Solution structure of the inner DysF domain of myoferlin and implications for limb girdle muscular dystrophy type 2b

Journal Article

Citation:

Patel, P.; Harris, R.; Geddes, S.M.; Strehle, E.M.; Watson, J.D.; Bashir, R.; Bushby, K.; Driscoll, P.C.; Keep, N.H. (2008)
Solution structure of the inner DysF domain of myoferlin and implications for limb girdle muscular dystrophy type 2b
Journal of Molecular Biology 379 (5), pp. 981-990

Solution structure of the inner DysF domain of myoferlin and implications for Limb Girdle Muscular Dystrophy Type 2B

Pryank Patel[1], Richard Harris[2], Stella M. Geddes[1], Eugen-Matthias Strehle[1,3], James D. Watson[4], Rumaisa Bashir[5], Katharine Bushby[3], Paul C. Driscoll[2], and Nicholas H. Keep[1],

1. Institute for Structural Molecular Biology and School of Crystallography, Birkbeck University of London, London WC1E 7HX UK

2. Institute for Structural Molecular Biology and Department of Biochemistry and Molecular Biology, University College London, London WC1E 6BT UK

3. Institute of Human Genetics, University of Newcastle, Newcastle-upon-Tyne NE1 3BZ UK

4. European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD,UK

5. School of Biological and Biomedical Sciences, South Road, Durham, DH1 3LE UK

Contact Nicholas Keep email n.keep@mail.cryst.bbk.ac.uk

Telephone 020-7631-6852 Fax 020-7631-6803

Running title

Structure of Inner DysF domain of myoferlin

Summary

Mutations in the protein dysferlin, a member of the ferlin family, lead to limb girdle muscular dystrophy type 2B and Myoshi myopathy. The ferlins are large proteins characterised by multiple C2 domains and a single C-terminal membrane spanning helix. However, there is sequence conservation in some of the ferlin family in regions outside the C2 domains. In one annotation of the domain structure of these proteins, an unusual internal duplication event has been noted where a putative domain is inserted in between the N and C terminal parts of a homologous domain. This domain is known as the DysF domain. Here we present the solution structure of the inner DysF domain of the dysferlin paralogue myoferlin, which has a unique fold held together by stacking of arginine and tryptophans, mutations in which lead to clinical disease in dysferlin.

Introduction

Limb girdle muscular dystrophy (LGMD) describes a range of progressive muscle wasting diseases, which begin with the arms and pelvic girdle and are subclassified based on the underlying genetic defect. Miyoshi myopathy [1] is an autosomal recessive muscle wasting disease that begins in the distal, posterior leg muscles. The dysferlin gene was identified by reverse genetics as the locus of mutations that lead to both autosomal recessive LGMD type 2B [2] and Myoshi myopathy [3]. Dysferlin was so named as it showed homology to the *C. elegans* gene *fer1*. Mutations in fer1 cause defects in spermatoza motility due to a failure of the immature spermatid to fuse with membranous organelles [4].

The ferlin family of proteins is characterised by a number of C2 domains, normally labelled A, B etc. starting from the N-terminus, and a C-terminal membrane spanning helix. C2 domains were originally characterised as the second homologous region amongst isoforms of Ca^{2+}-dependent protein kinase C and were subsequently identified in a wide range of proteins (for a review see [5]). C2 domains consist of around 130 residues, which form two four-stranded anti-parallel β sheets, with two different topologies known [6; 7]. C2 domains classically bind phospholipids in a Ca^{2+}-dependent manner, although some domains have been reported to bind to phospholipids independent of calcium ions and others have not had any binding properties demonstrated for them [8]. C2 domains also form protein-protein interactions and have been shown to have more than one interaction interface [9].

In humans the ferlin family consists of three characterised proteins; dysferlin, myoferlin [10] and otoferlin [11]. Myoferlin has 56% sequence identity to dysferlin and is

highly expressed in early differentiating myoblasts, in contrast to dysferlin, which is most highly expressed in mature myotubes [12; 13]. Dysferlin knockout mice show defects in Ca^{2+}-dependent sarcolemmal membrane repair and accumulate vesicles at the sarcolemma [14]. Microarray analysis of dysferlin deficient mice identified annexins as potential interaction partners and the interaction was confirmed experimentally [15]. Myoferlin knockout mice show reduced muscle size and smaller myofibres, although there is some increase in dysferlin to compensate [13]. Conversely myoferlin is upregulated in dystrophic and regenerating muscle and myoferlin and dysferlin can partially compensate for each other. Otoferlin is mutated in a non-syndromic form of deafness [11]. Otoferlin interacts with SNAP25 and SNARE proteins and probably acts as the calcium sensor that triggers membrane fusion at auditory cell synapses [16].

Domain annotation of the various ferlin family members varies significantly between the two main sequence based annotation databases Pfam [17] and SMART [18], although they do cross-refer to domains annotated by the other (Figure 1). The databases do not agree on the number of C2 domains. There are probably seven C2 domains in both myoferlin and dysferlin but both the 5^{th} and 7^{th} domains are only weakly predicted by both databases with E values > 0.001. In contrast the C2A domains of both dysferlin and myoferlin are classical C2 domains. The C2A domain of dysferlin has been shown experimentally to be a Ca^{2+}-dependent phospolipid binding domain[12], and the structure of the C2A domain of myoferlin (PDB 2dmh, unpublished) has regular type II topology (similar to phospholipase Cδ).

Outside the C2 domains Pfam predicts, for both myoferlin and dysferlin, a FerI domain that lies between the C2B and C2C domains, and FerA and FerB, which lie in the long stretch between C2C and C2D. SMART on the other hand does not itself contain these three predicted domains, but does propose that dysferlin, myoferlin and fer-1 contains two copies of the DysF domain. The DysF domains in dysferlin, myoferlin and fer1 exist as an internal duplication; an inner DysF domain is surrounded by the N-terminal and C-terminal (DysFN and DysFC) sections of a second outer domain (Fig. 1). The DysF domain was first reported in *Drosophila* [19]. Gene CG6468 was examined as it has multiple repeats of a β-propeller like domain. Much of the remainder of the protein was then found to be homologous to a region of dysferlin. However the internal duplication of the DysF domain is only seen in the ferlins and not in the β-propeller like proteins (which are found in a range of eukaryotes). SMART also finds the DysF domain in some yeast peroxisomal proteins that Pfam annotates with the much larger Pex24p domain of unknown function found exclusively in fungi. The DysF domain is not however found in otoferlin, and so its role is not general to all ferlin family members.

It is therefore of interest to ascertain whether the DysF domain of the ferlins is a structured domain, particularly as a number of disease causing mutations map to this region [20]. We present here the solution structure of the inner DysF domain of myoferlin, the first structure of a DysF domain, which has a novel fold.

Results

Structure of the DysF domain

The solution structure of human myoferlin residues 923-1040, with an additional 5 amino acids (IDPFT) arising from the vector on the N-terminus, has been solved. The 20 lowest energy conformers compatible with 2810 distance restraints, 87 one bond N-H residual dipolar coupling restraints, 48 hydrogen bond restraints and 109 dihedral angle restraints (see Table 1) have been deposited in the Protein Data Bank **(nxxx)** and the restraints in the BMRB (**BMRB**). The rmsd for these structures is 0.74 Å for backbone atoms and 1.14 Å for all atoms (Fig. 2a)

The structure of the DysF domain consists largely of two long β strands (residues 929-939 and 1018-1028) orientated antiparallel to each other. The intervening 77 residues form a long loop, which pack against both sides of the central beta hairpin. Several regions of this loop form short stretches of β strand which extend the central hairpin to a broken four-stranded sheet (948-954, 974-975, 981-982, 994-996) in an anti-parallel manner (Fig.2, Fig 3, Fig 4). The region annotated by SMART as DysFN supplies the first strand and around two thirds of the loop (red in Fig.2) and the region annotated as DysFC provides a third of the loop and the second strand. The point of insertion of the inner DysF domain in the gene duplication is between secondary structure elements. This means that both DysF domains can adopt the same fold, with the inner domain emerging from the loop half way along the beta sheet of the outer domain. It is not known whether the two domains are in fixed orientation or flexibly connected. The homologous inner DysF construct from human dysferlin (residues 935-1067) showed a dispersed HSQC (data not shown) but a full set of 3D

experiments was not collectable due to an apparently longer correlation time than would be expected for a protein of this size. This indicates that the myoferlin and dysferlin DysF domains are folded. They would be expected to be similar in structure as there is 57% sequence identity 72% sequence similarity between the inner DysF domains of dysferlin and myoferlin.

Neither Secondary Structure Matching (SSM) [21] nor DALI [22] find significant matches (highest Z score SSM 4.8 and DALI 4.3). A number of proteins have antiparallel strands of roughly the same length and orientation but the connection between the two strands in DysF seems to be unique.

The electrostatic surface of the myoferlin inner DysF domain is largely negatively charged (Fig 5) with one positively charged patch corresponding to the region between the side chains of K1004 and K1013 (region 1), a small slightly positively charged patch due to the side chain of K1029 (region 2) and a neutral strip running up one face with a slight positive charge corresponding to the side chain of R1021 (region 2). The cavities on the domain are small. Using SURFNET [23] on the core domain (residues with over 20 NOES per residue 927-1031) the largest gap region is only 612 $Å^3$. Although formally large enough to hold 20-30 atoms, normally cavities are much larger than the ligands. Only 12.5% of ligands of 7 atoms or more occupied a cleft of less than 1000 $Å^3$ [24], indicating that any molecule that binds to these sites is likely to be small maybe only an ion. There are three electrostatically negative cavities of over 550-650 $Å^3$ close to S939/T1017 (Region 4), D932/E933/D955 (Region 5) and D951/D959/K1004/R1020 (Region 6). The conservation is particularly high in regions 5 and 6 (over 75% of residues are present in 60% or more

of DysF domain sequences). Given the small size and the electrostatics, these could be metal ion binding sites although the C2 domains have been presumed to be the calcium sensors in membrane repair. The His-Tag cleaved protein is retained on a nickel column. However neither calcium or magnesium ions altered the HSQC, although zinc and nickel ions precipitated the sample.

Sequence conservation in DysF domains

The most highly conserved residues in the DysFN region (conserved in 65% of DysF sequences –SMART database) are a YENQ sequence and a number of tryptophans, which in myoferlin inner domain are 935YQNE and W946, W973, W975 and W980. In the DysFC region the conserved residues are GWxY (992GWEY in myoferlin inner domain) at the N-terminal end and RRRxWxR at the C-terminal end (1019 RRRRLVR in myoferlin inner domain) (see Fig. 3, 4a). Many of these conserved tryptophans and arginines form arginine/aromatic (R/W) stacking interactions, which is a well-known structural motif [25]. In the DysF domain, the best-defined stack consists of W946 and W993 which sandwich R1019. R1021 extends this stack. Indeed in other DysF domains this stack may be extended by a further W as the equivalent of L1023 in myoferlin is a W in two-thirds of DysF domains. The orientations of R1019 and R1021 are buttressed by E938 and Q936 respectively (Figure 4b). Further R/W stack motifs are also found for R1025 which is sandwiched between W975 and W980, R1020, which is sandwiched between W1008 and Tyr935 and the W973, R1027 pair. The tryptophans are extremely well defined in the NMR structure with an average of over 60 distant restraints per residue (47 is the lowest) and the arginines average 50 restraints per residue. Thus the structural determinants of the DysF domain are a combination of positively charged and aromatic residues.

A search was performed to look for other examples of R/W stacking using JESS [26]. Using the four residues W946, R1019, W993 and R1021 as a search motif, matches were found in 17 structures in the PDB all of which were growth hormone/interleukin receptors. These receptors have 6 extracellular domains, the second and third of which are required for ligand binding. The third domain contains a WSXWS motif, which is common to a broad range of growth factor receptors. In some of the receptors, such as gp130 [27], the residues in this motif form part of a stack of three tryptophans and three arginines that is involved in the structural integrity of the domain rather than ligand binding [28]. Intriguingly, a recent report implicates myoferlin in stabilising the VEGF receptor-2 in endothelial cells [29]. A further JESS search with residues W970, W980 and R1025 again found the growth factor receptors along with thrombospondin repeats (TSR) [30], which have 3 tryptophans sandwiching two arginines. TSR are extracellular domains, found in 41 human proteins with up to 18 repeats in each protein that mediate cell attachment, glycan binding and inhibit matrix metalloproteinases. The W/R motif in combination with disulphide bonds are thought to give stability to the TSR domain fold.

Disease causing mutations.

Surveys of patients find mutations throughout the dysferlin gene [31] and a single hotspot is not apparent. Of the 91 sites of missense entries in dysferlin in the Leiden Muscular Dystrophy Database (LMDD), 12 (13% of the total) map to the inner DysF domain, which contains 5.5% of the total protein sequence, so the DysF domain is more susceptible to mutation than average for the protein. These include some of the most common disease causing mutations R959W (reported 22 times) and W999C (reported 15 times). Disruption of the inner domain of dysferlin, particularly

perturbing the W/R stacking is clearly a cause of dysferlinopathy (Table 2; Fig.4c). Six of the 12 residues involved in the R/W stacking in the inner DysF domain of dysferlin are sites of disease causing mutations (W992R, W999C, R1038Q, R1039W, R1044S, R1046H), almost certainly resulting in an unfolded domain in the mutations. There is also a mutation in the W/R stacking of the outer domain W930C equivalent to W999C in the inner domain [20]. The polymorphism in the inner DysF region can be rationalised. R1022Q (not involved in R/W stacking) is a surface exposed side chain, which will tolerate hydrophilic substitution; it is H1003 in the myoferlin inner domain. Clearly disrupting the tight packing of the DysF domain by disrupting the W/R stacking causes dysferlinopathy almost certainly due to degradation of unfolded protein.

Discussion

The significant number of disease-causing mutations in the dysferlin gene that map to the inner DysF domain demonstrates that this domain must be correctly folded for dysferlin to be functional. However the similarities of the stacked R/W motifs between DysF and the growth hormone receptors and thrombospondin domains probably represent convergent evolution to a substructure that stabilises small β sheets rather than indicating a functional similarity.

How the internal duplication has been tolerated is clear. The insertion occurs in a loop, where extending the loop to include another domain is tolerated. There are mutations reported in the linker region in the Leiden database R1060T and a duplication of residues A1064 and E1065, although the latter is reported to be a polymorphism in the Italian population but not in East Asian controls [31]. This raises

the possibility of the relative orientations of the two DysF domains being important as the linker is sensitive to mutation. Domain insertion is not a rare event, especially with domains of this length – over 60% of inserted domains are less than 130 residues, and nearly 80% are less than 175 residues [32]. It is however rare for this event to occur in non-enzymatic proteins and for the parent and insert domains to be the same. One example is the PH domain complex of myosin-X, which has a PH domain inserted into another PH domain [33].

A number of binding partners of dysferlin have been reported (for a recent review see [34]). These include annexin A1 [15], which has been shown to be required for Ca2+-dependent membrane sealing [35], caveolin 3 [36], which is mutated in limb girdle muscular dystrophy type 1C, AHNAK [37] and beta-parvin [38]. Interaction of dysferlin with calpain 3m which is mutated in Limb Girdle Muscular Dystrophy Type 2A, has been implied from co-association studies [39]. The C-terminus of AHNAK binds to the C2A domain of dysferlin and myoferlin [37]. The first CH domain of beta-parvin binds to the intracellular juxtamembrane section of dysferlin, in fact the pathological mutation in the inner DysF domain of dysferlin W999C did not prevent binding to beta parvin ruling out disruption of the interaction with beta parvin as a source of the pathology [38].

The major candidate for DysF domain binding is caveolin-3, as the scaffold domain of caveolin-3 has been shown by phage display enrichment to bind W rich peptides with a motif WXWXXXXW [40]. There are conserved versions of this motif in the inner DysF domains of myoferlin and dysferlin starting at W973 of myoferlin [36]. In the outer domain the spacing is shorter between the second and third W. Although W973 and W975 makes a possible binding surface the third W980 is on the opposite face of

the folded DysF domain. This supports the possibility that interaction with caveolin-3 occurs when the DysF domain is in a denatured state, as an unfolded state much more resembles the peptides selected in the Couet study than the mostly buried tryptophans in the folded structure, which do not form a single binding surface. Unfolding of the DysF domain could be a mechanism for degrading dysferlin and myoferlin at the end of the repair cycle. Some metals (nickel and zinc) precipitate the domain and this could be via aggregation of an unfolded state.

Our structural studies of the inner DysF domain of myoferlin have allowed us to rationalise the molecular basis of a significant number of the missense mutations in dysferlin. The structure resolves some of the issues of annotation of domains in the ferlins and provides a framework for fully understanding the role of this domain in Limb Girdle Muscular Dystrophy.

Materials and Methods

Sample preparation

The DysF domain cDNA, encoding residues 923-1040 of Human Myoferlin, was cloned into the pET151-D vector (Invitrogen). The recombinant DysF domain was over-expressed in *Escherichia coli* strain BL21 Star (DE3), grown in minimal media containing $^{15}NH_4Cl$ as the sole nitrogen source, with either ^{12}C or ^{13}C-glucose as the sole carbon source, to produce the uniformly ^{15}N- or $^{15}N/^{13}C$-labelled protein. Protein expression was induced at an OD_{600} = 0.5 by the addition of IPTG to a final concentration of 0.1 mM for 6 hours. After nickel ion affinity chromatography, the sample was treated with TEV protease for 18 hours (4°C) to cleave the His_6-tag fragment. DysF was then purified by size-exclusion chromatography.

NMR Spectroscopy

Samples for NMR spectroscopy were prepared by exchanging ^{15}N or $^{15}N/^{13}C$ uniformly-labelled DysF protein into buffer containing 20 mM MES, 100 mM NaCl and 90% H_2O/10% D_2O or 100% D_2O, pH 6.5, at a final protein concentration of 1.3 mM. NMR spectra were acquired at 298 K (except where indicated) on Varian UnityPLUS and INOVA spectrometers (operating at nominal ^1H frequencies of 500 MHz and 600/800 MHz, respectively) each equipped with a triple resonance (^1H, ^{13}C, ^{15}N) probe including a Z-axis pulse field gradient coil. All NMR spectra were processed using NMRpipe/NMRDraw [41] and analysed using the CCPN Analysis package [42]. ^1H, ^{13}C and ^{15}N chemical shifts were referenced indirectly to sodium 2,2-dimethyl-2-silane-pentane-5-sulphonate (DSS).

Sequence-specific backbone resonance assignments were obtained by combining the data from the following 3D gradient sensitivity enhanced triple resonance experiments: HNCO, HNCA, HN(CO)CA, HN(CA)CO, HA(CA)NH, HA(CACO)NH, HNCACB [43] and CBCA(CO)NH [44]. An assigned ^1H-^{15}N-HSQC is shown in Figure 6. Side-chain resonance assignments were obtained from 3D ^{15}N-separated TOCSY-HSQC (60 ms mixing time) and 3D [^1H,^{13}C]-HCCH-TOCSY (16 ms) spectra. Distance restraints were derived from 3D ^{15}N and ^{13}C-edited NOESY-HSQC (100 ms mixing times). One-bond ^1H-^{15}N residual dipolar couplings were obtained from [^1H,^{15}N]-IPAP-HSQC spectra [45] acquired at 293 K in the presence and absence of 5% *n*-octyl-penta(ethylene glycol):octanol, 0.96:1 [46].

Structure Calculations

Inter-proton distance restraints were derived from cross-peaks of 3D [^{1}H,^{15}N]-NOESY-HSQC and [^{1}H,^{13}C]-NOESY-HSQC spectra. The cross-peaks of 3D [^{1}H,^{15}N]-NOESY-HSQC were grouped into four categories according to their relative peak intensities and were designated with the corresponding interproton distance restraint limits of 1.8-2.5 Å, 1.8-3.0 Å, 1.8-3.5 Å and 1.8-4.0 Å. The cross-peaks of 3D [^{1}H,^{13}C]-NOESY-HSQC were grouped into five categories according to their relative peak intensities and were designated with the corresponding interproton distance restraint limits of 1.8-2.5 Å, 1.8-3.0 Å, 1.8-3.5 Å, 1.8-4.0 Å and 1.8-5.0 Å. 0.5 Å per methyl group was added to the upper bound of the distance restraint for NOE cross-peaks that involved methyl groups.

The structure calculations were carried out within the CNS program [47] using the IUPAC PARALLHDGv5.3 parameter set and TOPALLHDGv5.3 energy function parameters for distance geometry calculations and simulated annealing [48; 49]. The ambiguous distance restraints were filtered iteratively, based upon the coordinates of an ensemble of intermediate conformers, and duplicate restraints were discarded. A total of 2810 NOE-derived inter-proton distance restraints were included in the final iterations of the structure calculations. Backbone torsion angle restraints for φ and ψ were derived from analysis of $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$ and ^{15}N, H_N chemical shift databases as implemented in the program TALOS [50].

Initial estimates of the value of the axial and rhombic components of the molecular alignment tensor were obtained with MODULE [51], -20.4 Hz and 0.33 Hz, respectively. These values were used as a starting point for a procedure that simultaneously refined the protein structure and ascertained the values of the

alignment tensor components using a grid search approach. The final values of the axial and component and rhombicity of the molecular alignment tensor used in the refinement stage were -19.687 Hz and 0.350 Hz, respectively. The final set of structures was obtained from water refinement calculations in CNS using modified RECOORD protocols [45 52]. The quality of the ensemble of structures was evaluated by the PROCHECK program [53]. The ensemble of structures and NMR restraints have been deposited with the PDB and BMRB (PDB2k20, BMRB 15718)

Acknowledgements

We thank Dr Fiona Norwood for helpful discussion. Pryank Patel was supported by a BBSRC studentship.

References

1. Miyoshi, K., Kawai, H., Iwasa, M., Kusaka, K. & Nishino, H. (1986). Autosomal recessive distal muscular dystrophy as a new type of progressive muscular dystrophy. Seventeen cases in eight families including an autopsied case. *Brain* **109 (Pt 1)**, 31-54.

2. Bashir, R., Britton, S., Strachan, T., Keers, S., Vafiadaki, E., Lako, M., Richard, I., Marchand, S., Bourg, N., Argov, Z., Sadeh, M., Mahjneh, I., Marconi, G., Passos-Bueno, M. R., Moreira Ede, S., Zatz, M., Beckmann, J. S. & Bushby, K. (1998). A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. *Nature Genetics* **20**, 37-42.

3. Liu, J., Aoki, M., Illa, I., Wu, C., Fardeau, M., Angelini, C., Serrano, C., Urtizberea, J. A., Hentati, F., Hamida, M. B., Bohlega, S., Culper, E. J., Amato, A. A., Bossie, K., Oeltjen, J., Bejaoui, K., McKenna-Yasek, D., Hosler, B. A., Schurr, E., Arahata, K., de Jong, P. J. & Brown, R. H., Jr. (1998). Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nature Genetics* **20**, 31-6.

4. Achanzar, W. E. & Ward, S. (1997). A nematode gene required for sperm vesicle fusion. *Journal of Cell Science* **110 (Pt 9)**, 1073-81.

5. Nalefski, E. A. & Falke, J. J. (1996). The C2 domain calcium-binding motif: structural and functional diversity. *Protein Science* **5**, 2375-90.

6. Sutton, R. B., Davletov, B. A., Berghuis, A. M., Sudhof, T. C. & Sprang, S. R. (1995). Structure of the first C2 domain of synaptotagmin I: a novel Ca2+/phospholipid-binding fold. *Cell* **80**, 929-38.

7. Essen, L. O., Perisic, O., Cheung, R., Katan, M. & Williams, R. L. (1996). Crystal structure of a mammalian phosphoinositide-specific phospholipase C delta. *Nature* **380**, 595-602.

8. Rizo, J. & Sudhof, T. C. (1998). C2-domains, structure and function of a universal Ca2+-binding domain. *The Journal of Biological Chemistry* **273**, 15879-82.

9. Lu, J., Machius, M., Dulubova, I., Dai, H., Sudhof, T. C., Tomchick, D. R. & Rizo, J. (2006). Structural basis for a Munc13-1 homodimer to Munc13-1/RIM heterodimer switch. *PLoS Biology* **4**, e192.

10. Britton, S., Freeman, T., Vafiadaki, E., Keers, S., Harrison, R., Bushby, K. & Bashir, R. (2000). The third human FER-1-like protein is highly similar to dysferlin. *Genomics* **68**, 313-21.

11. Yasunaga, S., Grati, M., Cohen-Salmon, M., El-Amraoui, A., Mustapha, M., Salem, N., El-Zir, E., Loiselet, J. & Petit, C. (1999). A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a nonsyndromic form of deafness. *Nature Genetics* **21**, 363-9.

12. Davis, D. B., Doherty, K. R., Delmonte, A. J. & McNally, E. M. (2002). Calcium-sensitive phospholipid binding properties of normal and mutant ferlin C2 domains. *The Journal of Biological Chemistry* **277**, 22883-8.

13. Doherty, K. R., Cave, A., Davis, D. B., Delmonte, A. J., Posey, A., Earley, J. U., Hadhazy, M. & McNally, E. M. (2005). Normal myoblast fusion requires myoferlin. *Development (Cambridge, England)* **132**, 5565-75.

14. Bansal, D., Miyake, K., Vogel, S. S., Groh, S., Chen, C. C., Williamson, R., McNeil, P. L. & Campbell, K. P. (2003). Defective membrane repair in dysferlin-deficient muscular dystrophy. *Nature* **423**, 168-72.

15. Lennon, N. J., Kho, A., Bacskai, B. J., Perlmutter, S. L., Hyman, B. T. & Brown, R. H., Jr. (2003). Dysferlin interacts with annexins A1 and A2 and mediates sarcolemmal wound-healing. *The Journal of Biological Chemistry* **278**, 50466-73.

16. Roux, I., Safieddine, S., Nouvian, R., Grati, M., Simmler, M. C., Bahloul, A., Perfettini, I., Le Gall, M., Rostaing, P., Hamard, G., Triller, A., Avan, P., Moser, T. & Petit, C. (2006). Otoferlin, defective in a human deafness form, is essential for exocytosis at the auditory ribbon synapse. *Cell* **127**, 277-89.

17. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J.,

Yeats, C. & Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Research* **32**, D138-41.

18. Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., Ponting, C. P. & Bork, P. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Research* **32**, D142-4.

19. Ponting, C. P., Mott, R., Bork, P. & Copley, R. R. (2001). Novel protein domains and repeats in Drosophila melanogaster: insights into structure, function, and evolution. *Genome Research* **11**, 1996-2008.

20. Therrien, C., Dodig, D., Karpati, G. & Sinnreich, M. (2006). Mutation impact on dysferlin inferred from database analysis and computer-based structural predictions. *Journal of the Neurological Sciences* **250**, 71-8.

21. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica* **60**, 2256-68.

22. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends in Biochemical Sciences* **20**, 478-80.

23. Laskowski, R. A. (1995). SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graphics* **13**, 323-30, 307-8.

24. Glaser, F., Morris, R. J., Najmanovich, R. J., Laskowski, R. A. & Thornton, J. M. (2006). A method for localizing ligand binding pockets in protein structures. *Proteins* **62**, 479-88.

25. Flocco, M. M. & Mowbray, S. L. (1994). Planar stacking interactions of arginine and aromatic side-chains in proteins. *Journal of Molecular Biology* **235**, 709-17.

26. Barker, J. A. & Thornton, J. M. (2003). An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis. *Bioinformatics (Oxford, England)* **19**, 1644-9.

27. Bravo, J., Staunton, D., Heath, J. K. & Jones, E. Y. (1998). Crystal structure of a cytokine-binding region of gp130. *The EMBO Journal* **17**, 1665-74.

28. Somers, W., Ultsch, M., De Vos, A. M. & Kossiakoff, A. A. (1994). The X-ray structure of a growth hormone-prolactin receptor complex. *Nature* **372**, 478-81.

29. Bernatchez, P. N., Acevedo, L., Fernandez-Hernando, C., Murata, T., Chalouni, C., Kim, J., Erdjument-Bromage, H., Shah, V., Gratton, J. P., McNally, E. M., Tempst, P. & Sessa, W. C. (2007). Myoferlin regulates vascular endothelial growth factor receptor-2 stability and function. *The Journal of biological chemistry* **282**, 30745-53.

30. Tan, K., Duquette, M., Liu, J. H., Dong, Y., Zhang, R., Joachimiak, A., Lawler, J. & Wang, J. H. (2002). Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication. *The Journal of Cell Biology* **159**, 373-82.

31. Kawabe, K., Goto, K., Nishino, I., Angelini, C. & Hayashi, Y. K. (2004). Dysferlin mutation analysis in a group of Italian patients with limb-girdle muscular dystrophy and Miyoshi myopathy. *European Journal of Neurology* **11**, 657-61.

32. Aroul-Selvam, R., Hubbard, T. & Sasidharan, R. (2004). Domain insertions in protein structures. *Journal of Molecular Biology* **338**, 633-41.

33. Berg, J. S., Derfler, B. H., Pennisi, C. M., Corey, D. P. & Cheney, R. E. (2000). Myosin-X, a novel myosin with pleckstrin homology domains, associates with regions of dynamic actin. *Journal of Cell Science* **113 Pt 19**, 3439-51.

34. Han, R. & Campbell, K. P. (2007). Dysferlin and muscle membrane repair. **19**, 409-416.

35. McNeil, A. K., Rescher, U., Gerke, V. & McNeil, P. L. (2006). Requirement for annexin A1 in plasma membrane repair. *The Journal of biological chemistry* **281**, 35202-7.

36. Matsuda, C., Hayashi, Y. K., Ogawa, M., Aoki, M., Murayama, K., Nishino, I., Nonaka, I., Arahata, K. & Brown, R. H., Jr. (2001). The sarcolemmal proteins dysferlin and caveolin-3 interact in skeletal muscle. *Human molecular genetics* **10**, 1761-6.

37. Huang, Y., Laval, S. H., van Remoortere, A., Baudier, J., Benaud, C., Anderson, L. V., Straub, V., Deelder, A., Frants, R. R., den Dunnen, J. T., Bushby, K. & van der Maarel, S. M. (2007). AHNAK, a novel component of the dysferlin protein complex, redistributes to the cytoplasm with dysferlin during skeletal muscle regeneration. *The FASEB journal* **21**, 732-42.

38. Matsuda, C., Kameyama, K., Tagawa, K., Ogawa, M., Suzuki, A., Yamaji, S., Okamoto, H., Nishino, I. & Hayashi, Y. K. (2005). Dysferlin interacts with affixin (beta-parvin) at the sarcolemma. *Journal of neuropathology and experimental neurology* **64**, 334-40.

39. Anderson, L. V., Harrison, R. M., Pogue, R., Vafiadaki, E., Pollitt, C., Davison, K., Moss, J. A., Keers, S., Pyle, A., Shaw, P. J., Mahjneh, I., Argov, Z., Greenberg, C. R., Wrogemann, K., Bertorini, T., Goebel, H. H., Beckmann, J. S., Bashir, R. & Bushby, K. M. (2000). Secondary reduction in calpain 3 expression in patients with limb girdle muscular dystrophy type 2B and Miyoshi myopathy (primary dysferlinopathies). *Neuromuscular disorders* **10**, 553-9.

40. Couet, J., Li, S., Okamoto, T., Ikezu, T. & Lisanti, M. P. (1997). Identification of peptide and protein ligands for the caveolin-scaffolding domain.

Implications for the interaction of caveolin with caveolae-associated proteins. *The Journal of biological chemistry* **272**, 6525-33.

41. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.

42. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, P., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* **59**, 687-696.

43. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. (1994). A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. *J Am Chem Soc* **116**, 11655-11666.

44. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *J Magn Reson* **103**, 203-216.

45. Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement ofJand Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J Magn Reson* **131**, 373-378.

46. Ruckert, M. & Otting, G. (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J Am Chem Soc* **122**, 7793-7797.

47. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallographica* **54**, 905-921.

48. Linge, J. P. & Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *J Biomol NMR* **13**, 51-9.

49. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. & Nilges, M. (2003). Refinement of protein structures in explicit solvent. *Proteins* **50**, 496-506.

50. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302.

51. Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. (2001). A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J Biomol NMR* **20**, 223-31.

52. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R. & Bonvin, A. M. (2005). RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**, 662-72.

53. Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**, 283-291.

54. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics (Oxford, England)* **15**, 305-8.

55. DeLano, W. L. (2002). The PyMOL Molecular Graphics System.

56. Cho, H. J., Sung, D. H., Kim, E. J., Yoon, C. H., Ki, C. S. & Kim, J. W. (2006). Clinical and genetic analysis of Korean patients with Miyoshi myopathy: identification of three novel mutations in the DYSF gene. *Journal of Korean Medical Science* **21**, 724-7.

57. de Luna, N., Gallardo, E., Soriano, M., Dominguez-Perles, R., de la Torre, C., Rojas-Garcia, R., Garcia-Verdugo, J. M. & Illa, I. (2006). Absence of dysferlin alters myogenin expression and delays human muscle differentiation "in vitro". *The Journal of Biological Chemistry* **281**, 17092-8.

58. Nguyen, K., Bassez, G., Bernard, R., Krahn, M., Labelle, V., Figarella-Branger, D., Pouget, J., Hammouda el, H., Beroud, C., Urtizberea, A., Eymard, B., Leturcq, F. & Levy, N. (2005). Dysferlin mutations in LGMD2B, Miyoshi myopathy, and atypical dysferlinopathies. *Human Mutation* **26**, 165.

59. Cagliani, R., Fortunato, F., Giorda, R., Rodolico, C., Bonaglia, M. C., Sironi, M., D'Angelo, M. G., Prelle, A., Locatelli, F., Toscano, A., Bresolin, N. & Comi, G. P. (2003). Molecular analysis of LGMD-2B and MM patients: identification of novel DYSF mutations and possible founder effect in the Italian population. *Neuromuscular Disorders* **13**, 788-95.

60. Cagliani, R., Magri, F., Toscano, A., Merlini, L., Fortunato, F., Lamperti, C., Rodolico, C., Prelle, A., Sironi, M., Aguennouz, M., Ciscato, P., Uncini, A., Moggio, M., Bresolin, N. & Comi, G. P. (2005). Mutation finding in patients with dysferlin deficiency and role of the dysferlin interacting proteins annexin A1 and A2 in muscular dystrophies. *Human Mutation* **26**, 283.

61. Nagashima, T., Chuma, T., Mano, Y., Goto, Y., Hayashi, Y. K., Minami, N., Nishino, I., Nonaka, I., Takahashi, T., Sawa, H., Aoki, M. & Nagashima, K. (2004). Dysferlinopathy associated with rigid spine syndrome. *Neuropathology* **24**, 341-6.

62. Tagawa, K., Ogawa, M., Kawabe, K., Yamanaka, G., Matsumura, T., Goto, K., Nonaka, I., Nishino, I. & Hayashi, Y. K. (2003). Protein and gene analyses

of dysferlinopathy in a large group of Japanese muscular dystrophy patients. *Journal of the Neurological Sciences* **211**, 23-8.

63. Takahashi, T., Aoki, M., Tateyama, M., Kondo, E., Mizuno, T., Onodera, Y., Takano, R., Kawai, H., Kamakura, K., Mochizuki, H., Shizuka-Ikeda, M., Nakagawa, M., Yoshida, Y., Akanuma, J., Hoshino, K., Saito, H., Nishizawa, M., Kato, S., Saito, K., Miyachi, T., Yamashita, H., Kawai, M., Matsumura, T., Kuzuhara, S., Ibi, T., Sahashi, K., Nakai, H., Kohnosu, T., Nonaka, I., Arahata, K., Brown, R. H., Jr., Saito, H. & Itoyama, Y. (2003). Dysferlin mutations in Japanese Miyoshi myopathy: relationship to phenotype. *Neurology* **60**, 1799-804.

Figure Legends

Figure1. Domain Structure and Sequence of the ferlins PFAM and SMART domain representations of human myoferlin.

Figure 2. Structure of myoferlin DysF domain a) Ensemble of 20 lowest energy structures of myoferlin inner DysF domain with the DysFN domain in red and the DysFC domain in green. b) Stereo pair of backbone of lowest energy structure Figure drawn with Pymol [55]

Figure 3. Sequence alignment of DysF domains of human dysferlin and myoferlin, *C.elegans* fer-1, yeast peroxisomal protein PEX30 and human β-propeller protein DKFZP434B0335. The rows correspond to the DysFN and DysFC regions. Secondary structure of the myoferlin inner domain is shown. Figure drawn using Espript [54].

Figure 4 Conserved and mutated residues a) Conserved residues in ball and stick. Arginines in blue, Tryptophans in yellow, tyrosines in red, other residues green. b) Close up of residues Q936, E938, W946, Y993, R1019, and R1021. c) Disease-causing mutations in ball and stick. Tryptophans and histidines are labelled in yellow, arginines in blue, tyrosines in red and alanines in green. This view is 180 degrees away from figure 5a to show the opposite face of the β-sheet. Figure drawn with Pymol [55]

Figure 5 Electrostatic surface of DysF domain (positive blue, negative red). Left and centre views approximately 180 degree rotation apart right view approximately 90 degrees with respect to central. Figure drawn with Pymol [55].

Figure 6. ^{1}H-^{15}N HSQC of inner DysF domain of human myoferlin showing assignments. Spectrum recorded in 20 mM MES, 100 mM NaCl and 90% H2O/10% D2O at 298K.

Figure1

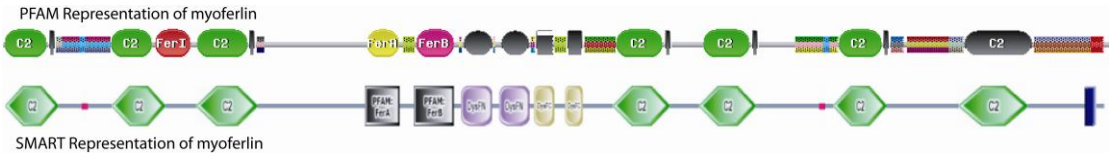

Figure 2

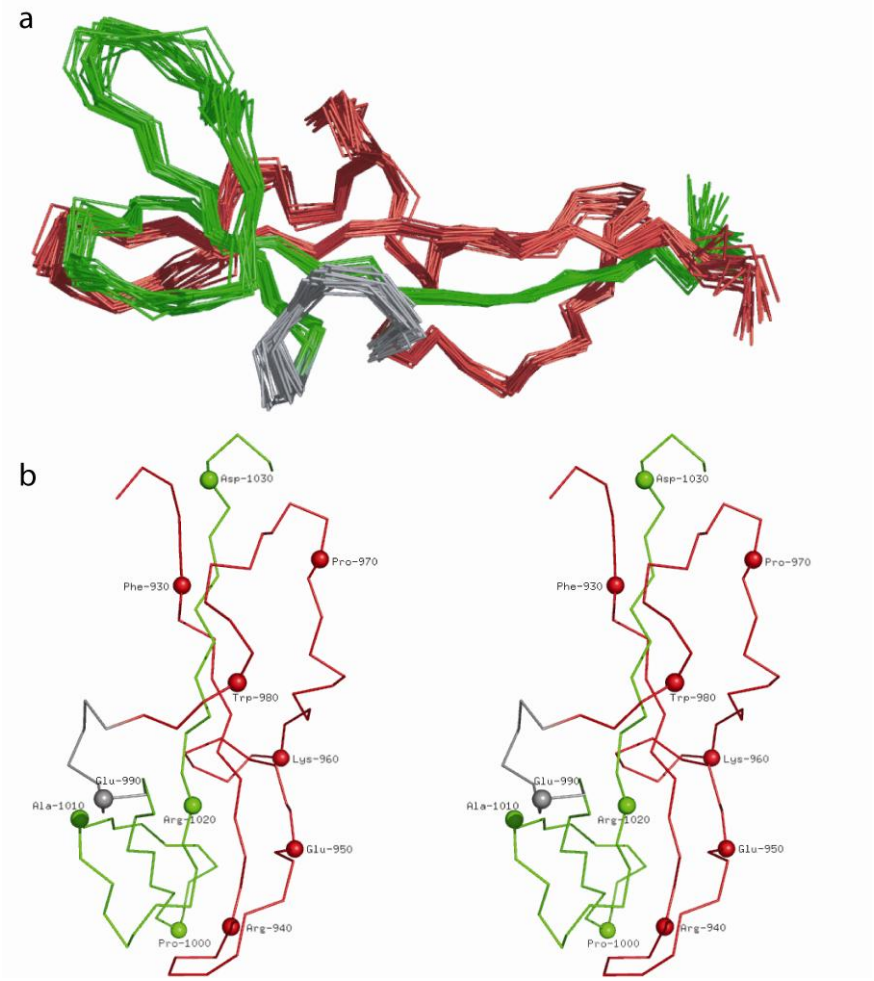

Figure 3

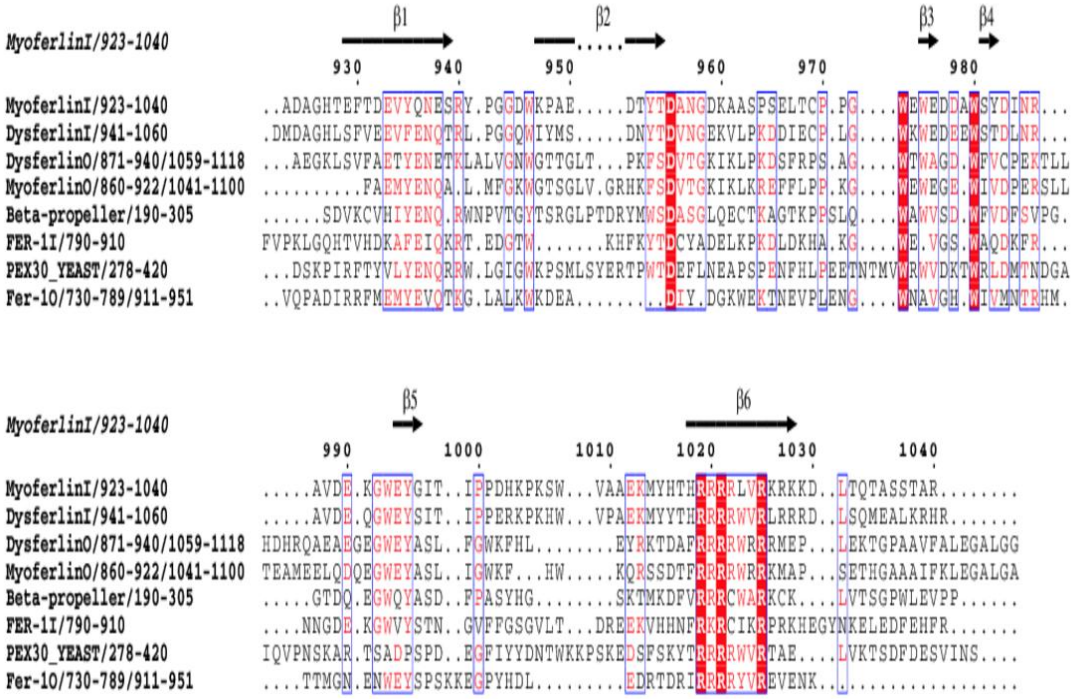
β1
β2
β3
β4
MyoferlinI/923-1040
930
940
950
960
970
980
MyoferlinI/923-1040 ..ADAGHTEFTDEVYQNESRY.PGGDWKPAE.....DTYTDANGDKAASPSELTCP.PG....WEWEDDAWSYDINR...
DysferlinI/941-1060 .DMDAGHLSFVEEVFENQTRL.PGGQWIYMS.....DNYTDVNGEKVLPKDDIECP.LG....WKWEDEEWSTDLNR...
DysferlinO/871-940/1059-1118 ...AEGKLSVFAETYENETKLALVGNWGTTGLT...PKFSDVTGKIKLPKDSFRPS.AG....WTWAGD.WFVCPEKTLL
MyoferlinO/860-922/1041-1100 .....FAEMYENQA.L.MFGKWGTSGLV.GRHKFSDVTGKIKLKREFFLPP.KG....WEWEGE.WIVDPERSLL
Beta-propeller/190-305 ......SDVKCVHIYENQ.RWNPVTGYTSRGLPTDRYMWSDASGLQECTKAGTKPPSLQ....WAWVSD.WFVDFSVPG.
FER-1I/790-910 FVPKLGQHTVHDKAFEIQKRT.EDGTW.....KHFKYTDCYADELKPKDLDKHA.KG....WE.VGS.WAQDKFR...
PEX30_YEAST/278-420 ...DSKPIRFTYVLYENQRRW.LGIGWKPSMLSYERTPWTDEFLNEAPSPENFHLPEETNTMVWRWVDKTWRLDMTNDGA
Fer-1O/730-789/911-951 ..VQPADIRRFMEMYEVQTKG.LALKWKDEA.....DIY.DGKWEKTNEVPLENG....WNAVGH.WIVMNTRHM.
β5
β6
MyoferlinI/923-1040
990
1000
1010
1020
1030
1040
MyoferlinI/923-1040 .....AVDE.KGWEYGIT..IPPDHKPKSW...VAAEKMYHTHRRRRLVRKRKKD..LTQTASSTAR.....
DysferlinI/941-1060 .....AVDE.QGWEYSIT..IPPERKPKHW...VPAEKMYYTHRRRRWVRLRRRD..LSQMEALKRHR.....
DysferlinO/871-940/1059-1118 HDHRQAEAEGEGWEYASL..FGWKFHL.....EYRKTDAFRRRRWRRRMEP...LEKTGPAAVFALEGALGG
MyoferlinO/860-922/1041-1100 TEAMEELQDQEGWEYASL..IGWKF...HW.....KQRSSDTFRRRRWRRKMAP...SETHGAAAIFKLEGALGA
Beta-propeller/190-305 .....GTDQ.EGWQYASD..FPASYHG.....SKTMKDFVRRRCWARKCK....LVTSGPWLEVPP.....
FER-1I/790-910 ....NNGDE.KGWVYSTN..GVFFGSGVLT...DREEKVHHNFRKRCIKRPRKHEGYNKELEDFEHFR.....
PEX30_YEAST/278-420 IQVPNSKAR.TSADPSPD..EGFIYYDNTWKKPSKEDSFSKYTRRRRWVRTAE....LVKTSDFDESVINS....
Fer-1O/730-789/911-951 ....TTMGN.ENWEYSPSKKEGPYHDL.....EDRTDRIRRRRYVREVENK..

Figure 4

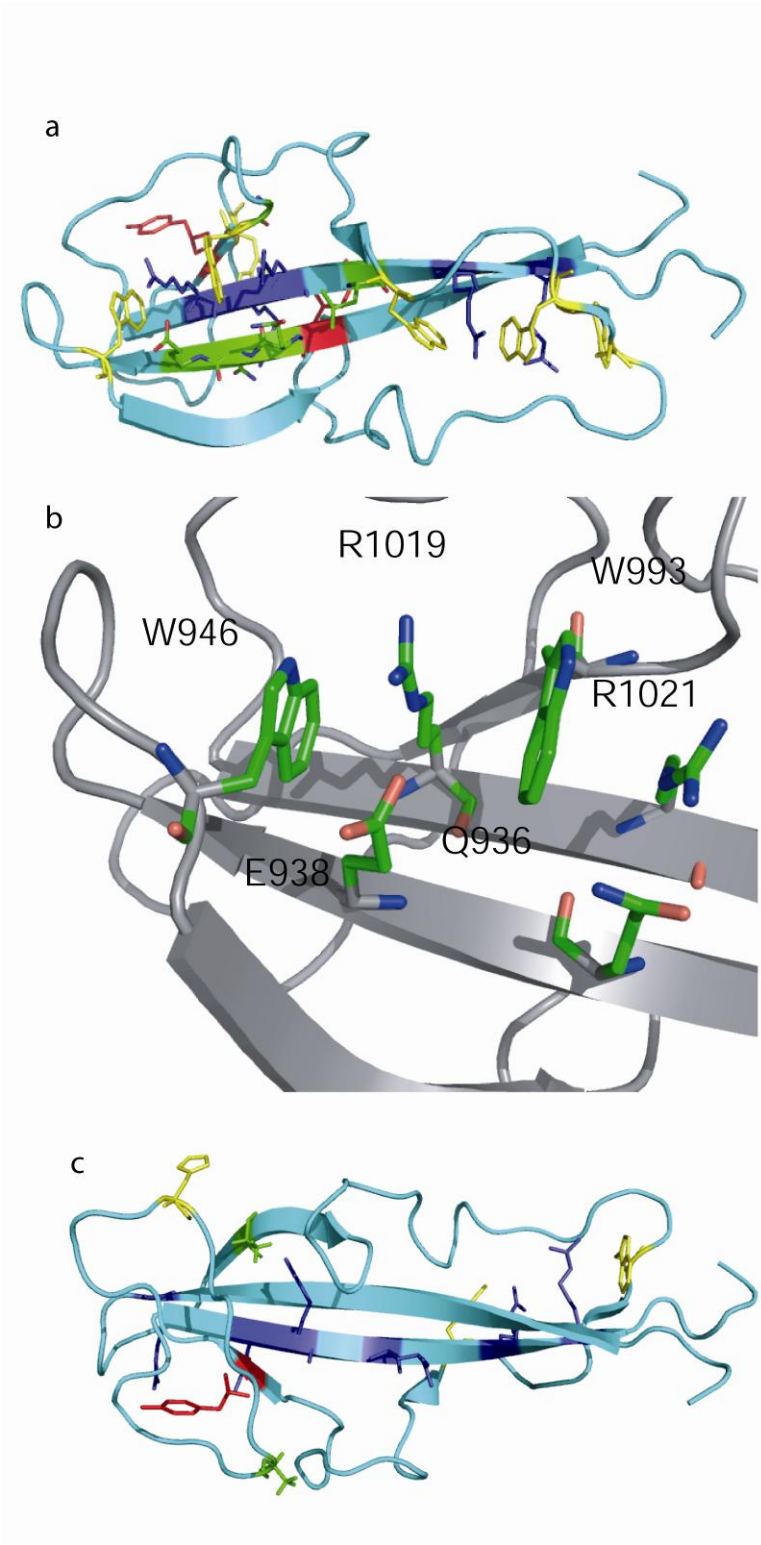

Figure 5

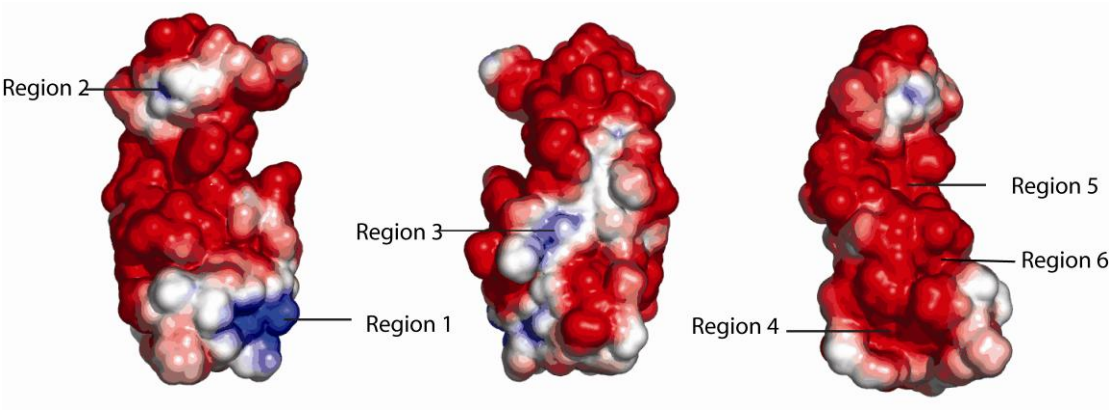

Figure 6