Functional protein networks unifying limb girdle muscular dystrophy

Morrée, A. de

Chapter 1
Protein networks unifying Limb Girdle Muscular Dystrophy

Introduction

Limb Girdle Muscular Dystrophy (LGMD) is a rare progressive disorder that mainly affects the skeletal muscles of the pelvic and shoulder girdle. LGMD is characterized by a period of good health until disease onset, which depending on the gene involved can start ranging from early teens, to midlife. After disease manifestation, LGMD progresses and within several years the muscle tissue shows severe wasting and increasing weakness, and patients often end wheelchair bound [32]. While in most cases the proximal muscles are first affected, with time other more distal muscles follow. LGMD is not deadly per se, but in many cases there is an increased risk for cardiac and respiratory failure [32].

Because of these risk factors it is important that a correct diagnosis is reached rapidly [32]. LGMD is genetically heterogeneous and can be transmitted in an autosomal dominant (LGMD1) and autosomal recessive (LGMD2) manner. For many LGMD loci the disease gene has been identified. Strikingly, there are at least 21 different genes that when mutated give rise to an LGMD phenotype (Table 1) [160]. Being this heteregoneous, diagnosis of LGMD is not straightforward. It has been estimated however, that it should be possible to reach a precise diagnosis in roughly 75% of the LGMD patients [32]. Diagnosis is based on clinical presentation of affected muscles, creatine kinase levels, and often western blotting and DNA analysis [32]. An additional complication with LGMD is that the genes involved are often causally linked to other muscle diseases as well. Phenotypes may be partially overlapping, resulting in a risk of misdiagnosis (for examples see below) [32].

With so many genes involved in a single phenotype LGMD is often considered a group of progressive muscle disorders [32,160]. However, while it is certainly possible to regard the distinct subtypes as separate disease entities, the possibility of a common mechanism underlying all forms of LGMD is worthwhile to explore; certainly in the light of a potential intervention strategy. Most LGMD genes are not solely expressed in muscle, yet their phenotype seems to be restricted predominantly to skeletal muscle. For most of the encoded proteins a function has been ascertained (Table 1) [160].

Based on protein function LGMD has been divided into major disease mechanisms: 1) a structural defect at the muscle membrane (Sarcoglycans [221]), 2) muscle membrane repair deficiency (Dysferlin [13]), 3) defects in sarcomere remodeling, cytoskeleton structure and cytoskeleton-membrane interactions (Calpain 3 (CAPN3) [17]) [118]. With the identification of ANO5 mutations in LGMD2L a possible fourth disease mechanism has been uncovered that regards ion channel dysfunction (see discussion, [25]). Closer inspection of these pathogenic mechanisms may give an idea of commonalities between the LGMD types

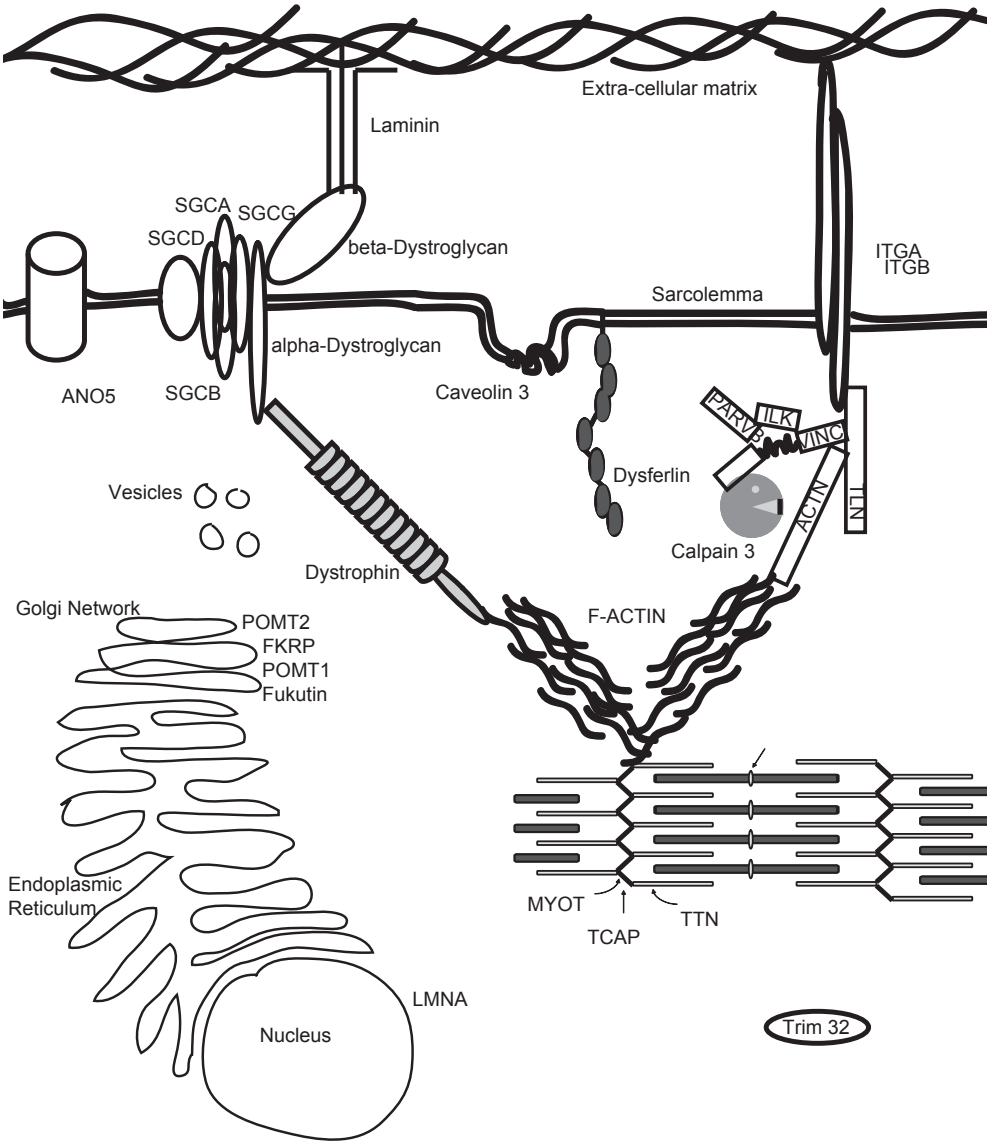

Figure 1: Schematic overview of the Dystrophin-glycoprotein anchoring complex. All identified LGMD proteins are localized in the cartoon.

LGMD and membrane stability

The skeletal muscle fibers provide contractile strength to the body. Such force is generated by the internal sarcomere. At the Z-disk the sarcomere connects to the internal Actin cytoskeleton. The filamentous Actin network is bound by the Dystrophin-glycoprotein complex (DGC) which links to the extracellular matrix and thereby allows for transduction of the mechanical force (Figure 1) [16]. Mutations in partners of the DGC have been identified in various muscle disorders, the most severe and prevalent disorder being Duchenne Muscular Dystrophy, which is caused by mutations in the DMD gene encoding for Dystrophin [115]. Dystrophin localizes to the intercellular side of the sarcolemma. It is a modular rod-like protein with an Actin binding domain at its N-terminus and a large C-terminus, connected by a large number of repeat domains. Dystrophin directly interacts with filamentous Actin and to a large protein complex at the sarcolemma. It thereby provides a molecular link between the internal Actin cytoskeleton and the extracellular matrix.

The central part of the anchoring complex is built by the transmembrane glycoprotein Dystroglycan. Dystroglycan is post-translationally cleaved. The resulting α-Dystroglycan is fully extracellular and interacts with extracellular matrix proteins such as Laminin. These interactions are mediated via glycosyl groups that are enzymatically added to Dystroglycan as it travels through the ER and Golgi. Interestingly, several mutations in the genes encoding of at least four different glycosylation enzymes have been causally linked to LGMD (Table 1). The extracellular α-Dystroglycan stays in complex with the transmembrane β-Dystroglycan. This protein interacts on the cytosolic side of the sarcolemma with Dystrophin.

The anchoring complex is stabilized in the sarcolemma by a group of proteins called Sarcoglycans, together with Sarcospan [221]. These proteins prevent extreme movement of the anchor when much force is generated by the muscle. The four different Sarcoglycans form a tight complex, and loss of one often results in secondary loss of the others, indicating that they depend on each other's presence for correct localization and function [221]. Mutations in all four Sarcoglycan genes have been identified in LGMD [221]. Intriguingly, α-Sarcoglycan contains a putative ATP binding site [221], suggesting additional functionalities for this protein.

LGMD and membrane repair

One of the best studied LGMD genes is Dysferlin (DYSF). The mutational spectrum for Dysferlin is extensive. Mutations are found throughout the gene without apparent hotspots and include nonsense, missense, and splice site mutations

(www.dmd.nl/dysf). Moreover, mutations in the Dysferlin gene can also cause other disease entities, such as Miyoshi Myopathy (MM [165]) and Distal Anterior Compartment Myopathy (DACM [126]. All of them are collectively referred to as "Dysferlinopathies" [160]. Intriguingly, it has been reported that a single Dysferlin mutation resulted in LGMD in one patient and MM in another [255]. The Dysferlinopathies are characterized by a late onset at age 17-25. Proteomic analysis indicated a decrease in glycolytic type I fiber marker proteins and a concomitant increase in oxidative type II marker proteins, suggesting differences in muscle fiber composition [70]. In addition, Dysferlinopathy patients have a higher number of immature fibers than healthy controls [45]. The muscle tissue is characterized by a strong inflammation of mainly monocytes and macrophages [193], and the disease is often misdiagnosed as polymyositis [193], an autoimmune disease of the muscle.

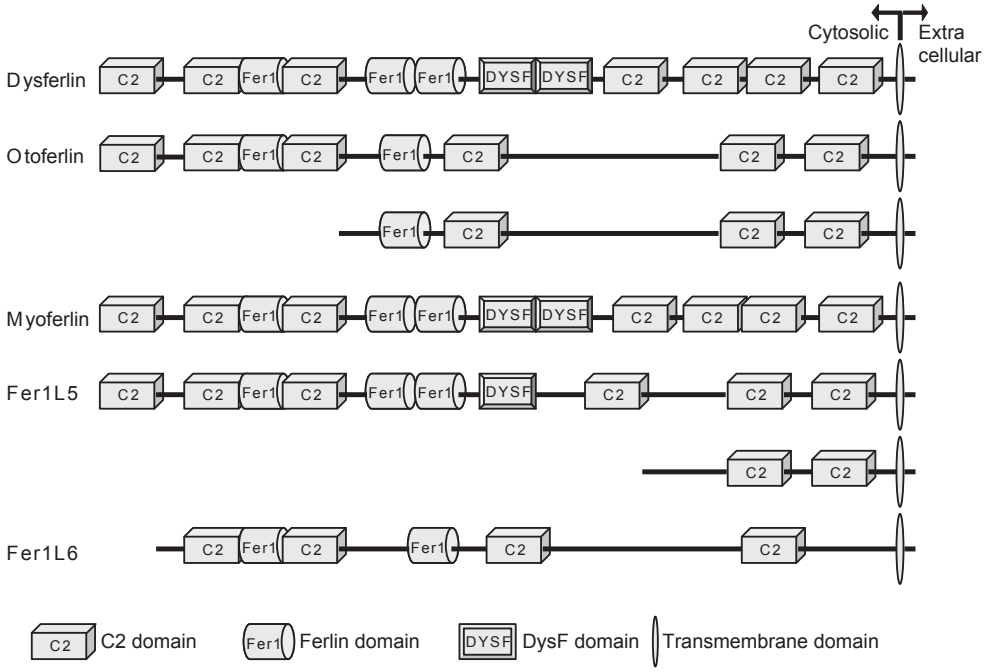

Figure 2: Schematic overview of the Dysferlin protein and its family members. Domain prediction is based on CCD tool (NCBI). The sequence N-terminal of the transmembrane domain is on the cytosolic side of the cell membrane. For Dysferlin, the extracellular sequence is ~11 amino acids only.

The Dysferlin gene encodes a 230 kDa protein that contains a single pass transmembrane domain at its C-terminus, and has seven C2 domains, three ferlin domains and two DysF domains (Figure 2). The larger part of Dysferlin is intracellular. The C2 domains are calcium-sensitive phospholipid-binding domains and are believed to be essential for Dysferlin function [243,244]. For the first C2 domain, C2A, calcium-dependent phospholipid-binding activity has been shown [243]. Through the C2 domains Dysferlin can also interact with other proteins, in certain cases also in a calcium-sensitive manner (AHNAK2 [119]). The function of the Ferlin and DysF domains are not clear. Based on structural experiments it is thought that mutations in this domain cause the protein to misfold [198], and such misfolding might affect the entire protein.

Cellular studies showed that Dysferlin is a vesicular protein that travels to and from the sarcolemma, and responds to changes in intracellular free calcium [13]. When the plasma membrane of cultured myotubes is ruptured, for instance by laser-mediated wounding, Dysferlin is rapidly recruited to the plasma membrane in a calcium-dependent manner, and accumulates at the site of the lesion [13]. This recruitment is essential for patch-fusion repair of the damaged membrane [12,13], and it has been hypothesized that a disturbed membrane repair capacity underlies the different Dysferlinopathies [214].

Dysferlin is part of the Ferlin family of proteins named after the *Caenorhabditis elegans* gene Ferlin (*fer1*) [15]. Ferlin is involved in vesicular function in *C. elegans* spermatids [253]. Mutations in Ferlin result in infertility [253]. Recently it was shown that these worms also have a muscle phenotype [150]. The human genome encodes six Ferlin-like (Fer1L) genes.

Otoferlin (Fer1L2) is expressed in hair cells of the inner ear, and participates in cellular transmission of sound [217]. Mutations in Otoferlin cause an autosomal recessive form of congenital deafness (DFNB9 MIM#601071) [187]. Otoferlin is expressed in a long and a short isoform (Figure 1) [262]. In humans, the long isoform is found in brain, while the short isoform is expressed in cochlea, vestibule and brain [27,262]. In mice, only the long isoform is observed, yet a mutant mouse model recapitulates the hearing impairment [166]. Otoferlin binds to Syntaxin 1a and Snap25 (both through C2F) and is involved in calcium-regulated vesicle trafficking [209]. In addition Otoferlin interacts with the L-type voltage gated calcium channel (via C2D domain) and aids in regulating neurotransmission in hair cells [209]. Interestingly, the C2 domains that have thus far been described in Otoferlin protein-protein interactions, are present in both Otoferlin isoforms.

Myoferlin (Fer1L3) is like Dysferlin expressed in skeletal muscle, but at a lower level [63]. Myoferlin is important for myoblast differentiation, as knockdown impairs the fusion of cultured myoblasts [75,76]. Myoferlin expression increases

with differentiation [75]. While Dysferlin is mostly found at the sarcolemma of mature skeletal muscle fibers, Myoferlin is predominantly found at peri-nuclear membranes [75]. No pathogenic mutations in Myoferilin have yet been described [63]. Knockdown of Myoferlin in endothelial cells attenuated membrane repair [21], suggesting that Myoferlin has a membrane repair function as well and that it might be able to substitute Dysferlin as a potential strategy for therapy. However, while the introduction of human Myoferlin in Dysferlin deficient mice yielded improved the membrane resealing capacity of isolated myoblasts, the general dystrophic phenotype was not rescued [2]. This suggests that Dysferlin has functions beyond membrane repair which cannot be readily substituted by Myoferlin.

The Fer1L4, Fer1L5 and Fer1L6 proteins have not yet been characterized.

Several mouse models have been described that have altered or absent Dysferlin expression [13,22]. These models recapitulate the human phenotype and show a mild progressive muscular dystrophy, and a disturbed membrane repair capability, both *in vivo* and in cell culture [13,22]. A high level (10-fold) of Dysferlin overexpression also causes a muscle pathology [93], different from membrane repair defects, with increased Endoplasmic Reticulum (ER) stress levels. Like Dysferlinopathy patient muscle tissue, Dysferlin deficient mouse models show strong immune infiltrate after massive muscle damage, either through repeated eccentric contractions [215] or through repeated injections with the snake venom notoxin [45,168].

It is believed that infiltration of immune cells in skeletal muscle tissue is required for normal physiology [9,231]. Throughout life and development there is ongoing communication between muscle tissue and immune cells. This communication is not only important for the clearing of pathogens and dead cells, but also seems to be important for maintenance and differentiation of skeletal muscle tissue [231]. When muscle fibers are damaged, immune cells (first neutrophils, and then monocytes/macrophages) are recruited to clean up the damaged cells that are beyond repair [9,231]. This is mediated by M1 pro-inflammatory macrophages. In a second phase of the immune response M2 contra-inflammatory macrophages are needed to stimulate regeneration [9]. It has been shown that soluble factors from such macrophages can activate satellite cells [252].

In the absence of Dysferlin the neutrophils and monocytes appear later in the damaged muscle tissue and stay around longer [45]. Both observations may be of importance to the dystrophic phenotype, and hint at a function of Dysferlin beyond membrane repair. No increased sensitivity of Dysferlin patients to infections has been documented. It could well be that Dysferlin serves a communicative role,

between monocyte and muscle.

Recent studies showed Dysferlin expression in cells of non-muscle lineage [6]. These cells include neuronal cells and monocytes [6]. While it is interesting to speculate that Dysferlin is a membrane repair protein in these cell types as well, this is less likely in the case of monocytes/macrophages. These cells destroy other cells, clear up debris and matrix proteins, and participate in cytokine signaling, and should one be damaged others will be recruited and take its place.

It was shown that monocytes from Dysferlin knockout mice, and Dysferlinopathy patients are more aggressive in phagocytosis assays, indicating altered function [193]. This might contribute to the phenotype, but is unlikely to explain all of the pathology. In recent experiments Dysferlin was introduced in Dysferlin knockout mouse muscle tissue, by transgenesis [185] and by AAV treatment [168]. Both studies reported a complete rescue of the contraction induced phenotype [168,185]. The presence of Dysferlin in monocytes however is indicative of Dysferlin functions other than membrane repair (see Chapter 7 for further reading). And indeed, at later ages the rescued mouse models still developed a mild phenotype [168], indicative of a role of non-muscle Dysferlin in the pathogenicity.

Proteomic studies have shown Dysferlin to interact with a number of proteins, many of which complement the membrane repair function of Dysferlin. Annexins A1 and A2 are ubiquitous calcium-sensitive membrane fusogens that interact with Dysferlin, and presumably aid in patching membrane tears (for further interactions see Chapter 2) [33,162,179]. Trim72 is a redox sensor that acts upstream of Dysferlin and is essential for Dysferlin vesicle nucleation [34,36]. A mutant mouse model for Trim72 has a mild muscular dystrophy [34], but to date no pathogenic mutations have been reported in humans. This is different for the Dysferlin binding partner Caveolin 3 [172]. Mutations in this protein cause LGMD1C [186]. Caveolin 3 is an essential component of caveolae, small invaginations of the membrane [85]. It participates in Dysferlin trafficking and endocytosis. Recent experimental data, showed that Caveolin 3 affects Dysferlin endocytotosis, and that several pathogenic Caveolin 3 mutations impaired this phenomenon [36,110,111]. Both proteins are also found at T-tubuli [142,143], where they may also serve a common role. In LGMD2B loss of Dysferlin often coincides with a secondary loss of Caveolin 3 [251]. Recently, CAPN3 was also identified in the Dysferlin protein complex [120]. Mutations in the cysteine protease CAPN3 cause LGMD2A [211], and like Caveolin 3 [251], CAPN3 is also often lost in LGMD2B patient tissue [7]. Interestingly, mutations in CAPN3 were considered to represent the third pathogenic mechanism in LGMD [118].

LGMD and structural stability

Skeletal muscle is very adaptive and continuously undergoes cycles of degeneration and regeneration. To do this it needs to remodel its cytoskeleton and contractile apparatus. The contractile apparatus is formed around the giant protein Titin (3500 kDa). Titin functions as a molecular ruler along which the contractile proteins are aligned [147]. Upon contraction-induced damage this giant protein complex needs to be disassembled, to remove toxic protein fragments, and to allow for rapid reassembly [17]. It is believed that the cysteine protease CAPN3 is centrally involved in this process.

CAPN3 is a member of the Calpain family of cysteine proteases which participate in limited proteolysis in response to free calcium levels [139]. It is the only disease causing member of the family, and its expression is limited to muscle [139]. Contrary to its ubiquitous family members it is non-responsive to the endogenous inhibitor Calpastatin (which in fact is a substrate [195]), and it is extremely sensitive to fluctuations in intracellular free calcium levels (within nanomolar range [191,195]). CAPN3 is an unstable protease that autolyses upon activation. Its estimated *in vitro* half-life is less than ten minutes [138] and consequently not much is known of its substrates, or function. Within skeletal muscle most of the CAPN3 is found in its full-length non-autolysed form of 94 kDa [139]. This form is believed to be proteolytically inactive [72]. It localizes for 90% to the sarcomere [17], and directly interacts with Titin [107,195]. It is hypothesized that the Titin interaction is used to store inactive CAPN3, to allow for local proteolysis upon activation [107,138]. Interestingly, mutations in Titin at the CAPN3 binding site also cause LGMD [104]. Elegant proof for this model comes from experiments with transgenic mice. Mice that overexpress human CAPN3 were crossed with mice that had mutations in the CAPN3 binding domain in Titin [122,195]. While the CAPN3 transgenic mouse itself had no phenotype, the Titin mutant mouse showed a mild muscular dystrophy. Crossing both mouse models aggravated the muscle phenotype, strongly indicating the importance of Titin in buffering CAPN3 activity [122,195].

Several CAPN3 transgenic mouse models have been described, including knock-out and overexpression (of wild-type, proteolyticaly inactive and constitutively active CAPN3) mice [14,152,153,237,240]. Unlike Dysferlin, overexpression of CAPN3 is without apparent phenotype [237], suggesting that the muscle can cope with extra copies of CAPN3. It has been estimated that the number of CAPN3 binding sites on Titin greatly exceeds the amount of wild-type CAPN3 protein molecules, possibly explaining the large buffering capacity of muscle for CAPN3 and the absence of an overt phenotype [17]. All the other models however, show a mild muscular dystrophy phenotype. A first direct clue for the function of CAPN3

was found in hind-limb suspension experiments [153]. Where wild-type muscle showed strong muscle atrophy, degeneration and sarcomere remodeling during the experiment, and clear regeneration thereafter, this process was impaired in CAPN3 deficient muscle [153]. Therefore, CAPN3 is considered to be essential for muscle atrophy and possibly degeneration and regeneration. In agreement with this, the very few substrates reported are all structural proteins [98,242], and ectopic overexpression of CAPN3 causes cell rounding and detachment, indicative of cytoskeleton remodeling [242] (see Chapter 3 and 4 for further reading).

Over 400 pathogenic CAPN3 mutations have been described and roughly one third of these are miss sense (www.dmd.nl/capn3). There is no clear mutational hotspot [151]. It has been estimated that roughly one third of all mutations does not impair the proteolytic activity [184]. Modeling experiments on the experimentally determined structure of Calpain 2 indicated that several mutations affect the autolytic activity of CAPN3, showing that a change in the activity time span might already be deleterious [87,135]. This resulted in a model where locally stored inactive CAPN3 (on Titin molecules [107]) allows for local proteolysis upon activation [17]. However, the activation signal remains elusive and biochemical support for this hypothesis is wanting [17] (see Chapter 4 for further reading).

Due to the identification of CAPN3 in the Dysferlin protein complex, the hypothesis was put forward that CAPN3 functions in membrane repair [118]. It has recently been shown that ubiquitous Calpains, Calpain 1 and 2, are essential components of the non-muscle membrane repair system [183]. However, membrane repair assays on CAPN3 deficient primary myoblasts showed no clear difference in membrane repair capacity [182], indicating that the link between LGMD2A and 2B is not that straightforward.

Linking LGMD proteins

In all forms of LGMD muscle tissue is correctly formed and functional. Concomitantly, in all three above described pathogenic mechanisms deregulation of muscle maintenance seems to be central, rather than muscle development. In all forms of LGMD an increased level of regeneration has been observed. This suggests that when muscle maintenance processes fail, damaged muscle fibers need to be replaced. Muscle has a high capacity for regeneration, but it is not unlimited [56,57].

Muscle tissue contains a large number of satellite cells: a self renewing population of mononucleated cells that can differentiate into myoblasts, and thus give rise to new myofibers [56,57]. The self renewing capacity of these cells is a limiting factor in muscle regeneration. A possible model for LGMD could be that fibers that fail in proper maintenance are removed and replaced as long as

possible. This stress on the satellite cell population results in a quicker loss of regenerative capacity [31]. When regeneration can no longer match the loss of existing fibers, and the rate of myofiber loss exceeds the rate of replenishment, LGMD disease onset would eventually follow.

In line with this model the earliest onset of disease is described for CAPN3 (LGMD2A) patients [28]. Extrapolating from the mouse models[152,153], these patients likely have an impaired muscle degeneration and regeneration capacity, making them especially prone to defects in muscle maintenance. Dysferlin patients (LGMD2B) can still regenerate their myofibers. Onset of LGMD2B is thus later than for CAPN3 patients.

To maintain correct muscle function specific molecular signaling pathways exist. Pathways that mechanically or chemically sense contraction and resulting damage, and that need to respond appropriately. The molecular interactions between Dysferlin and Caveolin 3 [172] and CAPN3 [120] provide a first indication for a connecting molecular network that underlies muscle maintenance and LGMD pathogenicity. Maintenance processes require rapid and clear-cut communication at a molecular level. An intriguing protein in such network might be AHNAK, which has been implicated in many different processes that are important to muscle, ranging from membrane repair, vesicle trafficking, excitation-contraction coupling, and (neuronal) remodeling [3,19,52,53,90,100,106,118,119,146,175,176,206,220,222] (for further reading on AHNAK see Chapters 3, 5 and 6). AHNAK consists of a large body of repeat units that fold into a B-propellor [146], and a N- and C-terminus that are intrinsically disordered [113,146,227]. Given the large number of hitherto described ANAK binding partners (>10) [5,99,177], and its partially disordered folding, we propose that AHNAK functions as a Hub protein, which are important central components of protein-protein networks [109].

We hypothesized that unraveling further molecular communications would provide access to a potential protein network that underlies and unifies all forms of LGMD. In Chapter 2 therefore a robust and reproducible method for uncovering protein-protein interactions is described for the LGMD protein Dysferlin. The uncovered interactions provide a new basis for extending the molecular networks in muscle maintenance and LGMD pathogenicity.

Table 1					
Disease	**MIM**	**Gene**	**Protein**	**Expression**	**Function**
LGMD1A	#159000	*MYOT*	Myotillin	Skeletal muscle, heart *	Sarcomere component
LGMD1B	#159001	*LMNA*	Lamin A/C	Ubiquitous	Nuclear lamina component
LGMD1C	#607801	*CAV3*	Caveolin 3	Skeletal muscle, heart, brain, adrenal gland	Essential component of caveolae
LGMD1D	%603511	*unknown*	Unknown		
LGMD1E	%602067	*Unknown*	Unknown		
LGMD1F	%608423	*Unknown*	Unknown		
LGMD1G	%609115	*Unknown*	Unknown		
LGMD2A	#253600	*CAPN3*	CAPN3	Skeletal muscle, heart *	Proteolytic regulation of the cytoskeleton
LGMD2B	#253601	*DYSF*	Dysferlin	Skeletal muscle, heart, monocytes, kidney, trophoblast	Membrane repair
LGMD2C	#253700	*SGCG*	γ-sarcoglycan	Skeletal muscle, heart, monocytes, mesenchym	Dystrophin glycoprotein complex stability
LGMD2D	#608099	*SGCA*	α-sarcoglycan	Skeletal muscle, heart, monocytes	Dystrophin glycoprotein complex stability
LGMD2E	#604286	*SGCB*	β-sarcoglycan	Skeletal muscle, heart	Dystrophin glycoprotein complex stability
LGMD2F	#601287	*SGCD*	δ-sarcoglycan	Skeletal muscle, heart, intestine	Dystrophin glycoprotein complex stability
LGMD2G	#601954	*TCAP*	Theletonin	Skeletal muscle and heart	Sarcomere component
LGMD2H	#254110	*TRIM32*	Trim32	Skeletal muscle, brain *	E3 ubiquitin ligase and microRNA regulator

LGMD2I	#607155	*FKRP*	Fukutin-Related Protein	Skeletal muscle, heart, thymus, kidney, placenta *	Glycosylation enzyme
LGMD2J	#608807	*TTN*	Titin	Skeletal muscle, heart, trophoblast	Sarcomere component
LGMD2K	#609308	*POMT1*	Protein O-Mannosyl transferase 1	Skeletal muscle, heart, cerebellum, lymph node, placenta *	Glycosylation enzyme
LGMD2L	#611307	*ANO5*	Tmem16	Skeletal muscle, heart, brain*	Calcium-dependent Chloride channel
LGMD2M	#611588	*FKTN*	Fukutin	Heart, intestine, trophoblast	Glycosylation enzyme
LGMD2N	*607439	*POMT2*	Protein O-Mannosyl transferase 2	Adrenal gland, thyroid gland, monocytes, kidney	Glycosylation enzyme

Table 1: Overview of LGMD subtypes. For each disease variant the protein and gene identifier is given, together with the MIM number, expression summary and protein function. The expression profile is based on proteinatlas, or when marked with * genecards. Protein function was gathered from PubMed references.