are sheathed with connective tissue and arranged in parallel longitudinally bundling them into units called fascicles. The nerves and the blood vessels along with the

connective tissue are located between these units. The fascicles with the interfascicle structures and tendons are enclosed in a connective tissue sheath continuous to the connective tissue around muscle fibers. This forms a muscle attached to the bones by its ends. (Silverthorn 2007)

Each muscle cell is long and cylindrically shaped with multiple nuclei located close to the cell surface. There can be up to several hundred nuclei in a single cell. Skeletal muscle cells are formed by the fusion of many embryonic stem cells into one. In adults, satellite cells stay dormant next to the functioning muscle fibers until activated to differentiate to muscle cells for growth and regeneration. Muscle fibers are composed of chains of contractile bundles called myofibrils. Myofibrils are composed of repeating basic contractile units called sarcomeres (see **Figure 2**). More detailed information about sarcomere structure and mechanism of contraction in **Chapter 2**. (Alberts, Johnson et al. 2008, Silverthorn 2007)

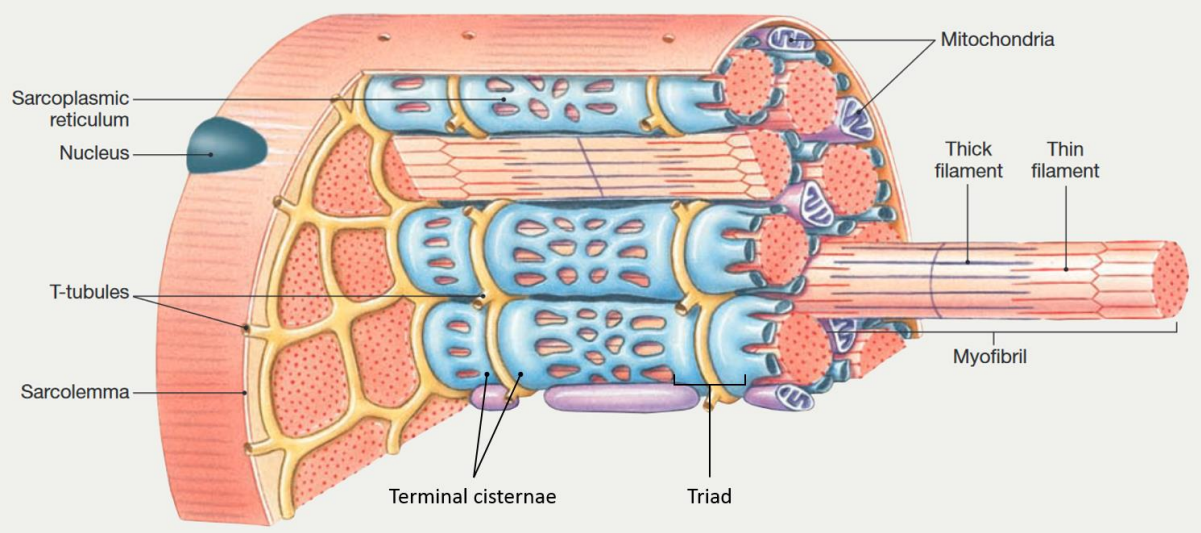


Figure 2. **Skeletal muscle fiber structure.** Modified (Silverthorn 2007).

Each myofibril is enclosed in endoplasmic reticulum specific for skeletal muscle cells called sarcoplasmic reticulum. Sarcoplasmic reticulum consists of tubules and enlarged end regions called the terminal cisternae (see **Figure 2**). Ca^{2+}-ATPases in sarcoplasmic reticulum enable the calcium-ions to be concentrated inside the lumen of sarcoplasmic reticulum or to be secluded from it. The release of Ca^{2+}-ions into the cytosol acts as a signal for the onset of a contraction. (Alberts, Johnson et al. 2008)

The muscle cell membrane called the sarcolemma surrounds muscle fibers and forms a network of transverse tubules called t-tubules. Therefore, lumen of t-tubules is continuous with the extracellular fluid. T-tubules are located adjacent to the terminal cisternae with which the t-tubules co-operate to conduct the signal for contraction. One t-tubule and two terminal

cisternae, one at each side, form a triad (see **Figure 2**). T-tubules enable the action potential to penetrate the muscle fiber evenly, making the signal reach terminal cisternae in muscle fiber almost simultaneously. This leads to a rapid contraction of myofibrils as a unit. Lack of t-tubules would lead to conduction of action potential through the cytosol, which would mean slower response and non-simultaneous contraction. (Silverthorn 2007)

Contraction requires big amounts of energy. For fast and effective energy provision, there are large quantities of mitochondria and glycogen granules in the muscle fiber cytoplasm, also called as sarcoplasm (see **Figure 2**). Majority of ATP needed for contraction is provided through oxidative phosphorylation in mitochondria. (Alberts, Johnson et al. 2008, Silverthorn 2007)

1.2. Cardiac muscle

Cardiac muscle, i.e. semi-striated muscle, is located singularly in the heart and its only task is to pump blood through the body. Cardiac muscle tissue cannot be controlled consciously, so the contraction of this muscle tissue is involuntary. It is controlled by the autonomic nervous system and hormones adjusting the contraction rate. The stimuli from these systems are transmitted to the natural pacemaker of the heart (sinus node), the region of cardiac muscle tissue in the right upper chamber of the heart. In the sinus nodes, electrical impulses are generated and then conducted to the other cardiac muscle cells stimulating them to contract. Because of this self-stimulating ability, cardiac muscle can be considered intrinsically controlled. (Alberts, Johnson et al. 2008, Silverthorn 2007)

Cardiac muscle tissue has features of both smooth and skeletal muscle. Similar to skeletal muscle fibers, cardiac muscle fibers are striated and follow sarcomere structure that forms fibrils. At the same, cardiac muscle fibers are shorter than skeletal muscle fibers, may be branched and have one nucleus per cell. The cardiac muscle cells are tightly linked to each other through special cell junctions called intercalated disks allowing electrochemical signals to pass quickly from one cell to another, so that the heart can beat as a unit. The combination of the branched structure of the cardiac muscle cells and the presence of intercalated discs makes this tissue strong enough to tolerate high blood pressures and their changes and to endure the strain of contracting throughout a lifetime.(Alberts, Johnson et al. 2008, Silverthorn 2007)

1.3. Smooth muscle

Visceral muscles are, as cardiac muscle, also the muscles that function involuntarily. They are responsible of, for example, bowel movement. They are controlled through the autonomic and enteric nervous systems. Smooth muscle tissue can be found around tubular visceral organs, such as an alimentary canal, the ureters and urinary bladder, the internal reproductive structures, the respiratory organs, and the arteries. This type of muscle tissue can also be found as arrector pili muscles of hair follicles and in the iris. Histologically, smooth muscle cells are relatively short, spindle-shaped singly nucleated cells that collocate close to each other and are connected with special intercellular gap junctions that form layers. (Alberts, Johnson et al. 2008)

Contraction of smooth muscles differs from cardiac or skeletal muscle contraction as it does not have a sarcomere structure. The contractile apparatus in smooth muscle cells is more loosely arranged with contractile fibers along the long axis of the cells and supporting fibers containing intermediate filaments. (Alberts, Johnson et al. 2008, Silverthorn 2007). According to one of the current models, contractile fibers containing actin and myosin are attached at one end to the plasma membrane and at the other end through cytoplasmic masses known as dense bodies to the intermediate filaments in the center of the cells. The supporting fibers are attached to contractile fibers at both ends. (Alberts, Johnson et al. 2008) Other models suggest that intermediate filaments and dense bodies form a cytoskeleton, actin is attached to dense bodies and myosin molecules are surrounded by the actin filaments (Silverthorn 2007). However, all the models are at a consensus that during contraction actin- and myosin filaments slide by each other leading to great shortening of the muscle cells. The filaments are attached to the plasma membrane at disc-like junctions that connect the adjacent cells allowing synchronized movement. Therefore, even though single smooth muscle cell movement is weak, their working together can produce powerful, long-lasting contractions leading to large movements, for example, pushing out a baby during childbirth. However, visceral muscles contract slower than striated muscles and the force is significantly smaller. (Alberts, Johnson et al. 2008, Silverthorn 2007)

2. Sarcomere structure and physiology

As mentioned in **Chapter 1.1.**, myofibrils are composed of sarcomeres, the basic contractile unit in striated muscle tissue, repeated along the whole length of the myofibril. The sarcomeres are formed from myofilaments: thick myosin and thin actin filaments. Thick and thin filaments

slide past each other during contraction, see **Figure 3** for structure. The elastic filaments are formed of titin, a large resilient protein. They run through the cores of thick filaments and anchor at the Z disc stabilizing the structure. Obscurin and myosin-binding protein-C (MyBP-C) are associated with elastic filament and are believed to modulate contraction and stabilize the sarcomeric structure. (Silverthorn 2007, Alberts, Johnson et al. 2008, Lin, Taejeong et al. 2017)

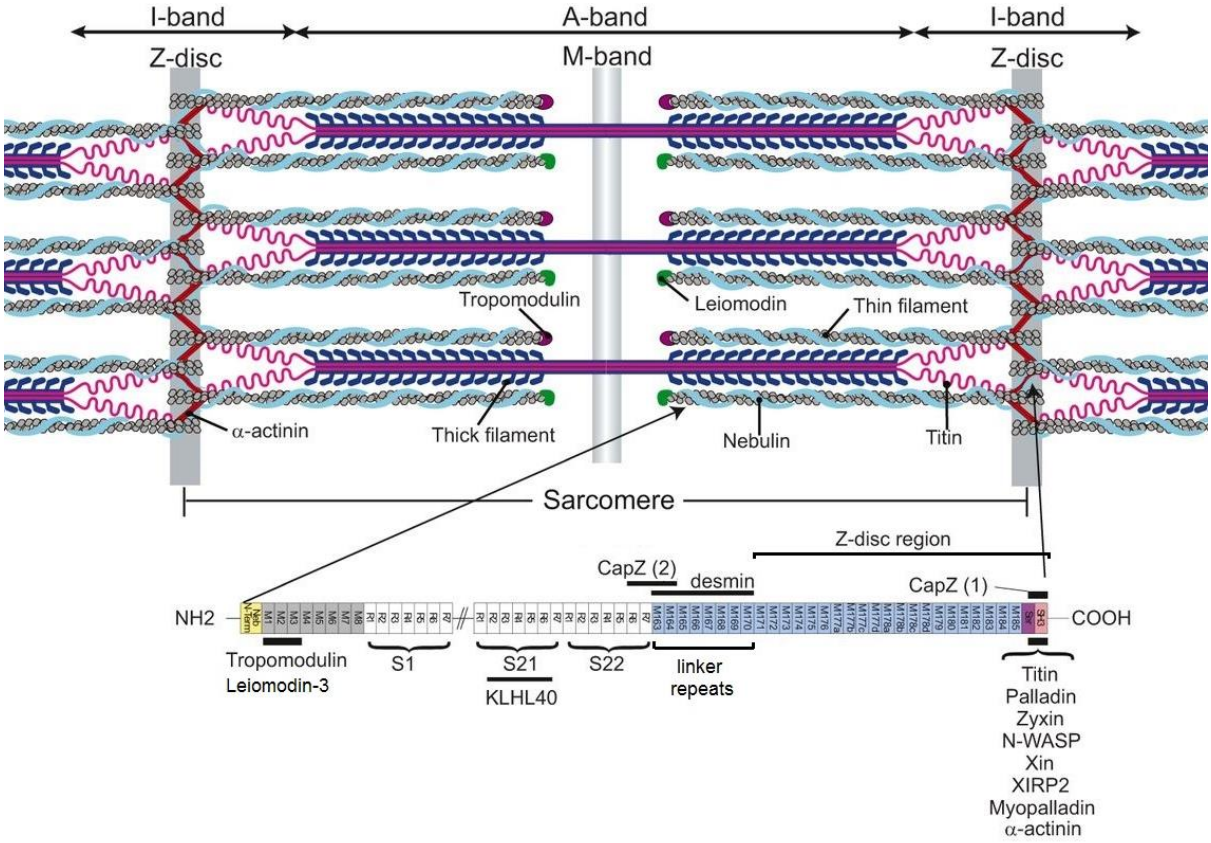


Figure 3. **Sarcomere and nebulin structures.** Nebulin is located in the thin filament. The N-terminus of nebulin almost reaches the end of the actin filament, whereas C terminus is anchored at Z-disc. Yellow box refers to the glutamic acid-rich domain in the N-terminus of nebulin. Lilac box refers to the serine rich domain and pink box refers to the SH3-domain, both located in the C-terminus. Between the termini, the protein structure of nebulin is modular, each small box represents one simple repeat. The majority of simple repeats form super repeats (SR). Grey boxes are simple repeats connecting N-terminus to SR region, while blue boxes are simple repeats connecting C-terminus to SRs. White boxes are simple repeats that form SR region (SRs here marked as S and a number). Modified (Chu, Gregorio et al. 2016).

A single sarcomere extends from one Z-line, or Z-disc, to another. Z-disc is a zigzag protein structure, which is composed of α-actinin and other proteins and serves as the attachment site for thin filaments and titin. I-bands form the light colored bands in the sarcomere if examined under the light microscope and they consist solely of thin filaments. Z-discs are located at midpoints of each I-band leaving half of the I-band belonging to one sarcomere and the other half to adjacent sarcomere. Due to the presence of thick filament along the entire length of A-

bands, they make up the dark bands of the sarcomere if examined under the light microscope. At the outer regions of the A-band thick and thin filament overlap, and in the middle only thick filaments are present. This middle region is called the H-zone. The M-line is a protein formation that divides the A-band and the H-zone in half. It acts as the attachment site for the thick filaments in the same manner as Z-disc acts for the thin filaments. During contraction, thin filaments and thick filaments slide past each other resulting in further overlapping of myofilaments and leading to shortening of the I-band and the H-zone (see **Figure 3**). (Luther 2009, Silverthorn 2007)

2.1. Thick filament structure

Thick filament is formed of myosin, which acts as a motor in sliding activity of thin filaments. Myosin is a protein superfamily that is classified into many classes and functional roles. In striated muscles, class II myosin is present. In a contraction, myosin (~220 kDa) binds to actin forming cross-bridges and, followed by calcium-induced conformational change, pulls the thin filaments across the thick filaments. (Lin, Taejeong et al. 2017, Silverthorn 2007)

Class II myosin contains two heavy chains, each of which is divided into heavy meromyosin fragment and light meromyosin fragment. The heavy meromyosin fragment is comprised of the head- (S1) and neck (S2) regions. The light meromyosin fragment (LMM) comprises the long tail region of the protein, which is connected through S2 to S1. The N-terminal subfragment S1 protrudes from thick filament at regular intervals of 14.3 nm and is the region that interacts with actin. The head also has a binding site for ATP, which is capable of converting the chemical energy of ATP into mechanical energy. The conformational change of this catalytic head is the driving force behind the contraction. For fine-tuning of each S1 fragment interacts with myosin light chains (MLCs) located within S1 and S2 fragments. The MLCs are formed of essential light chain (ELC) and regulatory light chain (RLC) and they regulate myosin and, possibly, MyBP-C function. From S2 towards the C-terminus tail region, the LMM, two heavy chains of a single myosin form α-helix coiled coil structure. (Lin, Taejeong et al. 2017, Frontera, Ochala 2015)

2.2. Thin filament structure

The basic structure of the thin filament consists of six proteins: actin, tropomyosin, troponins I, C and T, and nebulin. All these components interact closely with each other and the dynamics of the interactions during contraction are well studied. There is also a number of other proteins

that interact with the thin filament proteins modifying their properties and function. See thin filament structure in **Figure 4 (A)** (Silverthorn 2007, Alberts, Johnson et al. 2008)

A polymerized filamentous actin is the key component in the thin filament. The filamentous actin (F-actin) is polymerized from globular actin (G-actin) or actin monomer spontaneously. The two ends of actin are polymerized at a different rate if an excess of subunits is allowed to assemble. The fast-growing end is called a plus-end and the slowly growing a minus end. Actin filament is structurally a double α-helix that forms a coiled coil in the sarcomere. It serves as an anchor to other thin filament proteins and interacts with myosin during contraction. (Alberts, Johnson et al. 2008, Lin, Taejeong et al. 2017)

G-actin contains four subdomains SD1-SD4. Each half-helical turn of the actin filament is formed of seven monomers comprising of SD1/2. This region interacts with the opposing strand of seven monomers comprising of SD3/4. SD1/2 monomers can interact with other myofilament proteins like myosin, MyBP-C, and other proteins. More specifically, SD1 can interact with the myosin heads of the thick filament.(Lin, Taejeong et al. 2017)

In vertebrates, six genes can code actin and each gene has one isoform. *ACTA1* encodes actin α1 in skeletal muscles and *ACTC1* encodes cardiac actin α-protein. In smooth muscles *ACTA2* produces actin α2 and *ACTG2* produces actin γ2. Actin β and actin γ1 are generated ubiquitously by *ACTB*- and *ACTG1*-genes respectively. All actin genes are highly conserved and have little variation between species at the N-terminal region. (Perrin, Ervasti 2010, Lin, Taejeong et al. 2017)

Tropomyosin (Tm) is a 42 nm long and 37 kDa-sized α-helical protein. It occurs in dimers and forms coiled coil-structure that intertwines with F-actin. Tm is expressed from four different genes, *TPM1*, *TPM2*, *TPM3* and *TPM4*. Each gene can be alternatively spliced in several different ways yielding in up to 40 different isoforms (Geeves, Hitchcock-DeGregori et al. 2015). The isoforms are structurally similar, but there are variations in structural homology that suggest differences in capacity to regulate contraction. Tms can form both homodimers and heterodimers, but in striated muscles, Tms forms heterodimers. The dimers polymerize end-to-end, fitting into the major groove within the double α-helix of F-actin. Tms have two main functions. It acts as a stabilizer in the thin filament by interacting with tropomodulin and therefore reducing the depolymerization of thin filament components. Tm also has a major role in the regulation of contraction due to its role as a steric block of myosin binding sites on actin preventing the binding of filaments until induced. (Lin, Taejeong et al. 2017, Silverthorn 2007)

Troponin complex consists of three subunits, troponin C (TnC), troponin I (TnI) and troponin T (TnT). The main function of this complex is to detect and transduce the signal inducted to start a contraction. TnI and TnT can also serve as a biomarker to detect and assess the severity of cardiac injury. (Lin, Taejeong et al. 2017)

TnC is calcium-binding 18 kDa unit located in the N-terminus of the troponin complex. It has two isoforms in skeletal muscles similar in structure and function, but with slight variation in primary structure. The isoforms are called slow-skeletal TnC (ssTnC) and fast-skeletal TnC (fsTnC). In cardiac muscle, there is a special, structurally identical to ssTnC, a cardiac isoform of TnC (cTnC). TnC is formed of two globular domains at each terminus connected with an α-helix. The C-terminal globular domain has high Mg^{2+} and Ca^{2+} binding affinity *in vivo*, which means too low turnover for thin filament activation and is thought to serve as an anchor to the N-terminus of TnI. The N-terminal globular domain in TnC is the primary regulatory region that interacts with the C-terminal domain of TnI. The binding of Ca^{2+} to the N-terminus of TnC induces conformational change exposing a hydrophobic region, which leads to change in TnC-TnI-interaction. (Lin, Taejeong et al. 2017)

TnI is a 21 kDa unit in the troponin complex considered to play a role as an inhibitor of a sarcomeric activation in the absence of a Ca^{2+}-induced signal. In the inhibited state, the N-terminus of TnI binds loosely to TnC and TnT and the C-terminus binds strongly to actin. The interactions of TnI with mentioned thin filament subunits change upon the Ca^{2+} activation. It causes conformational change so that the C-terminus of TnI binds strongly to TnC and TnT loosening the bondage with actin. This conformational change is also thought to transduce the Ca^{2+}-induced signal to TnT. (Lin, Taejeong et al. 2017)

TnT, sized 36 kDa, is the largest unit of the troponin complex. It is formed of two helical domains, T2 at the C-terminus of TnT and T1 next to it at the center, and an unstructured flexible region called hypervariable domain at the N-terminus. The T2 domain interacts with TnI, TnC and the central region of Tm and the T1 domain interacts with the C-terminus of Tm. Hypervariable domain has no verified functions, but it is suggested that this domain allows TnT to adapt specifically for different structures. Originally, TnT was considered to have an exclusively structural role, but recent studies suggest that TnT has several phosphorylation sites regulating thin filament activation. When Ca^{2+} induced signal reaches TnT, it secedes from Tm allowing Tm to move. (Lin, Taejeong et al. 2017)

Nebulin is a member of a large nebulin family of actin-binding proteins along with N-RAP, nebulette, LASP-1, and LASP-2. They are expressed in many cell types and have various functions like regulation of thin filament length, sarcomeric structure and function. Nebulin is a gigantic 600-900 kDa protein predominantly expressed in skeletal muscle fibers. Nebulette (107 kDa) is the predominant isoform in cardiac muscles, where it has similar functions as nebulin has in skeletal muscle. More information about nebulin structure and interactions in the thin filament in **Chapter 3**. (Lin, Taejeong et al. 2017)

2.3. Contraction

To generate the force needed for movement three subsequent steps are needed in a muscle cell. First, the signal from the somatic motor neuron is received and transformed into an electrical signal in the muscle fiber. Second, excitation-contraction coupling occurs, which leads to the third step: a contraction-relaxation cycle. (Silverthorn 2007)

The signal from the motor neuron is transmitted to the muscle fibers through a chemical synapse called neuromuscular junction. When the action potential reaches the presynaptic terminal, it induces the release of Ca^{2+} to the presynaptic space in the motor neuron. The calcium ions bind to sensor proteins of the synaptic vesicles containing neurotransmitters, which sets off the fusion of vesicles with the plasma membrane and the release of neurotransmitter into the synaptic cleft. In vertebrates, the released neurotransmitter is acetylcholine (ACh). ACh molecules transfer across the synaptic cleft via diffusion and bind to the nicotinic acetylcholine receptors (nAChRs) on the sarcolemma. The bond between nAChR and ACh induces a conformational change in nAChR, which enables its ion channel properties and permits sodium ions to enter the postsynaptic terminal. This causes a depolarization in muscle fibers, an electrical signal that is conducted to all fiber regions via the T-tubule system. (Silverthorn 2007)

The impulse arriving at the triad through T-tubules sets off voltage sensitive subunits of dihydropyridine receptors (DHPRs) on the T-tubule allowing negligible calcium flow into a fiber. It also permits physical interaction between DHPRs and ryanodine receptors (RyRs) on adjacent sarcoplasmic reticulum, which triggers the release of large amounts of calcium into the sarcoplasm. Excitation is triggered by a high concentration of calcium in the cytosol. (Heiny, Meissner 2012)

The released cytosolic calcium binds to TnC inducing its conformational change, which transduces the signal to TnI. It also undergoes a conformational change and induces a change

in binding affinity of TnT to Tm. That change facilitates Tm shift from a blocked to a closed position on the myosin binding sites exposing them and allowing myosin heads to bind to F-actin forming cross-bridges. Myosin binding to actin shifts Tm position on actin further (see **Figure 4 (A, B1, B2)**). (Frontera, Ochala 2015)

After binding of myosin heads to actin, an inorganic phosphate is released from myosin head enabling the power stroke, a movement of myosin head, which pushes thin filaments past thick filaments (see **Figure 4 (B3)**). After the power stroke, ADP is released (see **Figure 4 (B4)**) and ATP is enabled to bind to myosin head leading to myosin detachment from actin (see **Figure 4 (B5)**). ATP is

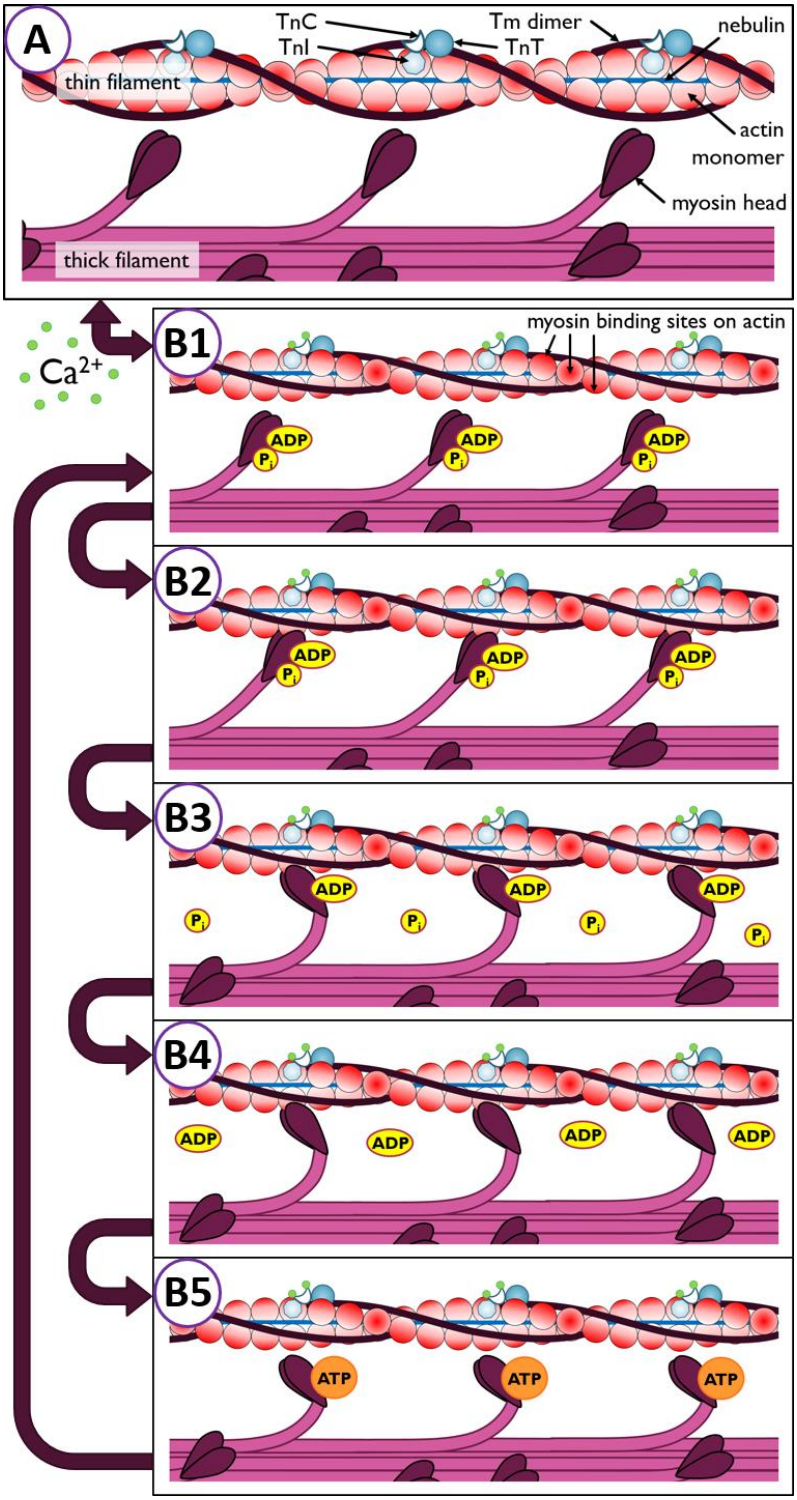


Figure 4. **Interaction between thick and thin filaments during contraction. (A)** At rest, cytosolic calcium-ion concentration is low, Tm sterically blocks myosin from binding to actin and filaments are unattached. Note: the exact positioning of nebulin in thin filament is unknown, in this figure nebulin idlocated in the middle of the filament. **(B1)** With the increase of calcium concentration in cytosol, Ca^{2+}-ions bind to TnC inducing conformational change when TnI and Tm dimer are pushed aside exposing myosin binding sites on actin monomers. **(B2)** Myosin heads attach to actin. **(B3)** A phosphate is released from myosin initiating power stroke and forcing the rotation of the myosin head. This movement pushes thin filament past it. **(B4)** After the power stroke ADP is released. **(B5)** ATP binds to its binding site on myosin leading to dissociation from actin. ATP is hydrolysed by ATPase activity of myosin to ADP and P_i, which stay attached to myosin.

hydrolyzed into ADP and inorganic phosphate due to myosin ATPase activity and myosin head returns to its original position ready to repeat the contraction cycle while the thin filament is activated. (Frontera, Ochala 2015)

3. Nebulin

3.1. The nebulin gene (*NEB*)

NEB is the nebulin coding gene located in chromosomal region 2q23 and is 249 kb in size. It consists of 183 exons out of which 17 can be alternatively spliced leading to numerous different nebulin isoforms. It was hypothesized that such a large variety of isoforms is needed due to their location in many different fiber types and muscles, and their varying functions during development and in adulthood. The smallest exon size is 42 bp and the biggest is 569 bp. Intron sizes also vary from a little below 100 bp up to several thousand bp in length. *NEB* is modular in structure, composed of repetitively occurring similar exons formed most probably through duplications due to transposable elements such as *Alu* and LINE. This area is super repeat (SR) coding area flanked by exons coding simple repeats and end domains. Translation of *NEB* starts from exon 3 and ends with the last exon, 183. There are three alternative splice patterns of exons 63-66, 143-144 and 167-177. See the structure of *NEB* in **Figure 6** (Chapter 4.1). (Donner, Sandbacka et al. 2004, McElhinny, Kazmierski et al. 2003)

Exons 63-66 are always expressed together. This means that the transcript would either contain all exons 63-66 or none of them. In this thesis, transcribed SR fragment with exons 63-66 is called S11a and without S11b. Exons 143 and 144 are exclusively spliced meaning that either exon 143 or 144 is included, but never both in the same transcript. In human fetus, exon 143 is expressed in brain tissue and 144 is expressed in skeletal muscles. In adults transcription of these exons varies, but both exons expression occurs in skeletal muscle fibers. In this thesis, the SR containing exon 143 is referred to as S21a and the SR containing exon 144 is referred to as S21b. Exons 167-177 are independently spliced, which leads to 121 different transcripts. These exons are located outside the SR area. (Donner, Sandbacka et al. 2004, McElhinny, Kazmierski et al. 2003, Ottenheijm, Granzier et al. 2012, Laitila, Hanif et al. 2012)

In addition, there is a 24 exon triplicated region formed of eight almost identical exons repeated thrice, exons 82-89, 90-97 and 98-105. The triplicate region locates in the middle of protein and is included in SR-part of nebulin. This region is thought to be the result of *Alu* and LINE transposable elements and two duplication events in a human ancestor (Kiiski, Lehtokari et al. 2015). The clusters of 8 exons of the triplicates differ from one another only by 1% of the

sequence, which makes studying this region rather difficult. It has been hypothesized that the triplicated region would have alternative splicing, but no experimental evidence of this exists to this date. (Laitila, Hanif et al. 2012, Donner, Sandbacka et al. 2004, McElhinny, Kazmierski et al. 2003)

Over 200 pathogenic nebulin variants have been identified and some of these variants lead to alterations in nebulin-actin interactions (Lehtokari, Kiiski et al. 2014a). This, however, cannot be effectively studied using bioinformatics tools because of the enormous size and repetitive modular structure of nebulin. Therefore, alterations in actin-nebulin interaction has to be analyzed functionally.

3.2. Protein structure

Nebulin, encoded by *NEB*, is a 600-900 kDa filamentous protein located in the thin filament, as mentioned before. The largest part of nebulin, approximately 90%, is covered with the repetitive area that stretches between unique end domains and consists of 30-35 amino acid long sequences of α-helical tandem repeats called simple repeats. Each simple repeat contains one actin-binding SDxxYK-motif. There are 185 simple repeats in nebulin, M1-M185, see their positioning shown in **Figure 6**. The SR area consists of 22 unique SRs (26 different SRs due to alternative splicing of exons), located from M9 to M162 (see **Figure 6**). A SR is formed of seven simple repeats, meaning that each SR contains seven actin-binding sites. In addition, each SR contains a conserved motif for tropomyosin binding, WLKGIGW. In SR 21 there is a binding site for kelch-like family member 40 (KLHL40). The general protein structure, position in sarcomere and interaction partners is shown in **Figure 3**. (Ottenheijm, Granzier et al. 2012, Labeit, Ottenheijm et al. 2010)

A glutamic acid-rich domain resides at the N-terminal end, which is connected with simple repeats M1-M8 to the SRs. This end of nebulin stretches towards the M-line. At the repeats M1-M3, leiomodin-3 and tropomodulin binding sites are located (see **Figure 3**). More detailed structure of the N-terminus is undetermined. (Ottenheijm, Granzier et al. 2012, Donner, Sandbacka et al. 2004)

Nebulin C-terminus ends with a serine-rich domain capped with Src-homolog-3-domain (SH3-domain) bound to Z-disc. These two domains serve as a binding site to numerous proteins including titin, xin actin-binding repeat-containing protein 2 (XIRP2), palladin, myopalladin, α-actinin, neural Wiskott-Aldrich syndrome protein (N-WASP), zyxin, xin, cysteine-rich protein (CSRP) and F-actin-capping protein subunit alpha-1 (CapZ). The C-terminus and SRs

are connected with 23 simple repeats long region, M163-M185. Z-disc region covers the simple repeats M171-M185. The repeats M163-M170 connect nebulin SR area to Z-disc region and contain binding sites for desmin and CapZ (see **Figure 3**). (Ottenheijm, Granzier et al. 2012, Chu, Gregorio et al. 2016)

3.3. Nebulin terminal interactions

The N terminus of nebulin is located in the proximity of the thin filament pointed end, which extends past nebulin N-terminus. As mentioned earlier, nebulin N-terminus has a high-affinity binding site for leiomodin-3 (LMOD3) and tropomodulin (Tmod), a thin filament capping protein that has two tropomyosin- and two actin-binding sites. LMOD3 has one tropomyosin-binding site and three actin-binding sites. Since thin filament pointed end extends past the N-terminus of nebulin, the interaction between Tmod and nebulin was suggested to be transient, present only early in myofibril assembly for thin filament assembly regulation. According to one of the hypothesis, nebulin recruits Tmod to the location specified by nebulin length during myofibril assembly. From this position Tmod can diffuse and cap thin filament. More studies are required to back up this hypothesis. (Chu, Gregorio et al. 2016, Ottenheijm, Granzier et al. 2012)

The C-terminus of nebulin is located in the Z-disc of the sarcomere and the section between SRs and Z-disc next to the C-terminus called linker repeats is also part of the C-terminal zone. This linker region interacts with desmin and CapZ. Desmin is a part of cytoskeleton structure called intermediate filament. The interaction between desmin and nebulin linker repeats leads to the participation of desmin in regulation of thin filament length, spacing between thin filaments in Z-disc and also, in myofibril alignment. CapZ is the actin end capping protein, which provides structural stability to the sarcomere. It binds to nebulin at the linker repeats and SH3-domain located in Z-disc. Based on the locations of CapZ binding sites on nebulin it was hypothesized that nebulin might cross-link two adjacent actin filaments at the Z-disc periphery. This is a model requiring thorough research to be confirmed.(Chu, Gregorio et al. 2016, Pappas, Bhattacharya et al. 2008)

Numerous proteins bind to the serine-rich domain and SH3-domain at the end of C-terminus. These proteins can be divided into three groups based on their function in association with nebulin. The first group partakes in the cytoskeletal organization and consists of α-actinin, myopalladin, palladin, CSRP, and zyxin. The second group participates in myofibrillogenesis and includes xin, XIRP2, and N-WASP. Their interaction with nebulin is transient, occurring

only during myofibril assembly and remodeling (Eulitz, Sauer et al. 2013). The third group consists of a single protein, the gigantic titin sized approximately 3.7 MDa. The nebulin C-terminal SH3-domain can bind to two proline-rich regions of titin, one of them located in Z-disc (Zis1) and the other in the I-band (PEVK). The Zis1 region of titin has the corresponding location with SH3 unlike PEVK region, but functionally this interaction is largely unclear. (Yamamoto, Vitiello et al. 2013)

3.4. Nebulin-actin interaction

Binding to actin is a dynamic process that can be assessed through different methods. One of these methods is actin co-sedimentation. This method is an *in vitro* assay that can help analyze binding of specific proteins or protein fragments to F-actin, measure their binding affinities. This method is based on the big size of polymerized actin, which sediments during centrifugation at high speed. If protein fragments bind to actin, they co-sediment with it during ultracentrifugation making it possible to determine the amount of bound protein compared to unbound at specific protein concentrations. (Srivastava, Barber 2008, Heier, Dickinson et al. 2017)

Among other proteins, nebulin is one of the proteins with actin-binding capacity. A number of studies were conducted to understand the nebulin-actin interactions concentrating on nebulin as whole protein and as nebulin fragments *in vitro*. The studies concluded that nebulin binds to actin with high affinity (Ottenheijm, Granzier et al. 2012, Labeit, Kolmerer 1995, McElhinny, Kazmierski et al. 2003). Nebulette studies also demonstrate nebulette binding to actin with high affinity, though it lacks the SR area present in nebulin. (Ogut, Hossain et al. 2003)

Only a few research attempts have been conducted to investigate the differences in actin-binding strength of nebulin fragments. The protein fragmentation approach was made possible only after determination of nebulin and *NEB* sequences. In the early studies, it was indicated that the actin-binding strength can vary from one simple repeat to another leading to different actin-binding strength of SRs (Jin, Wang 1991). This was confirmed by later studies. (Laitila, Lehtonen et al. 2019, Marttila, Hanif et al. 2014, Lehtonen 2017, Laitila, Hanif et al. 2012)

In order to study the actin-nebulin binding Laitila and Lehtonen with their colleagues created and described nebulin SR panel (Laitila, Lehtonen et al. 2019, Lehtonen 2017). The panel consists of 26 protein fragments corresponding to the SRs of nebulin, and it was created using minigene approach. Actin-binding strength of the SR panel was determined using actin co-sedimentation assay. There were significant differences in actin-binding strength observed

across the panel presented below (**Figure 5**). In addition to wild-type SRs, five variant-containing nebulin SR fragments were tested. (Laitila, Lehtonen et al. 2019, Lehtonen 2017)

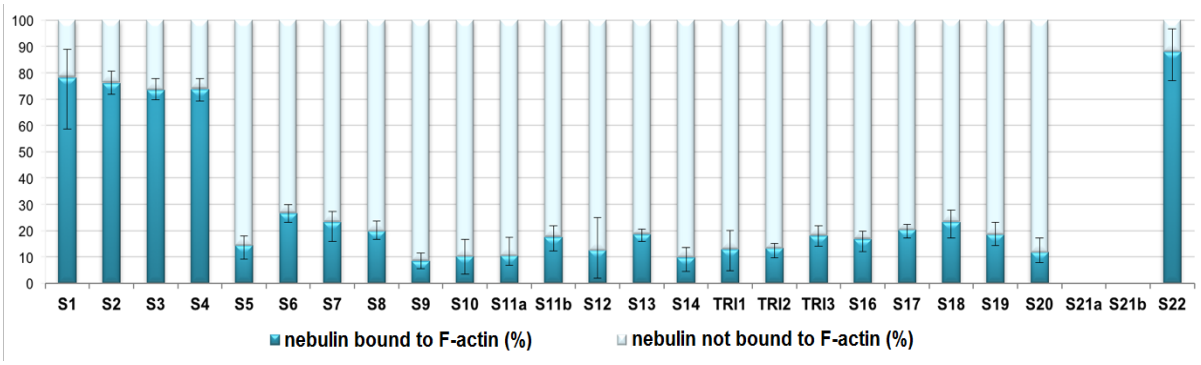


Figure 5. **Actin-binding strength of SRs.** Strong binding was observed at the ends of SR region of nebulin, in S1-4, and in S22. In the middle of nebulin, actin binding was weak. The extraction of the SRs S21A, and S21B from cells was unseccessful and therefore these SRs could not be tested for actin-binding strength. (Laitila, Lehtonen et al. 2019, Lehtonen 2017)

3.5. Other nebulin interactions in SR area

The KLHL40 binding site is located at the end of the SR area in nebulin, in S21. It was suggested that this interaction provides thin filament stability by blocking nebulin degradation and promoting the correct folding of the nebulin. KLHL40 was also demonstrated to interact with leiomodin-3 and regulate its protein levels. (Chu, Gregorio et al. 2016, Garg, O'Rourke et al. 2014)

The predicted tropomyosin binding site on each nebulin SR indicates the interaction between nebulin and Tm, and possibly also the TnT complex. The studies of this interplay show occurrence of interactions between Tm, Tn-complex and nebulin *in vitro* using nebulin fragments and nebulin-like motifs (Marttila, Hanif et al. 2014, Ogut, Hossain et al. 2003). The nebulin-like motifs were confirmed to bind to Tm, stronger to TnT and even stronger to Tm-TnT-complex. Addition of TnI to the Tm-TnT-complex showed a slight decrease in binding affinity of nebulin-like motifs to the complex. However, the affinity of nebulin-like motifs to actin was proved to decrease in presence of Tm-TnT-TnI-structure. On the other hand, the presence of nebulin-like motif increased Tm-Tn-complex affinity to actin. This result demonstrates the presence of nebulin and Tm binding domains on actin that are not mutually exclusive, though affected by interactions of each other. (Ogut, Hossain et al. 2003)

3.6. Functions

Nebulin has multiple tasks in skeletal muscle fibers. The main nebulin function was thought to be working as a scaffold for thin filament proteins. The periodicity of the SR region matches

to the structure of actin filaments, seven nebulin simple repeats and seven actin monomers per α-helical turn, which supported this hypothesis. This was shown not to be exact by knockout (KO) mouse studies. It was determined that the absence of nebulin leads to varying thin filament length, but not their ablation. This indicates that nebulin regulates thin filament minimal length and to contributes to final thin filament length by stabilizing it rather than being the sole determinant. (Lin, Taejeong et al. 2017)

Mouse KO studies revealed nebulin association with other functions such as regulation of skeletal muscle contraction. KO mice generate significantly less force, which can be due to decreased actin-myosin cross-bridge formation as a result of decreased thin filament length. It has also been shown that nebulin participates in calcium handling of the sarcoplasmic reticulum. (Chu, Gregorio et al. 2016, Ottenheijm, Granzier et al. 2012)

C-terminal interactions of nebulin and desmin were shown to be essential to the lateral alignment of myofibrils and the sufficient desmin assembly at Z-disc. Nebulin is also necessary for correcting Z-disc structure, nebulin KO mice had wider Z-discs and nemaline bodies typically found in nemaline myopathy patients (Tonino, Pappas et al. 2009). In addition, reduction of nebulin in myocytes decreases CapZ assembly, which leads to non-uniform organization of the barbed ends of the thin filament (Pappas, Bhattacharya et al. 2008). Surprisingly, deletion of the SH3-domain from nebulin C-terminus has very little effect on thin filament formation and function considering its long list interacting partners. The only consequences are increased susceptibility to contraction-induced injury and slightly decreased stress generation under isometric conditions (Yamamoto, Vitiello et al. 2013). (Chu, Gregorio et al. 2016)

Nebulin was also shown to be required for successful myofibrillogenesis. In particular, nebulin was suggested to initiate myofibrillar actin filament formation through nebulin interactions with N-WASP, xin, and XIRP2. Nebulin interactions with xin and XIRP2 also play a significant role in myofibrillar remodeling in adult muscle fibers. (Eulitz, Sauer et al. 2013, Chu, Gregorio et al. 2016)

4. Nemaline myopathy and other *NEB*-related disorders

Congenital myopathies are muscle disorders characterized by muscle weakness, poor muscle bulk, and dysmorphic features. Nemaline myopathy (NM) is one of the most common congenital myopathies. Variants in *NEB* were most commonly associated with NM, but

according to the more recent findings, the discovered *NEB* variants are associated with other disorders as well. These disorders are distal nebulin myopathy, a distal form of NM, core-rod myopathy, distal core-rod myopathy, and fetal akinesia. (Lehtokari, Kiiski et al. 2014b)

4.1. Nemaline myopathy (NM)

Nemaline myopathy (NM) is a rare congenital myopathy that affects primarily the skeletal muscles. It is, however, the most common congenital myopathy. Mutations in twelve genes have been identified to cause NM. Recessive *NEB* variants and *de novo* dominant *ACTA1* variants are the most frequent disease-causing variants. Hallmarks of this disorder are nemaline bodies in muscle tissue samples and muscle weakness that usually manifests in proximal muscles such as face, neck, and trunk as well as upper legs and arms, even though distal muscles can be affected as well. Nemaline bodies are rod-like formations in muscle fibers consisting of proteins located in thin filament and Z-disc. Other possible clinical features are depressed or absent deep tendon reflexes, difficulties feeding and swallowing, foot and joint deformities, scoliosis, respiratory insufficiency and decreased fetal movement. Respiratory insufficiency is the most common cause of death in NM patients. The majority of survived children past the age of two learn to walk independently. (Romero, Sandaradura et al. 2013, Donner, Sandbacka et al. 2004)

In 1999 the 70th European Neuromuscular Centre (ENMC) International Consortium on Nemaline Myopathy defined six clinical NM subtypes based on the phenotype severity and age of onset (Romero, Sandaradura et al. 2013). In order from most severe to the mildest, the subtypes are severe, intermediate, typical congenital, mild, adult-onset and other, such as Amish (Laitila, Lehtonen et al. 2019). Also, new NM types caused by splice variant in *TNNT3* and null mutation in *MYO18B* were revealed recently (Sandaradura, Bournazos et al. 2018, Alazami, Kentab et al. 2015). Though the myopathy caused by mutation in *MYO18B* was not yet concluded to be classified as "nemaline myopathy" due to variety of other symptoms not typical to NM (Laitila, Lehtonen et al. 2019). The symptoms defined by the severity and the age of onset for each NM form can overlap as the disease progresses. Clinical inclusion criteria for each form of NM are listed in **Table 1**. The exclusion criteria from the most common form of NM, typical NM, are contractures or fractures at birth, lack of spontaneous movements or respiration at birth or both, failure to sit or walk independently and use of wheelchair by the age of 11 (Lehtokari, Kiiski et al. 2014a).

Table 1. **Clinical inclusion criteria for six NM forms.** Modified (Lehtokari, Kiiski et al. 2014b).

Form of NM	Clinical inclusion criteria
Severe NM	Onset at or before birth: no spontaneous movements; no spontaneous respiration, or with severe contractures or fractures at birth
Intermediate NM	Infantile-onset: patient breathing and moving at birth, but unable to maintain respiratory independence, or to sit and walk independently; use of a wheelchair before the age of 11; contractures developing in early childhood
Typical congenital NM	Onset in infancy: typical distribution of muscle weakness (weakness most pronounced in facial, bulbar, and respiratory muscles, neck flexors, and limb-girdle muscles; initially proximal, later also distal limb involvement); motor milestones delayed but reached; slowly progressive or nonprogressive
Mild NM	Childhood or juvenile onset
Adult-onset NM	Adult-onset (from age 20 to 50)
Other types of NM	NM caused by recessive *TNNT1*- mutation (Amish): early childhood onset, originally described in the Old Order Amish population of Pennsylvania; NM caused by *TNNT3*-splice variant

4.2. Distal nebulin myopathy

Variants in *NEB* were long believed to cause only NM until two homozygous *NEB* missense variants were identified. In homozygous forms these variants caused a clinically milder phenotype of distal muscle weakness and absence of nemaline bodies in tissue samples. One of the missense variants, p.Ser6366Ile, increases nebulin-actin affinity and the other, p.Thr7382Pro, reduces nebulin-tropomyosin affinity (Marttila, Lehtokari et al. 2014). When these variants are found in patients as compound heterozygous form with more disruptive variants, their combination is known to cause NM. Hence, it was hypothesized that less disruptive variants could cause milder, nemaline rod free myopathies. (Lehtokari, Kiiski et al. 2014b)

4.3. Distal NM

After the discovery of distal nebulin myopathy, a distal NM was described. The main difference compared to distal nebulin myopathy is the presence of nemaline bodies in the diagnostic biopsy. A combination of heterozygotic *NEB* variants was identified in one Finnish patient, a French family, and a Hungarian family, all unrelated to each other. The combination of two mutations in each patient, one more disruptive i.e. frameshift and one less disruptive mutation i.e. missense mutation or splice site mutation, lead to a similar phenotype in all of the patients. Curiously, the same missense mutation, p.Ser6366Ile, is found in the Finnish patient with distal

NM and in patients with distal nebulin myopathy. This finding confirms that *NEB* variants that cause NM and distal myopathies form a clinical and histological continuum. (Lehtokari, Kiiski et al. 2014b, Lehtokari, Pelin et al. 2011)

4.4. Core-rod myopathy

Typical core-rod myopathy patients exhibit proximal muscle weakness and histological muscle biopsy findings show core-shaped structures in addition to nemaline rods. The presence of cores in muscle tissue is exclusively typical for central core myopathy. The severity of core-rod myopathy varies from mild to severe. The *NEB* variants causing core-rod myopathy follow autosomal recessive inheritance pattern. (Romero, Lehtokari et al. 2009, Lehtokari, Kiiski et al. 2014b)

4.5. Distal core-rod myopathy

Recessive *NEB* variants can also cause distal core-rod myopathy as a compound of heterozygous variants. Histologically the muscle tissue samples of the patients show definitive core and rod structures. Clinically distal core-rod myopathy resembles distal NM by distal muscle weakness and, in some patients, neck flexor and respiratory muscle weakness. Interestingly, one of the *NEB* variants present in distal core-rod myopathy, frameshift variant in exon 171, is shared by a core-rod myopathy patient and severe NM patient. (Lehtokari, Kiiski et al. 2014b)

4.6. Fetal akinesia

In studies of severe nemaline myopathies, recessive *NEB* variants with fetal akinesia deformation sequence (FADS) were identified that lead to fetal akinesia phenotype in affected infants. Fetal akinesia refers to a spectrum of disorders with different underlying pathologic mechanisms with unifying symptom of reduction or lack of fetal movement. Affected infants displayed characteristic fetal akinesia symptoms such as reduced fetal movement, polyhydramnios, arthrogryposis, rocker-bottom feet, talipes, cleft palate, low-set ears and a lack of spontaneous breathing at birth. (Ravenscroft, Sollis et al. 2011) Some patients were performed muscle biopsy on and were shown almost complete replacement of muscle with adipose tissue. (Lehtokari, Kiiski et al. 2014b)

4.7. Pathogenic nebulin variants

Nebulin is thought to tolerate variation well, which makes identification of variant pathogenicity more complicated. Only 7% of all the nebulin variants are hypothesized to be

pathogenic. In addition, variants are mostly scattered evenly across the *NEB* gene, the only exception being intron 32, where the variant frequency is notably higher. (Lehtokari, Kiiski et al. 2014a)

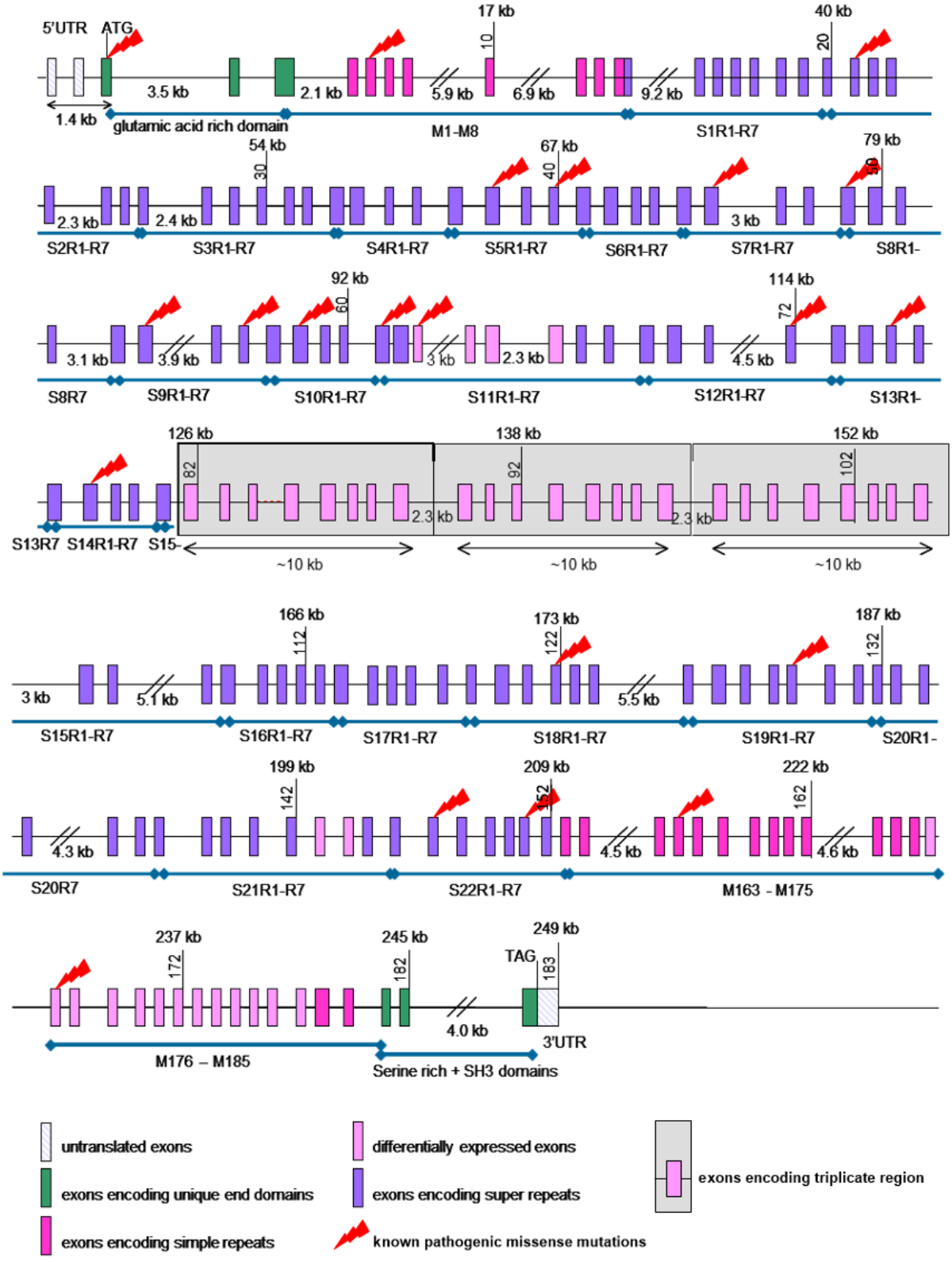


Figure 6. **Structure of *NEB* and its known pathogenic missense mutations.** Modified (Donner, Sandbacka et al. 2004).

In Helsinki, over 200 pathogenic *NEB* variants were identified by (Lehtokari, Kiiski et al. 2014b) using a variety of methods including single-strand conformation polymorphism, denaturing high-performance liquid chromatography, genomic or cDNA Sanger sequencing, genomic or exomic next-generation sequencing (NGS), copy-number variants detecting microarray, and targeted NGS panel of known neuromuscular genes. Majority of the *NEB*-variants presented in patients occurred as a combination of two different heterozygous variants. The rest of the arisen variants could either be in homozygous form only or in homozygous and heterozygous forms. The found variant types were splice-site mutations (34%), frameshift mutations due to small, less than 20 bp long, indels (32%), nonsense mutations (23%), missense mutations (7%) and large, over 1kb long, deletions and duplications (4%). The missense variants were speculated to be possibly pathogenic by affecting nebulin-actin or nebulin-tropomyosin interactions. These missense mutations have been marked in **Figure 6**. (Modified from (Donner, Sandbacka et al. 2004).

4.8. Other NM causing genes

Variants in 11 genes can lead to nemaline myopathy. The majority of these variants were found in *NEB* and *ACTA1*. Other identified variants were found in *TPM2, TPM3, TNNT1, TNNT3, CFL2, KBTBD13, KLHL40, KLHL41, LMOD3* and *MYPN.*

ACTA1 is a highly conserved gene located at 1q42.13 that encodes α-actin. Due to its conserved nature, mutations are not well tolerated in actin and NM patients often present severe phenotype. Known NM causing variants are evenly spread across exons and follow either autosomal recessive or *de novo* autosomal dominant inheritance pattern. Mosaicism can occur in autosomal dominant variants. Autosomal dominant variants are most commonly associated with NM. (Laing, Dye et al. 2009)

TPM2 is located in 9p13.3 and it encodes β-Tm. *TPM3* is located in 1q21.3 and it encodes α-Tm respectively. NM causing variants can follow both inheritance patterns, recessive or dominant. *TMP3* variants cause more severe phenotype in comparison with *TPM2* variants. All the NM causing variants are evenly dispersed along the genes causing many different malfunctions in the Tm proteins. Most frequently variants affect Tm binding to actin. (Marttila, Lehtokari et al. 2014)

TNNT1 is TnT coding gene located in 19q13.42. Recessive *TNNT1* variants can cause NM by disrupting the normal function of Tn complex in the sarcomere. The first discovered recessive

nonsense variant caused unique severe NM if homozygous. It was found in the Amish population. (Johnston, Kelley et al. 2000)

TNNT3 encodes troponin T (TnT), a thin filament structure protein, and it is located in chromosome 11. The splice variant, that can be the cause for severe congenital NM was previously associated only with distal arthrogryposis. However cDNA studies confirmed the splice variant as the likely cause also for NM. (Sandaradura, Bournazos et al. 2018)

CFL2 encodes a protein called cofilin-2 and is located at 14q13.1. Cofilin-2 is found in skeletal muscles and is thought to function as a stabilizer of the sarcomeric thin filament. Known NM causing *CFL2* variants are autosomal recessive. (Agrawal, Greenleaf et al. 2006)

KBTBD13, KLHL40, and *KLHL41* are located at 15q22.31, 3p22.1 and 2q31.1 respectively. All three genes encode proteins from the protein family containing kelch repeats and BTB domain. These structures assist protein interactions in the cell involved in transcription regulation, myofibril arrangement, and cytoskeleton modification. *KBTBD13* variants follow autosomal dominant inheritance pattern and *KLHL40* and *KLHL41* follow autosomal recessive inheritance pattern. (Adams, Kelso et al. 2000, Sambuughin, Yau et al. 2010)

LMOD3 is the leiomodin-3 coding gene located at 3p14.1. Leiomodin-3 is part of tropomodulin protein family and participates in the regulation of thin filament length and organization. It has been suggested that one of leiomodin-3 key roles is its function as a link between the thin filament and kelch protein family offering a common pathogenesis pathway for multiple different genetic forms of NM. Autosomal recessive *LMOD3* variants are associated with phenotypically severe NM. (Yuen, Sandaradura et al. 2014)

MYPN is a myopalladin coding gene and it is located in 10q21.3. Myopalladin is located in Z discs and I bands of a sarcomere and is able to bind to nebulin and actin. NM causing *MYPN* variants follow autosomal recessive inheritance mode and are biallelic loss-of-function mutations, which lead to phenotypically mild, slowly progressing NM with later age of onset. (Miyatake, Mitsuhashi et al. 2017)

5. 3D protein structure prediction of nebulin

There are three approaches in predicting 3D structure of a protein from the sequence, *ab initio* modeling, using threading methods and comparative modeling (CM). The approach is selected based on the availability of the template structures in PDB-databases. Structures that have no related proteins in PDB library must be built using *ab initio* modeling, which is time-consuming

and successful only for short sequences, below 120 amino acids long.(Wu, Skolnick et al. 2007) CM is based on evolutionary-related homology identified by sequence or sequence profile comparisons, based on which high-resolution models can be produced. Threading methods use similar approach as CM, comparing the query sequence to the 3D-structures of the evolutionarily unrelated proteins aiming to recognise similar folds as in the query sequence. (Roy, Kucukural et al. 2010, Zhang 2008)

Nebulin is a difficult target protein to research due to its size and highly repetitive sequence composition. This makes *ab initio*-modeling very time- and resourse-consuming method. Also, lack of data in PDB libraries makes its prediction complicated using only the CM and threading methods. Therefore, no single approach would be effective in studying nebulin.

5.1. I-TASSER

I-TASSER (**I**terative **T**hreading **ASSE**mbly **R**efinement) software implements a combined approach of 3D-structure prediction, using *ab initio*-modelling, threading methods and atomic-level structure refinement. Composite approaches, including I-TASSER, have been demonstrated to be advantageous in community-wide CASP experiments. (Roy, Kucukural et al. 2010)

The I-TASSER workflow is presented in **Figure 7.** The first step of I-TASSER software is threading. In this step, similar structures or structural motifs as the query are determined using PDB library. First, the query is aligned with non-redundant sequence database by PSI-BLAST, homologs are aligned and the sequence profile is created, which is then used to predict secondary structures by PSIPRED. The sequence profile and secondary structures are then run through PDB structure library using LOMETS, a meta-threading server that combines FUGUE, HHSEARCH, MUSTER, PROSPECT, PPA, SP3

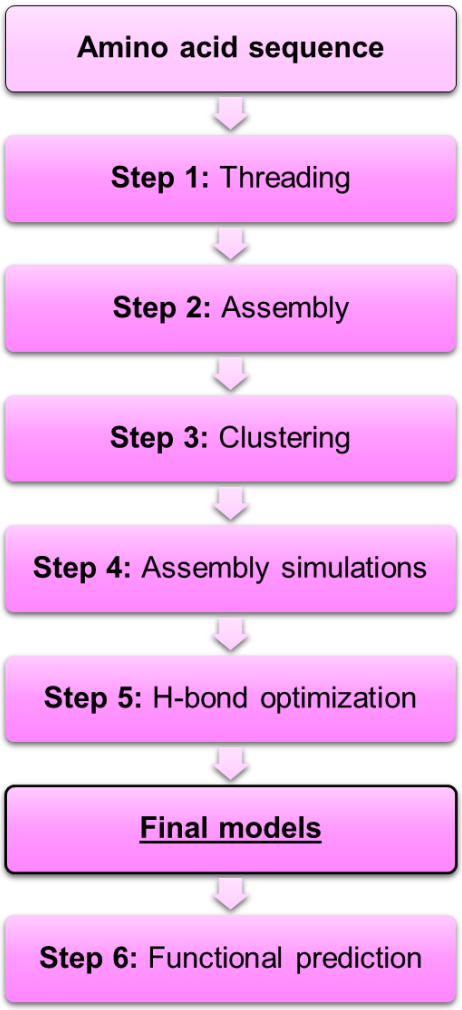


Figure 7. **Outline of I-TASSER protocol for protein structure- and function prediction.** The protocol outline is presented in steps.

and SPARKS threading programs. (Roy, Kucukural et al. 2010)

The second step, assembly, follows. Continuous template fragments that are well aligned with the query in the alignments are cut from the template structures and are used to assemble with the unaligned regions constructed by *ab initio* modeling. The assembly is performed using Monte Carlo simulation technique, which performs simulations at different temperatures in parallel. The low temperature structures are then clustered by SPICKER software in order to determine the clusters of structures with low free energy state.(Roy, Kucukural et al. 2010)

In the next step, assembly simulation, a simulation is performed starting from the previously determined clusters with the external constraints provided by LOMETS alignments and PDB structures. The aim is to remove steric tension and to refine the global topology. The assembly simulations, like after assembly, finish in clusters and the lowest free energy structures are used as input for program called REMO. It performs hydrogen-bond optimization and delivers the final models.(Roy, Kucukural et al. 2010)

In the final step, possible protein functions are predicted. The determination of the query protein functions is based on searching the structural matches in the protein structure-function libraries created for this purpose. As a result, for each query, known (GO) terms, ligand binding sites and whether the protein has enzymatic activity. (Zhang 2008)

The quality of the final predicted models is assessed with confidence score (C-score) and local accuracy score (L-score). C-score is determined based on the quality of threading alignments and the C-score above -1,5 is used as a cutoff to determine the correct topology. The L-score is defined as distance deviation for each amino acid between predicted and native structure. The average error of C-score is 0.08 and of L-score 2.21 Å. (Zhang 2008)

There are two ways of using I-TASSER, as a stand-alone software, I-TASSER suite, or via I-TASSER server provided by Zhang Lab (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The most important difference in Suite and Server versions is the lack of some threading programs in Suite, such as HHpred, SP3, FFAS, due to the copyright issues. The server usage offers three additional options, assigning additional restraints and templates, excluding some templates and specifying a secondary structure for specific residues. In addition to server options, the stand-alone software allows the user to specify the sequence identity cut off for homologous templates, the top template output number for each threading program, the number of output models, and simulation length defined in hours. It also gives an option to exclude homologous templates and to remove trajectory files containing low-temperature replicas in I-

TASSER simulations. Because of the heavy workload, only one task at a time can be run at the server whereas multiple tasks can be run at the same time locally using Suite. The runtime on the server is approximately two days for 500 amino acid long sequence. With stand-alone software similar runs can take from a day to over a week depending on parallelization and simulation time options. The runtime is heavily influenced by the sequence and the coverage of the template threads. Both server and stand-alone software are capable of producing high-quality models. (Roy, Kucukural et al. 2010, Zhang 2008)

Experimental part

6. Aims

In the NM patient database used for this study, there are 70 families with one mutation found, which was verified to be pathogenic. In 30 families there was additionally found a missense mutation. Missense variants are considered to be well tolerated in *NEB* and the bioinformatic tools usually fail to predict their impact on protein function and pathogenicity. Therefore, the pathogenicity of these variants must be assessed functionally. Current knowledge about protein interactions within the thin filament is incomplete and one of the hypothesized disease-causing mechanisms in *NEB*-related myopathies is altered interaction between nebulin and its binding partners. This study targets nebulin-actin interactions.

6.1. Aim 1: Corresponding variants in SRs with different actin-binding capacity

Due to the differences in actin-binding capacity of SRs compared to each other, one of the aims was to determine whether corresponding mutations in different SRs would have a similar or different effect on actin-binding capacity.

6.2. Aim 2: Missense variants in *NEB* and nebulin-actin interaction

Missense mutations found in *NEB* from nemaline myopathy patients can be pathogenic modifying factors or have no effect on nebulin function. For this thesis, seven missense variants were selected to study the effect of these mutations on actin-binding capacity compared to wild-type nebulin using the SR panel (see **Chapter 3.4.**).

7. Materials and methods

7.1. Workflow

To achieve research aims, the workflow shown in **Figure 8** was followed. The variants studied in this thesis, are presented in **Chapter 7.2**.

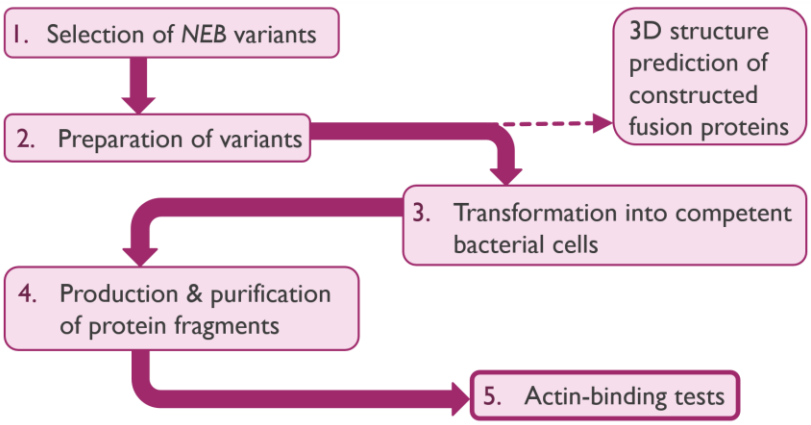


Figure 8. **Study workflow.**

For all nebulin variants and wild-type structures, 3D-models were predicted using I-TASSER suite and I-TASSER server from amino acid sequences. The best predicted models were visualized with PyMOL and their reliability was assessed based on the confidence- and local accuracy scores provided by software (see **Chapter 7.2.3**).

The wild-type SR fragments were inserted into the pGEX-4T-1 expression vector as described in **Chapter 7.4.1**. Selected variants were produced using site-directed mutagenesis into wild-type SR fragments with specifically designed "mutagenesis primers" (see **Chapter 7.4.2**). The mutagenized plasmids were transformed into DH5α-bacterial cells (see **Chapter 7.5.1**). The vectors with the right inserts were selected and verified by capillary sequencing (Sanger sequencing, https://www.fimm.fi/en/services/technology-centre/sequencing) described in **Chapter 7.5.2**. These vectors were transformed into BL21-cells and the sequence was confirmed once more with Sanger sequencing (see **Chapter 7.5.3**). All the variants were prepared in the same manner for this study.

BL21 transformants were cultured and SR-fragments were produced using IPTG-induced protein expression. The protein fragments were produced as GST-fusion proteins (see sequences in **Attachments**) and purified using glutathione agarose beads. Successful production was verified with SDS-PAGE and the protein fragments were eluted from the beads (see **Chapters 7.6.1-7.6.3**). The elutes were purified twice by ultracentrifugation following actin binding test kit protocol (see **Chapter 7.6.4**). The success of the elution step was verified with SDS-PAGE (see **Chapter 7.7.3**).

The actin binding experiments were performed as *in vitro* co-sedimentation assay described in **Chapter 7.7**. The protein bound to F-actin sediments with it into pellet during ultracentrifugation leaving unbound fraction in the supernatant. After ultracentrifugation step, both pellet and supernatant fractions were run separately on SDS-PAGE gels (see **Chapter 7.7.3**). The intensities of the bands were quantified from the gels with ImageJ and the percentage of bound nebulin fragments to F-actin was determined. The results were statistically analyzed in comparison with wild-type SR actin binding results performed previously (see **Chapter 7.8**).

7.2. Selection of *NEB* variants

7.2.1. Corresponding variants in SRs with different actin-binding capacity

Two SRs were selected; SR1 was picked as a strongly binding SR and SR7 as weakly binding SR. Two of the variants, S1-m-1 and S7-m-2, were selected from the NM patient database (see

Table 2). From these variants, corresponding variants were created; S1-m-2 corresponds to S7-m-2 and S7-m-1 corresponds to S1-m-1. These two variants were constructed by replacing three nucleotides and were not found in patients. Also, for S7-m-2 there was one heterozygoe carrier lised in Exac. Both studied deletions were not found in patients, but for S7-del-1 there was one heterozygote carrier listed in Exac. See data about the variants in **Table 2** and their visualized locations in **Figure 9**.

Table 2. **Variant data for the first experiment.** Locations of the mutations in cDNA reported according to GenBank ID NM_001164507.1, in its translation, GenBank ID NP_001157979.1, in relation to actin- and tropomyosin binding sites and whether the variant was found in a patient or in Exac is listed below.

Variant	*NEB* fragment	Mutations	Amino acid change, exon location	Relation to binding sites	Patient/ No Exac/ No
S1-m-1	ex 13-20, SR 1	c.1436A>G	p.Tyr479Cys, exon 16	Not at binding site	Patient No
S1-m-2	ex 13-20, SR 1	c.1480_1482delAAA insTGT	p.Lys494Cys exon 17	At the actin-binding site	No No
S1-del-1	ex 13-20, SR 1	c.1372_1377del	p.Tyr458_Lys459del exon 16	At the actin-binding site	No No
S7-m-1	ex 45-49, SR 7	c.5836_5838delGAG insTGT	p.Glu1946Cys exon 46	Not at binding site	No No
S7-m-2	ex 45-49, SR 7	c.5881C>T	p.Arg1961Cys exon 46	At the actin-binding site	Patient Exac
S7-del-1	ex 45-49, SR 7	c.5770_5775del	p.Tyr1924_Lys1925del exon 46	At the actin-binding site	No Exac

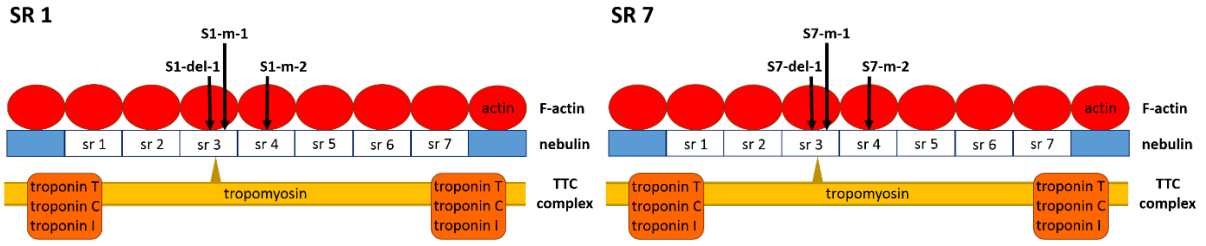


Figure 9. **A scheme of SR 1 and 7 shows the locations of the studied variants.** Detailed figures show the binding sites of actin and tropomyosin to nebulin SR. In the figure simple repeats are marked "sr", each sr has one binding site for actin monomer and each SR has one binding site to tropomyosin. Arrows point to the variant locations.

7.2.2. Missense variants in NEB and nebulin-actin interaction

For this experiment, seven variants in four SR were selected, three variants from the beginning of SR-panel (SR2 and SR4), from SRs with high actin-binding strength, and four from weakly binding SRs in the middle part of nebulin. One of the four tested constructs is a combination of two variants: construct S10-m-2&3 has both variants (S10-m-2 and S10-m-3) in it. One of the aims in this experiment was to examine the possible influence of mutations close to the

actin-binding site on actin-binding strength. See data about the variants in **Table 3** and their visualized locations in **Figure 10**.

***Table 3.* Variant data for the second experiment.** Locations of the mutations in cDNA reported according to GenBank ID NM_001164507.1, in its translation, GenBank ID NP_001157979.1, and in relation to actin and tropomyosin binding sites is listed below.

Variant	*NEB* fragment	Mutations	Amino acid change	Relation to binding sites
S2-m-1	ex 20-27, SR 2	c.2318A>G	p.Tyr773Cys	At the actin-binding site
S4-m-1	ex 33-37, SR 4	c.3416C>T	p.Thr1139Met	Not at binding site
S4-m-2	ex 33-37, SR 4	c.3765A>T	p.Gln1255His	At the actin-binding site
S10-m-2	ex 57-61, SR 10	c.8072G>A	p.Arg2691His	At the actin-binding site
S10-m-3	ex 57-61, SR 10	c.8114C>G	p.Ser2705Cys	Not at binding site
S10-m-2&3	ex 57-61, SR 10	c.8072G>A, c.8114C>G	p.Arg2691His, p.Ser2705Cys	At the actin-binding site
S16-m-1	ex 109-114, SR 16	c.17462G>A	p.Arg5821His	At the tropomyosin binding site

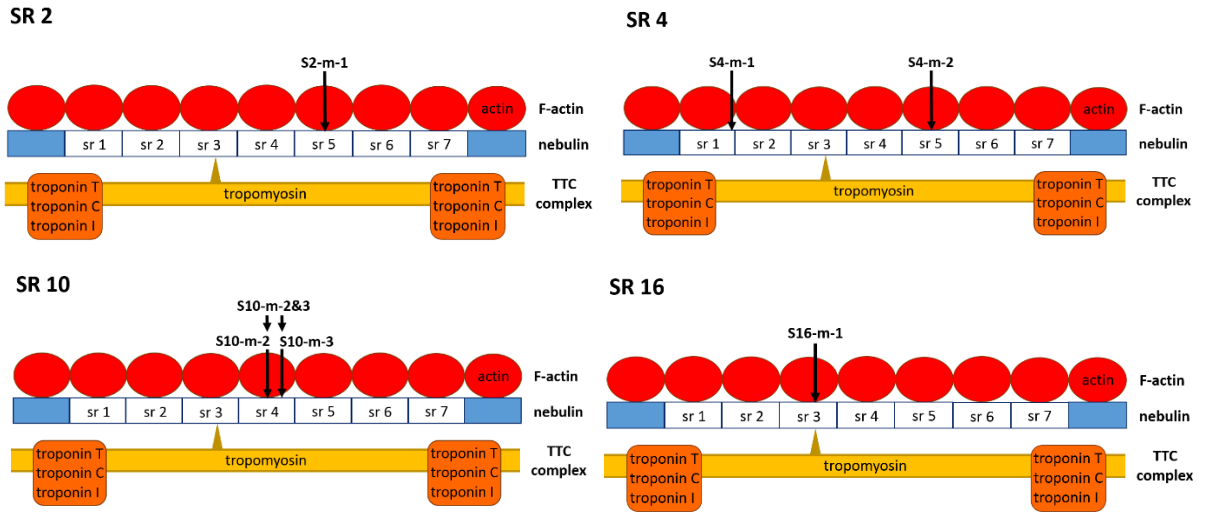


***Figure 10.* A scheme of SR 2, 4, 10 and 16 shows the locations of the variants.** Detailed figures show the binding sites of actin and tropomyosin to nebulin SR. In the figure simple repeats are marked "sr", each sr has one binding site for actin monomer and each SR has one binding site to tropomyosin. Arrows point to the variant locations.

7.2.3. 3D structure prediction with I-TASSER suite and I-TASSER server

The only input in I-TASSER 3D structure prediction is the amino acid sequence of the fusion proteins, which is determined from the DNA-sequence. The fusion protein constructs consist of super repeat fragments inserted at the end of GST gene in pGEX-4T-1-vector using either *EcoRI*- or *BamHI* restriction sites (see pGEX-4T-1 map in **Attachments**). The inserts for SRs 2 and 10 were synthesized in GenScript and the inserts contain only the required SR sequence. The rest of the structures were produced by reverse transcription from mRNA and were first cloned into pCR 2.1 TOPO-vector for selection of the correct inserts. The restriction of inserts

from TOPO-vector and its insertion into pGEX-4T-1 adds six nucleic acids to each end of the insert. Due to this production method, the inserts cover the required SRs with short redundant sequences at the ends of the SR areas. The sequences were determined and then translated to amino acid sequences using ExPASy (see **Online resources**, sequences in **Attachments**). The constructs were developed prior to this thesis.

In this thesis, the I-TASSER stand-alone software and the server version were used. I-TASSER server runs were carried out at default settings without any further specifications. The Suite was run on CSC server, Taito cluster (CSC, Kajaani, Finland). The files were transferred to the server using WinSCP software (SourceForge Media, La Jolla, CA). The software runs were specified to use all templates available with the default cutoff of 0.3 and to apply top 20 templates for threading programs (default setting). The I-TASSER simulations were defined to run in parallel, for a maximum of 60 hours. The number of final models was set to default 5 for the possibility of comparison and selection of the most accurate model. The removal of trajectory files was left at default. The output models were in.pdb files, general statistics for models were assembled in cscore-file and the L-scores in lscore.txt. The result files were transferred from CSC server with WinSCP.

The.pdb-files for each model were explored with PyMOL-software (Schrödinger, New York, NY, USA). Using this tool, the GST- and insert-parts, as well as the mutation points were color-coded and the image with the best representation of the protein structure was exported. The C-scores and the L-scores were compiled into Excel (Microsoft, Redmond, WA, USA). L-scores in I-TASSER output is stated for each amino acid of each of the predicted models. They were depicted and presented in this thesis in a scatter plot with their average to provide a detailed overview of the predicted models. The final models for each nebulin fragment were selected based on two criteria, the prediction accuracy determined by C- and L-scores, and similarity of models within the same SR.

7.3. Preparation of variants

7.3.1. Plasmid extraction

Wild-type SR constructs were cloned in pGEX-4T-1 expression vectors and transformed in *E.coli* BL21 bacterial cells previously.Small liquid cultures (5ml LB, Amp100) were started from the wild-type glycerol stocks overnight at +37°C. The plasmids were extracted from the liquid cultures with QIAprep® Spin Miniprep Kit (Qiagen, Hilden, Germany) following the Quick-Start protocol (February 2015). Only alterations to the protocol were that recommended

wash with PB-buffer was left out and the plasmids were eluted into 30 µl of sterile water instead of the Elution Buffer offered by the manufacturer. The plasmid concentration was determined using DeNovix DS-11 FX+ spectrophotometer (DeNovix Inc., Wilmington, DE, USA).

7.3.2. Site-directed mutagenesis

Next step was the mutagenesis reaction. All the desired variants were missense mutations or short in-frame deletions. The variants were created by incorporating the desired nucleotide in place of the original one in the central area of the primers, both forward and reverse. When the product was amplified using polymerase in PCR, the number of variant vectors increased exponentially. (Bachman 2013)

Site-directed mutagenesis reactions were performed using QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA) and the primers were designed using web-based QuikChange Primer Design Program provided by the manufacturer (see **Table 4**). The concentration of the ordered primers was by default 100 µM and they were diluted with sterile water to 100 ng/µl.

Table 4. **Primers for mutagenesis.** The variant nucleotide is marked bold and the triplet it is highlighted. The amino acid changes are also listed. In the case of S10-m-2&3, site-directed mutagenesis reaction was repeated once more to S10-m-3 with S10-m-2 primers.

Construct	Template	Primer	Sequence 5' to 3'	Amino acid change
S2-m-1	SR 2	F	cgaaacagcttagtgatctgaat**tg**caaagcaaaacatgaagg	p.Tyr773Cys
		R	ccttcatgttttgctttg**ca**attcagatcactaagctgtttcg	
S4-m-1	SR 4	F	gagaagacaaagtccaaatacaac**atg**ccccatgatatg	p.Thr1139Met
		R	catatcatggggc**at**gttgtatttggactttgtcttctc	
S4-m-2	SR 4	F	ggcaaaacagaacacgaag**cat**gtcagtgatatcttatacaag	p.Gln1255His
		R	cttgtataagatatcactgac**atg**cttcgtgttctgttttgcc	
S10-m-2	SR 10	F	gagtgaccatgtttacc**at**cagcacccagatcaat	p.Arg2691His
		R	attgatctgggtgctga**tg**gtaaacatggtcactc	
S10-m-3	SR 10	F	tttccagccttatggatt**gc**ataccaatggttttggc	p.Ser2705Cys
		R	gccaaaaccattggtatg**ca**atccataaggctggaaa	
S10-m-2&3	S10-m-3	F	gagtgaccatgtttacc**at**cagcacccagatcaat	p.Arg2691His p.Ser2705Cys
		R	attgatctgggtgctga**tg**gtaaacatggtcactc	
S16-m-1	SR 16	F	tgacttggaatggttgc**at**ggaattgggtggatgc	p.Arg5821His
		R	gcatccacccaattcca**tg**caaccattccaagtca	

Mutagenesis reaction was performed following the manufacturer's protocol. Once the reaction mix was prepared, the reaction program was run on the samples with a thermal cycler (Bio-Rad, Hercules, CA, USA) as listed in **Table 5**.

***Table 5.* Mutagenesis reaction mix and reaction program.** The concentration of all the primers was 100 ng/µl. As a template dsDNA were used plasmids extracted according to chapter 4.2.1., the amount of added template varied depending on the DNA concentration. Amount of ddH_2O was calculated to add up the total sample volume to 50 µl before adding the enzyme (QuikChange lightning enzyme was the last reagent to add into reaction mix).

Reaction mix		**Reaction program**			
Volume	Reagent	Segment	Cycles	Temperature	Time
5 µl	10 x reaction buffer	1	1	95 °C	2 minutes
1.25 µl	forward primer	2	18	95 °C	20 seconds
1.25 µl	reverse primer			60 °C	10 seconds
1 µl	dNTP			68 °C	2.5 minutes
1.5 µl	QuikSolution reagent	3	1	68 °C	5 minutes
2.1-5 µl	template dsDNA				
35-37.9 µl	ddH_2O				
1 µl	QuikChange Lightning Enzyme				

After the program was finished, 2 µl of *Dpn* I restriction enzyme was added into the reaction mix, mixed gently, spun down and incubated at 37 °C for 5 minutes.

7.4. Transformation into competent bacterial cells

Prior to vector transformation into expression bacterial cells, plasmids with high quality and successful mutation have to be selected and then verified. Therefore, the mutagenesis products were first transformed into *E.coli* DH5α-cells for screening.

7.4.1. DH5α transformation

The vectors were transformed into Library Efficiency® DH5α chemically competent *E.coli*-cells (Invitrogen, Carlsbad, CA, USA). DH5α cells were melted on ice for ten minutes, 50 µl of the cells were transferred to a separate tube and 2 µl of *Dpn* I treated mutagenesis product was added. The mixture was placed on ice for 30 minutes. After incubation, it was heat shocked for 45 seconds at +42°C in a water bath and subsequently placed on ice for five minutes. 200 µl of room temperature SOC medium was added to the mix and the cells were grown on a shaker at +37°C at 250 rpm for one hour. The bacteria were plated on pre-warmed LB-plate containing ampicillin (100 mg/ml) and grown overnight at +37°C.

Two transformant colonies were selected from each plate. A liquid culture (5ml LB, Amp100) and a pure culture on LB-plates with ampicillin (Amp100) were grown from each selected colony overnight at +37°C on a shaker. Plasmids were extracted and their concentrations were measured (see **Chapter 7.3.1**).

7.4.2. Verification of the sequence

Extracted plasmids (2µl) were digested with restriction enzymes specific to each SR fragment. To digest S10 variants, *Bam*HI restriction enzyme was used, and *Eco*RI was used to digest S1, S2, S4, S7 and S16 variants. The digestion was performed using FastDigestion-products (Thermo Fisher Scientific, Waltham, MA, USA), incubating reaction mix (see **Table 6**) for at least 20 minutes at +37°C.

Table 6. **Digestion mix.**

Reagent	Volume (µl)
10X FastDigest Green Buffer	2
DNA (up to 1 µg)	2
Restriction enzyme	1
dH_2O	17-X
Total	20

The digestion products were run on 1% agarose gels (Agarose Molecular Grade, Bioline) with added Midori Green Advanced DNA Stain. The picture was taken with Nippon-software (Nippon Genetics EUROPE GmbH, Düren, Germany).

The verified plasmids were sesuenced at FIMM Technology Center Sequencing Laboratory (Helsinki, Finland) using ABI3730xl DNA Analyzer-device (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was done in both directions, forward and reverse, for better coverage.

Table 7. **Plasmid-primer mix for sequencing.**

Component	Volume (µl)
pGEX-4T-1 forward primer (5 µM)	1.6
pGEX-4T-1 reverse primer (5 µM)	1.6
Plasmid (150-300 ng)	5

The sequencing results were examined using Sequencher 5.0 software. After confirming the right sequence, a glycerol stock was prepared.

7.4.3. Transformation into BL21

Vectors extracted from DH5α cells and with confirmed sequence were transformed into One Shot® BL21 (DE3) chemically competent *E.coli* cells. The protocol of vector transformation into BL21-cells was almost the same as into DH5α (see **Chapter 7.4.1**). The volume of plasmid sample used in transformation was reduced to 1 µl. Successful transformation was verified as described in **Chapter 7.4.2**.

7.4.4. Glycerol stock preparation

For each construct glycerol stock was prepared. Glycerol stocks were made by mixing 600µl of bacterial culture with 300µl of 99.5% glycerol. The stocks were stored at -80°C.

7.5. Production and purification of protein fragments

7.5.1. Protein production

Protein production was initialized with 10ml LB liquid cultures from glycerol stocks at ampicillin concentration 100 µg/ml. The bacteria were cultured overnight on a shaker at +37°C, then transferred into bigger culture volume (100 ml LB) with double the ampicillin concentration (200 µg/ml) and grown further on the shaker at the same temperature until the OD_{600} value of the cultures was between 0.5 and 0.8. When the desired OD_{600} value was reached cultures were cooled on ice to room temperature and nebulin protein fragment production was induced by IPTG (final concentration 0.45 mM) for 3h to overnight at room temperature on a shaker (250 rpm). After that bacterial cells were pelleted by centrifugation at 5750 x g for 10 minutes at +20°C (Thermo Electron Corporation, Sorvall™ RC 6 Plus, Asheville, NC, USA). The pellets were either stored at -20°C or placed on ice for protein extraction.

7.5.2. Protein extraction

Protein extraction and the steps thereafter were performed on ice to minimize protein degradation. The bacterial pellets were re-suspended in 1xPBSwith added 1x protease inhibitors (Thermo Scientific, 88665 Pierce Protease Inhibitor Mini Tablets, Rockford, IL, USA) to block any proteases extracted from the cells alongside the target protein.

Cells were lysed in two steps; first physically by sonicating and then chemically with lysozyme and Triton X-100. Sonication was performed in 10 second periods with 20-second intervals to let the samples cool down after sonication and preserve the protein from degrading. For chemical lysis lysozyme was added to the lysate in a final concentration of 0.25 mg/ml and

Triton X-100 in a final concentration of 0.5-percentage. Chemical lysis proceeded for an hour at a rotator, after which the lysates were clarified by centrifugation at 15 000 x g for 15-30 minutes. The clear, soluble proteins containing supernatants were separated from cell debris containing pellets.

7.5.3. Affinity chromatography with glutathione agarose beads

The produced nebulin fragment variants are GST-fusion proteins (see **Chapter 4**) and due to the GST-tag bind with high affinity to glutathione enabling purification with Protino® Glutathione Agarose 4B beads (affinity chromatography). The beads were washed with 1xPBS and 50 µl (75% slurry) per original culture was added per 4 ml of clarified lysate. The fusion proteins were allowed to attach to the beads overnight in a rotator at +8°C, and unbound proteins were washed away once with 1xPBS-Triton X-100 1% solution and twice with 1xPBS. Five microliters of these washed beads were run in SDS-PAGE-gel to verify successful production of the fusion protein. The remaining beads were then either stored at -20°C in 1xPBS or proceeded straight to elution step.

Protein fragments attached to glutathione beads can be competed away by a high concentration of glutathione. For the elution of protein fragments elution buffer (50 mM Tris-HCl pH 8.0 and 10 mM or 50 mM glutathione with 1x Pierce Protease Inhibitor (Sigma-Aldrich, T-1503 Trizma® base, St. Louis, MO, USA; Acros Organics, Glutathione 98%, NJ, USA; Thermo Scientific, 88665 Pierce Protease Inhibitor Mini Tablets, Rockford, IL, USA in respectively order). Elution buffer with stronger glutathione concentration (50 mM) was used to weakly produced protein fragments to maximize the protein concentration of the sample. The elution buffer was filtered with 0.45 µm filter before use. The required volume of elution buffer equals half the volume of the used 75% bead slurry. The elution buffer is added to the beads and nebulin fragments are eluted overnight in a rotator at +8°C.

7.5.4. Protein purification

The elutes were purified from the beads by centrifuging them several times at 5750 x g, for 3 minutes at a time at +4°C until the elutes were clear. The next step was ultracentrifugation. The aim of this purification step was to dispose of the finer impurities and sedimented proteins. The elutes were centrifuged twice at 150 000 x g for 60 minutes at +4°C (Beckman Coulter, Optima™ MAX Ultracentrifuge, Brea, CA, USA). The purified protein samples were run on SDS-PAGE-gels to verify successful protein purification and to determine approximate protein

concentration (see **Chapter 6.6.3**). These purified protein samples were the protein samples tested in actin co-sedimentation assay.

7.6. Actin binding tests

7.6.1. Actin co-sedimentation assay

Nebulin binding tests to actin were performed as pull-down assays using commercial Actin Binding Protein Biochem Kit™ Muscle Actin (Cytoskeleton, Inc., Denver, CO, USA). The nebulin fragments were let to bind to F-actin for 30 minutes and then ultracentrifuged at high speed. All the F-actin with bound nebulin fragments co-sedimented into pellets and all the unbound protein stayed in the supernatant (see **Figure 11**). The pellets were re-suspended in sterile water forming pellet sample. The supernatant and pellet samples were run on the same SDS-PAGE-gel for comparable quantification. Each variant was tested separately. Actin co-sedimentation assay for each variant SR was repeated 5-10 times to minimize random variation (see **Table 11** and **Table 12**). As a control, the samples without F-actin were run following the same protocol. This step was done for each actin binding test to verify the absence of pellet and therefore any proteins in it. Tested protein fragments are GST-fusion proteins. To verify that GST does not bind actin, actin binding test was earlier performed on GST-protein following the same protocol.

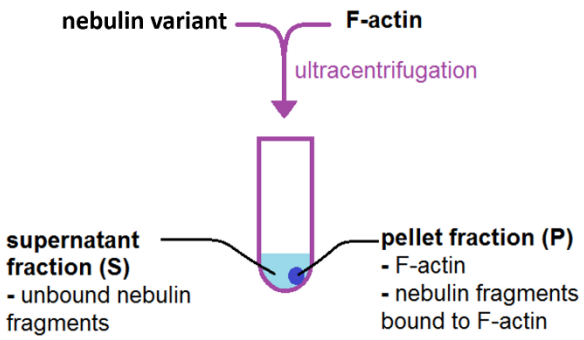


Figure 11. **Actin co-sedimentation assay.** During ultracentrifugation sample separates to the fraction bound to F-actin in pellet and unbound in supernatant.

Before the binding test, the polymerised F-actin stock was produced according to Manufacturer's protocol. 250 µg of lyophilized actin from rabbit skeletal muscle was re-suspended in 250 µl of General Actin Buffer (GAB) (5 mM Tris-HCl pH 8.0 and 0.2 mM $CaCl_2$) and the resuspension was incubated on ice for 30 minutes to ensure total resuspension. Actin was polymerized by adding 25 µl of Actin Polymerization Buffer (APB) (10x stock 100 mM Tris, pH 7.5 500 mM KCl, 20 mM $MgCl_2$ and 10 mM ATP) and polymerization proceeded for one hour at RT. After the polymerization actin was stored on +4°C. The final concentration of the F-actin stock was 21 µM. For control binding tests F-actin buffer was prepared by mixing APB and GAB in 1:10.

In the actin binding test, 10 µl of test protein was mixed with 40 µl of F-actin stock. After the incubation samples were centrifuged at 150 000 x g for 90 minutes at +24°C (Beckman Coulter, Optima™ MAX Ultracentrifuge, Brea, CA, USA). The supernatant fraction (50 µl) was transferred to a clean Eppendorf tube. The pellet was re-suspended in 30 µl of sterile water. The resulting fractions were run separately on SDS-PAGE-gels (see **Chapter 7.6.3**).

7.6.2. Actin affinity assay

To increase the reliability of the results of the previously described actin co-sedimentation assay, the affinity (i.e., the dissociation equilibrium constant) of actin binding to S16 and its variant S16-m-1 was measured. Therefore, actin co-sedimentation assay was performed for SR fragments with different concentrations. The actin-binding strength was standardized using nebulin protein fragment concentration, plotted with Hill slope fit and then the dissociation constant was determined.

F-actin stock and F-actin buffer were prepared in the same way and the same amount reagents were added to the binding reaction (see **Chapter 7.6.1**). Nebulin fragment concentrations for the S16 and S16-m-1 samples were determined from elution gels (see **Chapter 7.1**) and were diluted to five different concentrations (see **Table 8**). The dilutions, 10 µl of each, were added to the actin binding reaction, resulting in nebulin fragment concentrations listed in **Table 8** (Nebulin concentration in the actin binding reaction) with the total reaction volume being 50 µl. The control actin binding was performed at each nebulin fragment concentration.

Table 8. **Nebulin fragment concentrations for S16 and S16-m-1 at different dilutions.** The concentrations of the original samples after elution for both S16 and S16-m-1 were determined to be 3 ng/µl (see **Chapters** 4.4. and 5.1.). Nebulin stocks for each fragment were diluted from these samples.

	Nebulin stock concentration	Nebulin concentration in the actin binding reaction
C1	0.1125 ug/ul	0.4144410 uM
C2	0.225 ug/ul	0.8288819 uM
C3	0.45 ug/ul	1.6578000 uM
C4	0.9 ug/ul	3.3155000 uM
C5	1.8 ug/ul	6.6311000 uM

After 30 minutes incubation (binding reaction), 10 µl of each sample (total sample) was transferred to a clean Eppendorf tube. The remaining 40 µl sample was ultracentrifuged according to actin co-sedimentation assay protocol (see **Chapter 6.6.1**). After ultracentrifugation, 40 µl of supernatant fraction was separated from the pellet. The pellet was

re-suspended in 40 µl of sterile water. Workflow presented in **Figure 12**. Supernatant and pellet fractions, as well as total samples, were run on the SDS-PAGE gels (see **Chapter 7.6.**).

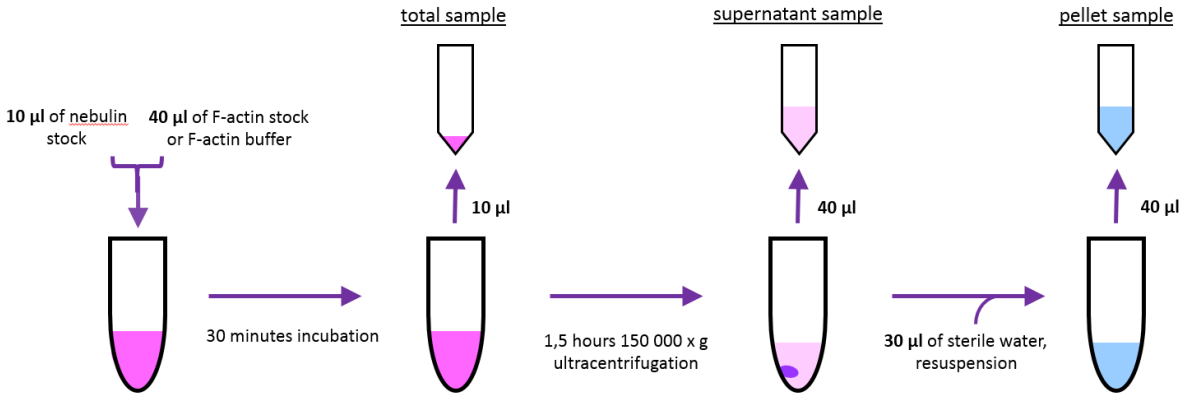


Figure 12. **Workflow of actin affinity assay.** The workflow was followed for each dilution of both S16 and S16-m-1 nebulin stock. It is similar to the idea of actin co-sedimentation assay with one additional step, total sample acquisition.

7.6.3. SDS-PAGE

To separate proteins based only on their size, sample buffer containing sodium dodecyl sulphate (SDS) and 2-Mercaptoethanol (βMe) was added to the samples to denature and give charge the proteins according to the length of their amino acid chain.

Sample preparation was different in each step. If protein samples were run to verify successful protein production or elution, taken sample was diluted with elution buffer to 10 µl volume and then diluted to double the volume (20 µl) with 2x Laemmli with βMe (Bio-Rad, 2x Laemmli Sample Buffer, USA; Bio-Rad 2-Mercaptoethanol, Hercules, CA, USA). The pellet samples from actin co-sedimentation assay were diluted in 1:1 ratio with 2x Laemmli with βMe, and the supernatant samples were diluted 5:1 with 5x Lane Marker Reducing Sample Buffer (Thermo Fisher Scientific, Waltham, MA, USA). Total-samples from actin affinity assay were diluted with 30 µl of MQ and then with 10 µl of 5x Lane Marker Reducing Sample Buffer. All Total-samples were run on the same gel. Both the pellet, as well as supernatant samples from actin affinity assay were diluted with 10 µl 5x Lane Marker Reducing Sample Buffer and run on the gel next to each other. In this setup the volumes of the samples acquired from actin affinity assay are equal, 50µl, and the quantifications are comparable to each other. All the samples were denatured before the run for five minutes at +95 °C.

20 µl of the samples were run on 12% SDS-PAGE gels (Bio-Rad, Mini-PROTEAN® TGX™ Precast Gels, Hercules, CA, USA). The volume of the protein marker run on the gel was 8 µl

(Bio-Rad, Precision Plus Protein™ Dual Color Standards, Hercules, CA, USA). All the gels with agarose beads samples and elution samples, as well as S2-m-1, S16, and S16-m-1 actin binding samples were run at 200V for 40 minutes. The rest of the actin binding samples were run first at 200V for 40 minutes and then for 30 to 45 minutes more at 150V to increase the resolution of the bands, and gels were washed with MQ (thrice for five minutes) and then stained in 50 ml of Coomassie stain (Bio-Rad, Bio-Safe™ Coomassie G-250 Stain, Hercules, CA, USA). The gels were destained with MQ and scanned with Odyssey device (LI-COR, Lincoln, NE, USA).

7.7. Data analysis

Raw data were obtained from SDS-PAGE-gel pictures using ImageJ software (National Institutes of Health, Bethesda, MD, USA). Few of the default settings were changed. Unit of length used in quantifications was manually changed into pixels, picture type to 8-bit and background was subtracted to be light. The area, integrated density and mean grey value were selected in measurement setting options. All the quantification data was stored in Excel (Microsoft, Redmond, WA, USA). Each band from the gels was quantified thrice and the mean of these values was considered the actual quantification for the band.

Protein concentrations were quantified from elution gel by drawing a standard curve from BSA samples with known protein concentration. The BSA standard samples and the elutes were quantified together each round. The acquired concentrations in micrograms per microliters were afterward converted into micromoles using GraphPad QuickCalc software that is available online (GraphPad Software, La Jolla, CA, USA; see **Online Resources**).

Actin-binding strength was determined by measuring the amount of nebulin fragment in supernatant and pellet fractions from the same sample at the same time. In total, the nebulin fraction in supernatant plus the fraction in pellet form 100% and percentage of the bound nebulin to F-actin can be determined.

To quantify the affinity, amount of nebulin fragments in total- and pellet samples, as well as the amount of actin were quantified in each group all five concentrations in one round. From data acquired from Total-samples standard curve was determined. Using this information, quantified amount of nebulin bound to F-actin was converted to concentration of nebulin fragments in pellet fraction and further its concentration per micromole of actin in the sample. This ratio was plotted against the unbound nebulin fragments concentration, which was calculated using the concentrations of bound and total nebulin fragments. The Hill slope was

fitted to the plot with GraphPad Prism-software (Graphpad Software, Inc., La Jolla, CA, USA). The dissociation constant, as well as other parameters, were calculated with this software.

Statistical analysis for actin co-sedimentation assay results was also performed using GraphPad Prism-software. Statistical significance of the change in actin-binding strength between variants and wild-type SRs was calculated with non-parametric one-way ANOVA test. Mann-Whitney-test was used to calculate the same between one variant and corresponding wild-type SR. (Hart 2000, Marusteri, Bacarea 2010)

8. Results

8.1. 3D structure prediction

The 3D structure was predicted for each variant and corresponding wild-type nebulin fragment selected in **Chapter 7.2.1 and Chapter 7.2.2**, following the protocol presented in **Chapter 7.2.3**. The compilation of the 3D structure images is presented in **Figure 13** predicted by I-TASSER suite and in **Figure 14** predicted by I-TASSER server.

The quality of each model is described with C-score in **Tables 9** and **10**. The C-score ranges from -4.04 to 1.49. L-scores for the models are assembled in **Attachment 4**. The L- scores range between 4Å and 20Å in general, being lower in the GST-region of the fusion proteins and higher in the regions of nebulin fragment.

Table 9. **C-scores for the 3D models of nebulin structures predicted by I-TASSER suite.** The wt stands for wild-type structure.

Construct	C-score	Construct	C-score	Construct	C-score	Construct	C-score
GST	1.49	S2 wt	-1.16	S7 wt	-1.23	S10-m-2	-0.76
S1 wt	-0.55	S2-m-1	-0.65	S7-m-1	-0.73	S10-m-3	-0.78
S1-m-1	-2.56	S4 wt	-0.79	S7-m-2	-2.91	S10-m-2&3	-2.88
S1-m-2	-2.88	S4-m-1	-0.81	S7-del-1	-2.4	S16 wt	-0.99
S1-del-1	-0.98	S4-m-2	-1.39	S10 wt	-2.5	S16-m-1	-1.17

Table 10. **C-scores for the 3D models of nebulin structures predicted by I-TASSER server.** The wt stands for wild-type structure.

Construct	C-score	Construct	C-score	Construct	C-score	Construct	C-score
GST	1.13	S2 wt	-3.75	S7 wt	-0.5	S10-m-2	-3.79
S1 wt	-1.00	S2-m-1	-3.86	S7-m-1	-1.42	S10-m-3	-3.93
S1-m-1	-0.96	S4 wt	-1.01	S7-m-2	-0.52	S10-m-2&3	-4.04
S1-m-2	-0.88	S4-m-1	-0.68	S7-del-1	-0.8	S16 wt	-0.96
S1-del-1	-1.01	S4-m-2	-1.05	S10 wt	-4.01	S16-m-1	-1.05

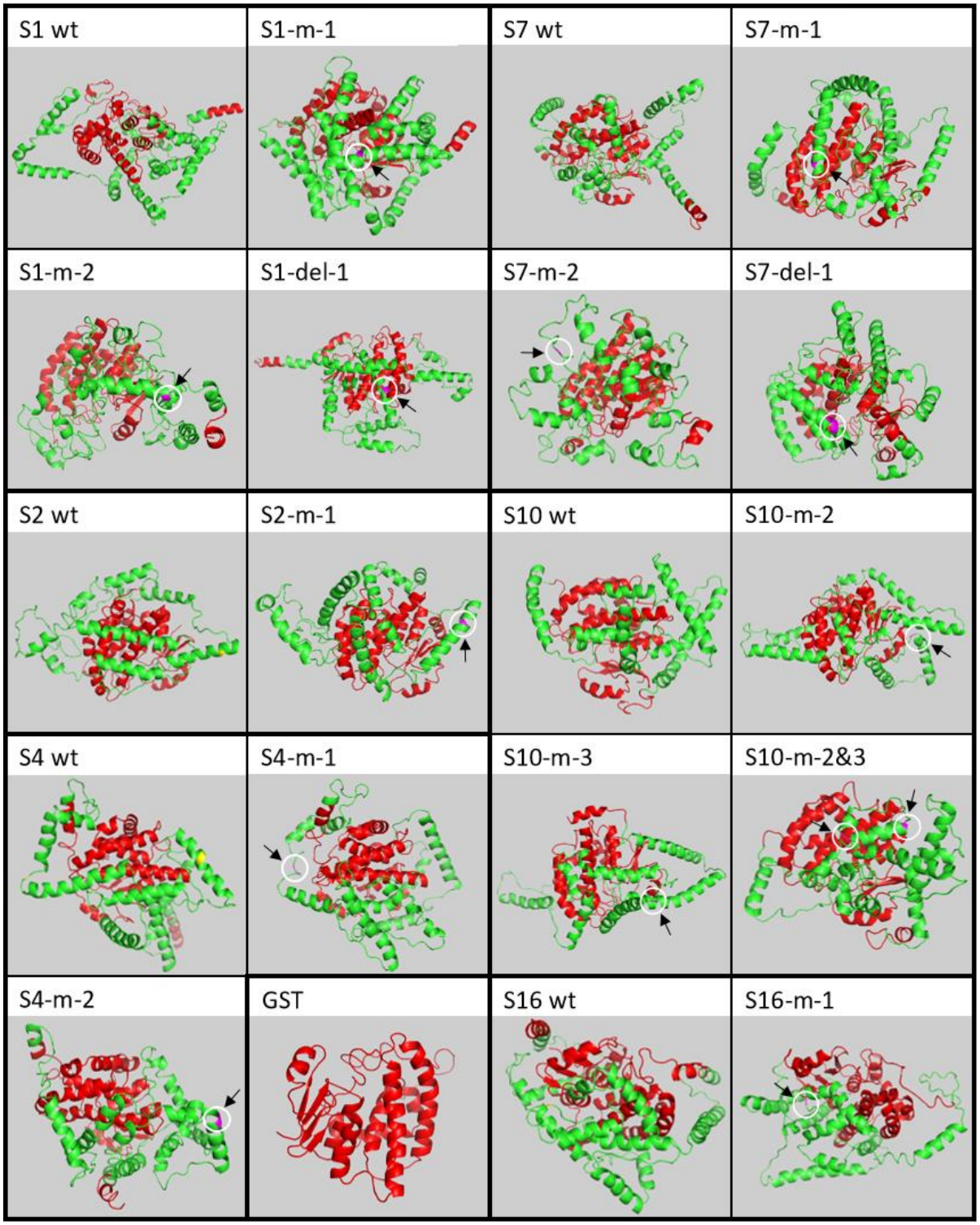


Figure 13. **Best models predicted by I-TASSER suite.** The best prediction selected based on C-scores and L-scores. The GST-part of the fusion protein is colored red and the nebulin fragment part is colored green. The mutation points are marked with magenta color, encircled and pointed out with arrows. All the structures are globular, with GST more tightly folded in the center of the structures and alpha-helical nebulin fragments are folded around it.

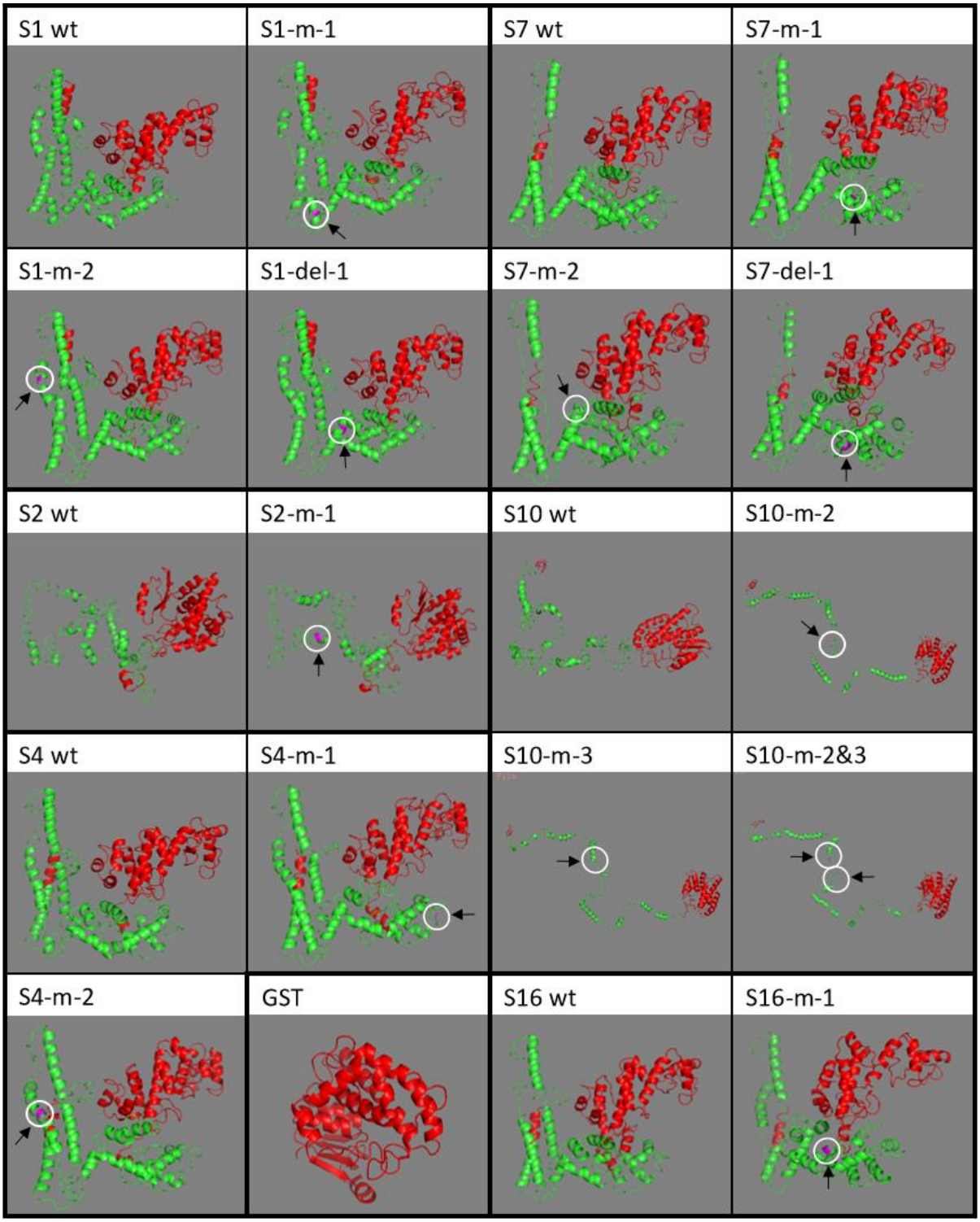


Figure 14. **Best models predicted by I-TASSER suite.** The best prediction selected based on C-scores and L-scores. The GST-part of the fusion protein is colored red and the nebulin fragment part is colored green. The mutation points are marked with magenta color, encircled and pointed out with arrows. All the structures are globular, with GST more tightly folded in the center of the structures and alpha-helical nebulin fragments are folded around it. In case of S10 wild-type and variants, the structure is significantly more linear according to the I-TASSER server prediction. Other nebulin fragment folds seem to be folded more separately from GST.

All the nebulin fragment structures in **Figures 13 and 14** appear to be globular according to the I-TASSER suite and I-TASSER server prediction. Generally, a folded GST is predicted to be the core of the structures, around which nebulin fragments are arranged in short periods of α-helices aiming for the lowest energy fold. The models presented in the **Figures 13** and **14** were selected to be the best out of the suggested I-TASSER runs output. In these models, cluster size is minimal compared to other models.

8.2. Production and purification

Variants were successfully prepared and the sequences were verified as described in Materials and methods. The following results for production and actin co-sedimentation assays of wild-type SR constructs were performed before this thesis and used in data analysis. All required fusion proteins were successfully produced and purified (see **Chapter 7.5.**). The gel pictures of eluted and purified proteins are compiled in **Figure 15**.

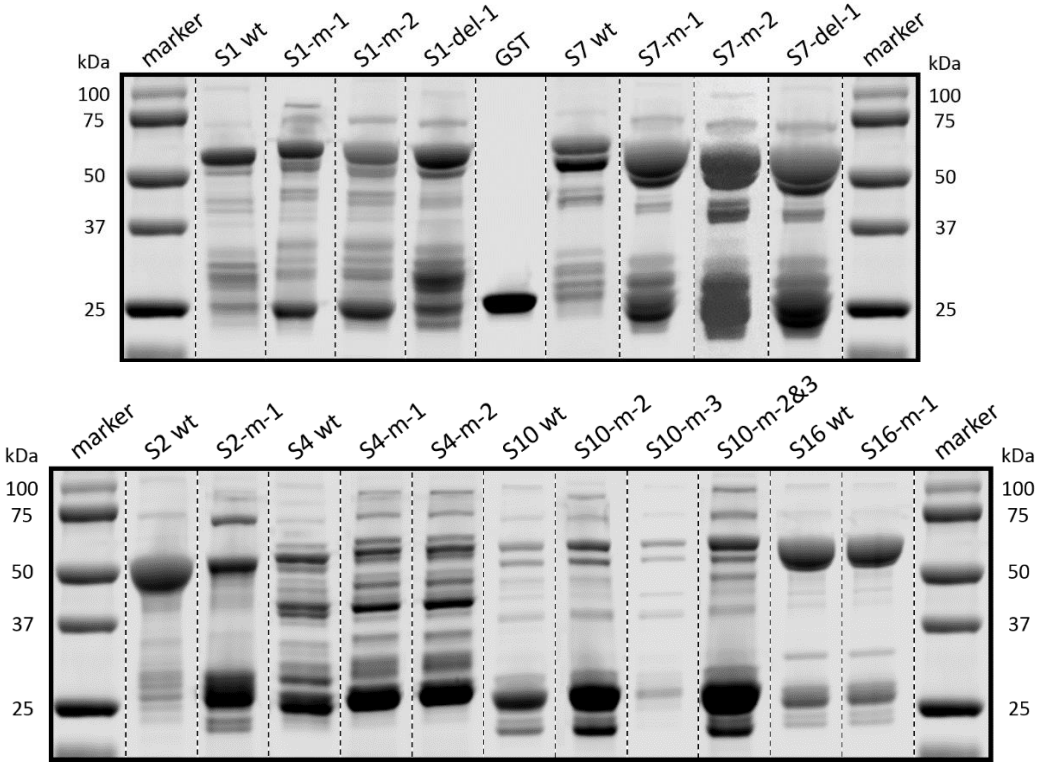


Figure 15. **Compilation of SDS-PAGE-gel pictures of the produced nebulin fragments.** The proteins in this figure were eluted from glutathione agarose beads and purified prior to the SDS-PAGE-gel run. Wild-type SR fragments marked as wt. The samples of wild-type nebulin fragments S4, S7 and S10 for SDS-PAGE gel run were diluted with 1xPBS in 1:2 ratio.

Table 10. **Concentrations of produced SR fragments.** The concentrations were determined for eluted protein samples as described in **Chapter 6.7.** Wild-type nebulin fragments marked as wt. All concentrations are calculated to be the concentrations of protein stocks that were used for actin binding tests.

Protein fragment	Concentration (µM)	Protein fragment	Concentration (µM)
S1 wt	10.24	S7 wt	15.51
S1-m-1	13.86	S7-m-1	35.45
S1-m-2	17.62	S7-m-2	26.25
S1-del-1	22.84	S7-del-1	45.22
S2 wt	18.98	S10 wt	8.65
S2-m-1	8.99	S10-m-2	4.82
S4 wt	4.46	S10-m-3	1.50
S4-m-1	2.71	S10-m-2&3	7.33
S4-m-2	2.81	S16 wt	55.58
		S16-m-1	54.57

As **Figure 15** shows, protein fragments of the same SRs seem to run in a similar pattern, both wild-type and variants. The amount of the produced protein, on the other hand, varies both within and between SRs. The protein concentrations were quantified from the SDS-PAGE-gels of eluted and purified protein fragments (see **Chapter 7.7.**) to numerically verify the variation in protein concentrations, see **Table 10**.

8.3. Actin binding tests

8.3.1. Controls for actin binding tests

Controls were run in the same manner as test samples in actin co-sedimentation assay (see **Chapter 7.6.1**). The results are presented in **Figure 16**. Control 1 shows that GST alone does not sediment into a pellet. Control 3 demonstrates opposite behavior for F-actin alone by sedimenting almost totally to pellet. Note the F-actin degradation pattern resulting in the faint band sized slightly above 25 kDa control ladder band. It is not to be confused with GST in Control 2. Control 2 verifies that presence of both proteins in the same sample does not affect their individual behavior during ultracentrifugation. Therefore GST and F-actin do not interact with

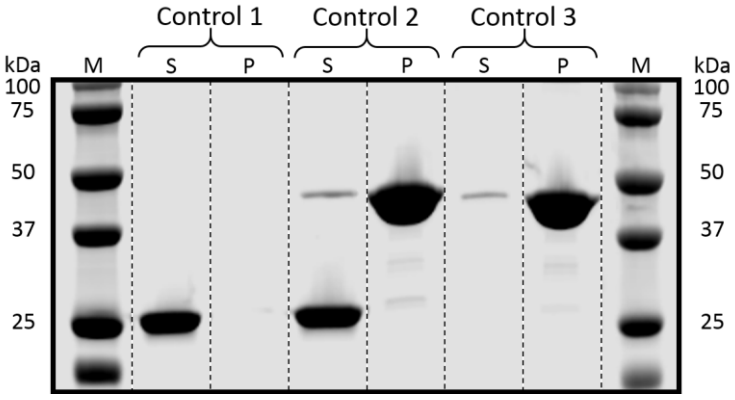


Figure 16. **Control for actin binding tests.** S refers to supernantant fraction and P to pellet fraction of the ultracentrifuged sample. M stands for marker. In Control 1 only GST (sized ~25kDa) containing sample was run, in Control 3 only actin (sized ~42kDa) containing sample was run, and in Control 2 sample containing both GST and actin was run.

each other. Also, all of the produced SR fragments stay in supernatant fraction during ultracentrifugation (see **Figure 17**).

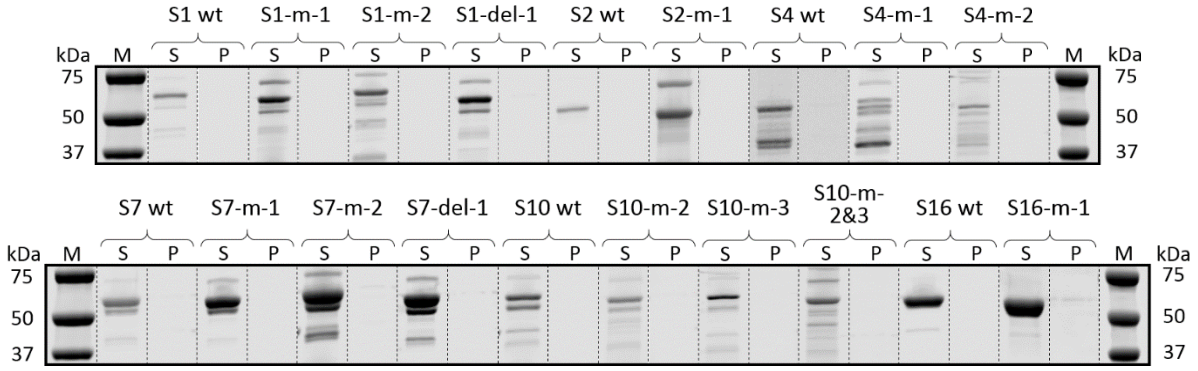


Figure 17. **Controls for SR fragments.** S refers to supernatant fraction and P to pellet fraction of the ultracentrifuged sample. M stands for the marker, wt for wild-type. The GST-SR fusion protein band is located at ~60kDa. The rest of the bands seen in this compilation represent degraded parts of the required fusion protein.

8.3.2. Actin co-sedimentation assay: corresponding variants in SR1 and SR7

The primary results, SDS-PAGE-gel pictures, for corresponding variants in SR1 and SR7 are compiled in **Figure 18.** Actin co-sedimentation assay was repeated 5-10 times for each variant following the workflow in Materials and methods (see **Chapter 7.6.1.**). The pictures of the previous results for wild-type actin co-sedimentation assay also presented in **Figure 18**.

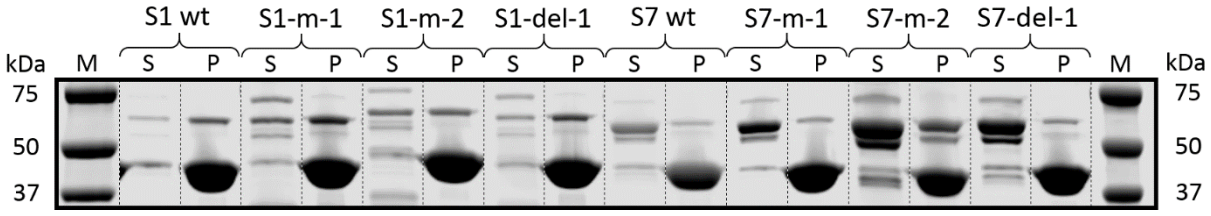


Figure 18. **Compilation of actin co-sedimentation assay SDS-PAGE-gel pictures for SR1 and SR7 variant and wild-type protein fragments.** S refers to supernatant fraction and P to pellet fraction of the ultracentrifuged sample. M stands for the marker, wt for wild-type. Actin band is sized ~42kDa.

It is clear from **Figure 18** that wild-type SR1 and its variants have strong actin-binding capacity. Wild-type SR7 and its variants bind weakly to actin. Moreover, there seems to be variation between variants and wild-type nebulin fragments of the same SR.

The results for variants were quantified from the SDS-PAGE-gel pictures and statistically analyzed as described in **Chapter 7.7**. The result data for SR 1 and 7 variants are summarized in **Table 11** and visualized as histograms in **Figure 19**.

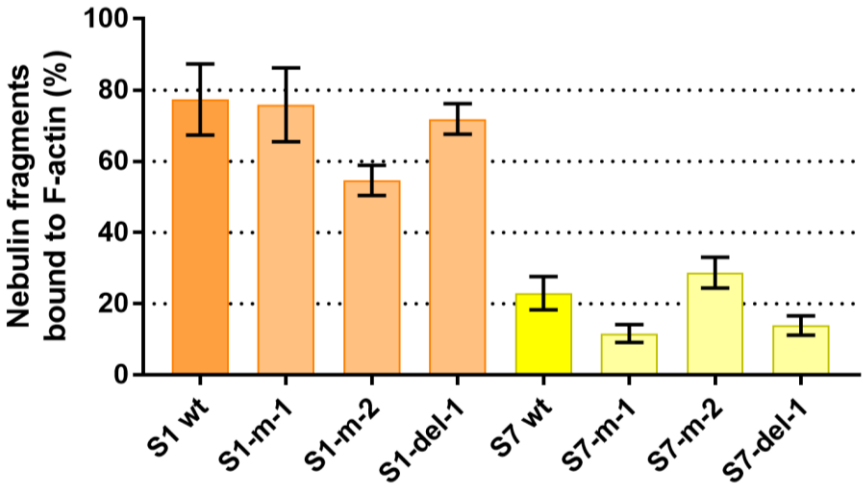


Figure 19. **Actin-binding strength of nebulin SR 1 and 7 (S1 and S7) variants each compared to the corresponding wild-type nebulin fragment.**

Table 11. **Statistical data of actin co-sedimentation assays for corresponding variants in SRs 1 and 7.** Each variant was tested against wild-type nebulin fragment of the same SR using Mann-Whitney's test. Significance was set to be P<0.05. Also, the numbers of co-sedimentation assay repeats are listed.

Nebulin fragments	Descriptive statistics			Mann-Whitney test		Repeats
	Mean	Standard deviation	Standard error of mean	Significant difference	P-value	
S1 wt	77.35	10.01	3.165	-	-	10
S1-m-1	75.88	10.39	4.244	No	0.8949	6
S1-m-2	54.65	4.252	1.901	Yes	0.0013	5
S1-del-1	71.86	4.288	1.356	No	0.0753	10
S7 wt	22.98	4.667	1.905	-	-	6
S7-m-1	11.67	2.512	0.8374	Yes	0.0008	9
S7-m-2	28.74	4.286	1.75	No	0.0931	9
S7-del-1	13.94	2.71	0.9032	Yes	0.0016	9

According to the results presented, S1-m-1, S1-del-1, and S7-m-2 did not affect the actin-binding strength of nebulin fragments. The other variants, S1-m-2, S7-m-1, and S7-del-1, significantly decreased the actin-binding strength of the SRs.

8.3.3. Actin co-sedimentation assay: missense variants in SRs 2, 4, 10 and 16

The primary results for wild-type nebulin fragments and variants in SR2, SR4, SR10 and SR16 are compiled in **Figure 20**. Actin co-sedimentation assay was repeated 5-9 times for each variant. SDS-PAGE-gels were analyzed (see **Chapter 7.7**) and the resulting data was presented in the histogram in **Figure 21** and **Table 12**.

From these results, it can be observed that S2-m-1 and S4-m-2 have weaker actin-binding capacity than the corresponding wild-type nebulin fragments. The S10-m-2 and S10-m-2&3 on the other hand bind more strongly to actin than wild-type S10.

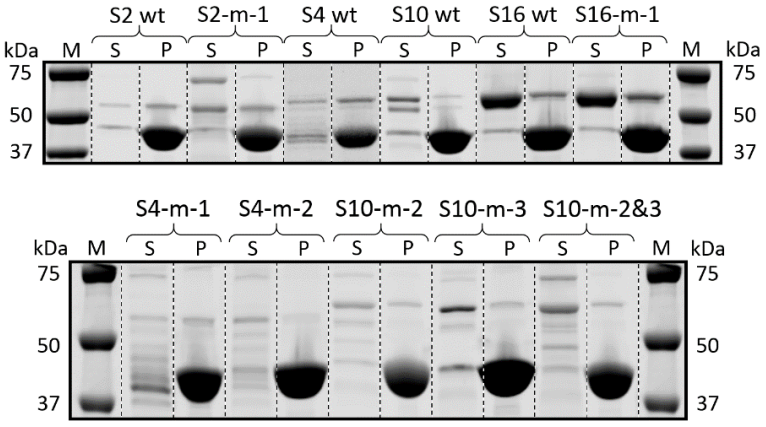


Figure 20. **Compilation of actin co-sedimentation assay SDS-PAGE-gel pictures for SRs 2, 4, 10 and 16 variant and wild-type protein fragments.** The gels for SR4 and SR10 variants were run longer than the rest of the variants and were compiled separately. S refers to supernantant fraction and P to pellet fraction of the ultracentrifuged sample. M stands for the marker, wt for wild-type. Actin band is sized ~42kDa.

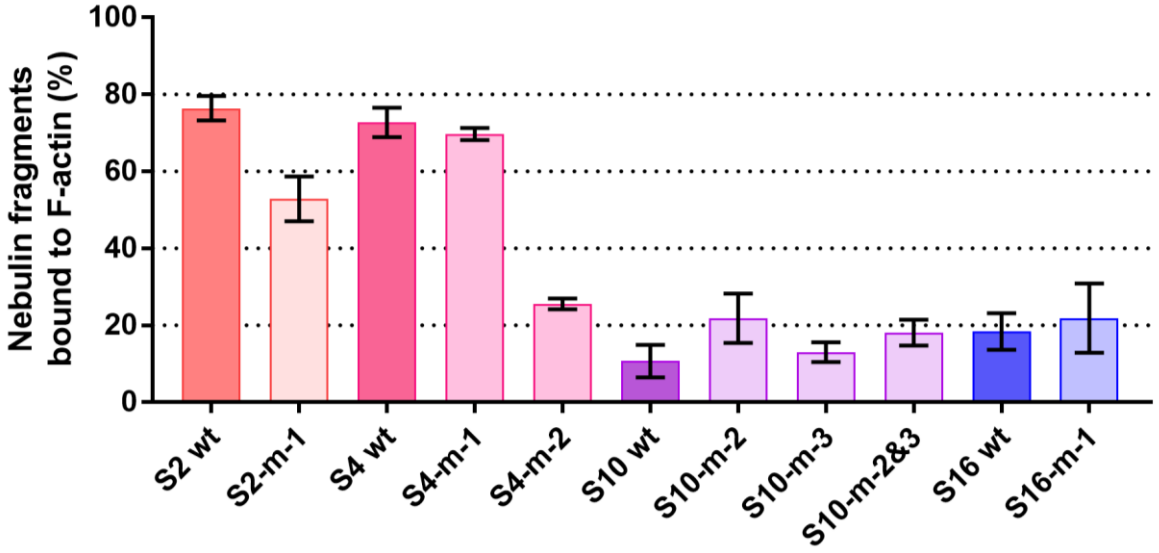


Figure 21. **Actin-binding strength of nebulin SR 2, 4, 10 and 16 (S2, S4, S10, and S16) variants each compared to the corresponding wild-type nebulin fragment.**

Table 12. **Statistical data of actin co-sedimentation assays for variants in SRs 2, 4, 10 and 16.** Each variant was tested against wild-type nebulin fragment of the same SR using Mann-Whitney's test. Significance was set to be P<0.05.

Nebulin fragment	Descriptive statistics			Mann-Whitney test		Repeats
	Mean	Standard deviation	Standard mean error	Significant difference	P-value	
S2 wt	76.44	3.169	1.417	-	-	5
S2-m-1	52.91	5.803	2.369	Yes	0.0043	6
S4 wt	72.77	3.806	1.702	-	-	5
S4-m-1	69.73	1.547	0.6316	No	0.1775	6
S4-m-2	25.59	1.437	0.5866	Yes	0.0043	6
S10 wt	10.72	4.249	1.416	-	-	9
S10-m-2	21.85	6.373	2.253	Yes	0.0016	8
S10-m-3	13.02	2.591	1.058	No	0.3884	6
S10-m-2&3	18.13	3.383	1.381	Yes	0.0028	6
S16 wt	18.42	4.749	1.679	-	-	8
S16-m-1	21.93	8.998	3.401	No	0.3969	7

8.3.4. Actin affinity assay

The actin affinity assay was done for two nebulin fragments, S16 and its variant, S16-m-1. Proteins were produced and purified according to Materials and methods (see **Chapter 7.6.2.**). The results were quantified from SDS-PAGE-gel pictures, see compilation below in Figure.

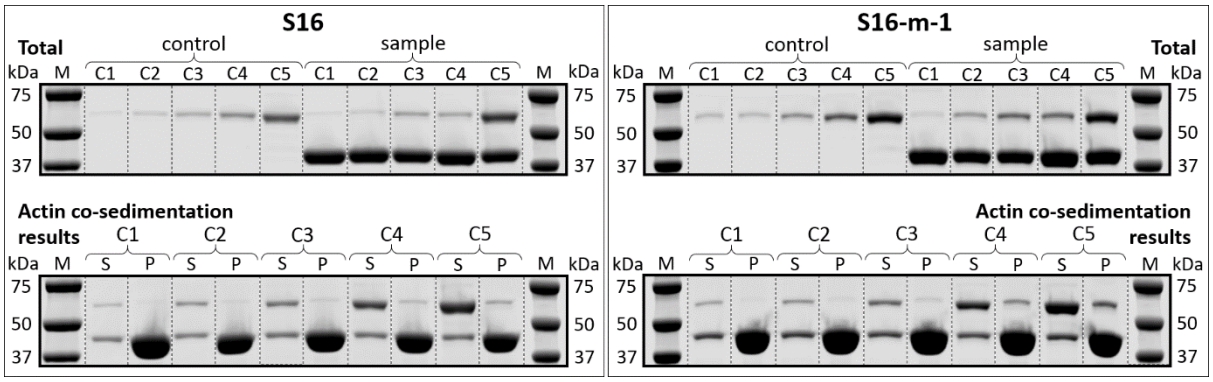


Figure 22. **Compilation of SDS-PAGE-gel pictures for S16 and S16-m-1 actin affinity determination.** Actin band is sized ~42kDa.

Protein fragments seem to co-sediment with F-actin in its usual manner, binding weakly to it. One surprising result can be noticed already at this point: in both S16 and S16-m-1 Total-samples containing F-actin amount of nebulin fragments at second (C2) and third (C3) strongest concentrations seems to be the same. The amount of SR-fragment in C3 was diluted from C2 sample to be half as much as in C2 sample (see **Table 8**). This result was also observed in the standard curves below (**Figure 23**).

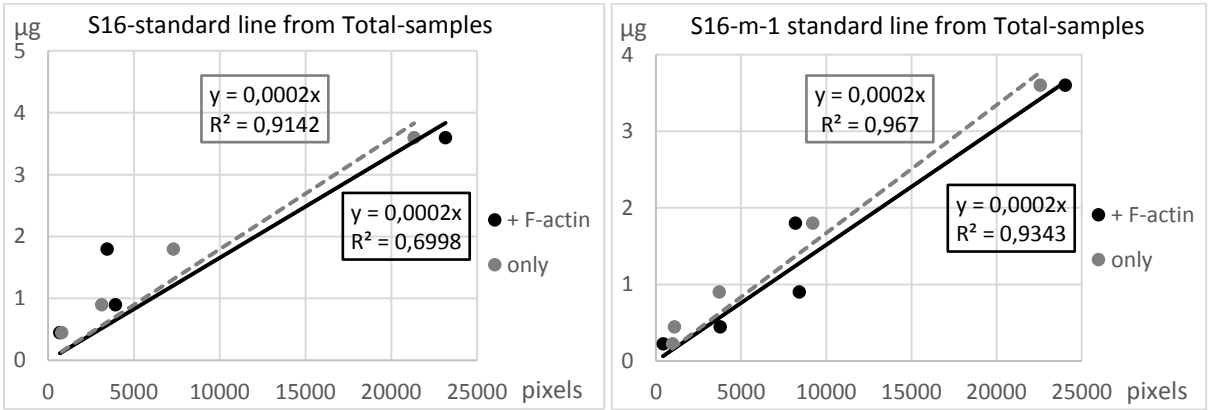


Figure 23. **Standard lines for S16 and S16-m-1 concentration per pixel.** Drawn based on the intensity of gel bands from Total-samples. Orange lines were drawn using control samples of total, containing nebulin only. Blue lines were drawn using the gel pictures of total samples with F-actin.

Using standard curves, S16, and S16-m-1 results were quantified and described in **Chapter 7.7**. The data was fit to Hill equation and plotted bound nebulin fragment (µM/µM F-actin) against free nebulin fragment protein (see **Figure 24**).

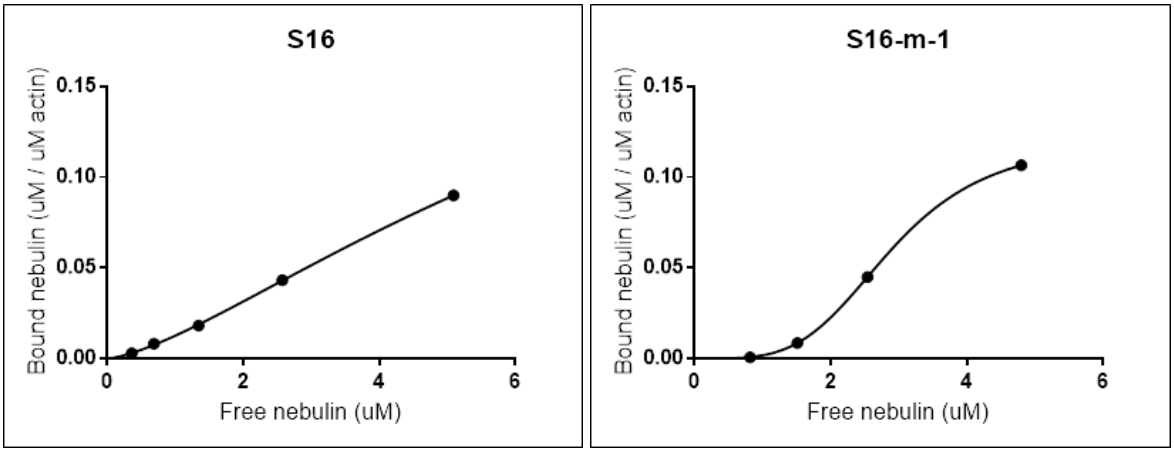


Figure 24. **Bound S16 and S16-m-1 to F-actin plotted against the free nebulin in the sample and fitted to Hill equation.**

The shape of the S16-m-1 curve is more sigmoidal in comparison to S16, barely reaching the plateau-phase of the slope. The S16 curve is more linear than sigmoidal according to **Figure 24**, with no indication of approaching plateau-phase at the end of the curve. Statistics for the fit are presented in **Table 13** and visualized in **Figure 25**.

Table 13. **Statistics for the nonlinear fit of S16 and S16-m-1 describing the specific binding of the nebulin fragments to F-actin with Hill slope.** K_d is dissociation constant, B_{max} is the maximum specific binding and h is Hill slope or Hill coefficient. Std. error refers to standard errors in K_d, B_{max}, and h calculated for best-fit. 95% CI are 95% confidence intervals.

	S16			**S16-m-1**		
	B_{max} (µM/ µM actin)	**h**	**K_d** (µM)	**B_{max}** (µM/ µM actin)	**h**	**K_d** (µM)
Best-fit	0.2643	1.434	8.083	0.1212	3.945	2.905
Std. error	0.04794	0.06746	1.706	0.0006226	0.04543	0.01253
95% CI	0.1599 → +∞	1.154 → 1.744	4.453 → +∞	0.1146 → 0.131	3.415 → 4.584	2.774 → 3.105

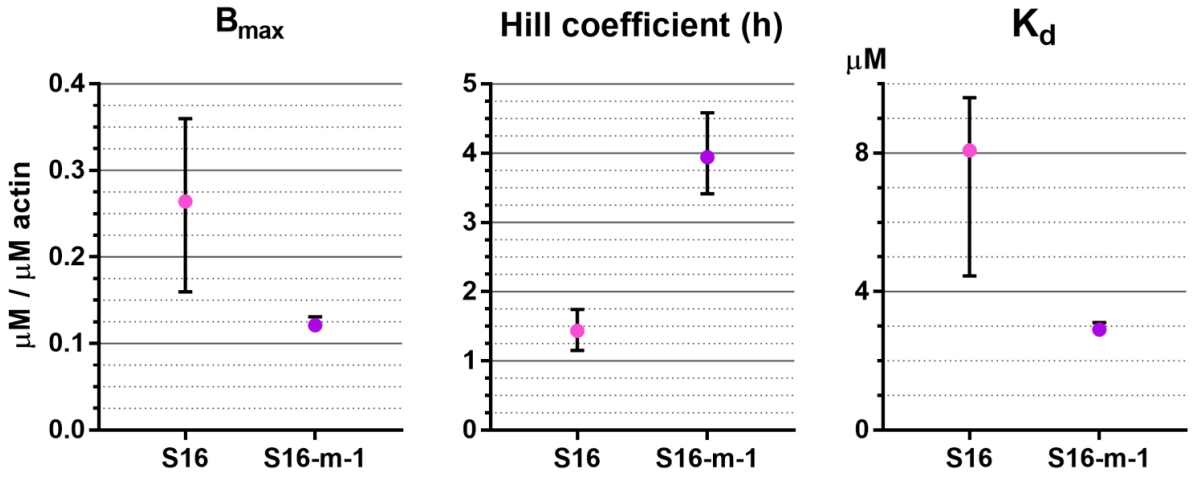


Figure 25. **K_d, B_{max} and Hill coefficient (h) visualized with their standard deviations.**

The resulted for S16 are significantly different from S16-m-1, which can be observed both from **Table 13** and **Figure 25**. The error bars in **Figure 25** are large for S16 K_d- and B_{max}-values, and the 95% confidence intervals (CI) tabulated in **Table 13** are even larger. Nevertheless, the 95% CIs for S16 results do not cross with 95% CIs for S16-m-1 results, nor do the error bars.

9. Discussion

9.1. Production of the variants

The produced nebulin fragments show significant differences in the produced quantity on SDS-PAGE-gels although the production procedure remained the same for all SR fragments. This result was verified by quantification and presented in **Table 10**. The production was repeated from two to several times for each SR fragment, and the produced amount of the protein did not change significantly between the repeats. The variations in protein concentrations between SRs may be caused by sequence differences, which can lead to the formation of inclusion bodies and their sedimentation with the cell debris during protein purification (see **Chapter 7.5.2**).

During the process of protein fragment purification, small quantities of the product are inevitably lost. In cases of S4 variants and wild-type structure, which are weakly produced, any loss of product can have an impact on actin-binding strength quantification. The protein concentrations were determined with ImageJ software and it has a drawback of analysing the gel pictures two-dimensionally whereas the SDS-PAGE-gels are three-dimensional. This can be the cause for slight errors in protein concentration quantifications. Also, the gels were commercial, but variation between them is possible. These error sources apply also for actin binding tests.

Varying protein concentrations should not affect the results of the actin binding tests. In actin co-sedimentation assay results, the sedimented part of nebulin fragments was expressed as a percentage of all the amount of nebulin fragment in the assay. However, it was empirically deduced that the nebulin fragment concentration affects the actin-binding strength in some cases. For example, in case of S16 wild-type protein and its variant, the nebulin fragments bind stronger to actin in smaller concentrations. This observation can be explained by the high productivity of S16 wild-type protein and its variant, which leads to excess of fusion protein compared to F-actin in the actin binding reaction. In this case the amount of fusion protein bound to F-actin is constant and the amount of unbound fusion protein varies affecting the ratio

between bound and unbound proteins. In the actin affinity assays the concentrations of the wild type structure and its variant were set to be identical.

The fragments were produced in *E.coli* cells. Therefore they did not undergo any posttranslational modifications due to the lack of enzymes in the cells, which may have led to slight changes in structure and function of the produced fragments compared to their native structure. The structure may have also be altered by the GST-part of the produced protein fragments (see **Chapter 9.4**).

Also, without posttranslational modifications the proteins are prone to degradation, which can be seen in SDS-PAGE-gel pictures as bands additional to ~60kDa fusion protein bands (**see Figure 15**). The extensive protein degradation pattern in wild-type S4 and S10, and their variants complicated the quantification from the SDS-PAGE-gels. The fusion protein band and the first smaller size product band are too far from each other to be analysed as a single band, but too close to have no to effect on the other bands quantification. This source of possible error was minimized by gel run optimization: after standard run at 200 V for 40 minutes, the samples were run for 30-45 minutes more at 150 V.

9.2. Analysis of actin co-sedimentation assay results

The mechanism of the pathogenesis in NM caused by missense mutations is largely unknown. One of the hypothesis is that missense mutations cause changes in nebulin-actin binding, which leads to muscle weakness and aggregation of nemaline bodies.

Following the first aim (**see Chapter 6**), the actin-binding strength of variants in two SRs, S1 and S7, with corresponding missense mutations and six nucleotide deletions was determined. The missense mutations, p.Tyr479Cys (S1-m-1) and p.Arg1961Cys (S7-m-2), as well as the deletion, p.Tyr458_Lys459del (S1-del-1), did not cause significant changes in actin-binding strength of the nebulin fragments. The missense mutations were not located at actin-binding sites, but the deletion in S1 was. The missense mutations, p.Lys494Cys (S1-m-2) and p.Glu1946Cys (S7-m-1), and deletion, p.Tyr1924_Lys1925del (S7-del-1), caused decrease in actin-binding strength of the nebulin fragments. All of these three mutations were at the actin-binding sites. These results suggest that the location of the mutations has an effect: the mutations outside of the actin-binding sites does not seem to affect nebulin-actin interactions. However, location of the mutation at actin-binding site does not ensure changes in actin-binding strength of the nebulin fragment. Another probable factor in the effect the mutation can have on actin-

binding strength is the varying actin-binding capacity between SRs. S1 binds strongly to actin and S7 is much weaker, which makes alteration at each actin-binding site more significant.

Following the second aim, actin-binding strength of seven variants was determined. Six variants contained a single missense mutation, S2-m-1 (p.Tyr773Cys), S4-m-1 (p.Thr1139Met), S4-m-2 (p.Gln1255His), S10-m-2 (p.Arg2691His), S10-m-3 (p.Ser2705Cys) and S16-m-1 (p.Arg5821His), and the seventh variant, S10-m-2&3 contained two mutations, p.Arg2691His and p.Ser2705Cys. S16-m-1 was located at a tropomyosin binding site in the proximity of actin-binding site, S4-m-1 and S10-m-3 not at any known binding sites, and the rest were located at actin-binding sites. According to the results, all of the mutations outside actin-binding sites had no effect on actin-binding strength, which is consistent with the previous results for S1 and S7 variants. S2-m-1 and S4-m-2 actin-binding strength decreased, and S10-m-2 and S10-m-2&3 actin-binding strength increased. S10-m-2&3 actin-binding strength increased less than variant S10-m-2. This indicates that the mutations with close locations can modify the function of one another. Further exploration is required in order to understand the reasons behind increase and decrease of actin-binding strength caused by mutations as well as the possible cooperativity between the adjacent mutations.

According to the ClinVar database, S2-m-1 mutation is most likely benign. There are five entries stating that p.Tyr773Cys missense mutation is most likely benign due to conservative change, position at poorly conserved location and inconsistency with pathogenic occurrence. There is high occurrence frequency observed, especially in the Asian population (0.0249) indicating that this is a benign SNV. Also, three entries in the ClinVar database support that p.Arg2691His in S10-m-2 and S20-m-2&3 is a benign SNV for the same reasons as for the S2-m-1 mutation with only difference is that this mutation has high occurrence frequency especially among African population (0.006) instead of Asian population. However, both variants were found in NM patients and the presence of these mutations altered the actin-binding strength of the super repeats. This leaves an open question whether these mutations can have a significant altering effect on NM pathogenesis.

The variant S16-m-1 is located in tropomyosin binding site that is close to the actin-binding site (**see Figure 10**). Therefore, there is a possibility that this variant can affect tropomyosin-binding strength of the nebulin as well as actin-binding strength. According to co-sedimentation assay results, there was no significant difference in actin-binding strength between the wild-type and the variant. However, the error bars are large and for the definitive

answer, the actin-binding affinities of the nebulin fragments need to be determined. Also, the variation in actin-binding strength depending on the nebulin fragment concentration leads to the necessity of dissociation constant determination.

There were three general sources of error in the actin co-sedimentation assays: presence of GST as a part of protein fragment, complexity of quantification with ImageJ and protein degradation of the produced nebulin fragments. GST-part of the fusion proteins does not bind to F-actin and was not removed. However it may affect the folding, which could affect the results. ImageJ-quantification is a source of error in data analysis also due to the smear from the actin band which this sensitive software registers and complicates quantification of the nebulin bands. In S4-variants the protein degradation is more intensive compared to other produced super repeats and the bands are located very close to each other, which complicated quantification with ImageJ further and may affect the final results.

9.3. Analysis of actin affinity assay results

Using actin co-sedimentation assay it was determined that p.Arg5821His mutation in S16 (S16-m-1) causes slightly higher actin-binding strength, but due to large error bars, the difference could not be considered significant. The results produced by the actin affinity assay showed much more notable difference in nebulin-actin interactions of S16 and S16-m-1 proteins compared to each other.

The results were concluded based on the calculated values of three parameters. The dissociation constant, K_d, represents ligand concentration that binds to half the receptor sites at equilibrium. The lower K_d is, the higher is the binding affinity. In the performed assay K_d was significantly lower in S16-m-1, which supports the possibility of S16-m-1 binding more strongly to F-actin than S16, which was suggested by co-sedimentation assay results.

The Hill coefficient (h) can give an indication of a number of active binding sites. When h equals one, there is a single active binding site in the tested molecule. If $h>1$, the value for this coefficient roughly corresponds to the number of positively cooperative binding sites. And if $h<1$, it means that the binding sites are negatively cooperative. The Hill coefficient in S16-m-1 is higher than in S16, which means higher positive cooperativity between binding sites and therefore higher binding affinity to F-actin.

B_{max}-values represent the maximum number of specific binding sites on nebulin in ratio to actin in each assay. According to the assay results, B_{max}-value for S16 is higher than for S16-m-1, so S16 has more binding sites specific for actin.

In actin affinity determination, the main parameter is dissociation constant K_d, and Hill coefficient and B_{max} provide additional information about the binding sites. So, according to the results from this affinity assay, there are more specific actin-binding sites in S16, but the sites in S16-m-1 are more cooperative and posess higher binding affinity. However, the observed results are preliminary because the assay was performed just once. It makes these preliminary results very interesting, but they are not conclusive and need to be repeated for definitive result. Also, S16 did not reach the plateau in the **Figure 24** probably due to not optimal protein concentrations, which enlarges the error bars and makes the K_d- and B_{max}-values less accurate. All in all, the affinity assay clarified previous actin co-sedimentation assay results proving to be more exact and catching smaller differences, as well as providing additional information.

9.4. 3D-models

The prediction of the GST-nebulin fragment -fusion proteins were performed in order to visualize the structure of the produced proteins. The output models represent the structures of the tested SRs and their variants as they would fold in the actin binding tests, without posttranslational modifications.

Both, I-TASSER Suite and Server performed well and gave similar predictions as an output with C-scores higher than -1,5 which indicates correct general topology, and with average L-scores below 10Å. In all I-TASSER predictions the GST-part of the fusion protein was globular and well defined based on low l-scores (see **Appendix 4.**) The folding of the nebulin-part of the fusion protein varied depending on SR and the software used, Suite or Server. In some cases the structures were more linear, in other less linear. Nebulin-parts folded always in a similar pattern, around the GST-part making the overall structure of the fusion proteins globular. This indicates that GST may play a role in fusion protein folding, and could affect the capacity of nebulin fragments to interact with F-actin. However, the effect should be minimal due to its location around the GST-core and not the other way around. General similarity between wild-type SR-fragments and their variants make the detection of structural differences possible in the future. However, currently there are several limitations which makes drawing conclusions impossible from the acquired results.

One of the most important drawbacks in accurate structure prediction of nebulin fragments is lack of knowledge of its 3D-structure, especially the extensive repetitive SR area. This affected the certainty of the software to make the correct prediction. This showed in C-scores of the predicted models. The GST alone had a score of around 1.5 whereas the fusion proteins had significantly lower scores.

The I-TASSER server performed well, but it has very little options in how the software can be run. However, more threading programs such as SP3, FFAS and others that are not available at default in the Suite, allowed wider diversity of possible structures. This resulted in higher L-scores of the Server predictions because the software could recognize more folding options and was less certain of the correct one.

In order to accurately predict smaller changes in structures, the software needs to be optimized for the best prediction and many more threading programs need to be included into the software. The lack of the threading programs in Suite led to very few variants of possible structures and to insensitivity to single amino acid changes in sequences. Based on the results, the standalone I-TASSER with the default settings and scarce nebulin structure data in databases cannot be used for predicting fine differences between SR-fragments and their varinats. The results from Suite and Server in comparison with each other show the big effect of differences in threading programs. The plasticity of the Suite allow the software to be modified and optimized for more accurate predictions than available with default Suite and Server settings.

9.5. Future prospects

The actin affinity assays were conducted as a test of new method for quantification of nebulin fragment-actin-binding strength and the results need to be verified. Also, the method needs to be optimised for the curve to reach the plateau-phase for the results to be more reliable.

SRs could be modelled without GST for full understanding of the effect GST has on the protein folding. Also, the structure for all SRs could be predicted to monitor structural differences and structural homology between SRs. The predicted structures can indicate the position and folding of nebulin in thin filament.

The importance of testing more missense variants found in NM patients is now more definitive because of the change some of the tested variants caused in actin-binding strength. The effect

multiple missense mutations can have on actin-binding strength is also poorly understood and should be explored further.

10. Conclusions

According to the predicted 3D-structures of GST-nebulin-fusion proteins, the GST-part of the protein folds into a globular structure. The nebulin fragment folds around the GST, depending on the SR sequence. For full understanding of nebulin 3D-structure, the structure of the nebulin fragments need to be predicted without the GST acting as a core.

The results of actin co-sedimentation assays for S1- and S7-variants indicate no correlation between corresponding variants in different SR with different actin-binding strength. Instead they show the tendency of having the opposite effect on actin-binding strength. This demonstrates that variation in each SR can have various effects on actin-binding capacity depending on a SR. Missense variants in nebulin (see **Chapter 8.3.4.**) were demonstrated to have effect on actin-binding strength and can possibly affect muscle strength and be the cause of NM in patients.

The quantified actin affinities for S16 and S16-m-1 clarifies the results from co-sedimentation assay and the effect missense mutation has on actin affinity of S16. The variant binds to F-actin with higher affinity. Actin affinity assay shows promising results in terms of more precise quantification of actin-nebulin binding strength. With some adjustments, this method could be optimized and be widely used in future assays.

Acknowledgements

I would like to express my deepest gratitude to my supervisors Katarina Pelin, Vilma-Lotta Lehtokari and Johanna Lehtonen, who guided and helped me through the whole period of making this thesis. Thank you for your positive attitude, support and endless patience.

My gratitude also belongs to our group leaders, Carina Wallgren-Pettersson and Katarina Pelin, for making doing this thesis in their group possible. Also, I would like to thank to all the group members, Marilotta Turunen, Lydia Sagath, Sampo Koivunen, Kirsi Kiiski and Giorgia Bennici.

I would also like to thank Folkhälsan research center and the colleagues from there, you made the atmosphere in laboratory light and effortless, and working among you was a pleasure.

Sabita Kawan, thank you for your inexhaustible optimism and friendly encouragement.

The last, but not the least, thank you, my dearest family and friends for supporting me through the whole time towards the goal.

References

ADAMS, J., KELSO, R. and COOLEY, L., 2000. The kelch repeat superfamily of proteins: propellers of cell function. *Trends in cell biology,* **10**(1), pp. 17-24.

AGRAWAL, P.B., GREENLEAF, R.S., TOMCZAK, K.K., LEHTOKARI, V., WALLGREN-PETTERSSON, C., WALLEFELD, W., LAING, N.G., DARRAS, B.T., MACIVER, S.K., DORMITZER, P.R. and BEGGS, A.H., 2006. Nemaline Myopathy with Minicores Caused by Mutation of the CFL2 Gene Encoding the Skeletal Muscle Actinâ€"Binding Protein, Cofilin-2. *American Journal of Human Genetics,* **80**(1), pp. 162-167.

ALAZAMI, A.M., KENTAB, A.Y., FAQEIH, E., MOHAMED, J.Y., ALKHALIDI, H., HIJAZI, H. and ALKURAYA, F.S., 2015. A novel syndrome of Klippel-Feil anomaly, myopathy, and characteristic facies is linked to a null mutation in *MYO18B. J Med Genet,* **52**(6), pp. 400.

ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K. and WALTER, P., 2008. *Molecular biology of the cell.* 5 edn. United States of America: Garland Science.

BACHMAN, J., 2013. *Chapter Ninteen - Site-Directed Mutagenesis.* Academic Press.

BELIZÁRIO, J.E., FONTES-OLIVEIRA, C., BORGES, J.P., KASHIABARA, J.A. and VANNIER, E., 2016. Skeletal muscle wasting and renewal: a pivotal role of myokine IL-6. *SpringerPlus,* **5**, pp. 619.

CHU, M., GREGORIO, C.C. and PAPPAS, C.T., 2016. Nebulin, a multi-functional giant. *The Journal of experimental biology,* **219**(2), pp. 146.

DONNER, K., SANDBACKA, M., LEHTOKARI, V., WALLGREN-PETTERSSON, C. and PELIN, K., 2004. *Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts.*

EULITZ, S., SAUER, F., PELISSIER, M., BOISGUERIN, P., MOLT, S., SCHULD, J., ORFANOS, Z., KLEY, R.A., VOLKMER, R., WILMANNS, M., KIRFEL, G., VAN DER VEN, P.,F.M. and FÃ¼RST, D.,O., 2013. Identification of Xin-repeat proteins as novel ligands of the SH3 domains of nebulin and nebulette and analysis of their interaction during myofibril formation and remodeling. *Molecular biology of the cell,* **24**(20), pp. 3215-3226.

FRONTERA, W.R. and OCHALA, J., 2015. Skeletal Muscle: A Brief Review of Structure and Function. *Calcified tissue international,* **96**(3), pp. 183-195.

GARG, A., O'ROURKE, J., LONG, C., DOERING, J., RAVENSCROFT, G., BEZPROZVANNAYA, S., NELSON, B.R., BEETZ, N., LI, L., CHEN, S., LAING, N.G., GRANGE, R.W., BASSEL-DUBY, R. and OLSON, E.N., 2014. KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy. *The Journal of clinical investigation,* **124**(8), pp. 3529-3539.

GEEVES, M.A., HITCHCOCK-DEGREGORI, S. and GUNNING, P.W., 2015. A systematic nomenclature for mammalian tropomyosin isoforms. *Journal of muscle research and cell motility,* **36**(2), pp. 147-153.

HART, A., 2000. Mann-Whitney test is not just a test of medians: differences in spread can be important. *BMJ : British Medical Journal,* **323**(7309), pp. 391-393.

HEIER, J.A., DICKINSON, D.J. and KWIATKOWSKI, A.V., 2017. Measuring Protein Binding to F-actin by Co-sedimentation. *Journal of Visualized Experiments : JoVE*, (123), pp. 55613.

HEINY, J.A. and MEISSNER, G., 2012. Chapter 45 - Excitation—Contraction Coupling in Skeletal Muscle. In: N. SPERELAKIS, ed, *Cell Physiology Source Book (Fourth Edition)*. San Diego: Academic Press, pp. 783-800.

JIN, J.P. and WANG, K., 1991. Cloning, expression, and protein interaction of human nebulin fragments composed of varying numbers of sequence modules. *Journal of Biological Chemistry*, **266**(31), pp. 21215-21223.

JOHNSTON, J.J., KELLEY, R.I., CRAWFORD, T.O., MORTON, D.H., AGARWALA, R., KOCH, T., SCHÃ¤FFER, A.,A., FRANCOMANO, C.A. and BIESECKER, L.G., 2000. A Novel Nemaline Myopathy in the Amish Caused by a Mutation in Troponin T1. *American Journal of Human Genetics*, **67**(4), pp. 814-821.

KIISKI, K., LEHTOKARI, V., LÃ¶YTYNOJA, A., AHLSTÃ©N, L., LAITILA, J., WALLGREN-PETTERSSON, C. and PELIN, K., 2015. A recurrent copy number variation of the NEB triplicate region: only revealed by the targeted nemaline myopathy CGH array. *European Journal of Human Genetics*, **24**(4), pp. 574-580.

LABEIT, S. and KOLMERER, B., 1995. *The complete primary structure of human nebulinand its correlation to muscle structure.*

LABEIT, S., OTTENHEIJM, C.A.C. and GRANZIER, H., 2010. Nebulin, a major player in muscle health and disease. *The FASEB Journal*, **25**(3), pp. 822-829.

LAING, N.G., DYE, D.E., WALLGRENÂ€• PETTERSSON CARINA, GABRIELE, R., NICOLE, M., SUZANNE, L., WINDER, T.L., LOCHMÃ¼LLER HANNS, CLAUDIO, G., MITRANIÂ€• ROSENBAUM STELLA, DARREN, T., SPARROW, J.C., BEGGS, A.H. and NOWAK, K.J., 2009. Mutations and polymorphisms of the skeletal muscle Î±â€• actin gene (ACTA1). *Human mutation*, **30**(9), pp. 1267-1277.

LAITILA, J., HANIF, M., PAETAU, A., HUJANEN, S., KETO, J., SOMERVUO, P., HUOVINEN, S., UDD, B., WALLGREN-PETTERSSON, C., AUVINEN, P., HACKMAN, P. and PELIN, K., 2012. Expression of multiple nebulin isoforms in human skeletal muscle and brain. *Muscle & nerve*, **46**(5), pp. 730-737.

LAITILA, J., LEHTONEN, J., LEHTOKARI, V., SAGATH, L., WALLGREN-PETTERSSON, C., GRÖNHOLM, M. and PELIN, K., 2019. A nebulin super-repeat panel reveals stronger actin binding toward the ends of the super-repeat region. *Muscle & nerve*, **59**(1), pp. 116-121.

LEHTOKARI, V., KIISKI, K., SANDARADURA, S.A., LAPORTE, J., REPO, P., FREY, J.A., DONNER, K., MARTTILA, M., SAUNDERS, C., BARTH, P.G., DEN DUNNEN, J.,T., BEGGS, A.H., CLARKE, N.F., NORTH, K.N., LAING, N.G., ROMERO, N.B., WINDER, T.L., PELIN, K. and WALLGREN-PETTERSSON, C., 2014a. Mutation Update: The Spectra of Nebulin Variants and Associated Myopathies. *Human mutation*, **35**(12), pp. 1418-1426.

LEHTOKARI, V., KIISKI, K., SANDARADURA, S.A., LAPORTE, J., REPO, P., FREY, J.A., DONNER, K., MARTTILA, M., SAUNDERS, C., BARTH, P.G., DEN DUNNEN, J.,T., BEGGS, A.H., CLARKE, N.F., NORTH, K.N., LAING, N.G., ROMERO, N.B., WINDER, T.L., PELIN, K. and WALLGREN-PETTERSSON, C., 2014b. Mutation Update: The Spectra of Nebulin Variants and Associated Myopathies. *Human mutation*, **35**(12), pp. 1418-1426.

LEHTOKARI, V., PELIN, K., HERCZEGFALVI, A., KARCAGI, V., POUGET, J., FRANQUES, J., PELLISSIER, J.F., FIGARELLA-BRANGER, D., VON DER HAGEN, M., HUEBNER, A., SCHOSER, B., LOCHMÜLLER, H. and WALLGREN-PETTERSSON, C., 2011. *Nemaline myopathy caused by mutations in the nebulin gene may present as a distal myopathy.*

LEHTONEN, J.M., 2017. *Nebuliinin supertoistojen sitoutuminen aktiiniin ja tunnettujen varianttien vaikutus nebuliini-aktiini-interaktioon. (translation: Nebulin super repeat binding to actin and the effects of known variants in nebulin on nebulin-actin interaction)*, University of Helsinki.

LIN, B.L., TAEJEONG, S. and SAKTHIVEL, S., 2017. *Myofilaments: Movers and Rulers of the Sarcomere*.

LUTHER, P.K., 2009. The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. *Journal of muscle research and cell motility*, **30**(5-6), pp. 171-185.

MARTTILA, M., HANIF, M., LEMOLA, E., NOWAK, K.J., LAITILA, J., GRÃ¶NHOLM, M., WALLGREN-PETTERSSON, C. and PELIN, K., 2014. Nebulin interactions with actin and tropomyosin are altered by disease-causing mutations. *Skeletal Muscle*, **4**, pp. 15-15.

MARTTILA, M., LEHTOKARI, V., MARSTON, S., NYMAN, T.A., BARNERIAS, C., BEGGS, A.H., BERTINI, E., CEYHAN-BIRSOY, O., CINTAS, P., GERARD, M., GILBERT-DUSSARDIER, B., HOGUE, J.S., LONGMAN, C., EYMARD, B., FRYDMAN, M., KANG, P.B., KLINGE, L., KOLSKI, H., LOCHMÃ¼LLER, H., MAGY, L., MANEL, V., MAYER, M., MERCURI, E., NORTH, K.N., PEUDENIER-ROBERT, S., PIHKO, H., PROBST, F.J., REISIN, R., STEWART, W., TARATUTO, A.L., DE VISSER, M., WILICHOWSKI, E., WINER, J., NOWAK, K., LAING, N.G., WINDER, T.L., MONNIER, N., CLARKE, N.F., PELIN, K., GRÃ¶NHOLM, M. and WALLGREN-PETTERSSON, C., 2014. Mutation Update and Genotypeâ€“Phenotype Correlations of Novel and Previously Described Mutations in TPM2 and TPM3 Causing Congenital Myopathies. *Human mutation*, **35**(7), pp. 779-790.

MARUSTERI, M. and BACAREA, V., 2010. *Comparing groups for statistical differences: How to choose the right statistical test?* .

MCELHINNY, A.S., KAZMIERSKI, S.T., LABEIT, S. and GREGORIO, C.C., 2003. *Nebulin: The Nebulous, Multifunctional Giant of Striated Muscle.*

MIYATAKE, S., MITSUHASHI, S., HAYASHI, Y.K., PUREVJAV, E., NISHIKAWA, A., KOSHIMIZU, E., SUZUKI, M., YATABE, K., TANAKA, Y., OGATA, K., KURU, S., SHIINA, M., TSURUSAKI, Y., NAKASHIMA, M., MIZUGUCHI, T., MIYAKE, N., SAITSU, H., OGATA, K., KAWAI, M., TOWBIN, J., NONAKA, I., NISHINO, I. and MATSUMOTO, N., 2017. Biallelic Mutations in *MYPN*, Encoding Myopalladin, Are Associated with Childhood-Onset, Slowly Progressive Nemaline Myopathy. *The American Journal of Human Genetics*, **100**(1), pp. 169-178.

OGUT, O., HOSSAIN, M.M. and JIN, J., 2003. Interactions between Nebulin-like Motifs and Thin Filament Regulatory Proteins. *Journal of Biological Chemistry*, **278**(5), pp. 3089-3097.

OTTENHEIJM, C.A.C., GRANZIER, H. and LABEIT, S., 2012. The Sarcomeric Protein Nebulin: Another Multifunctional Giant in Charge of Muscle Strength Optimization. *Frontiers in Physiology*, **3**, pp. 37.

PAPPAS, C.T., BHATTACHARYA, N., COOPER, J.A. and GREGORIO, C.C., 2008. Nebulin Interacts with CapZ and Regulates Thin Filament Architecture within the Z-Disc. *Molecular biology of the cell*, **19**(5), pp. 1837-1847.

PERRIN, B.J. and ERVASTI, J.M., 2010. The actin gene family: Function follows isoform. *Cytoskeleton,* **67**(10), pp. 630-634.

QAISAR, R., BHASKARAN, S. and VAN REMMEN, H., 2016. *Muscle fiber type diversification during exercise and regeneration.*

RAVENSCROFT, G., SOLLIS, E., CHARLES, A.K., NORTH, K.N., BAYNAM, G. and LAING, N.G., 2011. Fetal akinesia: review of the genetics of the neuromuscular causes. *J Med Genet,* **48**(12), pp. 793.

ROMERO, N.B., SANDARADURA, S.A. and CLARKE, N.F., 2013. **Recent advances in nemaline myopathy**. *Current Opinion in Neurology,* **26(5)**, pp. 519-520-526.

ROMERO, N.B., LEHTOKARI, V.-., QUIJANO-ROY, S., MONNIER, N., CLAEYS, K.G., CARLIER, R.Y., PELLEGRINI, N., ORLIKOWSKI, D., BAROIS, A., LAING, N.G., LUNARDI, J., FARDEAU, M., PELIN, K. and WALLGREN-PETTERSSON, C., 2009. CORE-ROD MYOPATHY CAUSED BY MUTATIONS IN THE NEBULIN GENE. *Neurology,* **73**(14), pp. 1159.

ROY, A., KUCUKURAL, A. and ZHANG, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols,* **5**(4), pp. 725-738.

SAMBUUGHIN, N., YAU, K.S., OLIVÃ©, M., DUFF, R.M., BAYARSAIKHAN, M., LU, S., GONZALEZ-MERA, L., SIVADORAI, P., NOWAK, K.J., RAVENSCROFT, G., MASTAGLIA, F.L., NORTH, K.N., ILKOVSKI, B., KREMER, H., LAMMENS, M., VAN ENGELEN, B.G.M., FABIAN, V., LAMONT, P., DAVIS, M.R., LAING, N.G. and GOLDFARB, L.G., 2010. Dominant Mutations in KBTBD13, a Member of the BTB/Kelch Family, Cause Nemaline Myopathy with Cores. *The American Journal of Human Genetics,* **87**(6), pp. 842-847.

SANDARADURA, S.A., BOURNAZOS, A., MALLAWAARACHCHI, A., CUMMINGS, B.B., WADDELL, L.B., JONES, K.J., TROEDSON, C., SUDARSANAM, A., NASH, B.M., PETERS, G.B., ALGAR, E.M., MACARTHUR, D.G., NORTH, K.N., BRAMMAH, S., CHARLTON, A., LAING, N.G., WILSON, M.J., DAVIS, M.R. and COOPER, S.T., 2018. Nemaline myopathy and distal arthrogryposis associated with an autosomal recessiveÂ TNNT3Â splice variant. *Human mutation,* **39**(3), pp. 383-388.

SILVERTHORN, D.U., 2007. *Human physiology: An integrated approach.* 6 edn. San Francisco: Pearson/Benjamin Cummings.

SRIVASTAVA, J. and BARBER, D., 2008. Actin Co-Sedimentation Assay; for the Analysis of Protein Binding to F-Actin. *Journal of Visualized Experiments : JoVE,* (13), pp. 690.

TONINO, P., PAPPAS, C.T., HUDSON, B.D., LABEIT, S., GREGORIO, C.C. and GRANZIER, H., 2009. Reduced myofibrillar connectivity and increased Z-disk width in nebulin-deficient skeletal muscle. *Journal of cell science,* **123**(3), pp. 384-391.

WU, S., SKOLNICK, J. and ZHANG, Y., 2007. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biology,* **5**, pp. 17-17.

YAMAMOTO, D.L., VITIELLO, C., ZHANG, J., GOKHIN, D.S., CASTALDI, A., COULIS, G., PIASER, F., FILOMENA, M.C., EGGENHUIZEN, P.J., KUNDERFRANCO, P., CAMERINI, S., TAKANO, K., ENDO, T., CRESCENZI, M., LUTHER, P.K.L., LIEBER, R.L., CHEN, J. and BANG, M., 2013. The nebulin SH3 domain is dispensable for normal skeletal muscle structure but is required for effective active load bearing in mouse. *Journal of cell science,* **126**(23), pp. 5477-5489.

YUEN, M., SANDARADURA, S.A., DOWLING, J.J., KOSTYUKOVA, A.S., MOROZ, N., QUINLAN, K.G., LEHTOKARI, V., RAVENSCROFT, G., TODD, E.J., CEYHAN-BIRSOY, O., GOKHIN, D.S., MALUENDA, J., LEK, M., NOLENT, F., PAPPAS, C.T., NOVAK, S.M., DÂ€™AMICO, A., MALFATTI, E., THOMAS, B.P., GABRIEL, S.B., GUPTA, N., DALY, M.J., ILKOVSKI, B., HOUWELING, P.J., DAVIDSON, A.E., SWANSON, L.C., BROWNSTEIN, C.A., GUPTA, V.A., MEDNE, L., SHANNON, P., MARTIN, N., BICK, D.P., FLISBERG, A., HOLMBERG, E., VAN, D.B., LAPUNZINA, P., WADDELL, L.B., SLOBODA, D.D., BERTINI, E., CHITAYAT, D., TELFER, W.R., LAQUERRIÃ¨RE, A., GREGORIO, C.C., OTTENHEIJM, C.A.C., BÃ¶NNEMANN, C.,G., PELIN, K., BEGGS, A.H., HAYASHI, Y.K., ROMERO, N.B., LAING, N.G., NISHINO, I., WALLGREN-PETTERSSON, C., MELKI, J., FOWLER, V.M., MACARTHUR, D.G., NORTH, K.N. and CLARKE, N.F., 2014. Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *The Journal of clinical investigation*, **124**(11), pp. 4693-4708.

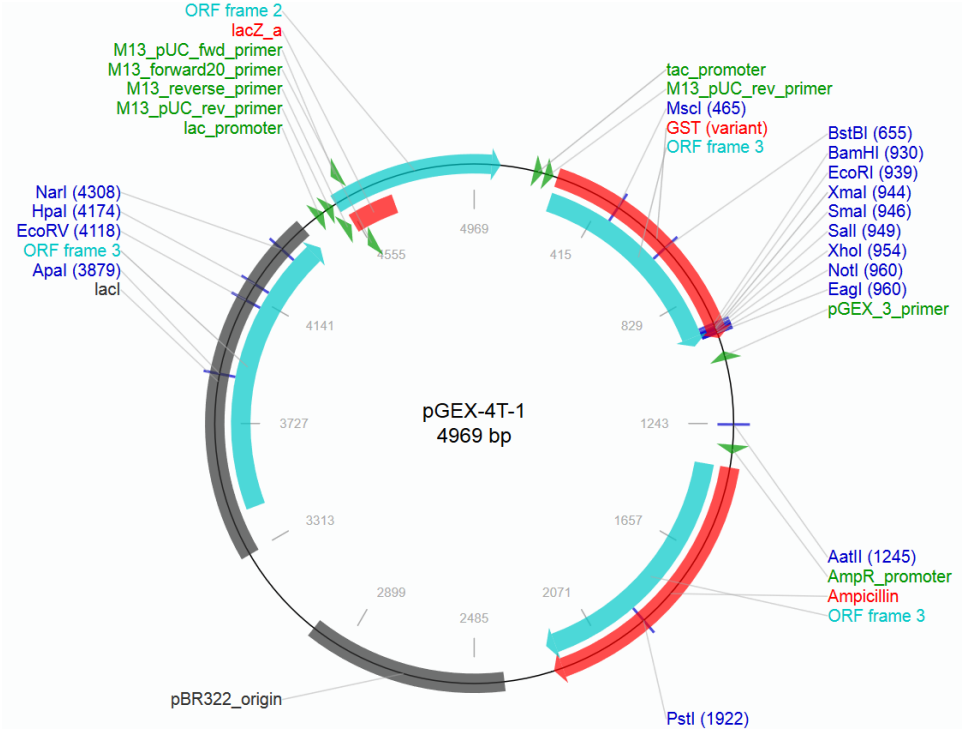
ZHANG, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**(1), pp. 40.

Online resourses

1. GraphPad QuickCalc: https://www.graphpad.com/quickcalcs/ 27.06.2019
2. ExPASy: https://web.expasy.org/translate/ 27.06.2019
3. ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/ 27.06.2019

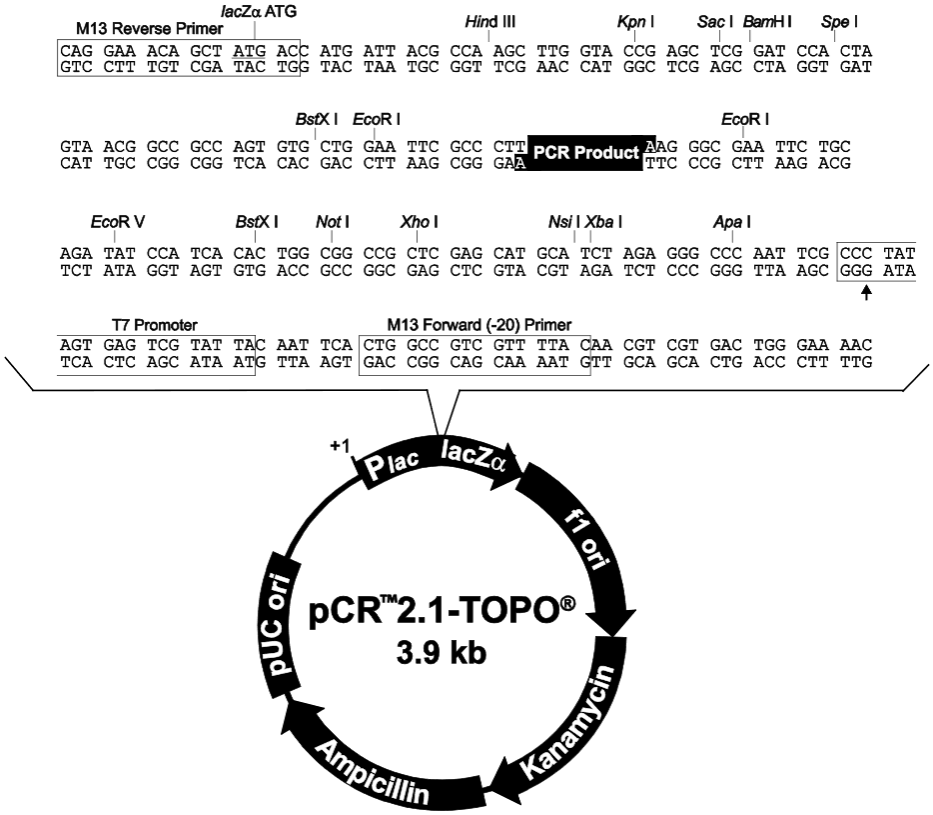
Appendix

1. pGEX-4T-1 map



Source: https://www.addgene.org/browse/sequence_vdb/2873/

2. pCR™ 2.1 TOPO map



Source: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/topota_man.pdf

3. Amino acid sequences of GST-SR fusion proteins

Super repeat area in **bold**.

Mutated amino acids marked as **white text on black background**.

Deletion point marked as ▮.

Insert amino acid sequence marked in blue.

GST-S1

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALFMQTETPEYKMNKKAGVAASKVKYKEDYEKNKGKADYNVL**PASENPQLRQLKAAGDALSDKLYKENYEK**
TKAKSINYCETPKFKLDTVLQNFSSDKKYKDSYLKDILGHYVGSFEDPYHSHCMKVTAQNSDKNYKAEYEEDRGK
GFFPQTITQEYEAIKKLDQCKDHTYKVHPDKTKFTQVTDSPVLLQAQVNSKQLSDLNYKAKHESEKFKCHIPPDT
PAFIQHKVNAYNLSDNLYKQDWEKSKAKKFDIKVDAIPLLAAKANTKNTSDVMYKKDYEKNKGKMIGVLSINDDP
KMLHSLKAEFPGRLERPHRD

GST-S1-m-1

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALFMQTETPEYKMNKKAGVAASKVKYKEDYEKNKGKADYNVL**PASENPQLRQLKAAGDALSDKLYKENYEK**
TKAKSINYCETPKFKLDTVLQNFSSDKKYKDSYLKDILGHYVGSFEDPYHSHCMKVTAQNSDKNYKAEYEEDRGK
GFFPQTITQECEAIKKLDQCKDHTYKVHPDKTKFTQVTDSPVLLQAQVNSKQLSDLNYKAKHESEKFKCHIPPDT
PAFIQHKVNAYNLSDNLYKQDWEKSKAKKFDIKVDAIPLLAAKANTKNTSDVMYKKDYEKNKGKMIGVLSINDDP
KMLHSLKAEFPGRLERPHRD

GST-S1-m-2

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALFMQTETPEYKMNKKAGVAASKVKYKEDYEKNKGKADYNVL**PASENPQLRQLKAAGDALSDKLYKENYEK**
TKAKSINYCETPKFKLDTVLQNFSSDKKYKDSYLKDILGHYVGSFEDPYHSHCMKVTAQNSDKNYKAEYEEDRGK
GFFPQTITQEYEAIKKLDQCKDHTYCVHPDKTKFTQVTDSPVLLQAQVNSKQLSDLNYKAKHESEKFKCHIPPDT
PAFIQHKVNAYNLSDNLYKQDWEKSKAKKFDIKVDAIPLLAAKANTKNTSDVMYKKDYEKNKGKMIGVLSINDDP
KMLHSLKAEFPGRLERPHRD

GST-S1-del-1

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALFMQTETPEYKMNKKAGVAASKVKYKEDYEKNKGKADYNVL**PASENPQLRQLKAAGDALSDKLYKENYEK**
TKAKSINYCETPKFKLDTVLQNFSSDKKYKDSYLKDILGHYVGSFEDPYHSHCMKVTAQNSDKN▮AEYEEDRGKG
FFPQTITQEYEAIKKLDQCKDHTYKVHPDKTKFTQVTDSPVLLQAQVNSKQLSDLNYKAKHESEKFKCHIPPDTP
AFIQHKVNAYNLSDNLYKQDWEKSKAKKFDIKVDAIPLLAAKANTKNTSDVMYKKDYEKNKGKMIGVLSINDDPK
MLHSLKAEFPGRLERPHRD

GST-S2

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG

SPEF**INDDPKMLHSLKVAKNQSDRLYKENYEKTKAKSMNYCETPKYQLDTQLKNFSEARYKDLYVKDVLGHYVGS**
MEDPYHTHCMKVAAQNSDKSYKAEYEEDKGKCYFPQTITQEYEAIKKLDQCKDHTYKVHPDKTKFTAVTDSPVLL
QAQLNTKQLSDLNYKAKHEGEKFKCHIPADAPQFIQHRVNAYNLSDNVYKQDWEKSKAKKFDIKVDAIPLLAAKA
NTKNTSDVMYKKDYEKSKGKMIGALSEFPGRLERPHRD

GST-S2-m-1

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEF**INDDPKMLHSLKVAKNQSDRLYKENYEKTKAKSMNYCETPKYQLDTQLKNFSEARYKDLYVKDVLGHYVGS**
MEDPYHTHCMKVAAQNSDKSYKAEYEEDKGKCYFPQTITQEYEAIKKLDQCKDHTYKVHPDKTKFTAVTDSPVLL
QAQLNTKQLSDLNCKAKHEGEKFKCHIPADAPQFIQHRVNAYNLSDNVYKQDWEKSKAKKFDIKVDAIPLLAAKA
NTKNTSDVMYKKDYEKSKGKMIGALSEFPGRLERPHRD

GST-S4

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALAKGKMVGFQS**LQDDPKLVHYMNVAKIQSDREYKKDYEKTKSKYNTPHDMFNVVAAKKAQDVVSNVNYKH**
SLHHYTYLPDAMDLELSKNMMQIQSDNVYKEDYNNWMKGIGWIPIGSLDVEKVKKAGDALNEKKYRQHPDTLKFT
SIVDSPVMVQAKQNTKQVSDILYKAKGEDVKHKYTMSPDLPQFLQAKCNAYNISDVCYKRDWYDLIAKGNNVLGD
AIPITAAKASRNIASDYKYKEAYEKSKGKHVGFRSLQDDPKLVHYMNVAKLQKAEFPGRLERPHRD

GST-S4-m-1

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALAKGKMVGFQS**LQDDPKLVHYMNVAKIQSDREYKKDYEKTKSKYNMPHDMFNVVAAKKAQDVVSNVNYKH**
SLHHYTYLPDAMDLELSKNMMQIQSDNVYKEDYNNWMKGIGWIPIGSLDVEKVKKAGDALNEKKYRQHPDTLKFT
SIVDSPVMVQAKQNTKQVSDILYKAKGEDVKHKYTMSPDLPQFLQAKCNAYNISDVCYKRDWYDLIAKGNNVLGD
AIPITAAKASRNIASDYKYKEAYEKSKGKHVGFRSLQDDPKLVHYMNVAKLQKAEFPGRLERPHRD

GST-S4-m-2

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALAKGKMVGFQS**LQDDPKLVHYMNVAKIQSDREYKKDYEKTKSKYNTPHDMFNVVAAKKAQDVVSNVNYKH**
SLHHYTYLPDAMDLELSKNMMQIQSDNVYKEDYNNWMKGIGWIPIGSLDVEKVKKAGDALNEKKYRQHPDTLKFT
SIVDSPVMVQAKQNTKHVSDILYKAKGEDVKHKYTMSPDLPQFLQAKCNAYNISDVCYKRDWYDLIAKGNNVLGD
AIPITAAKASRNIASDYKYKEAYEKSKGKHVGFRSLQDDPKLVHYMNVAKLQKAEFPGRLERPHRD

GST-S7

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALKHIGFRS**LEDDPKLVHFMQVAKMQSDREYKKGYEKSKTSFHTPVDMLSVVAAKKSQEVATNANYRNVIH**
TYNMLPDAMSFELAKNMMQIQSDNQYKADYADFMKGIGWLPLGSLEAEKNKKAMEIISEKKYRQHPDTLKYSTLM
DSMNMVLAQNNAKIMNEHLYKQAWEADKTKVHIMPDIPQIILAKANAINMSDKLYKLSLEESKKKGYDLRPDAIP
IKAAKASRDIASDYKYKYNYEKGKGKMVGFRSLEDDPKLVHSMQVAKMQSDREKAEFPGRLERPHRD

GST-S7-m-1

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALKHIGFRS**LEDDPKLVHFMQVAKMQSDREYKKGYEKSKTSFHTPVDMLSVVAAKKSQEVATNANYRNVIH**
TYNMLPDAMSFELAKNMMQIQSDNQYKADYADFMKGIGWLPLGSLEACKNKKAMEIISEKKYRQHPDTLKYSTLM
DSMNMVLAQNNAKIMNEHLYKQAWEADKTKVHIMPDIPQIILAKANAINMSDKLYKLSLEESKKKGYDLRPDAIP
IKAAKASRDIASDYKYKYNYEKGKGKMVGFRSLEDDPKLVHSMQVAKMQSDREKAEFPGRLERPHRD

GST-S7-m-2

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALKHIGFRS**LEDDPKLVHFMQVAKMQSDREYKKGYEKSKTSFHTPVDMLSVVAAKKSQEVATNANYRNVIH**
TYNMLPDAMSFELAKNMMQIQSDNQYKADYADFMKGIGWLPLGSLEAEKNKKAMEIISEKKYCQHPDTLKYSTLM
DSMNMVLAQNNAKIMNEHLYKQAWEADKTKVHIMPDIPQIILAKANAINMSDKLYKLSLEESKKKGYDLRPDAIP
IKAAKASRDIASDYKYKYNYEKGKGKMVGFRSLEDDPKLVHSMQVAKMQSDREKAEFPGRLERPHRD

GST-S7-del-1

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALKHIGFRS**LEDDPKLVHFMQVAKMQSDREYKKGYEKSKTSFHTPVDMLSVVAAKKSQEVATNANYRNVIH**
TYNMLPDAMSFELAKNMMQIQSDNQ_**ADYADFMKGIGWLPLGSLEAEKNKKAMEIISEKKYRQHPDTLKYSTLMD**
SMNMVLAQNNAKIMNEHLYKQAWEADKTKVHIMPDIPQIILAKANAINMSDKLYKLSLEESKKKGYDLRPDAIPI
KAAKASRDIASDYKYKYNYEKGKGKMVGFRSLEDDPKLVHSMQVAKMQSDREKAEFPGRLERPHRD

GST-S10

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
S**IEDDPKMMWSMHVAKIQSDREYKKDFEKWKTKFSSPVDMLGVVLAKKCQTLVSDVDYKNYLHQWTCLPDQSDVI**
HARQAYDLQSDNLYKSDLQWLKGIGWMTSGSLEDEKNKRATQILSDHVYRQHPDQFKFSSLMDSIPMVLAKNNAI
TMNHRLYTEAWDKDKTTVHIMPDTPEVLLAKQNKVNYSEKLYKLGLEEAKRKGYDMRVDAIPIKAAKASRDIASE
FKYKEGYRKQLGHHIGARAGSPEFPGRLERPHRD

GST-S10-m-2

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
S**IEDDPKMMWSMHVAKIQSDREYKKDFEKWKTKFSSPVDMLGVVLAKKCQTLVSDVDYKNYLHQWTCLPDQSDVI**
HARQAYDLQSDNLYKSDLQWLKGIGWMTSGSLEDEKNKRATQILSDHVYHQHPDQFKFSSLMDSIPMVLAKNNAI
TMNHRLYTEAWDKDKTTVHIMPDTPEVLLAKQNKVNYSEKLYKLGLEEAKRKGYDMRVDAIPIKAAKASRDIASE
FKYKEGYRKQLGHHIGARAGSPEFPGRLERPHRD

GST-S10-m-3

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
S**IEDDPKMMWSMHVAKIQSDREYKKDFEKWKTKFSSPVDMLGVVLAKKCQTLVSDVDYKNYLHQWTCLPDQSDVI**
HARQAYDLQSDNLYKSDLQWLKGIGWMTSGSLEDEKNKRATQILSDHVYRQHPDQFKFSSLMDCIPMVLAKNNAI
TMNHRLYTEAWDKDKTTVHIMPDTPEVLLAKQNKVNYSEKLYKLGLEEAKRKGYDMRVDAIPIKAAKASRDIASE
FKYKEGYRKQLGHHIGARAGSPEFPGRLERPHRD

GST-S10-m-2&3

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
S**IEDDPKMMWSMHVAKIQSDREYKKDFEKWKTKFSSPVDMLGVVLAKKCQTLVSDVDYKNYLHQWTCLPDQSDVI**
HARQAYDLQSDNLYKSDLQWLKGIGWMTSGSLEDEKNKRATQILSDHVYHQHPDQFKFSSLMDCIPMVLAKNNAI
TMNHRLYTEAWDKDKTTVHIMPDTPEVLLAKQNKVNYSEKLYKLGLEEAKRKGYDMRVDAIPIKAAKASRDIASE
FKYKEGYRKQLGHHIGARAGSPEFPGRLERPHRD

GST-S16

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALKQKGHYVGTLT**ARDDNKIRWALIADKLQNEREYRLDWAKWKAKIQSPVDMLSILHSKNSQALVSDMDYR**
NYLHQWTCMPDQNDVIQAKKAYELQSDNVYKADLEWLRGIGWMPNDSVSVNHAKHAADIFSEKKYRTKIETLNFT
PVDDRVDYVTAKQSGEILDDIKYRKDWNATKSKYTLTETPLLHTAQEAARILDQYLYKEGWERQKATGYILPPDA
VPFVHAHHCNDVQSELKYKAEHVKQKGHYVGVPTMRDDPKLVWFEHAKAEFPGRLERPHRD

GST-S16-m-1

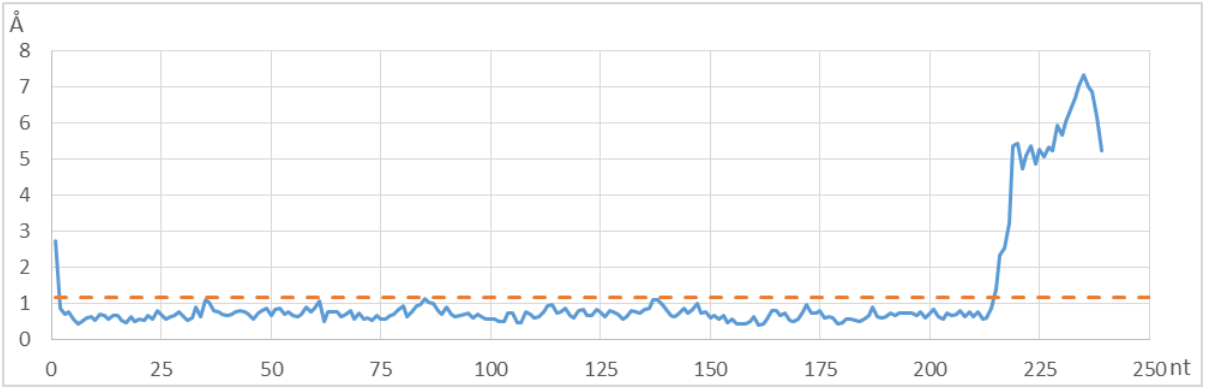
MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI
ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRG
SPEFALKQKGHYVGTLT**ARDDNKIRWALIADKLQNEREYRLDWAKWKAKIQSPVDMLSILHSKNSQALVSDMDYR**
NYLHQWTCMPDQNDVIQAKKAYELQSDNVYKADLEWLHGIGWMPNDSVSVNHAKHAADIFSEKKYRTKIETLNFT
PVDDRVDYVTAKQSGEILDDIKYRKDWNATKSKYTLTETPLLHTAQEAARILDQYLYKEGWERQKATGYILPPDA
VPFVHAHHCNDVQSELKYKAEHVKQKGHYVGVPTMRDDPKLVWFEHAKAEFPGRLERPHRD

4. Estimated L-scores of predicted 3D-models

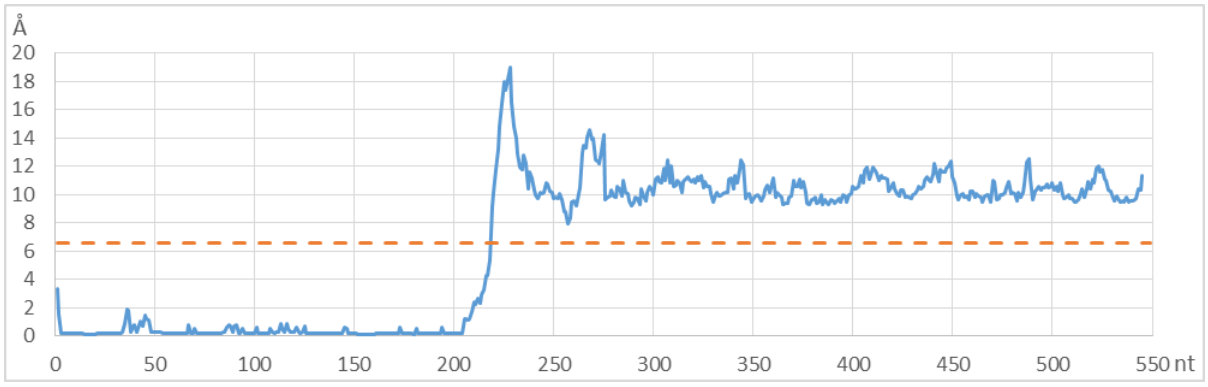
L-scores depicted in Ångtröms for each nucleotide (blue line). Average marked in orange.

L-scores for I-TASSER suite predictions:

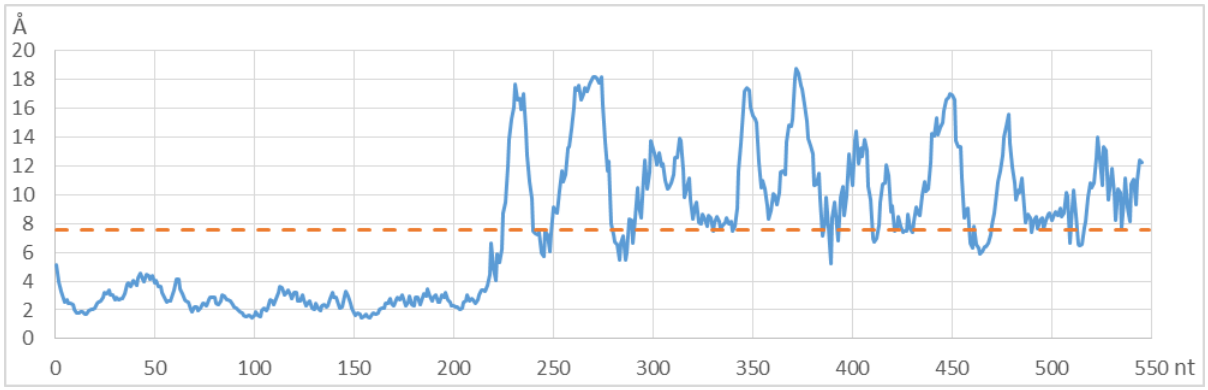
GST



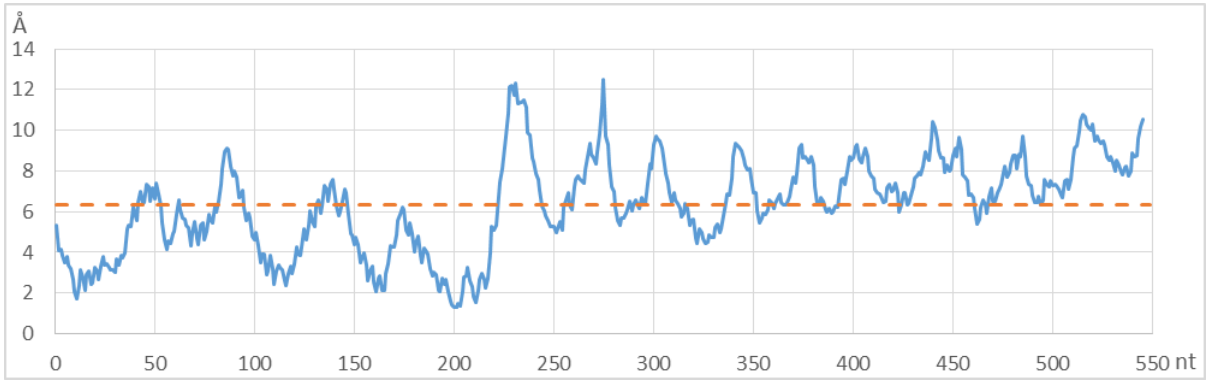
S1 wt



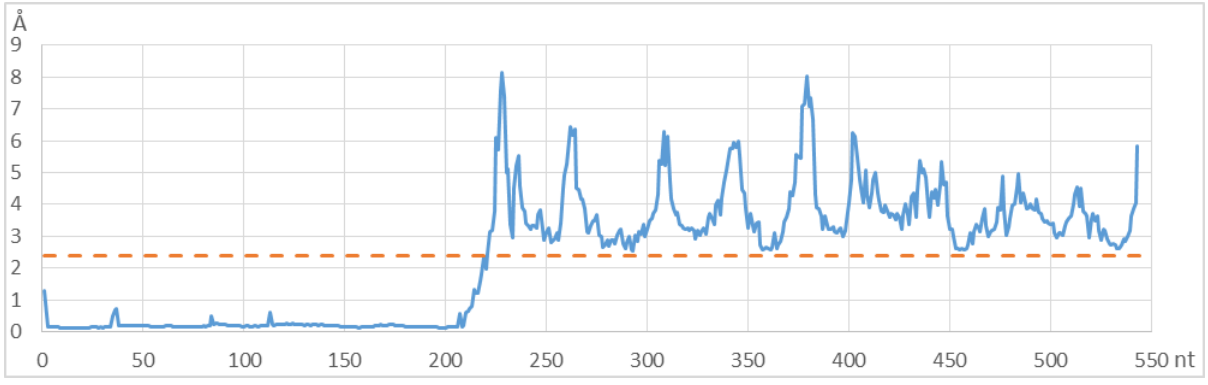
S1-m-1



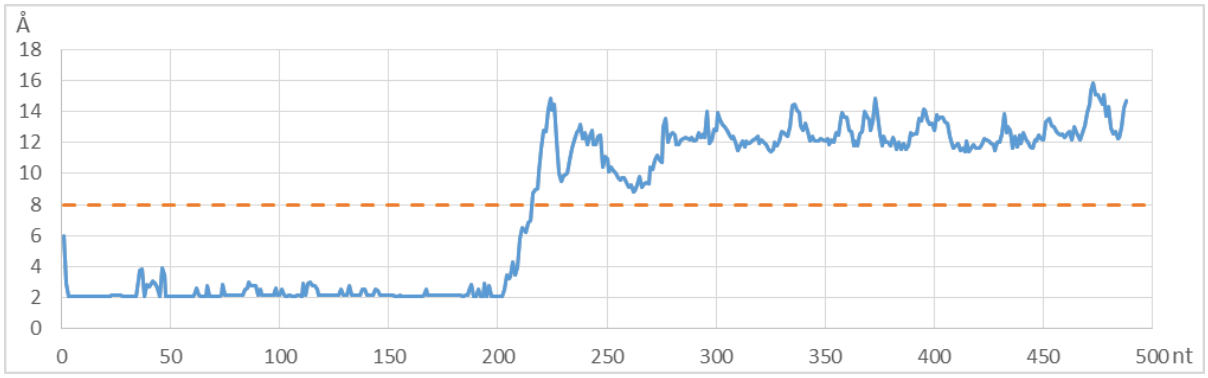
S1-m-2



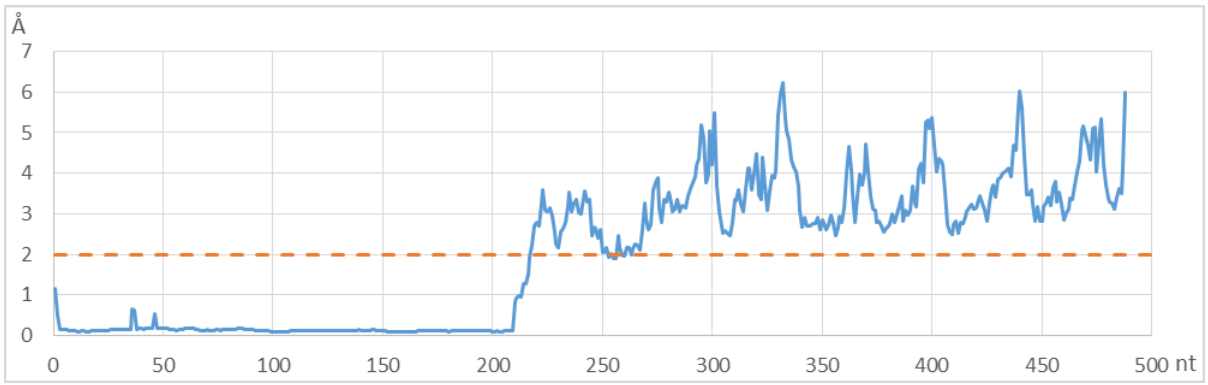
S1-del-1



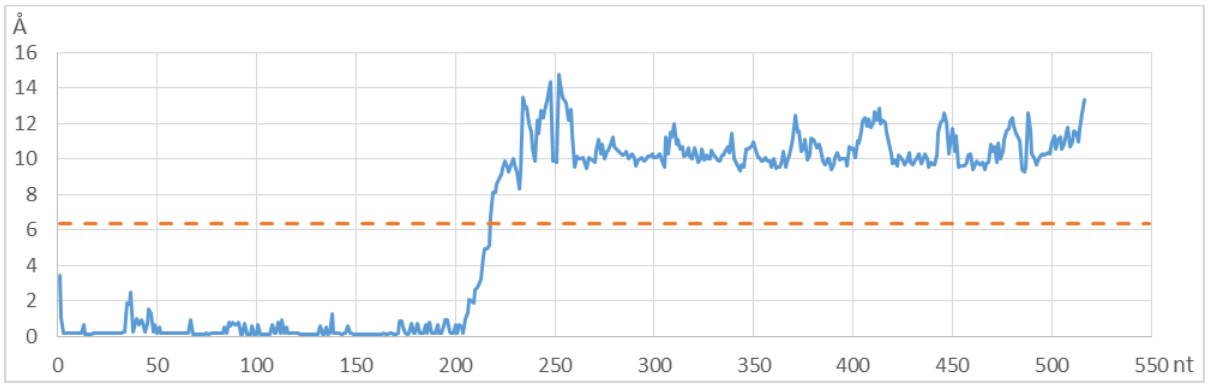
S2 wt



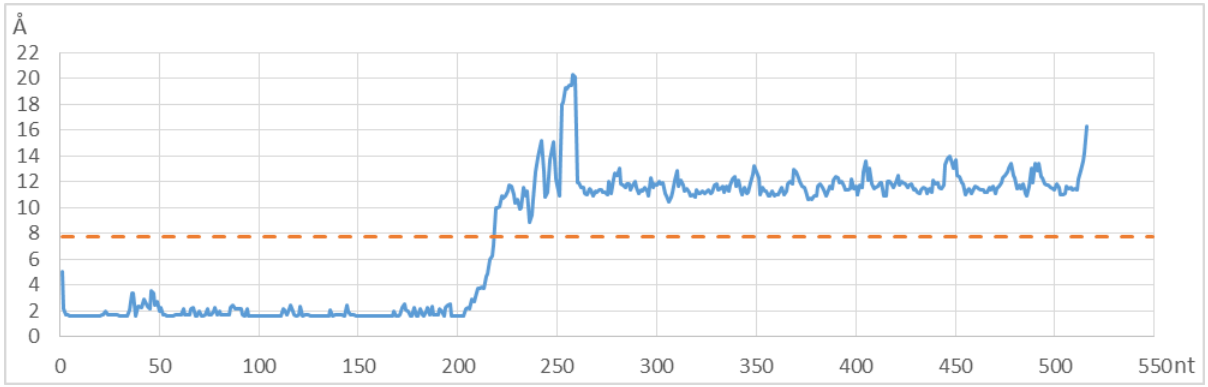
S2-m-1



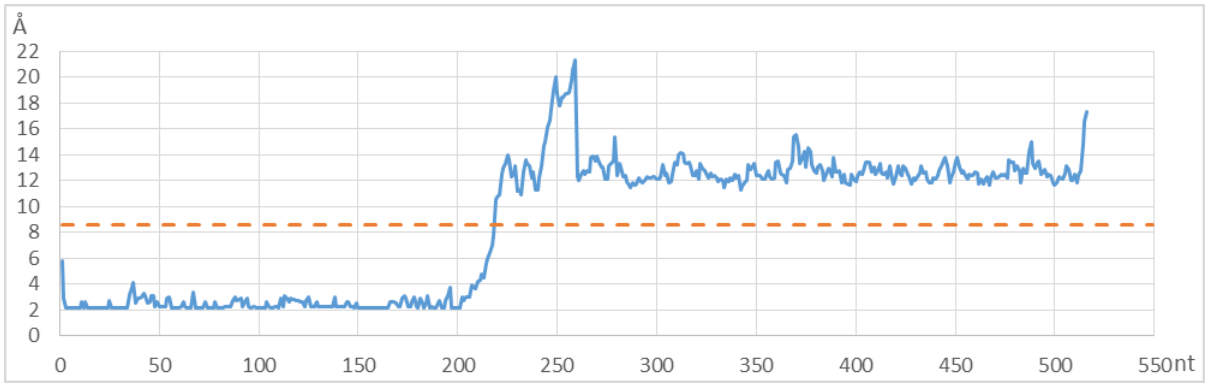
S4 wt



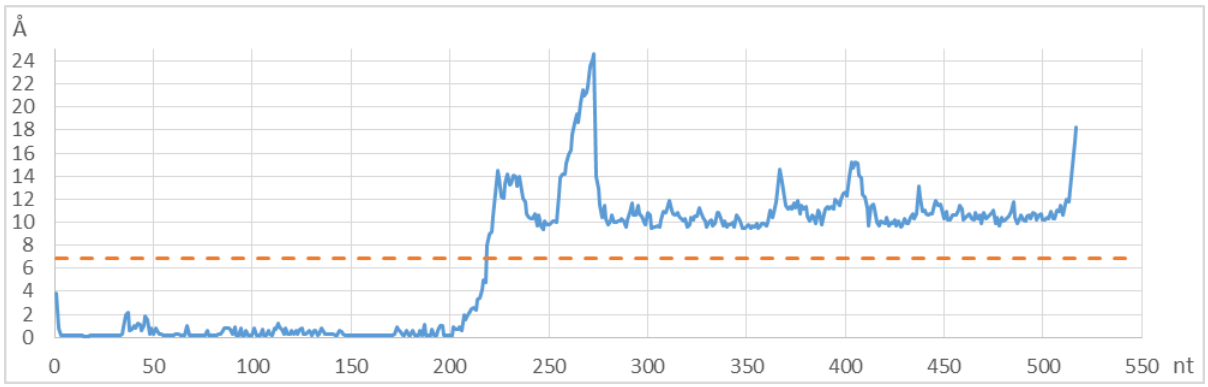
S4-m-1



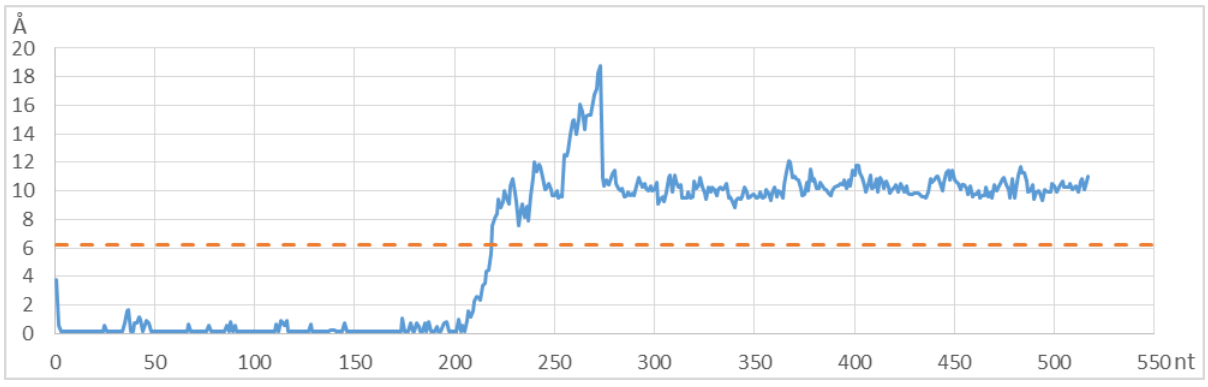
S4-m-2



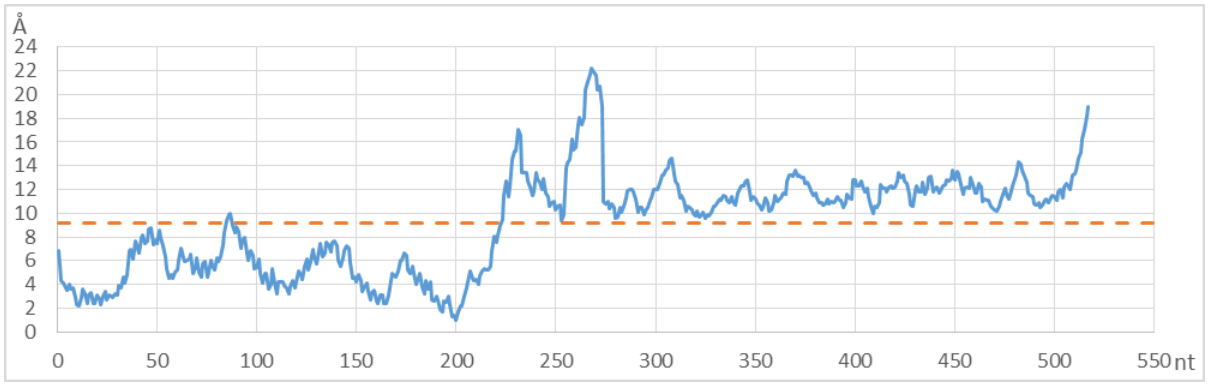
S7 wt



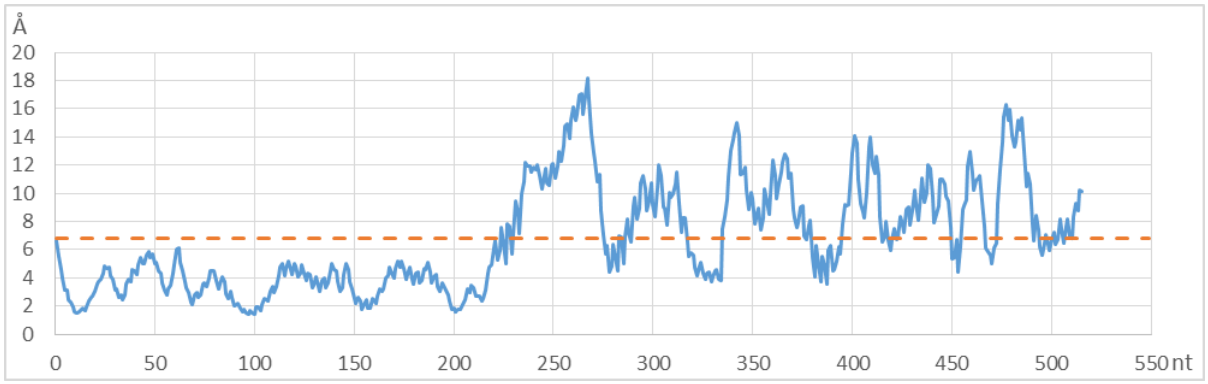
S7-m-1



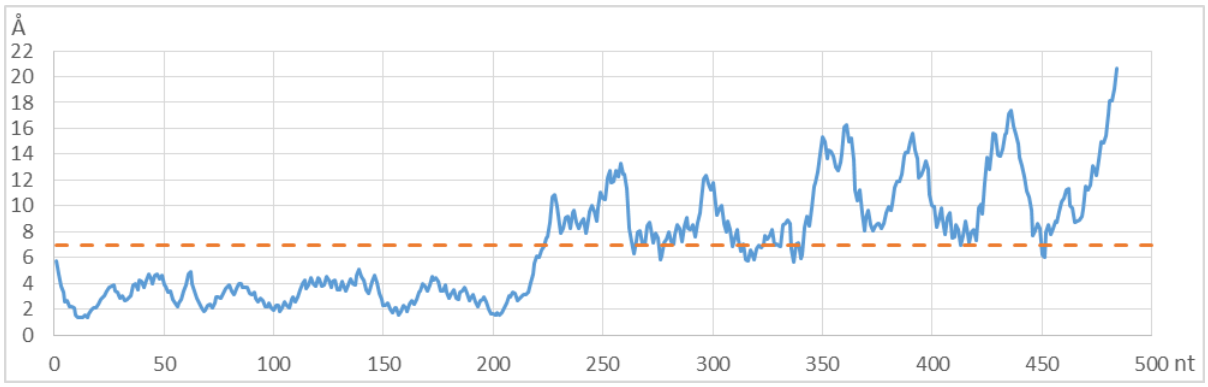
S7-m-2



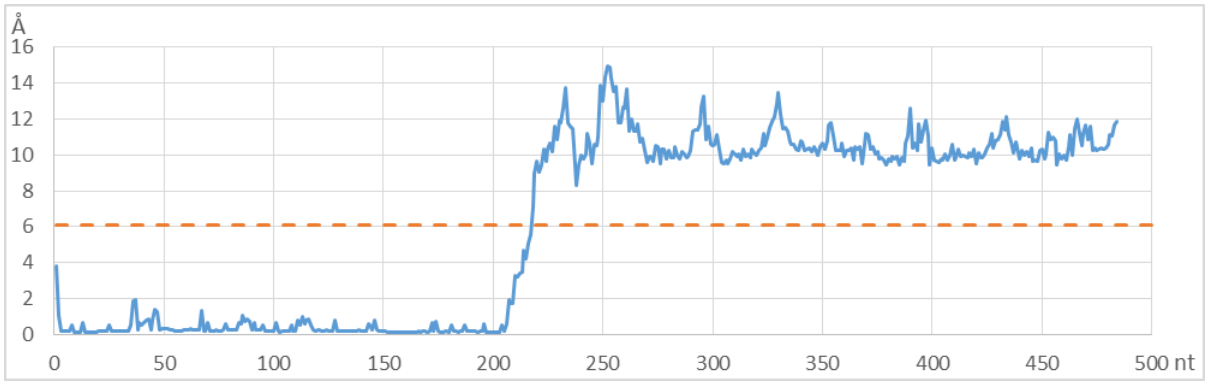
S7-del-1



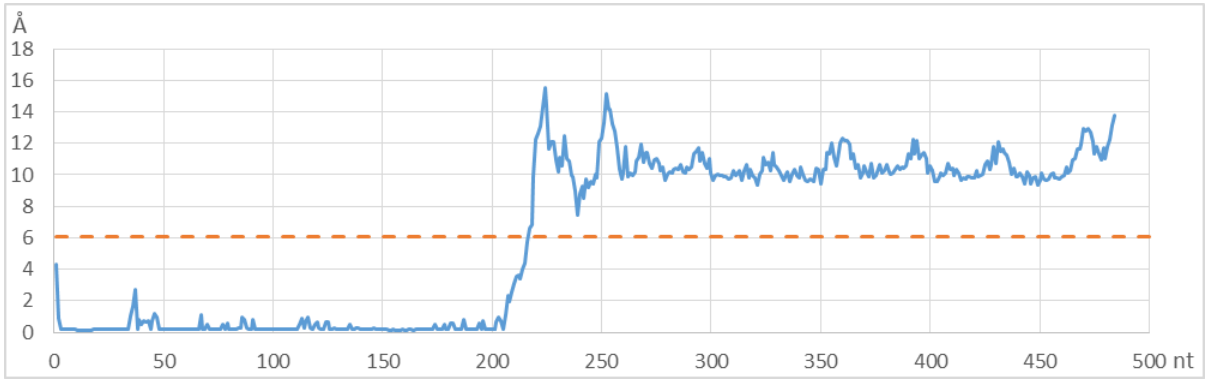
S10 wt



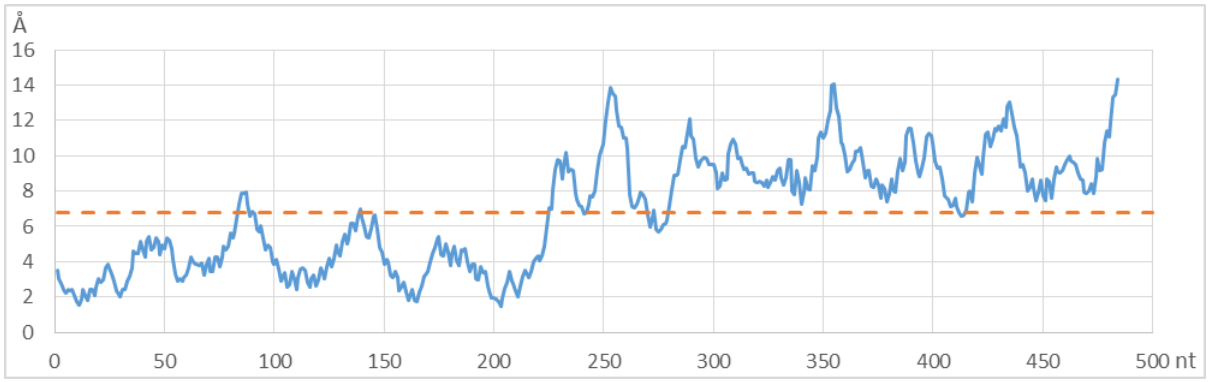
S10-m-2



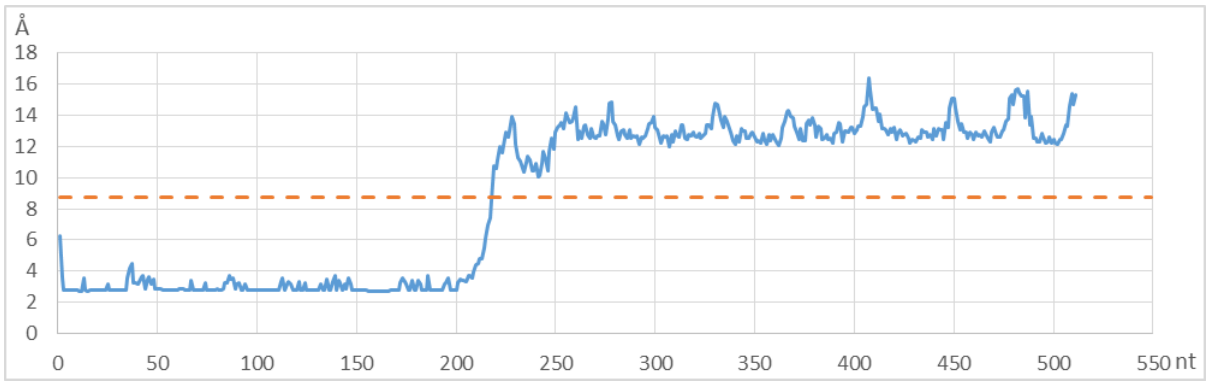
S10-m-3



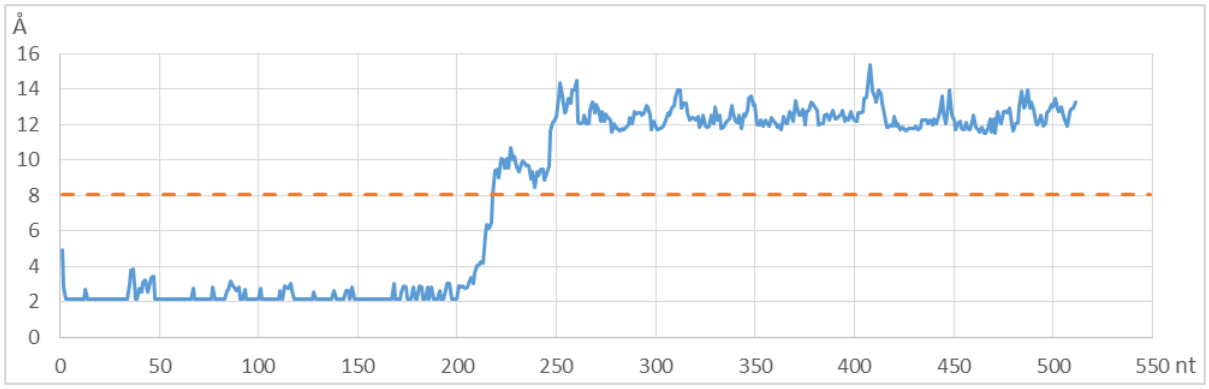
S10-m-2&3



S16 wt

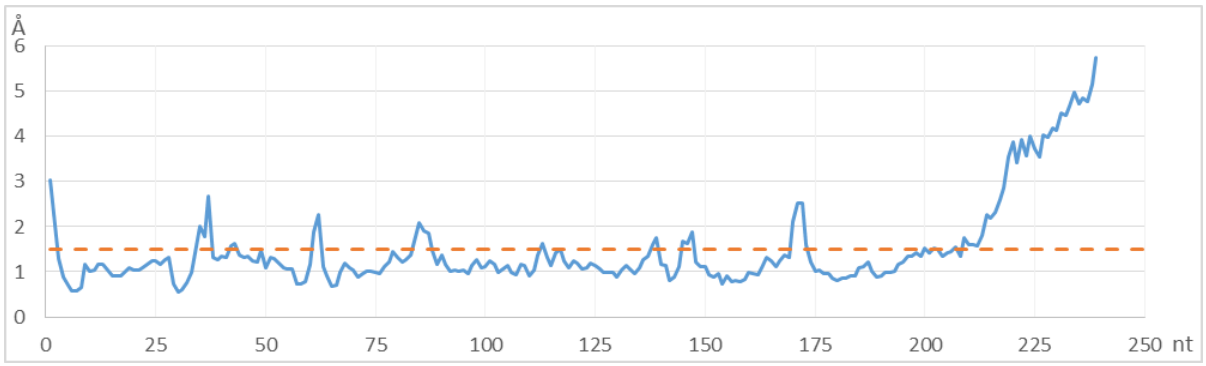


S16-m-1

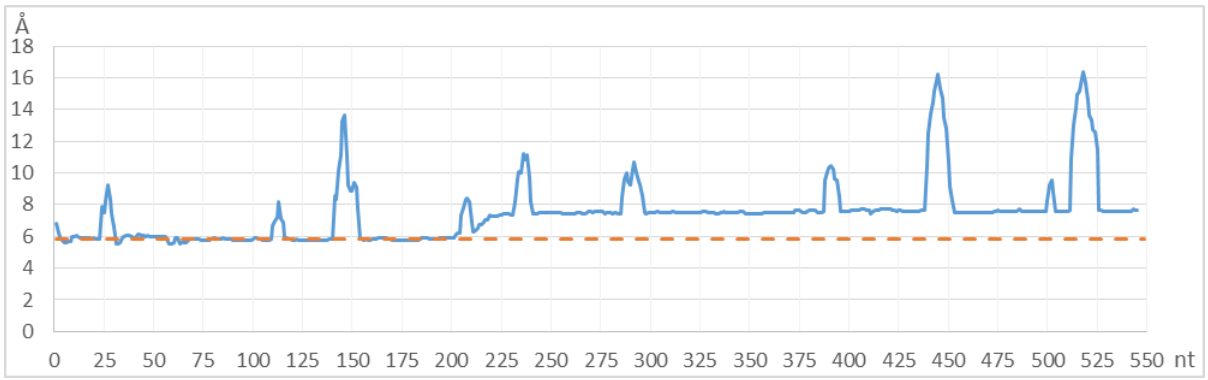


L-scores for I-TASSER server predictions:

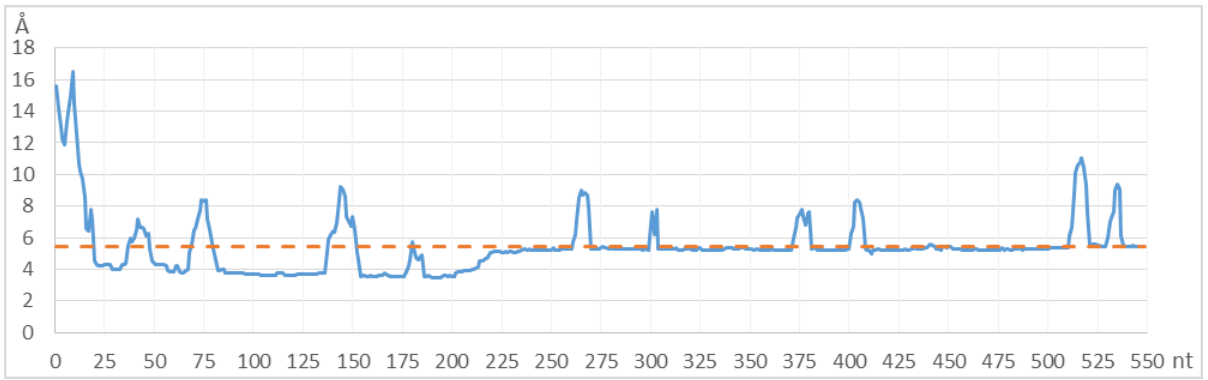
GST



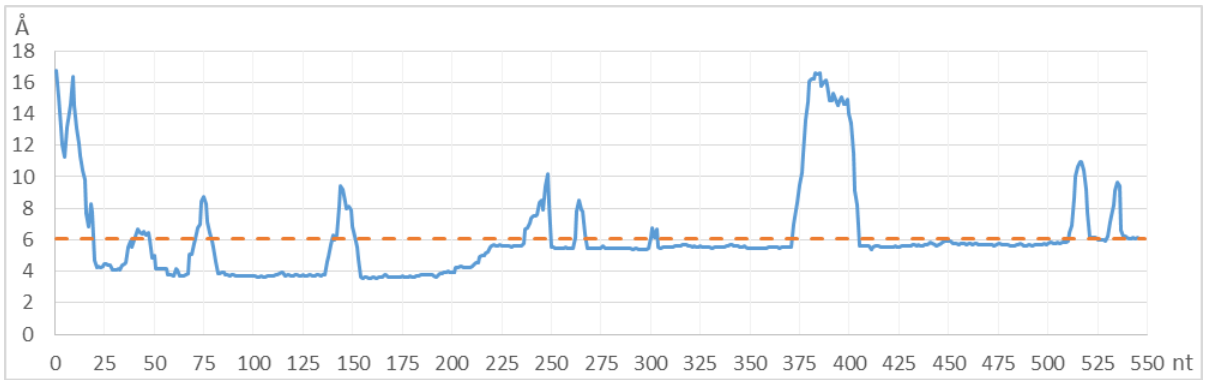
S1 wt



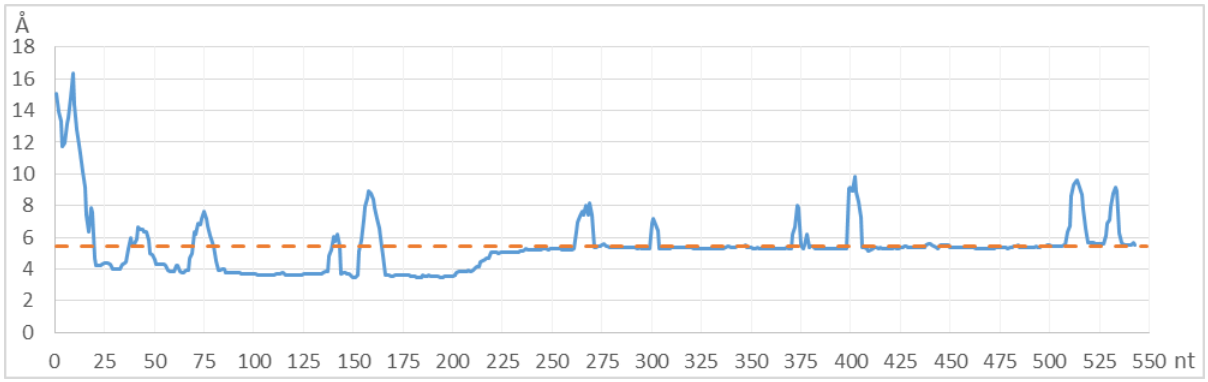
S1-m-1



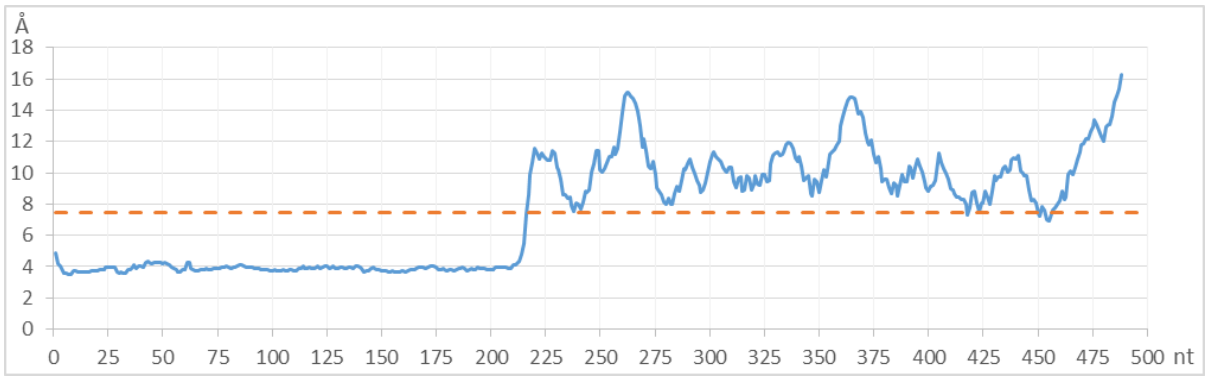
S1-m-2



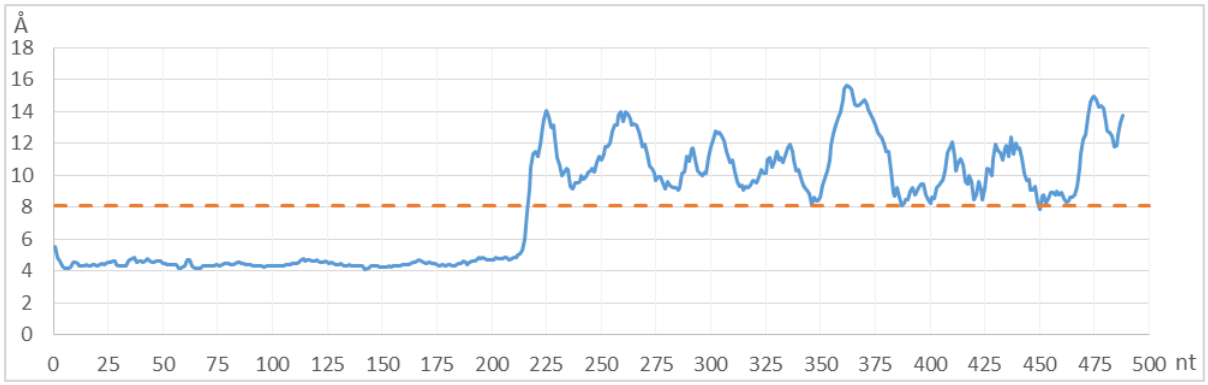
S1-del-1



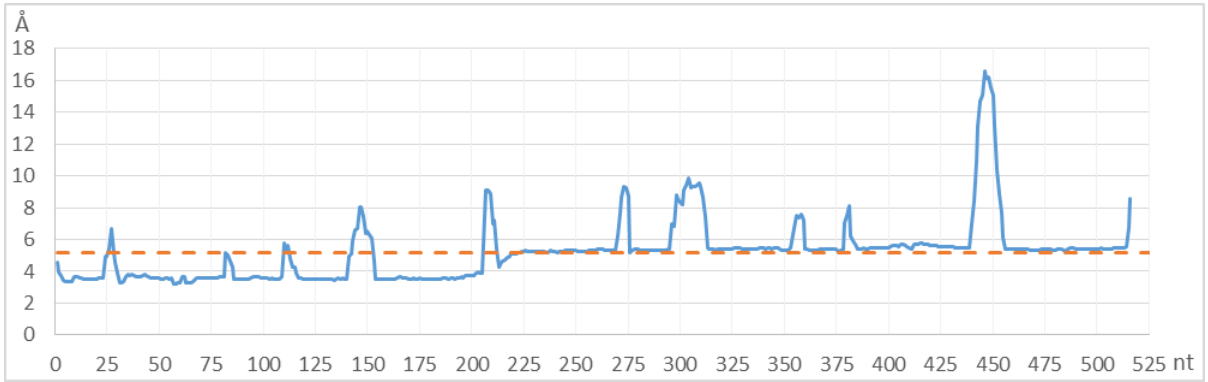
S2 wt



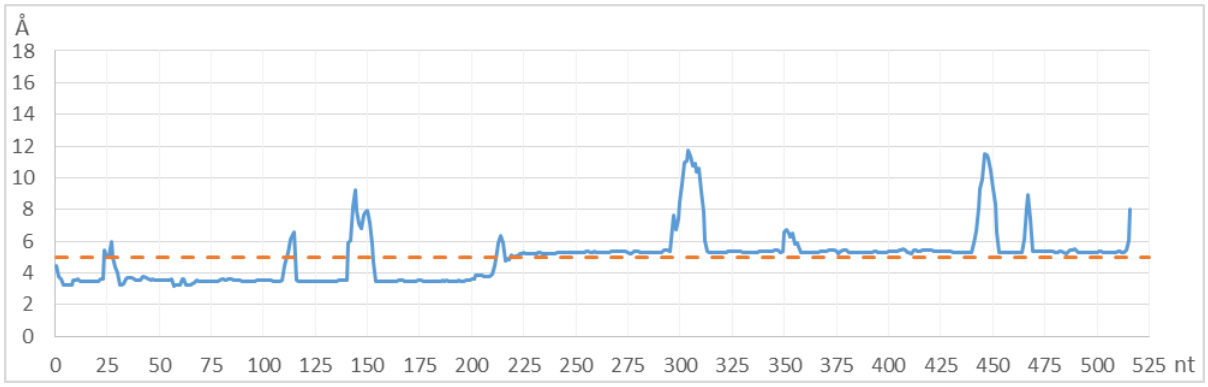
S2-m-1



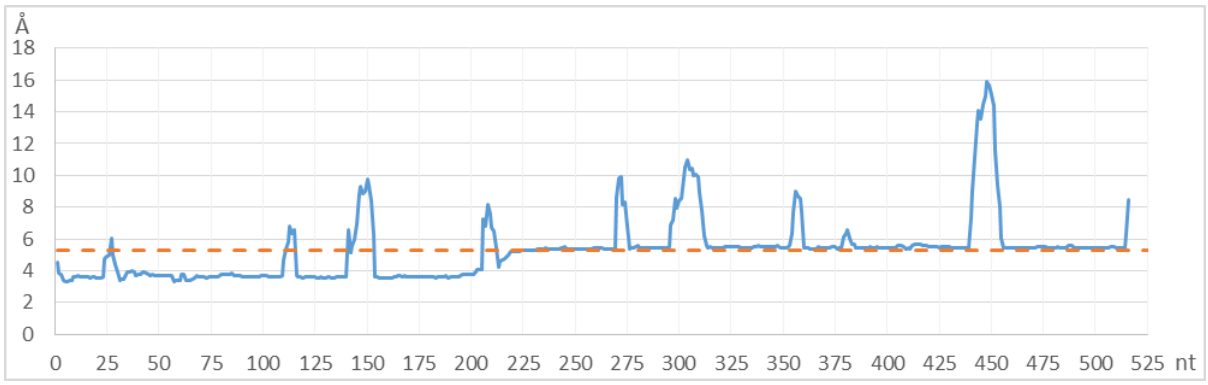
S4 wt



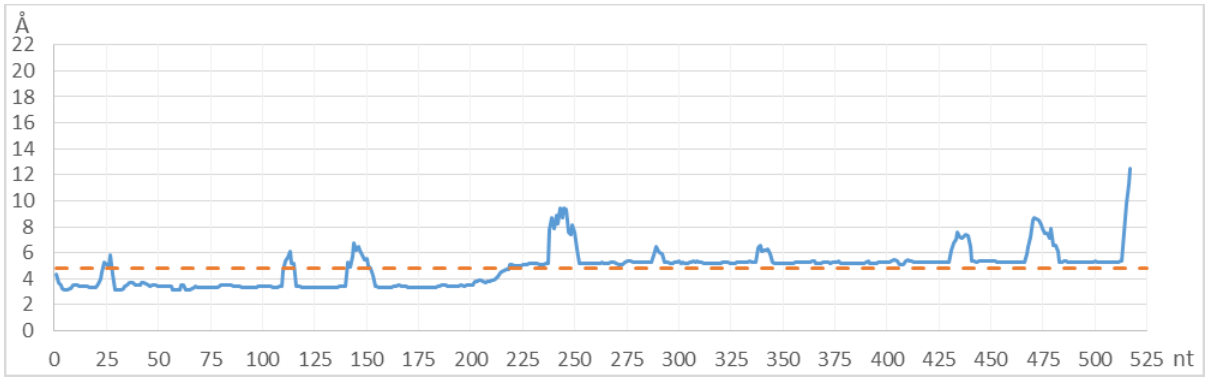
S4-m-1



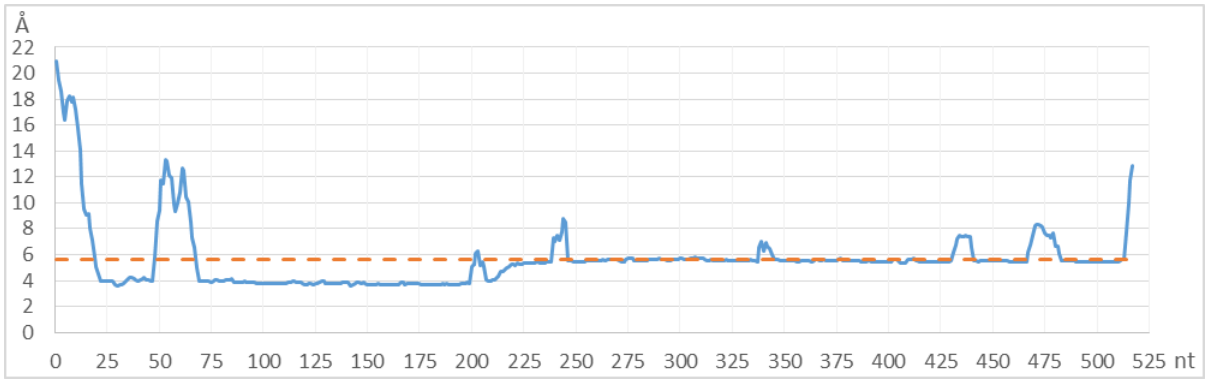
S4-m-2



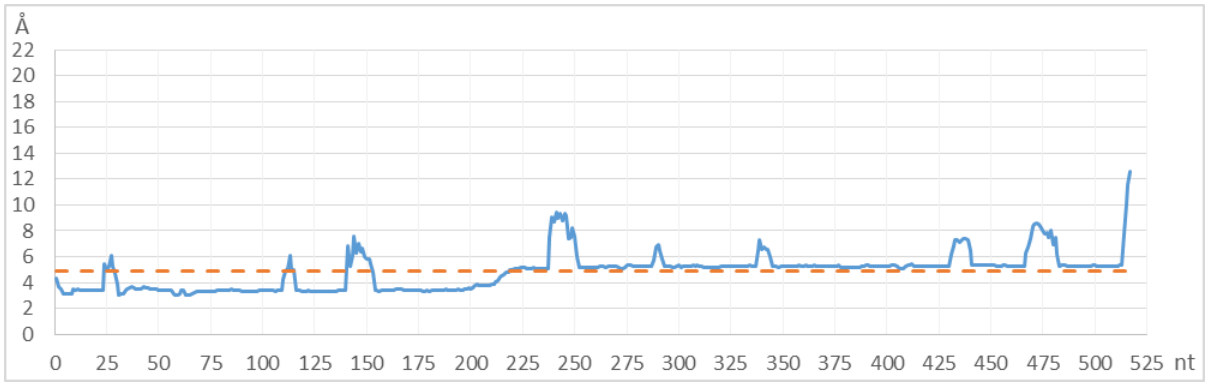
S7 wt



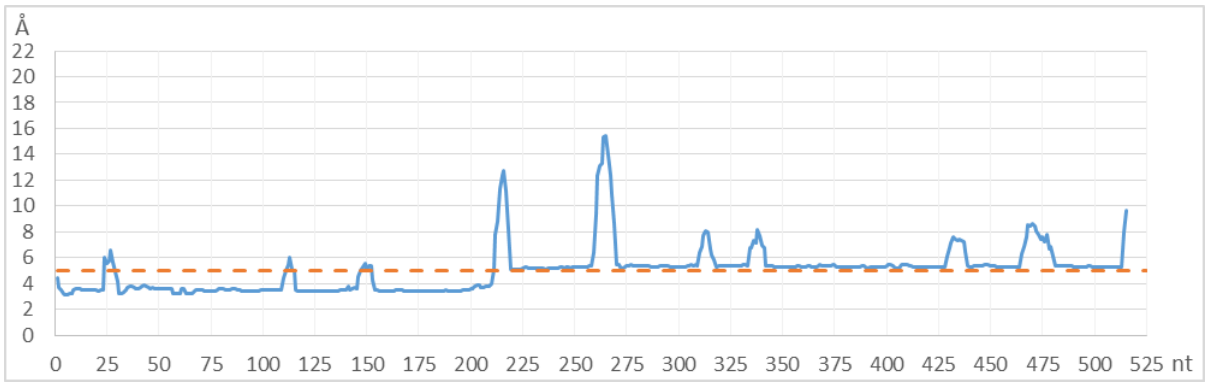
S7-m-1



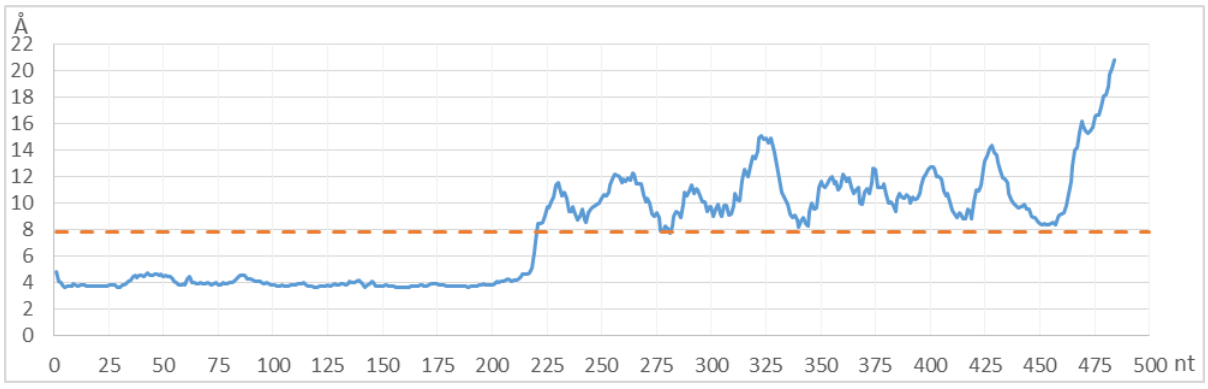
S7-m-2



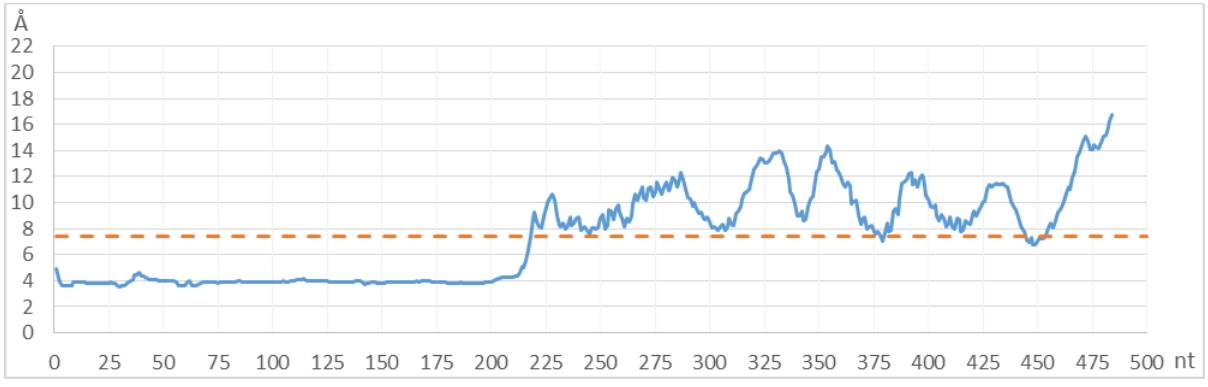
S7-del-1



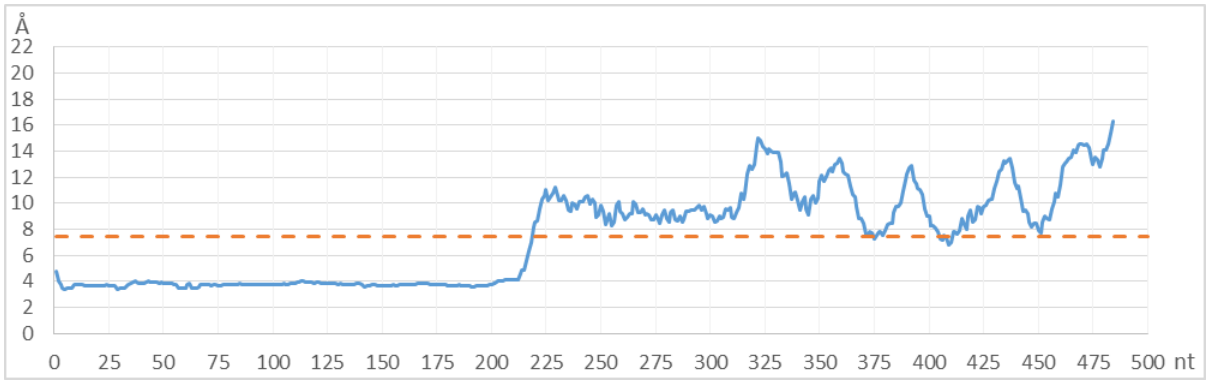
S10 wt



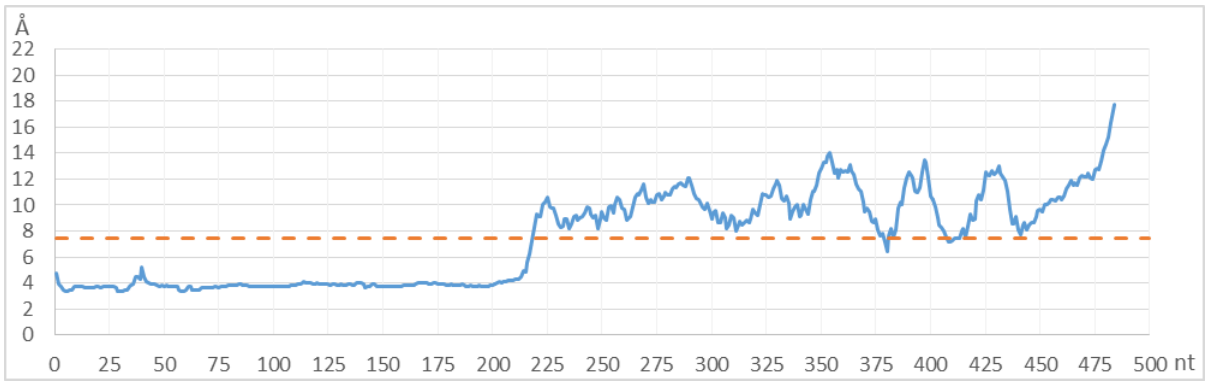
S10-m-2



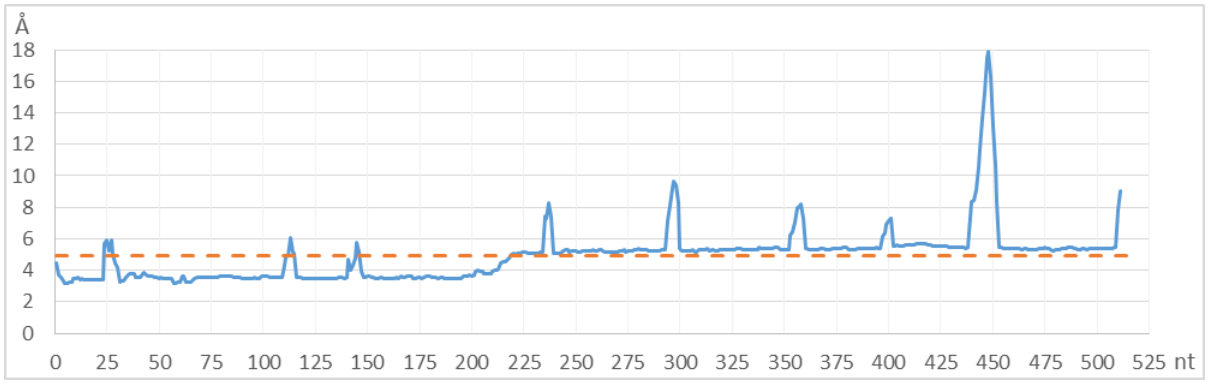
S10-m-3



S10-m-2&3



S16 wt



S16-m-1