Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy

Jing Liu[1,11], Masashi Aoki[1], Isabel Illa[2], Chenyan Wu[3], Michel Fardeau[4], Corrado Angelini[5], Carmen Serrano[2], J. Andoni Urtizberea[4], Faycal Hentati[6], Mongi Ben Hamida[6], Saeed Bohlega[7], Edward J. Culper[7], Anthony A. Amato[8], Karen Bossie[1], Joshua Oeltjen[1], Khemissa Bejaoui[1], Diane McKenna-Yasek[1], Betsy A. Hosler[1], Erwin Schurr[9], Kiichi Arahata[10], Pieter J. de Jong[3] & Robert H. Brown Jr[1]

Miyoshi myopathy (MM) is an adult onset, recessive inherited distal muscular dystrophy that we have mapped to human chromosome 2p13. We recently constructed a 3-Mb P1-derived artificial chromosome (PAC) contig spanning the MM candidate region. This clarified the order of genetic markers across the MM locus, provided five new polymorphic markers within it and narrowed the locus to approximately 2 Mb. Five skeletal muscle expressed sequence tags (ESTs) map in this region. We report that one of these is located in a novel, full-length 6.9-kb muscle cDNA, and we designate the corresponding protein 'dysferlin'. We describe nine mutations in the dysferlin gene in nine families; five are predicted to prevent dysferlin expression. Identical mutations in the dysferlin gene can produce more than one myopathy phenotype (MM, limb girdle dystrophy, distal myopathy with anterior tibial onset).

Introduction

Autosomal recessive muscular dystrophies constitute a genetically heterogeneous group of disorders[1–3]. Most are characterized by weakness and atrophy of proximal muscles, although in rare cases symptoms first arise in distal muscles. One type of distal myopathy, termed 'Miyoshi myopathy'[4] (MM), is characterized by weakness that initially affects the gastrocnemius muscle during early adulthood. In MM, serum levels of muscle enzymes such as creatine kinase are elevated in the early stage of the disease.

We initially mapped the MM locus to chromosome 2p12–14 between genetic markers *D2S292* and *D2S286* (ref. 5). Subsequently, our genetic mapping in MM families placed the MM locus between markers *GGAA-P7430* and *D2S2109* (ref. 6). At approximately the same time, Bushby and colleagues localized limb-girdle muscular dystrophy (LGMD-2B) to the same genetic interval[7–9]. Recently, two reports have described large, inbred kindreds whose members include both MM and LGMD2B patients[10,11]. In these studies, the disease gene(s) for both MM and LGMD2B mapped to the same genetic interval as the MM and LGMD2B loci. Moreover, individuals in both pedigrees with MM or LGMD2B phenotypes shared the same haplotypes. This raises the possibility that the two diseases may arise from the same gene defect and that the disease phenotypes are modified by additional factors.

To clone the MM/LGMD2B gene(s), we constructed a 3-Mb PAC contig spanning the entire MM/LGMD2B candidate region[12]. This high-resolution PAC contig resolved the marker order discrepancies in previous studies[6,8,13]. The physical size of the PAC contig also indicated that the previous minimal size estimation based on YAC mapping data was underestimated. Thus, it was important to isolate new informative markers to narrow the genetic interval, and our PAC contig provided a source for such markers. The unambiguously ordered polymorphic markers, including the five new markers generated in this study, allowed us to further refine the genetic interval for the MM gene. Moreover, the PAC contig served as a framework for establishing a transcription map to identify genes in the MM locus that are expressed in skeletal muscle, and are therefore candidates for MM/LGMD2B genes. We report here the cloning of a novel gene, dysferlin, and its mutation in patients with MM, LGMD2B and distal myopathy with onset in the anterior tibial muscles.

Results

Polymorphic markers spanning the MM locus

We first sought to identify new polymorphic markers within the PAC contig across the MM locus by hybridization with labelled oligonucleotides containing common repeat sequences. Seventeen different groups of overlapping PACs contained these repeat sequences, including some PACs with previously identified repeat markers. For example, three groups contained known markers *GGAA-P7430* (GGAA repeat), *D2S1394* (GATA repeat) and *D2S1398* (GGAA repeat)[14,15]. We did not attempt to isolate new repeat markers from these PACs and they were not further

[1]*Day Neuromuscular Research Laboratory, Room 6627, Building 149, 13th Street, Navy Yard, Charlestown, Massachusetts 02129, USA.* [2]*Neuromuscular Disease Section, Neurology Department, Hospital Sant Pau, U.A.B., Barcelona, Spain.* [3]*Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, Elm and Charlton Streets, Buffalo, New York 14263, USA.* [4]*Institut de Myologie, Association Française contre les Myopathies, Hôpital de la Salpetriere, 75651 Paris Cedex 13, France.* [5]*Neuromuscular Center, Department of Neurology, University of Padova, Padova, Italy 35128.* [6]*Service Neurologie, La Rabta, Tunis, Tunisia.* [7]*Department of Neurosciences, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia.* [8]*Department of Medicine/Neurology, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78284-7883, USA.* [9]*McGill Center for the Study of Host Resistance, Research Institute, Montreal General Hospital, Montreal, Quebec, Canada H3G 1A4.* [10]*Department of Neuromuscular Research, National Institute of Neuroscience, NCNP, Tokyo 187-8502, Japan.* [11]*Current address: PhageTech, CP 387 Place du Parc, Montreal, Quebec, Canada H2W 2N9. Correspondence should be addressed to R.H.B. (e-mail: brown@helix.mgh.harvard.edu).*

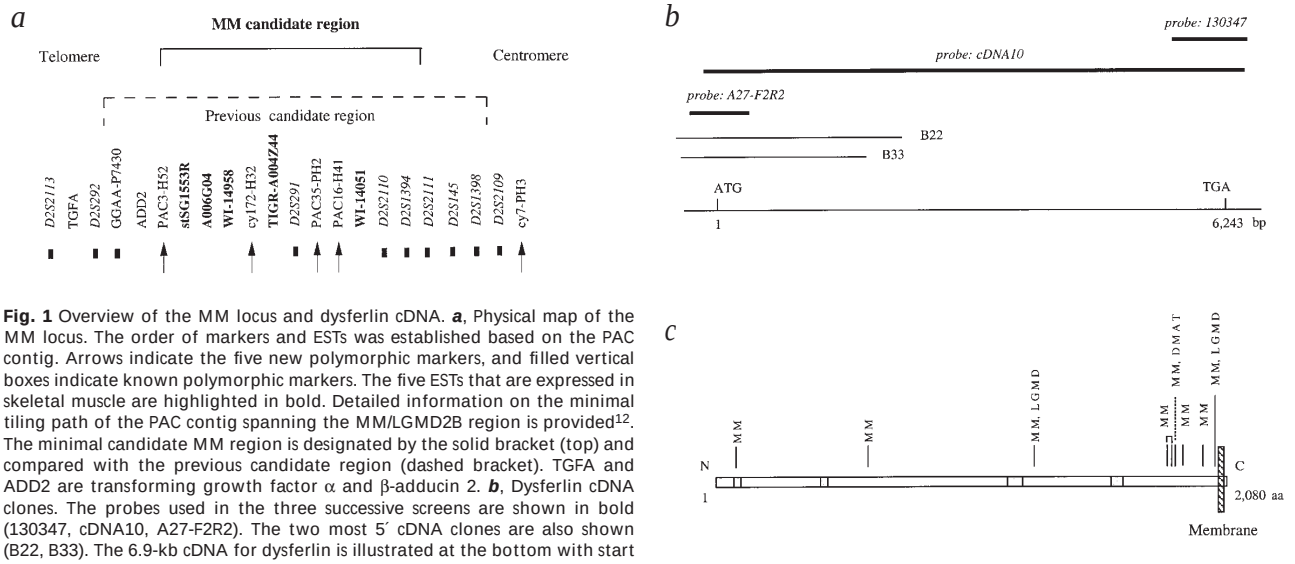

Fig. 1 Overview of the MM locus and dysferlin cDNA. ***a***, Physical map of the MM locus. The order of markers and ESTs was established based on the PAC contig. Arrows indicate the five new polymorphic markers, and filled vertical boxes indicate known polymorphic markers. The five ESTs that are expressed in skeletal muscle are highlighted in bold. Detailed information on the minimal tiling path of the PAC contig spanning the MM/LGMD2B region is provided[12]. The minimal candidate MM region is designated by the solid bracket (top) and compared with the previous candidate region (dashed bracket). TGFA and ADD2 are transforming growth factor α and β-adducin 2. ***b***, Dysferlin cDNA clones. The probes used in the three successive screens are shown in bold (130347, cDNA10, A27-F2R2). The two most 5′ cDNA clones are also shown (B22, B33). The 6.9-kb cDNA for dysferlin is illustrated at the bottom with start and stop codons as shown. ***c***, Predicted dysferlin protein. The predicted protein is represented by the rectangle. The locations of four C2 domains are indicated by stippled boxes, and the putative transmembrane region is hatched. Vertical lines above the cDNA denote the positions of the mutations (Table 2); the associated labels indicate the phenotypes (MM, Miyoshi myopathy; LGMD, limb girdle muscular dystrophy; DMAT, distal myopathy with anterior tibial onset).

analysed. Similarly, we excluded seven groups of PACs that contained known CA repeat markers. Seven groups of PACs that contained unidentified repeats remained, and for each group the PAC containing the smallest insert was selected for subcloning. Subclones were re-screened and positive clones were sequenced to identify repeats. In total, seven new repeat sequences were identified within the MM/LGMD2B PAC contig. Of these, five were polymorphic in the population tested. The information for these five markers is summarized (Table 1). Based on the PAC contig[12], we were able to place our five new markers and ten previously published polymorphic markers in an unambiguous order (Fig. 1*a*).

We analysed these markers in a large, consanguineous MM family[5,6]. MM is a recessive condition, so the locus can be defined by identifying regions of the genome that show homozygosity in affected individuals. Conversely, because of the high penetrance of this adult-onset condition, unaffected adult individuals are not expected to be homozygous by descent across the region. Analysis of haplotype homozygosity in this pedigree indicates that the disease gene lies between markers *D2S2111* and PAC3-H52 (data not shown). Based on our PAC mapping data, the physical distance for this interval is approximately 2.0 Mb. No recombination events were detected between four informative markers (markers cy172-H32 to PAC16-H41) and the disease locus in family MM-21 (Fig. 1*a*).

Identification of five genes expressed in muscle

Previously we mapped 22 ESTs and two genes (transforming growth factor α (TGFA) and β-adducin (ADD2)) to the MM/LGMD2B PAC contig[12] (Fig. 1*a*). PCR analysis of skeletal muscle cDNA indicates that five of these ESTs (A006G04, stSG1553R, WI-14958, TIGR-A004Z44 and WI-14051) map within the minimal genetic MM interval of MM and are expressed in skeletal muscle. We selected cDNA probes corresponding to each of these five ESTs for northern-blot analysis (Fig. 2). The tissue distribution, signal intensity and transcript size detected by the five cDNA probes varied. Probes corresponding to ESTs stSG1553R, TIGR-A004Z44 and WI-14958 detected signal in skeletal muscle. In addition, the cDNA corresponding to TIGR-A004Z44 detected a 3.8-kb brain-specific transcript instead of the 8.5-kb message that was present in other tissues. These five ESTs probably correspond to different genes, because the corresponding cDNA probes used for northern analysis derive from the 3′ end of messages, map to different

Table 1 • New polymorphic markers mapped to the MM/LGMD2B region

Marker	Repeat	Primers (5′→3′)	Annealing T_m (°C)	Size in PAC (bp)	No. of alleles[a]	Het[b]
PAC3-H52	CA	GATCTAACCCTGCTGCTCACC CTGGTGTGTTGCAGAGCGCTG	57	138	10	0.82
Cy172-H32[c]	CCAT	CCTCTCTTCTGCTGTCTTCAG TGTGTCTGGTTCACCTTCGTG	56	199	7	0.72
PAC35-PH2	CAT	TCCAAATAGAAATGCCTGAAC AGGTATCACCTCCAAGTGTTG	56	161	5	0.30
PAC16-H41	Complex	TACCAGCTTCAGAGCTCCCTG TTGATCAGGGTGCTCTTGG	58	280	4	0.41
Cy7-PH3	AAGG	GGAGAATTGCTTGAACCCAG TGGCTAATGATGTTGAACATTT	56	211	4	0.32

[a]Observed in 50 unrelated Europeans. [b]Heterozygosity index. [c]Located in intron 2 of the dysferlin gene.

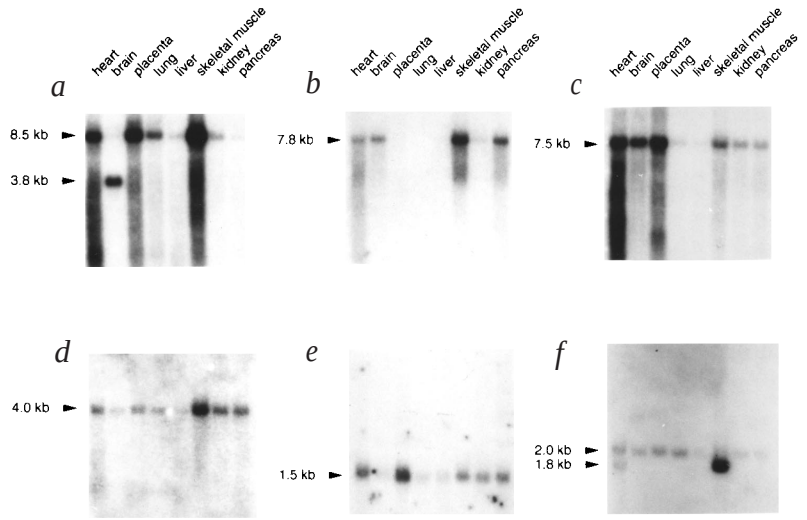

Fig. 2 Northern blot of five novel ESTs expressed in skeletal muscle. The same northern blot of poly A+ RNA from eight tissues (Clontech) was sequentially hybridized with: ***a***, the dysferlin probe 130347 (EST TIGR-A004Z44); ***b***, 172575 (EST WI-14958); ***c***, 48106 (EST WI-14051); ***d***, 184080 (EST stSG1553R); ***e***, 510138 (EST A006G04); and ***f***, β-actin as a control. RNA size markers are (kb): 9.5, 7.5, 4.4, 2.4 and 1.35.

positions in the MM/LGMD2B contig (Fig. 1*a*) and differ in their expression patterns.

Current database analysis suggests that three of these ESTs (stSG1553R, WI-14958 and WI-14051) do not match any known proteins[16] (see also http://www.ncbi.nlm.nih.-gov/UniGene/). A006G04 has weak homology with a protein sequence of unknown function that derives from *Caenorhabditis elegans*. TIGR-A004Z44 is only homologous to subdomains present in protein kinase C. The five genes corresponding to the ESTs are candidate MM/LGMD2B gene(s) because they are expressed in skeletal muscle and map within the minimal genetic interval of the MM/LGMD2B gene(s).

A full-length candidate MM gene

EST TIGR-A004Z44 has a strong skeletal muscle signal on the northern blot (Fig. 2) and it is also bracketed by genetic markers that show no recombination with the disease phenotype in family MM-21 (Fig. 1*a*). We therefore elected to clone and analyse the corresponding transcript as a candidate MM gene. Using the Unigene database, we identified a cDNA IMAGE clone (130347, 979 bp) that contained EST TIGR-A004Z44. With the insert of this cDNA clone, we screened a human skeletal muscle cDNA library and identified 50 cDNA clones. Clone cDNA10 contained the largest insert (approximately 6.5 kb; Fig. 1*b*), and was further subcloned by independent *Bam*HI and *Pst*I digestion. Both ends of all isolated cDNA and cDNA10 subclones were sequenced and the data subjected to sequence contig analysis. Two additional screens, first with the insert of cDNA10 and then with a 683-bp PCR product (A27-F2R2) amplified from the 5′ end of the cDNA contig, identified 87 additional cDNA clones. Clones B22 and B33 extended the 5′ end by 94 and 20 bp, respectively. The compiled sequence allowed us to generate a sequence of 6.9 kb (with 10-fold average coverage). Although we have not further extended the 5′ end of the gene to the 8.5 kb predicted by northern analysis (Fig. 2), we identified in this 6.9-kb sequence an open reading frame (ORF) of 6,243 bp (Fig. 3*a*). This ORF is preceded by an in-frame stop codon and begins with the sequence cgcaagcATGCTG; five of the first 7 bp are consistent with the Kozak consensus sequence for a start codon[17,18]. An alternative start codon, in the same frame, +75 bp downstream, appears less likely as a start site (GAGACGATGGGG). Thus, we believe that the entire coding region of this candidate gene has been identified, as represented by our 6.9-kb sequence contig.

a

```
   1 MLRVHILYAE NVHTPDTDIS DAYCSAVFAG VKKRTKVIKN SVNPVWNEGF
  51 EWDLKGIPLD QGSELHVVVK DHETMGRNRF LGEAKVPLRE VLATPSLSAS
 101 FNAPLLDTKK QPTGASLVLQ VSYTPLPGAV PLFPPPTPLE PSPTLPDLDV
 151 VADTGGEEDT EDQGLTGDEA EPFLDQSGGP GAPTTPRKLP SRPPPHYPGI
 201 KRKRSAPTSR KLLSDKPQDF QIRVQVIEGR QLPGVNIKPV VKVTAAGQTK
 251 RTRIHKGNSP LFNETLFFNL FDSPGELFDE PIFITVVDSR SLRTDALLGE
 301 FRMDVGTIYR EPRHAYLRKW LLLSDPDDFS AGARGYLKTS LCVLGPGDEA
 351 PLERKDPSED KEDIESNLLR PTGVALRGAH FCLKVFRAED LPQMDDAVMD
 401 NVKQIFGFES NKKNLVDPFV EVSFAGKMLC SKILEKTANP QWNQNITLPA
 451 MFPSMCEKMR IRIIDWDRLT HNDIVATTYL SMSKISAPGG EIEEEPAGAV
 501 KPSKASDLDD YLGFLPTFGP CYINLYGSPR EFTGFPDPYT ELNTGKGEGV
 551 AYRGRLLLSL ETKLVEHSEQ KVEDLPADDI LRVEKYLRRR KYSLFAAFYS
 601 ATMLQDVDDA IQFEVSIGNY GNKFDMTCLP LASTTQYSRA VFDGCHYYYL
 651 PWGNVKPVVV LSSYWEDISH RIETQNQLLG IADRLEAGLE QVHLALKAQC
 701 STEDVDSLVA QLTDELIAGC SQPLGDIHET PSATHLDQYL YQLRTHHLSQ
 751 ITEAALALKL GHSELPAALE QAEDWLLRLR ALAEEPQNSL PDIVIWMLQG
 801 DKRVAYQRVP AHQVLFSRRG ANYCGKNCGK LQTIFLKYPM EKVPGARMPV
 851 QIRVKLWFGL SVDEKEFNQF AEGKLSVFAE TYENETKLAL VGNWGTTGLT
 901 YPKFSDVTGK IKLPKDSFRP SAGWTWAGDW FVCPEKTLLH DMDAGHLSFV
 951 EEVFENQTRL PGGQWIYMSD NYTDVNGEKV LPKDDIECPL GWKWEDEEWS
1001 TDLNRAVDEQ GWEYSITIPP ERKPKHWVPA EKMYYTHRRR RWVRLRRRDL
1051 SQMEALKRHR QAEAEGEGWE YASLFGWKFH LEYRKTDAFR RRRWRRRMEP
1101 LEKTGPAAVF ALEGALGGVM DDKSEDSMSV STLSFGVNRP TISCIFDYGN
1151 RYHLRCYMYQ ARDLAAMDKD SFSDPYAIVS FLHQSQKTVV VKNTLNPTWD
1201 QTLIFYEIEI FGEPATVAEQ PPSIVVELYD HDTYGADEFM GRCTCQPSLE
1251 RMPRLAWFPL TRGSQPSGEL LASFELIQRE KPAIHHIPGF EVQETSRILD
1301 ESEDTDLPYP PPQREANIYM VPQNIKPALQ RTAIEILAWG LRNMKSYQLA
1351 NISSPSLVVE CGGQTVQSCV IRNLRKNPNF DICTLFMEVM LPREELYCPP
1401 ITVKVIDNRQ FGRRPVVGQC TIRSLESFLC DPYSAESPSP QGGPDDVSLL
1451 SPGEDVLIDI DDKEPLIPIQ EEEFIDWWSK FFASIGEREK CGSYLEKDFD
1501 TLKVYDTQLE NVEAFEGLSD FCNTFKLYRG KTQEETEDPS VIGEFKGLFK
1551 IYPLPEDPAI PMPPRQFHQL AAQGPQECLV RIYIVRAFGL QPKDPNGKCD
1601 PYIKISIGKK SVSDQDNYIP CTLEPVFGKM FELTCTLPLE KDLKITLYDY
1651 DLLSKDEKIG ETVVDLENRL LSKFGARCGL PQTYCVSGPN QWRDQLRPSQ
1701 LLHLFCQQHR VKAPVYRTDR VMFQDKEYSI EEIEAGRIPN PHLGPVEERL
1751 ALHVLQQQGL VPEHVESRPL YSPLQPDIEQ GKLQMWVDLF PKALGRPGPP
1801 FNITPRRARR FFLRCIIWNT RDVILDDLSL TGEKMSDIYV KGWMIGFEEH
1851 KQKTDVHYRS LGGEGNFNWR FIFPFDYLPA EQVCTIAKKD AFWRLDKTES
1901 KIPARVVFQI WDNDKFSFDD FLGSLQLDLN RMPKPAKTAK KCSLDQLDDA
1951 FHPEWFVSLF EQKTVKGWWP CVAEEGEKKI LAGKLEMTLE IVAESEHEER
2001 PAGQGRDEPN MNPKLEDPRR PDTSFLWFTS PYKTMKFILW RRFRWAIILF
2051 IILFILLLFL AIFIYAFPNY AAMKLVKPHS
```

b

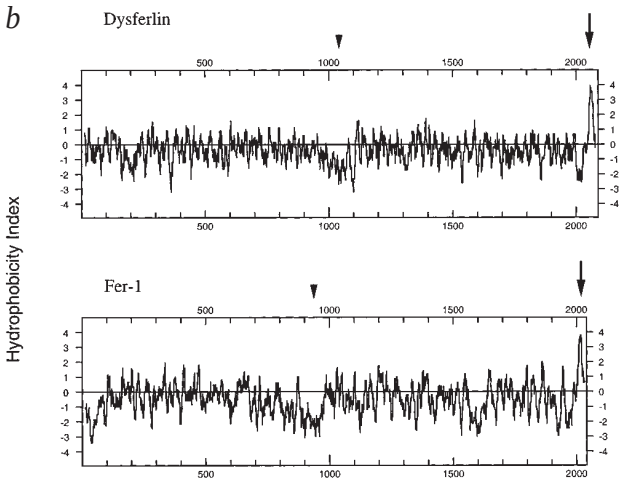

Fig. 3 Dysferlin amino acid sequence and hydrophobicity. ***a***, Sequence of the predicted 2,080 aa of dysferlin. The predicted membrane-spanning residues are in bold at the C terminus (2047–2063). Partial C2 domains are underlined. Bold, underlined sequences are putative nuclear targeting residues. Possible membrane retention sequences are enclosed in a box. ***b***, Comparison of the Kyle-Doolittle hydrophobicity plots reveals similarity of dysferlin protein and fer-1. On the Y-axis, increasing positivity corresponds to increasing hydrophobicity. Both proteins have a single, highly hydrophobic stretch at the C-terminal end (arrow). Both share regions of relative hydrophilicity at approximately residue 1,000 (arrowhead).

Table 2 • Mutations in dysferlin in distal myopathy and LGMD

Family	Ethnicity	Phenotype[a]	Exon	Mutation	BP change	Consequence
Homozygous						
59	Arabic	MM	3	537insA	insertion of A at 537	Frameshift
67	French	MM	20	Q605X	CAG→TAG at 2186	Nonsense
71	Spanish	DMAT	50	5966delG	deletion of G at 5966	Frameshift
75	Spanish	MM	50	5966delG	deletion of G at 5966	Frameshift
Heterozygous						
8	English	MM	49	E1883X	GAG→TAG at 5870	Nonsense
			52	6319+1G to A	Altered 5′splice site	Altered protein
56	Italian	MM, LGMD	36	I1298V	ATC→GTC at 4265	Missense
			54	R2042C	CGT→TGT at 6497	Missense
50	English	MM	50	H1857R	CAT→CGT at 5943	Missense
58	English	MM	51	6071,2del AG	deletion of AG at 6071,2	Frameshift

[a]MM, Miyoshi myopathy; DMAT, distal myopathy with anterior tibial onset; LGMD, limb girdle muscular dystrophy.

Identification of mutations in Miyoshi myopathy

To determine whether this 6.9-kb cDNA is mutated in MM we determined the genomic organization of the corresponding gene and identified the adjoining intronic sequence at each of the 55 exons encoding the cDNA (M.A. *et al*., in preparation). We then used single strand conformational polymorphism analysis (SSCP) to screen each exon in patients from 12 MM families. In patients for whom muscle biopsies were available, we also used RT-PCR to prepare cDNA for the candidate gene from the muscle biopsy specimen. This cDNA was amplified with a series of seven spanning primers, followed by 32 nested primers. Putative mutations identified in this way were confirmed by direct sequencing from genomic DNA using exon-specific intronic primers. These approaches identified nine different mutations in eight families (Table 2). Six of the nine mutations are predicted to alter translation of dysferlin, either through nonsense or frameshift changes; four of these five are homozygous and thus expected to result in loss of full-length dysferlin protein. For each mutated exon in these patients, we screened at least 50 control DNA samples (100 chromosomes) to determine the frequencies of the sequence variants. When possible, the parents and siblings of affected individuals were also screened to verify that defined mutations were appropriately co-inherited with the disease in each pedigree (Fig. 4). In two families (50 and 58; Table 2) heterozygous mutations were identified in one allele (respectively a missense mutation and a 2-bp deletion). We presume that we have not detected mutations in the other allele (or in three of our screened MM families) because (i) the mutant and normal SSCP products are indistinguishable, (ii) the defect in the affected allele deletes an entire exon, so that the SSCP (derived from the single, normal allele) appears normal, or (iii) the mutation lies outside of coding sequence (that is, in the promoter or a regulatory region of an intron). The disease-associated mutations were not detected in our controls and thus did not appear to be common polymorphisms.

Discussion

Our studies have defined a novel human skeletal muscle gene and provided evidence that it is implicated in human muscular dystrophy. The 6,243-bp ORF of this candidate MM gene is predicted to encode 2,080 aa (Fig. 1*c* and Fig. 3). At the amino acid level, this protein is homologous to the *C. elegans* protein fer-1 (27% identical, 57% identical or similar)[19,20]. In the absence of a well-defined function or set of homologues for the protein, we suggest that this dystrophy-associated, fer-1-like protein be designated 'dysferlin' (gene symbol *DYSF*).

The fer-1 protein was originally identified through molecular genetic analysis of a class of fertilization-defective *C. elegans* mutants in which spermatogenesis is abnormal[19]. The mutant fer-1 spermatozoa have defective mobility and show imperfect fusion of membranous organelles[20,21]. Like fer-1, dysferlin is a large protein with an extensive, highly charged hydrophilic region and a single predicted membrane-spanning region at the C terminus (Fig. 3*b*). There is a membrane retention sequence 3′ to the membrane-spanning stretch, indicating that the protein may be preferentially targeted to either endoplasmic or sarcoplasmic reticulum, probably as a type II protein (that is, with the NH_2 end and most of the following protein located within the cytoplasm; Fig. 1*c*). Several nuclear membrane targeting sequences are predicted within the cytoplasmic domain of the protein (http://psort.nibb.ac.jp/form.html).

It is of particular interest that the cytoplasmic component of this protein contains four motifs homologous to C2 domains. C2 domains are intracellular protein modules composed of 80–130 amino acids[22]. Originally identified in a calcium-dependent iso-

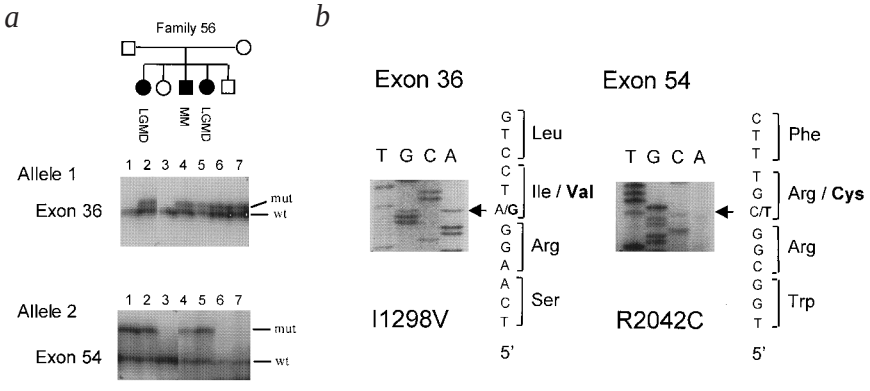

Fig. 4 Representative pedigree with dysferlin mutations. ***a***, Each member of the pedigree is illustrated above the corresponding SSCP analysis. For each affected individual (solid symbols) shifts are evident in alleles 1 and 2, corresponding respectively to exons 36 and 54. As indicated, the allele 1 and 2 variants are transmitted respectively from the mother and the father. The two affected daughters in this pedigree have the LGMD phenotype, whereas their affected brother has a pattern of weakness suggestive of MM. ***b***, The base pair sequence of exons 36 and 54 shows mutations A4265G (I1298V) and C6497T (R2042C).

form of protein kinase C (ref. 23), C2 domains are present in numerous proteins. These domains often arise in approximately homologous pairs described as double C2 or DOC2 domains. One DOC2 protein, DOC2α, is brain-specific and highly concentrated in synaptic vesicles[24], whereas another, DOC2β, is ubiquitously expressed[25]. Many C2 modules can fold to bind calcium, thereby initiating signalling events such as phospholipid binding. At distal nerve terminals, for example, the synaptic vesicle protein synaptotagmin has two C2 domains that, upon binding calcium, permit this protein to interact with syntaxin, triggering vesicle fusion with the distal membrane and neurotransmitter release[26].

The four dysferlin C2 domains are located at aa 32–82, 431–475, 1160–1241 and 1582–1660 (Figs 1*c* and 3*a*). It is through these regions that dysferlin has homology to any proteins other than fer-1. Each of these segments in dysferlin is considerably smaller than a typical C2 domain. Moreover, these segments are more widely separated in comparison with the paired C2 regions in synaptotagmin, DOC2α and β and related C2-positive proteins. For this reason, it is difficult to predict whether the four, relatively short C2 domains in dysferlin function analogously to conventional C2 modules. That dysferlin might, by analogy with synaptotagmin, signal events such as membrane fusion is suggested by the fact that fer-1 deficient worms show defective membrane organelle fusion within spermatozoa[21].

There are several reasons why we believe that mutations in the dysferlin gene cause muscular dystrophy. First, we have detected nine different mutations in the dysferlin gene in patients with distal and proximal muscular dystrophy, and these mutations are inherited in pedigrees in a mendelian manner; affected individuals receive one affected allele from each parent. Second, six of these mutations are predicted to prevent expression of the corresponding dysferlin protein, for example, four families (59, 67, 71 and 75) have homozygous defects that are predicted to produce a truncated or absent dysferlin protein. The mutation is identical in two families (1-bp deletion at codon 1865) of Spanish origin, and studies are now underway to determine if this reflects a founder effect. In family 8, the mutation in one allele is predicted to produce a stop codon (E1183X), whereas the change in the other allele alters a predicted splice site. Third, other mutations are predicted to cause missense changes, although none of the 50 control samples assayed revealed these missense mutations. That is, the missense variants do not appear to be polymorphisms (defined as changes that occur in 1% or more of the chromosomes from normal individuals).

Although the present study has focused on MM (which begins distally in the calf muscles), our data indicate that there is phenotypic heterogeneity among different individuals with the same dysferlin gene mutation. Thus, in one pedigree the same mutation produced MM in a brother and LGMD (which begins in the proximal hip and leg muscles) in his sisters with the same mutation. Moreover, two other families, one with MM and another with familial distal weakness affecting the anterior tibial muscles, also have an identical mutation. Despite the divergent distributions of muscle weakness, all of the patients share two features: childhood or early adult onset and pronounced elevations of serum creatine kinase levels.

Further investigation of dysferlin will be instructive from several perspectives. First, because the protein is novel, it will be informative to study its normal physiological properties and function, and to conduct parallel analyses of dysferlin isoforms in other tissues. Second, known defects causing common, non-myotonic muscular dystrophies fall into three broad categories: (i) loss of integrity of the muscle cell membrane (for example, defects in dystrophin, dystrophin-associated proteins[2,3,27,28] and α7-integrin[29]); (ii) altered enzymatic function of cytosolic, skeletal muscle calpain[30]; and (iii) altered energy generation[31]. Preliminary data indicate that membrane proteins and energy generation are normal in patients with dysferlin gene mutations (Table 2, family 71; I.I., pers. comm.). Thus, the dysferlinopathies will probably represent a new, fourth category of muscular dystrophy. Third, dysferlin is the first distal muscular dystrophy gene. (Although genetic analysis has implicated the titin gene in Udd distal dystrophy in Scandanavia, the gene defects have not yet been identified[32].) A detailed understanding of the dysferlinopathies should elucidate factors that determine whether the gradient of muscle degeneration is proximally or distally predominant. Genetic analysis of individuals or subsets of families that share the same dysferlin gene mutation but have different distributions of weakness should help define genetic factors that modify the associated muscle phenotype.

Methods

Identification of PAC clones containing repeat sequences. DNA from the PAC clones spanning the MM/LGMD2B region[12] was spotted onto Hybond N+ membranes (Amersham). The filters were hybridized independently with the following γ-^{32}P (Du Pont) labelled repeat sequences: 1, $(CA)_{15}$; 2, pool of $(ATT)_{10}$, $(GATA)_8$ and $(GGAA)_8$; 3, pool of $(GAAT)_8$, $(GGAT)_8$ and $(GTAT)_8$; and 4, pool of $(AAG)_{10}$ and $(ATC)_{10}$. Hybridization and washing of the filters were carried out at 55 °C following standard protocols[33].

Identification of repeat sequences and repeat typing. Miniprep DNA of PAC clones containing repeat sequences were digested with *Hin*dIII and *Pst*I and ligated into pBluescript II KS (+) vector (Stratagene) digested with the same endonucleases. Filters of the PAC subclones were hybridized to the γ-^{32}P labelled repeats that detected the respective PACs. For clones with an insert size greater than 1 kb whose repeat sequences could not be identified by a single round of sequencing, the inserts were further subcloned by digestion with *Hae*III and ligation in *Eco*RV-digested pZero-2.1 vector (Invitrogen). Miniprep DNA of the positive subclones were subjected to manual dideoxy sequencing with Sequenase (US Biochemicals). Primer pairs spanning the repeat sequences were selected using the computer program Oligo (version 4.0, National Biosciences). Primer sequences are shown (Table 1). All oligonucleotides were synthesized by Integrated DNA Technologies. PCR typing of the repeat markers followed described protocols[5].

Analysis of ESTs in skeletal muscle. Marathon-ready™ (Clontech) skeletal muscle cDNA (2 μl at approximately 0.1 ng/μl) was used as template in a PCR (10 μl) reaction for analysis of muscle expression of ESTs. The PCR conditions were the same as for the PCR typing of repeat markers.

cDNA clones (130347, 48106, 172575, 184080 and 510138) corresponding to the five ESTs that are expressed in muscle (respectively TIGR-A004Z44, WI-14051, WI-14958, stSG1553R and A006G04) were selected from the UniGene database (http:/www.ncbi.nlm.nih.gov/UniGene/) and obtained (Genome Systems). The cDNA probes were used to screen the MM/LGMD2B PAC filters to confirm that they mapped to the expected position in the MM/LGMD2B contig. A northern blot (Clontech) of multiple human tissues was sequentially hybridized to the five cDNA probes and a control β-actin cDNA at 65 °C following standard hybridization and washing protocols[33]. Between hybridizations, probes were removed by boiling the blot at 95–100 °C for 4–10 min with 0.5% SDS. The blot was then re-exposed for 24 h to confirm the absence of previous hybridization signals before proceeding with the next round of hybridization.

cDNA library screening. Approximately 1×10^6 recombinant clones of a λgt11 human skeletal muscle cDNA library (Clontech) were plated and screened following standard techniques[33]. The initial library screening was performed using the insert released from cDNA clone 130347 that contained EST TIGR-A0044Z44, corresponding to the 3′ end of the gene. Positive phages were plaque purified and phage DNA was isolated according to standard procedures[33]. The inserts of the positive clones were released by *Eco*RI digestion of phage DNA and subsequently subcloned into the *Eco*RI site of pBluescript II (KS+) vector (Stratagene).

DNA sequencing and contig analysis. To achieve complete coverage of the cDNA sequence, one of the positive clones (cDNA10) was digested indepen-

dently with *Bam*HI and *Pst*I and subcloned into pBluescript vector. Miniprep DNA of cDNA clones and subclones of cDNA10 were prepared using the Qiagen plasmid Miniprep kit. Sequencing was carried out from both ends of each clone using the SequiTherm EXCEL™ long-read DNA sequencing kit (Epicenter), fluorescent-labelled M13 forward and reverse primers and an LI-COR sequencer. Assembly of cDNA contigs and sequence analysis were performed using Sequencher software (Gene Codes).

Identification of exon-intron boundaries and mutation detection. To identify exon-intron boundaries within the gene, PAC DNA was extracted with the standard Qiagen-Mini Prep protocol. Direct sequencing was performed with DNA Sequence System (Promega) using ^{32}P end-labelled primers[34]. Exon-intron boundaries were identified as the sites where genomic and cDNA sequences diverged.

Total genomic DNA (20 ng) from immortalized lymphocyte cell lines was used as a template for PCR amplification analysis of each exon using primers located in the adjacent introns. SSCP analysis was performed as described[35]. In patients for whom muscle biopsies were available, mRNA was isolated using RNA-STAT-60™ (Tel-Test) and first-strand cDNA was synthesized from total RNA (1–2 μg) with MMLV reverse transcriptase and random hexamer primers (Life Technologies). This product (3 μl) was used for PCR amplification. Seven sets of primers were designed for muscle cDNA and overlapping cDNA fragments suitable for SSCP analysis were amplified. After initial denaturation at 94 °C for 2 min, amplification was performed using 30 cycles at 94 °C for 30 s, 56 °C for 30 s and 72 °C for 60 s. The sequences of polymorphisms detected by SSCP analysis were determined by the dideoxy termination method using the Sequenase kit (US Biochemicals). In some instances, the base-pair changes predicted corresponding changes in restriction endonuclease recognition sites. Such alterations in restriction endonuclease sites were verified by digesting the relevant PCR products with the appropriate restriction endonucleases.

Primers used for SSCP screening and exon sequencing are as follows: exon 3 F3261, 5′–TCTCTTCTCCTAGAGGGCCATAG–3′ and R3261, 5′–CTGTTCCTCCCCATCGTCTCATGG–3′; exon 20 F3121, 5′–GCTCCTCCCGTGACCCTCTG–3′ and R3121, 5′–GGGTCCCAGCCAGGAGCACTG–3′; exon 36 F2102, 5′–CCCCTCTCACCATCTCCTGATGTG–3′ and R2111, 5′–TGGCTTCACCTTCCCTCTACCTCGG–3′; exon 49 F1081, 5′–TCCTTTGGTAGGAAATCTAGGTGG–3′ and R1081, 5′–GGAAGCTGGACAGGCAAGAGG–3′; exon 50 F1091, 5′–ATATACTGTGTTGGAAATCTTAATGAG–3′ and R1091, 5′–GCTGGCACCACAGGGAATCGG–3′; exon 51 F1101, 5′–CTTTGCTTCCTTGCATCCTTCTCTG–3′ and R1101, 5′–AGCCCCCATGTGCAGAATGGG–3′; exon 52 F1111, 5′–GGCAGTGATCGAGAAACCCGG–3′ and R1111, 5′–CATGCCCTCCACTGGGGCTGG–3′; exon 54 F1141, 5′–GGATGCCCAGTTGACTCCGGG–3′ and R1141, 5′–CCCCACCACAGTGTCGTCAGG–3′. The sequence primers used to amplify all exons and the cDNA are available upon request.

GenBank accession number. Dysferlin cDNA and protein, AF075575.

Acknowledgements

The Cecil B. Day Investment Company and the Muscular Dystrophy Association funded many aspects of this project. R.B. is supported by NIH grants 5PO1AG12992, 5RO1N834913A and 5PO1NS31248. J.L. was funded by a post-doctoral fellowship of the Medical Research Council of Canada. M.A. received a stipend from the Muscular Dystrophy Association and the Pierre L de Bourgknecht Fund for ALS Research. P. de J. is supported by NIH (NIHGRI) grant 5RO1 HG1165 and U.S. Department of Energy grant DE-FCO3-96ER62294. E.S. is funded by the Canadian Genome Analysis and Technology Program. B.A.H. and K.J.B. were recipients respectively of NRSA fellowship IF32N10064 and a stipend from the Deater Foundation. I.I. and C.S. received grants from the Fondo Investigationes Sanitarias (F.I.S. 96/0863 and 98/5065 B.A.E.). C.A. is funded by grants from Telethon 916C and C19 and from a Finalized Research Grant 559/01/95 of Veneto Region. K.A. is supported by the Ministry of Health and Welfare (Japan) grants for the Center of Excellence and Brain Sciences. J.U. and M.F. are supported by the AFM, which funded family DNA collections in France; M.F. is also funded by Inserm (U153). We are grateful to L. Hayward for discussions of the predicted properties of dysferlin.

Received 10 July; accepted 5 August, 1998.

1. Hoffman, E. Clinical and histopathological features of abnormalities of the dystrophin-based membrane cytoskeleton. *Brain Pathol.* **6**, 49–61 (1996).
2. Brown, R.J. Dystrophin-associated proteins and the muscular dystrophies. *Annu. Rev. Med.* **48**, 457–466 (1997).
3. Bonnemann, C., McNally, E. & Kunkel, L. Beyond dystrophin: current progress in the muscular dystrophies. *Curr. Opin. Pediatr.* **8**, 569–582 (1996).
4. Miyoshi, M., Kawai, H., Iwasa, M., Kuska, K. & Nishino, H. Autosomal recessive distal muscular dystrophy as a new type of progressive muscular dystrophy. *Brain* **109**, 31–54 (1986).
5. Bejaoui, K. *et al.* Linkage of Miyoshi myopathy (distal autosomal recessive muscular dystrophy) locus to chromosome 2p12-13. *Neurology* **45**, 768–772 (1995).
6. Bejaoui, K. *et al.* Genetic fine mapping of the Miyoshi myopathy locus and exclusion of eight candidate genes. *Neurogenetics* **1**, 189–196 (1998).
7. Bashir, R. *et al.* A gene for autosomal recessive limb-girdle muscular dystrophy maps to chromosome 2p. *Hum. Mol. Genet.* **3**, 455–457 (1994).
8. Bashir, R. *et al.* Genetic and physical mapping at the limb-girdle muscular dystrophy locus (LGMD2B) on chromosome 2p. *Genomics* **33**, 46–52 (1996).
9. Passos-Bueno, M. *et al.* Confirmation of the 2p locus for the mild autosomal recessive limb-girdle muscular dystrohy gene (LGMD2B) in three families allows refinement of the candidate region. *Genomics* **27**, 192–195 (1995).
10. Weiler, T. *et al.* Limb-girdle muscular dystrophy and Miyoshi myopathy in an aboriginal Canadian kindred map to LGMD2B and segregate with the same haplotype. *Am. J. Hum. Genet.* **59**, 872–878 (1996).
11. Illarioshkin, S. *et al.* Refined genetic location of the chromosome 2p-linked progressive muscular dystrophy gene. *Genomics* **42**, 345–348 (1997).
12. Liu, J. *et al.* Generation of a 3-Mb PAC contig spanning the Miyoshi myopathy/limb girdle muscular dystrophy (MM/LGMD2B) locus on chromosome 2p13. *Genomics* **49**, 23–29 (1998).
13. Hudson, T. *et al.* An STS-based map of the human genome. *Science* **270**, 1945–1954 (1995).
14. Hudson, T. *et al.* Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* **13**, 622–629 (1992).
15. Gastier, J. *et al.* Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. *Hum. Mol. Genet.* **4**, 1829–1836 (1995).
16. Schuler, G. *et al.* A gene map of the human genome. *Science* **274**, 540-546 (1996).
17. Kozak, M. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8133 (1987).
18. Kozak, M. The scanning model for translation: an update. *J. Cell Biol.* **108**, 229–241 (1989).
19. Argon, Y. & Ward, S. *Caenorhabditis elegans* fertilization-defective mutants with abnormal sperm. *Genetics* **96**, 413–433 (1980).
20. Achanzar, W. & Ward, S. A nematode gene required for sperm vesicle fusion. *J. Cell Sci.* **110**, 1073–1081 (1997).
21. Ward, S., Argon, Y. & Nelson, G. Sperm morphogenesis in wild-type and fertilization-defective mutants of *Caenorhabditis elegans*. *J. Cell Biol.* **91**, 26–44 (1981).
22. Rizo, J. & Südhof, T. C2-domains, structure and function of a universal Ca+-binding domain. *J. Biol. Chem.* **273**, 15882–15897 (1998).
23. Nishizuka, Y. The molecular heterogeneity of protein kinase C and its implications for cellular regulation. *Nature* **334**, 661–665 (1988).
24. Orita, S. *et al.* Doc2: a novel brain protein having two repeated C2-like domains. *Biochem. Biophys. Res. Comm.* **206**, 439–448 (1995).
25. Sakaguchi, G., Orita, S., Maeda, M., Igarashi, H. & Takai, Y. Molecular cloning of an isoform of Doc2 having two C2-like domains. *Biochem. Biophys. Res. Comm.* **217**, 1053–1061 (1995).
26. Sudhof, T. & Rizo, J. Synaptotagmins: C2-domain proteins that regulate membrane traffic. *Neuron* **17**, 379–388 (1996).
27. Henry, M. & Campbell, K. Dystroglycan: an extracellular matrix receptor linked to the cytoskeleton. *Curr. Opin. Cell Biol.* **8**, 625–631 (1996).
28. Straub, V. & Campbell, K. Muscular dystrophies and the dystrophin-glycoprotein complex. *Curr. Opin. Neurol.* **10**, 168–175 (1997).
29. Hayashi, Y. *et al.* Mutations in the integrin α7 gene cause congenital myopathy. *Nature Genet.* **19**, 94–97 (1998).
30. Richard, I. *et al.* Mutations in the proteolytic gene calpain 3 cause limb girdle muscular dystrophy type 2A. *Cell* **81**, 27–40 (1995).
31. Schon, E.A. & DiMauro, S. Mitochondrial DNA mutations and pathogenesis. *J. Bioenerg. Biomembr.* **29**,131–149.
32. Haravuori, H. *et al.* Assignment of the tibial muscular dystrophy locus to chromosome 2q31. *Am. J. Hum. Genet.* **62**, 620-626 (1998).
33. Sambrook, J., Fritsch, F. & Maniatis, T. *Molecular Cloning: A Library Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).
34. Benes, V., Kilger, C., Voss, H., Pääbo, S. & Ansorge, W. Direct primer walking on P1 plasmid DNA. *Biotechniques* **23**, 98–100 (1997).
35. Aoki, M. *et al.* Mutations in the glutamate transporter EAAT2 gene do not cause abnormal EAAT2 transcripts in amyotrophic lateral sclerosis. *Ann. Neurol.* **43**, 645–653 (1998).