Cell, Vol. 81, 27–40, April 7, 1995, Copyright © 1995 by Cell Press

Mutations in the Proteolytic Enzyme Calpain 3 Cause Limb-Girdle Muscular Dystrophy Type 2A

Isabelle Richard,* Odile Broux,* Valérie Allamand,* Françoise Fougerousse,* Nuchanard Chiannilkulchai,* Nathalie Bourg,* Lydie Brenguier,* Catherine Devaud,* Patricia Pasturaud,* Carinne Roudaut,* Dominique Hillaire,* Maria-Rita Passos-Bueno,† Mayana Zatz,† Jay A. Tischfield,‡ Michel Fardeau,§ Charles E. Jackson,‖ Daniel Cohen,# and Jacques S. Beckmann*#

*Généthon
1, rue de l'Internationale
91000 Evry
France
†Department de Biologia
Instutio de Biociências
Universidad de São Paulo
São Paulo 05508-900
Brazil
‡Department of Medical and Molecular Genetics
Indiana University School of Medicine
Indianapolis, Indiana 46202-5251
§Institut National de la Santé et de la Recherche Médicale
Unité 153
Centre National de la Recherche Scientifique
Unité A614
17, rue du Fer à Moulin
75005 Paris
France
‖Henry Ford Hospital
Detroit, Michigan 48202
#Fondation Jean Dausset
Centre d'Etudes du Polymorphisme Humain
27, rue Juliette Dodu
75010 Paris
France

Summary

Limb-girdle muscular dystrophies (LGMDs) are a group of inherited diseases whose genetic etiology has yet to be elucidated. The autosomal recessive forms (LGMD2) constitute a genetically heterogeneous group with *LGMD2A* mapping to chromosome 15q15.1-q21.1. The gene encoding the muscle-specific calcium-activated neutral protease 3 (CANP3) large subunit is located in this region. This cysteine protease belongs to the family of intracellular calpains. Fifteen nonsense, splice site, frameshift, or missense calpain mutations cosegregate with the disease in LGMD2A families, six of which were found within La Réunion island patients. A digenic inheritance model is proposed to account for the unexpected presence of multiple independent mutations in this small inbred population. Finally, these results demonstrate an enzymatic rather than a structural protein defect causing a muscular dystrophy, a defect that may have regulatory consequences, perhaps in signal transduction.

Introduction

The term limb–girdle muscular dystrophy (LGMD) was first proposed by Walton and Nattrass (1954) as part of a classification of muscular dystrophies. LGMD is characterized by progressive symmetrical atrophy and weakness of the proximal limb muscles and elevated serum creatine kinase. The symptoms usually begin during the first two decades of life, and the disease gradually worsens, often resulting in loss of walking ability 10 or 20 years after onset (Bushby, 1994). Yet the precise nosological definition of LGMD still remains unclear. Consequently, various neuromuscular diseases such as facioscapulohumeral, Becker muscular dystrophies, and especially spinal muscular atrophies have occasionally been classified under this diagnosis. These issues highlight the difficulty in undertaking an analysis of the molecular and genetic defect(s) involved in this pathology.

Both autosomal dominant and recessive transmission have been reported, the latter being more common with an estimated prevalence of 10^{-5} (Emery, 1991). The localization of a gene on chromosome 15 (*LGMD2A*, MIM 253600; Beckmann et al., 1991) has provided proof for the genetic basis of one form of recessive LGMDs. Subsequent genetic analyses confirmed this chromosome 15 localization (Young et al., 1992; Passos-Bueno et al., 1993). The latter group also demonstrated genetic heterogeneity of this disease, while a recent study localized the *LGMD2B* gene to chromosome 2 (Bashir et al., 1994). Yet there is evidence that at least one other locus is involved, since genetic heterogeneity was demonstrated in the inbred Indiana Amish LGMD2 kindreds (Allamand et al., 1995a).

The nonspecific nosological definition, the relatively low prevalence, and genetic heterogeneity of this disorder limit the number of families that can be used to restrict the genetic boundaries of the *LGMD2A* interval. No cytogenetic abnormalities have been reported. Immunogenetic studies of the dystrophin-associated protein complex (Matsumara et al., 1993) and cytoskeletal or extracellular matrix proteins (e.g., Tomé et al., 1994) failed to demonstrate any deficiency. In addition, there is no known specific physiological feature or animal model to suggest a candidate gene. Thus, there was no alternative to a positional cloning strategy.

Detailed genetic and physical maps of the 15q15.1-q21.1 *LGMD2A* region were established. Construction and analysis of a 10–12 Mb yeast artificial chromosome (YAC) contig (Fougerousse et al., 1994) permitted the mapping of 33 polymorphic markers within this interval and to narrow the *LGMD2A* region to between *D15S514* and *D15S222*. Furthermore, extensive analysis of linkage disequilibrium suggested a likely position for the gene in the proximal part of the contig (Allamand et al., 1995b).

cDNA selection (Tagle et al., 1993) for muscle-expressed sequences encoded by this interval led to the identification of five known genes and 10 newly expressed sequences,

Table 1. PCR Primers for the Localization of the *CANP3* Gene

Primer Name	Primer Sequence (5'–3')	Position within the cDNA	PCR Product Size (Base Pairs)		Annealing Temperature (in Celsius)
			cDNA	Genomic DNA	
CANP3-in2.a	ATGGAGCCAACAGAACTGAC	341–360	108	1758	58
CANP3-in2.m	GTATGACTCGGAAAAGAAGGT	428–448			
CANP3-in13.a	TAAGCAAAAGCAGTCCCCAC	1893–1912	64	1043	58
CANP3-in13.m	TTGCTGTTCCTCACTTTCCTG	1936–1956			
CANP3-6a3.a	GTTTCATCTGCTGCTTCGTT	2342–2361	130	818	56
CANP3-6a3.m	CTGGTTCAGGCATACATGGT	2452–2471			
CANP3-ex1ter.a	TTCTTTATGTGGACCCTGAGTT	218–239	76	76	55
CANP3-ex1ter.m	ACGAACTGGATGGGGAACT	275–293			

all potential candidate genes (Chiannilkulchai et al., 1995). One of these, previously cloned by Sorimachi et al. (1989), appeared also to be a functional candidate gene, as it encodes a muscle-specific protein, CANP3 (named for calpain large polypeptide L3), which belongs to the calpain family (or calcium-activated neutral protease [CANP]; EC 3.4.22.17).

Calpains are nonlysosomal intracellular cysteine proteases (Murachi, 1989; Suzuki and Ohno, 1990; Croall and Demartino, 1991). The mammalian calpains include two ubiquitous proteins, CANP1 and CANP2, as well as two stomach-specific proteins (Sorimachi et al., 1993a) and CANP3. The ubiquitous enzymes consist of heterodimers with distinct large subunits associated with a common small subunit (Murachi, 1989), all of which are encoded by different genes (Ohno et al., 1989). The association of tissue-specific large subunits with a small subunit has not yet been demonstrated. The large subunits of calpains can be subdivided into four domains (Ohno et al., 1984). Domains I and III, whose functions remain unknown, show no homology with known proteins. The former, however, might be important for the regulation of the proteolytic activity (Imajoh et al., 1986). Domain II shows similarity with other cysteine proteases, which share histidine, cysteine, and asparagine residues at their active sites (Sorimachi et al., 1989). Domain IV comprises four EF hand structures that are potential calcium-binding sites. In addition, three unique regions with no known homology are present in the muscle-specific CANP3 protein, namely NS, IS1, and IS2, the latter containing a nuclear translocation signal (Sorimachi et al., 1989). These regions may be important for the muscle-specific function of CANP3.

We have determined the genomic organization of the human *CANP3* gene, which consists of 24 exons and extends over 40 kb, 35 kb of which have been sequenced. A systematic screening of this gene in LGMD families led to the identification of 15 different mutations, establishing that mutational events in *CANP3* are responsible for

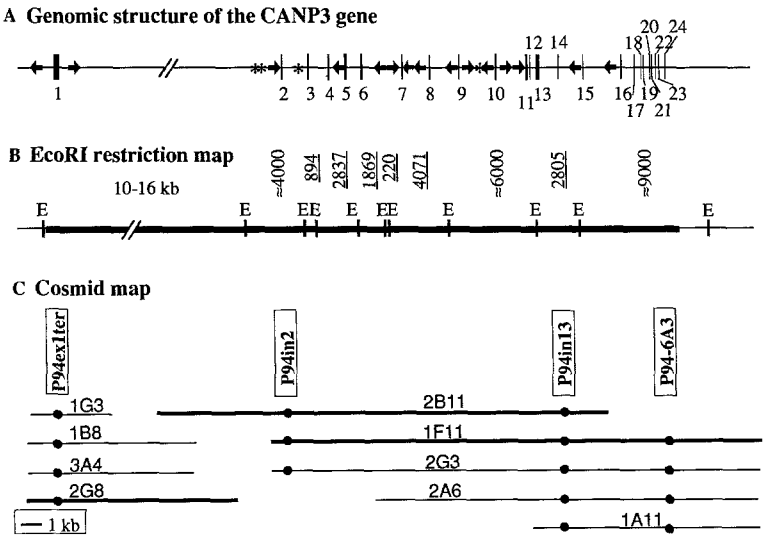

Figure 1. Genomic Organization of the *CANP3* Gene

(A) The *CANP3* gene covers a 40 kb region of which 35 kb were sequenced. Introns and exons are drawn to scale, exons being indicated by numbered vertical bars. Intron 1 is the largest one and remains to be fully sequenced as well as intron 8. Positions of intragenic microsatellites are indicated by asterisks. Closed and cross-hatched arrows indicate the orientation of Alu and *MER2* repeat sequences, respectively.

(B) EcoRI restriction map. An EcoRI (E) restriction map was established with cosmids from this region. The location of the *CANP3* gene is indicated by a closed bar. The size of the corresponding fragments are indicated and underlined when determined by sequence analysis.

(C) Cosmid map of the *CANP3* gene region. Cosmids, from a library constructed by subcloning YAC 774G4, are presented as lines. Overlaps are based on sequence-tagged site information from a cosmid contig established by I. R. et al. (unpublished data). Dots on lines indicate positive sequence-tagged sites, which are boxed in rectangles. A minimum of three cosmids covers the entire gene.

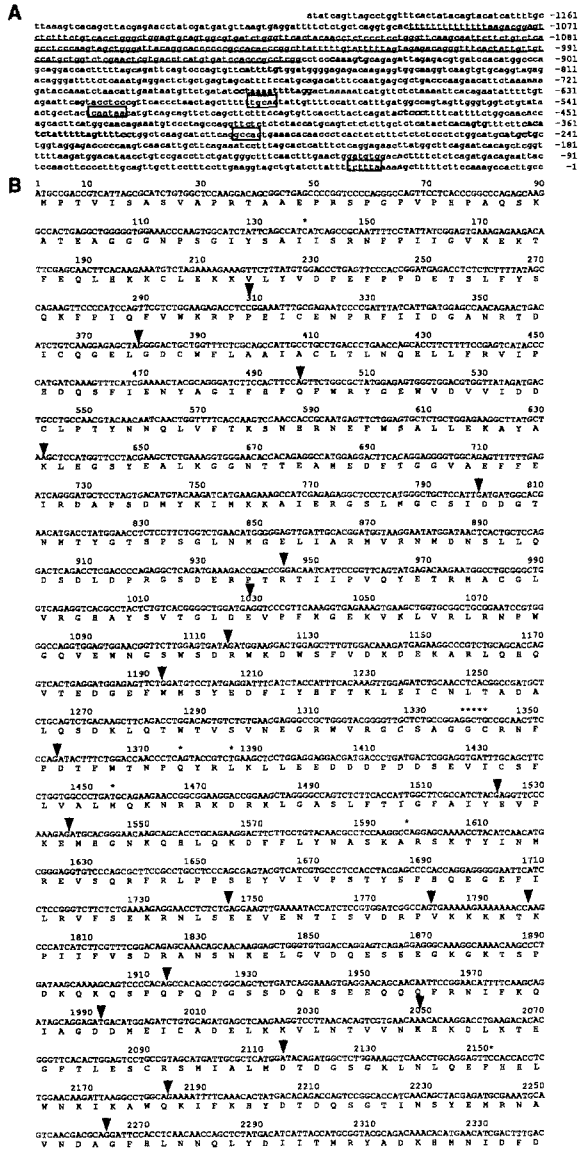

Figure 2. Sequence of the Human *CANP3* cDNA and Flanking 5′ and 3′ Genomic Regions

(A) and (C) show the polyadenylation signal and putative CAAT and TATAA sites, which are boxed. The putative Sp1 (position −477 to −472), MAF2-binding sites (position −364 to −343), and CArG box (position −685 to −672) are in bold. The Alu sequence present in the 5′ region is underlined. (B) shows the corresponding amino acids, indicated below the sequence. The coding sequence between the ATG initiation codon and the TGA stop codon is 2466 bp, encoding for an 821 amino acid protein. The adenine in the first methionine codon has been assigned position 1. Locations of introns within the *CANP3* gene are indicated by arrowheads. Nucleotides that differ from the previously published ones are indicated by asterisks.

LGMD2A. Finally, this demonstrates a case of muscular dystrophy resulting from an enzymatic rather than a structural defect.

Results

Localization of *CANP3* within the *LGMD2A* Interval

cDNA capture using YACs from the *LGMD2A* interval allowed the identification of 15 positional candidate genes. *CANP3* was one of two transcripts that showed muscle-specific expression as evidenced by Northern blot analysis (Chiannilkulchai et al., 1995). The *CANP3* gene had previously been localized to chromosome 15 (Ohno et al., 1989). Primers (Table 1; Figure 1) designed from different parts of the published human cDNA sequence (Sorimachi et al., 1989) were used to position the gene in a region previously defined as 15q15.1-q21.1 (Richard et al., 1994) and, more precisely, on three YACs (774G4, 926G10, and 923G7) localized in this region, between *D15S512* and *D15S488*, in a candidate region suggested by linkage disequilibrium studies (Allamand et al., 1995b).

These primers were also used to screen a cosmid library derived from YAC 774G4 (I. R. et al., unpublished data). Five cosmids were identified. Experiments with different primer pairs (Table 1) established that these cosmids cover all *CANP3* exons except exon 1 and that a second group of four cosmids contain this exon (Figure 1C). A minimal set of three overlapping cosmids (2G8–2B11–1F11) covers the entire gene. DNA from these cosmids was used to construct an EcoRI restriction map of this region (Figure 1B).

The *CANP3* Gene Sequence

Most of the sequences were obtained through shotgun sequencing of partial digests of cosmid 1F11 subcloned in M13 and Bluescript vectors and by walking with internal primers. The sequence assembly was in agreement with the restriction map of the cosmids. Sequences of exon 1 and adjacent regions were obtained by sequencing cosmid DNA or polymerase chain reaction (PCR) products from human genomic DNA. The first and eighth introns are still not fully sequenced, but there is evidence that they may be between 10 and 16 kb and about 2 kb in length, respectively (based on hybridization of restriction fragments; data not shown). The entire gene spans more than 40 kb.

The determined *CANP3* sequence completes the published human cDNA sequence (Sorimachi et al., 1989). It contains the missing 129 bases corresponding to the N-terminal 43 amino acids (Figure 2B). It also differs from the published sequence at 11 positions, three of which occur at third base positions of codons and preserve the encoded amino acid sequence. The other eight differences lead to changes in amino acid composition (Figure 2). As these different exons were sequenced repeatedly on at least 15 distinct genomes, we are confident that the sequence shown in Figure 2 represents an authentic

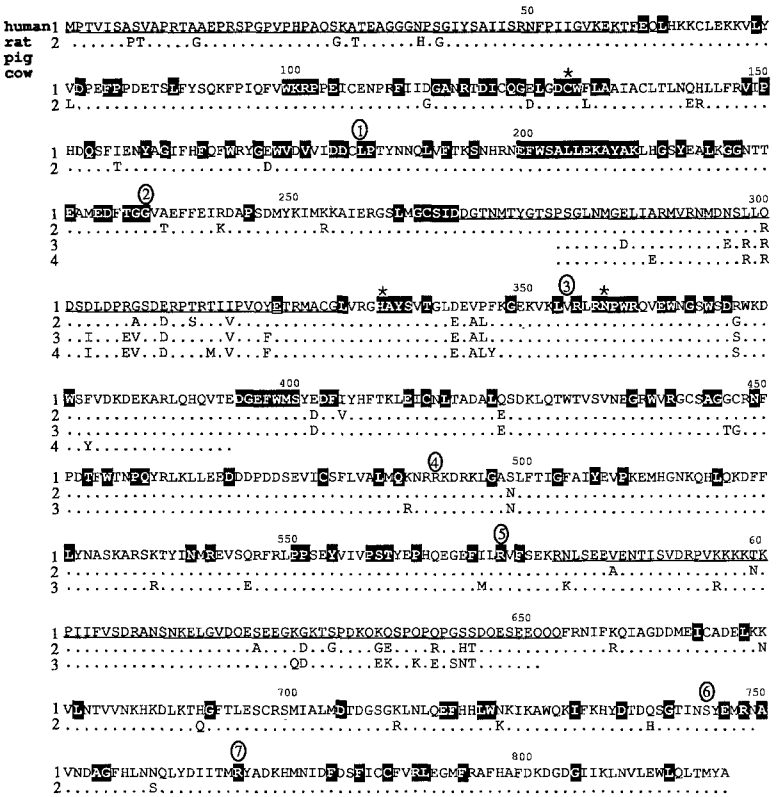

Figure 3. Alignments of Amino Acid Sequences of the Muscle-Specific Calpains

The human CANP3 protein is shown on the first line. The three muscles-specific sequences (NS, IS1, and IS2) are underlined, and the three active site residues are indicated by asterisks. The second line corresponds to the rat sequence (SwissProt accession number P16259). The third and fourth lines show the deduced amino acid sequences encoded by pig and cow expressed sequence tags (GenBank/EMBL accession numbers U05678 and U07858, respectively). Amino acids that are conserved among all known calpains are boxed in black. A period indicates that the same amino acid is present in the homologous sequence, and letters refer to the variant amino acid. Positions of missense mutations are given as encircled numbers above the corresponding amino acid.

sequence rather than minor polymorphic variants. Furthermore, these modifications increase the local similarity with the rat CANP3 amino acid sequence (Sorimachi et al., 1989), although the overall similarity is still 94%.

The ATG numbered 1 in Figure 2B is presumed to be the translation initiation site based on homology with the rat *CANP3* and is within a sequence with five nucleotides out of eight in common with the Kosak consensus sequence (Kosak, 1984). Putative CAAT and TATA boxes were observed 590 or 324 (CAAT) and 544 or 33 bp (TATA) upstream of the initiating ATG codon (Bucher, 1990). PCR amplification on muscle cDNA libraries excluded the downstream TATA site since it was contained within the amplified products (data not shown). A GC-box binding the Sp1 protein (Dynan and Tjian, 1983) was identified at position −477. Consensus sequences corresponding to potential muscle-specific regulatory elements were identified (Figure 2A). These include a myocyte-specific enhancer-binding factor 2 (MEF2)-binding site (Gosset et al. 1989), a CArG box (serum response elements observed in muscle-specific genes and in genes inducible by growth hormones; Minty and Kedes, 1986), and four E boxes (contact sites for basic–helix-loop-helix proteins found in members of the MyoD family; Blackwell and Weintraub, 1990). The functional significance of these putative transcription factor–binding sites in the regulation of *CANP3* gene expression remains to be established.

Two AAUAAA were identified, 520 and 777 bp downstream of the TGA stop codon. The sequencing of a partial *CANP3* cDNA containing a poly(A) tail demonstrated that the first AAUAAA is the polyadenylation signal. The latter is embedded in a region well conserved with the rat *CANP3* sequence and is followed after 4 bp by a GT cluster, present in most genes 3′ of the polyadenylation site (Birnstiel et al., 1985). The 3′ untranslated region of the *CANP3* mRNA is 565 bp long. Considering the most likely promoter location, the predicted length of the cDNA should therefore be approximately 3550 bp.

The DNA and protein sequences of the human *CANP3* gene show high degree of conservation with other calpains (Figure 3), in particular rat (SwissProt accession numbers J05121 and P16259), bovine (GenBank accession number U07858), and porcine (GenBank accession number U05678) homologous sequences. High local similarities between the human and rat DNA sequences are even observed at the 5′ (75%) or in different parts of the 3′ untranslated regions (over 60%) (data not shown), suggestive of evolutionary pressures in their untranslated regions.

Genomic Organization of the *CANP3* Gene

A comparison of the published *CANP3* human cDNA (Sorimachi et al., 1989) with the corresponding genomic sequence led to the identification of 24 exons ranging in length from 12 bp (exon 13) to 309 bp (exon 1), with a mean size of 100 bp (see Figure 1A). The size of introns ranges from 86 bp to about 10–16 kb for intron 1. The intron–exon boundaries (Table 2) exhibit close adherence

Table 2. Sequences at the Intron–Exon Junctions

Splice Donor Site	Score (%)	Intron	Score (%)	Splice Acceptor Site	Exon
					Exon 1 (309 bp)→
...CTCCGgtgagt...	88.5	←Intron 1→	99.0	...tttttgtttcacagGAAAT...	Exon 2 (70 bp)→
...GCTAGgtagga...	83.5	←Intron 2→	90.0	...gtgtctgcctgcagGGGAC...	Exon 3 (119 bp)→
...TCCAGgtgagg...	92	←Intron 3→	81.5	...acgcttctgtgcagTTCTG...	Exon 4 (134 bp)→
...GCTAAgtaagc...	82	←Intron 4→	81.5	...atcctctctctaagGCTCC...	Exon 5 (169 bp)→
...TTGATgtaagt...	87	←Intron 5→	79.5	...ccatcgggcctcagGATGG...	Exon 6 (144 bp)→
...CCCGGgtgtgt...	77.5	←Intron 6→	91	...ttactgctctacagACAAT...	Exon 7 (84 bp)→
...ATGAGgtaagc...	94	←Intron 7→	78.5	...tctgtgtgcttaagGTCCC...	Exon 8 (86 bp)→
...GATAGgtaggt...	89	←Intron 8→	91.5	...cattttcccaccagATGGA...	Exon 9 (78 bp)→
...TTCTGgtgagt...	88	←Intron 9→	92	...ttccaacctctcagGATGT...	Exon 10 (161 bp)→
...CCCAGgtggga...	80	←Intron 10→	68.5	...ttctgggggtgcagATACT...	Exon 11 (170 bp)→
...ACGAGgtgtgt...	85.5	←Intron 11→	86	...tgtttcttctcaagGTTCC...	Exon 12 (12 bp)→
...AAGAGgtatag...	70	←Intron 12→	87	...tccccatctctcagATGCA...	Exon 13 (209 bp)→
...TCTGAgtgagt...	76.5	←Intron 13→	97	...tgtattcctcacagGGAAG...	Exon 14 (37 bp)→
...CAGTGgtgagt...	89	←Intron 14→	93.5	...cttttcttatgcagAAAAA...	Exon 15 (18 bp)→
...CCAAGgtaggt...	89	←Intron 15→	87	...cctcctctctccagCCCAT...	Exon 16 (114 bp)→
...CACAGgtgtct...	80	←Intron 16→	88	...ttgtgcctccacagCCACA...	Exon 17 (78 bp)→
...GAGATgtgagt...	84	←Intron 17→	92.5	...cccttcctcctcagGACAT...	Exon 18 (58 bp)→
...CAAACgtgagt...	83	←Intron 18→	90	...ctccatcccccagACAAG...	Exon 19 (65 bp)→
...TGGATgtatcc...	56	←Intron 19→	88	...cctccctcctccagACAGA...	Exon 20 (69 bp)→
...GGCAGgtggga...	80	←Intron 20→	94	...ttttctattgccagAAATA...	Exon 21 (79 bp)→
...CGCAGgtgctg...	66	←Intron 21→	91	...ggtcccctccacagGATTC...	Exon 22 (117 bp)→
...GTTCAgtaagt...	79	←Intron 22→	93.5	...gcattctttcacagGAGCT...	Exon 23 (59 bp)→
...TGGAGgtaaag...	81	←Intron 23→	79	...gggacttctttcagTGGCT...	Exon 24 (27 bp)→

A score expressing adherence to the consensus was calculated for each size according to Shapiro and Senapathy (1987). Sequences of exons and introns are in uppercase and lowercase, respectively. Sizes of exons are given in parenthesis.

to 5′ and 3′ splice site consensus sequences (Shapiro and Senapathy, 1987). When the genomic sequence was submitted to GRAIL analysis (Uberbacher and Mural, 1991), 11 exons were correctly recognized, four were not identified, six were inadequately defined, and two were too small to be recognized (data not shown).

As already noted, the *CANP3* gene has three unique sequence blocks, NS (amino acid residues 1–61), IS1 (residues 267–329), and IS2 (residues 578–653). It is interesting to note that each IS sequence, as well as the nuclear translocation signal (residues 595–600) inside IS2, is essentially flanked by introns (Figure 4). The exon–intron organization of the human *CANP3* is similar to that reported for the chicken *CANP* (the only other large subunit calpain gene whose genomic structure is known; Emori et al., 1986).

Four microsatellite sequences were identified (see Figure 1A). Two of them are in the distal part of the first intron: an $(AT)_{14}$ and a previously identified nonpolymorphic mixed-pattern microsatellite, *D15S498* (Fougerousse et al., 1994). A $(TA)_7(CA)_4(GA)_{13}$ was identified in the second intron, and genotyping of 64 unrelated Centre d'Etudes du Polymorphisme Humain (CEPH) individuals revealed two alleles (with frequencies of 0.10 and 0.90). The fourth microsatellite is a mixed $(CA)_n(TA)_m$ repeat present in the ninth intron. The latter and the $(AT)_{14}$ repeat have not been investigated for polymorphisms. One *MER2* repeat and 14 members of the Alu family were identified in the *CANP3* gene (see Figure 1A), which has, thus, on average one Alu element per 2.5 kb.

Expression of the *CANP3* Gene

The pattern of tissue specificity was investigated by Northern blot hybridization. There is no evidence for the existence of an alternatively spliced form of *CANP3*, although

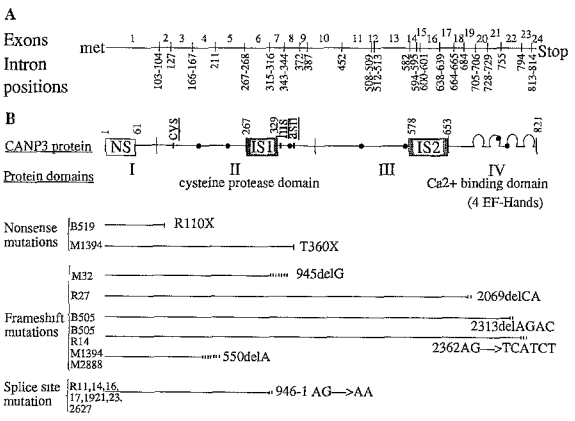

Figure 4. Distribution of the Mutations along CANP3 Protein

(A) Positions of the 23 introns are indicated by vertical bars. Numbers written below refer to the corresponding amino acids.

(B) Schematic representation of the CANP3 protein, its four domains (I, II, III, and IV), and muscle-specific sequences (NS, IS1, and IS2). The three active site residues are underlined. The positions of missense mutations are indicated by dots. The effects of nonsense and frameshift mutations are illustrated as truncated lines, representing the extent of protein synthesized. The out-of-frame sequence is shown by hatched lines. Names of the families carrying these mutations are indicated to the left of the lines.

Table 3. PCR Primers for the Analysis of the *CANP3* Gene

Primer Name	Primer Sequences (5′–3′)	Amplified Region	PCR Product Size (in Base Pairs) (Genomic DNA)	Annealing Temperature (in Celsius)
CANP3-pro.a	TTCAGTACCTCCCGTTCACC	Promoter	296	59
CANP3-pro.m	GATGCTTGAGCCAGGAAAAC			
CANP3-ex1.a	CTTTCCTTGAAGGTAGCTGTAT	Exon 1	438	60
CANP3-ex1.m	GAGGTGCTGAGTGAGAGGAC			
CANP3-ex2.a	ACTCCGTCTCAAAAAAATACCT	Exon 2	239	57
CANP3-ex2.m	ATTGTCCCTTTACCTCCTGG			
CANP3-ex3.a	TGGAAGTAGGAGAGTGGGCA	Exon 3	354	58
CANP3-ex3.m	GGGTAGATGGGTGGGAAGTT			
CANP3.ex4.a	GAGGAATGTGGAGGAAGGAC	Exon 4	292	59
CANP3-ex4.m	TTCCTGTGAGTGAGGTCTCG			
CANP3-ex5.a	GGAACTCTGTGACCCCAAAT	Exon 5	325	56
CANP3-ex5.m	TCCTCAAACAAAACATTCGC			
CANP3-ex6.a	GTTCCCTACATTCTCCATCG	Exon 6	315	57
CANP3-ex6.m	GTTATTTCAACCCAGACCCTT			
CANP3-ex7.a	AATGGGTTCTCTGGTTACTGC	Exon 7	333	56
CANP3-ex7.m	AGCACGAAAAGCAAAGATAAA			
CANP3-ex8.a	GTAAGAGATTTGCCCCCCAG	Exon 8	321	58
CANP3-ex8.m	TCTGCGGATCATTGGTTTTG			
CANP3-ex9.a	CCTTCCCTTCTTCCTGCTTC	Exon 9	173	56
CANP3-ex9.m	CTCTCTTCCCCACCCTTACC			
CANP3-ex10.a	CCTCCTCACCTGCTCCCATA	Exon 10	251	56
CANP3-ex10.m	TTTTTCGGCTTAGACCCTCC			
CANP3-ex11.a	TGTGGGGAATAGAAATAAATGG	Exon 11	355	57
CANP3-ex11.m	CCAGGAGCTCTGTGGGTCA			
CANP3-ex12.a	GGCTCCTCATCCTCATTCACA	Exon 12	312	61
CANP3-ex12.m	GTGGAGGAGGGTGAGTGTGC			
CANP3-ex13.a	TGTGGCAGGACAGGACGTTC	Exon 13	337	60
CANP3-ex13.m	TTCAACCTCTGGAGTGGGCC			
CANP3-ex14.a	CACCAGAGCAAACCGTCCAC	Exon 14	230	61
CANP3-ex14.m	ACAGCCCAGACTCCCATTCC			
CANP3-ex15.a	TTCTCTTCTCCCTTCACCCT	Exon 15	225	57
CANP3-ex15.m	ACACACTTCATGCTCTCTACCC			
CANP3-ex16.a	CCGCCTATTCCTTTCCTCTT	Exon 16	331	56
CANP3-ex16.m	GACAAACTCCTGGGAAGCCT			
CANP3-ex17.a	ACCTCTGACCCCTGTGAACC	Exon 17	270	61
CANP3-ex17.m	TGTGGATTTGTGTGCTACGC			
CANP3-ex18.a	CATAAATAGCACCGACAGGGA	Exon 18	258	59
CANP3-ex18.m	GGGATGGAGAAGAGTGAGGA			
CANP3-ex19.a	TCCTCACTCTTCTCCATCCC	Exon 19	159	57
CANP3-ex19.m	ACCCTGTATGTTGCCTTGG			
CANP3-ex20.a	GGGGATTTTGCTGTGTGCTG	Exon 20–21	333	61
CANP3-ex20.m	ATTCCTGCTCCCACCGTCTC			
CANP3-ex22.a	CACAGAGTGTCCGAGAGGCA	Exon 22	282	57
CANP3-ex22.m	GGAGATTATCAGGTGAGATGCC			
CANP3-ex23.a	CAGAGTGTCCGAGAGGCAGGG	Exon 22–23	608	61
CANP3-ex23.m	CGTTGACCCCTCCACCTTGA			
CANP3-ex24.a	GGGAAAACATGCACCTTCTT	Exon 24	375	58
CANP3-ex24.m	TAGGGGGTAAAATGGAGGAG			
CANP3-pA.a	ACTAACTCAGTGGAATAGGG	Polyadenylation signal	413	56
CANP3-pA.m	GGAGCTAGGATAGCTCAAT			

this cannot be excluded. A transcript of about 3.4–3.6 kb was detected in skeletal muscle mRNA (Chiannilkulchai et al., 1995). This size therefore further supports the position −544 as the functional TATA box.

Mutation Screening

CANP3 fulfils both positional and functional criteria to be a candidate gene for LGMD2. We therefore systematically screened 38 LGMD families for the presence of nucleotide changes in *CANP3* using a combination of heteroduplex (Keen et al., 1991) and direct sequence analyses.

PCR primers were designed to specifically amplify the exons and splice junctions as well as the regions containing putative CAAT and TATA boxes and the polyadenylation signal (Table 3). PCR products made on DNA of LGMD patients were then subjected either to heteroduplex analysis or direct sequencing, depending on whether the mutation, based on haplotype analysis, was expected to be heterozygous or homozygous, respectively. It was occasionally necessary to clone the PCR products to identify the mutations precisely (i.e., for microdeletions or insertions and for some heterozygotes). Disease-associated mutations are summarized in Table 4, and their position

Table 4. *CANP3* Mutations in LGMD2A Families

Exon	Families (Réunion Haplotype)	Nucleotide Position	Nucleotide Change	Amino Acid Position	Effect of Mutation	Restriction Site Change	Protein Domain	Mutation
2	B519*	328	CGA→TGA	110	Arg→stop	—	II	R110X
4	M42	545	CTG→CAG	182	Leu→Gln	—	II	L182Q
4	M1394, M2888	550	CAA→CA	184	Frameshift	—	II	550ΔA
5	M35, M37	701	GGG→GAG	234	Gly→Glu	—	II	G234E
6	M32	945	CGG→CG	315	Frameshift	Without SmaI	II	945ΔG
6–7	R11*, R14, R16*, R17, R19*, R21*, R26*, R27 (I, II)	946-1	G→A	—	Abberant splicing	—	II	946-1 G→A
8	M2407*	1061	GTG→GGG	354	Val→Gly	—	II	V354G
8	M1394	1079	TGG→TAG	360	Trp→stop	Without BstnI, without EcoRI, without ScrfI	II	T360X
11	M2888	1468	CGG→TGG	490	Arg→Trp		III	R490T
13	R12* (VII)	1715	CGG→CCAG	572	Arg→Gln	Without MspI	III	R572Q
19	R27 (VIII)	2069–2070	Deletion AC	690	Frameshift		IV	2069ΔCA
21	R14, R17 (III)	2230	AGC→GGC	744	Ser→Gly	Without AluI	IV	S744G
22	A*, B501*, M32	2306	CGG→CAG	769	Arg→Gln	—	IV	R769Q
22	B505	2313–2316	Deletion AGAC	771–772	Frameshift	—	IV	2313ΔAGAC
22	B505, R14 (IV)	2362–2363	AG→TCATCT	788	Frameshift	—	IV	2362AG→TCATCT

The first letter of the family code refers to the origin of the population (B, Brazil; M, metropolitan France; R, Isle of La Réunion; A, Amish; the corresponding haplotypes for La Réunion island families are numbered in parentheses). Families that are homozygous for mutation are indicated by asterisks. Positions are numbered on the basis of the cDNA and protein sequences starting from ATG and the first methionine residue, respectively. The mutated nucleotides are underlined.

along the protein is shown in Figure 4. Each mutation was confirmed by heteroduplex analysis, sequencing of both strands in several members of the family or enzymatic digestion when the mutation modified a restriction site. Segregation analyses of the mutations, performed on DNAs from all available members of the families, confirmed that these sequence variations are on parental chromosomes carrying a *LGMD2A* mutation. To assess the possibility that missense substitutions might be polymorphisms, their presence was systematically tested in a control population: none of the mutations was seen among 120 control chromosomes from the CEPH reference families.

Chromosome 15 Ascertained Families

The initial screening for causative mutations was performed on families from the La Réunion island (Beckmann et al., 1991), from the Old Order Amish of northern Indiana (Young et al., 1992), and from Brazil (Passos-Bueno et al., 1993).

La Réunion Island Families

Genealogical studies and the geographic origin of the families from La Réunion were suggestive of a single founder effect. Genetic analyses are, however, inconsistent with this hypothesis as the families present haplotype heterogeneity, with at least six different carrier chromosomes (Allamand et al., 1995b). In the course of this work, distinct mutations corresponding to six haplotypes have been identified (Table 4). Yet these mutations are absent in some of the minor haplotypes. Thus, some of them must still contain at least one other unidentified mutation.

In family R14, exons 13, 21, and 22 showed evidence for sequence variation upon heteroduplex analysis (Figure 5). Sequencing of the associated PCR products revealed a polymorphism in exon 13, a A→G mutation in exon 21 transforming Ser-744 to glycine in the second EF hand (see Figure 4), and a AG→TCATCT frameshift mutation in exon 22 causing premature termination at nucleotide 2400 where an in-frame stop codon occurs (see Figure 4). The exon 21 mutation and the polymorphism in exon 13 form an haplotype that is also encountered in family R17.

Affected individuals in family R12 are homozygous over the entire *LGMD2A* interval (Allamand et al., 1995b). Sequencing of the PCR products of exon 13 revealed a G→A transition at base 1715 of the cDNA resulting in a substitution of glutamine for Arg-572 (Figure 6) inside domain III, a residue that is highly conserved throughout all known calpains. This mutation, detectable by loss of an MspI restriction site, is present only in this family and in no other examined LGMD2A families or unrelated controls (data not shown).

In family R27, heteroduplex analysis followed by se-

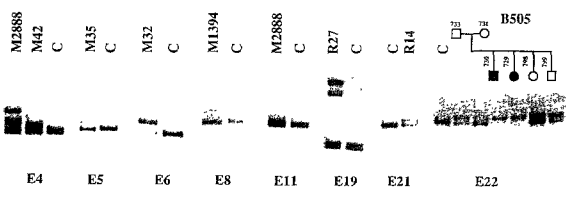

Figure 5. Representative Mutations Identified by Heteroduplex Analysis

Examples of mutations detected by heteroduplex analysis. Lanes C represent control samples. Exon numbers (E) are indicated below the gels. Pedigree B505 displays the segregation of two different mutations in exon 22.

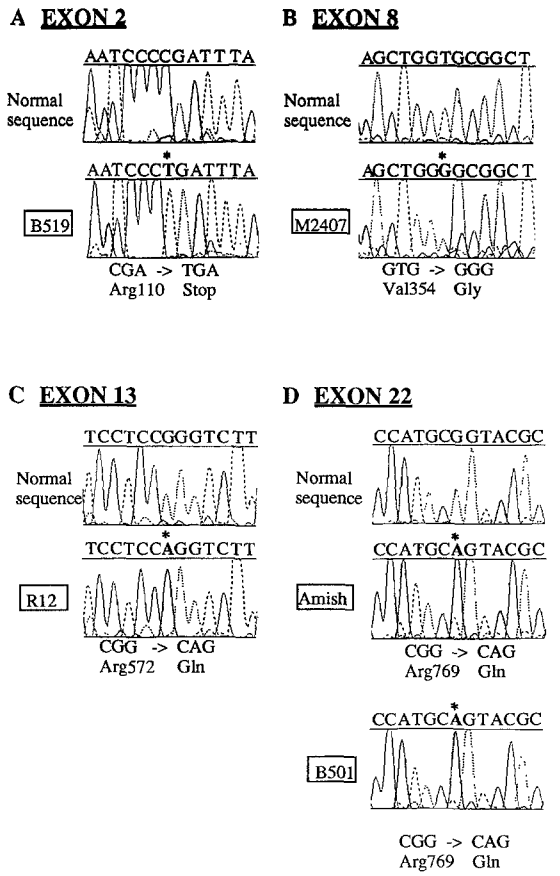

Figure 6. Sequence of Homozygous Mutations

Sequences from a healthy control are shown above the mutant sequences present in exons 2 (A), 8 (B), 13 (C), and 22 (D). Asterisks indicate the position of mutated nucleotides. The consequences on codon and amino acid residues are indicated at the left of each panel together with the name of the family.

quencing of the PCR products of an affected child revealed a 2 bp deletion in exon 19 (see Figure 5; Table 4). One AC out of three is missing at this position of the sequence, producing a stop codon at position 2069 of the cDNA sequence (see Figure 4).

Since both mutation detection enhancer heteroduplex analysis and direct sequencing failed to reveal the mutations in the two major haplotypes (I and II), an attempt was made to uncover aberrant splicing of the muscle-specific calpain mRNA from these patients by taking advantage of the illegitimate transcription in lymphocytes (Chelly et al., 1989). The PCR products corresponding to exons 3–9 showed the presence of a band about 400 bases longer than expected. Fine PCR mapping localized the mutation at the junction of exons 6 and 7, while sequencing of the corresponding products (Figure 7) revealed an acceptor splice site mutation (AG→AA), resulting in the utilization of an alternative site within intron 6, 391 bp upstream of exon 7. The same mutation was seen in haplotypes I and II (Table 4), even though they differ in 14 out of 28 markers. It is still unclear whether these are related to one another or a coincidence of two independent recurrent mutational

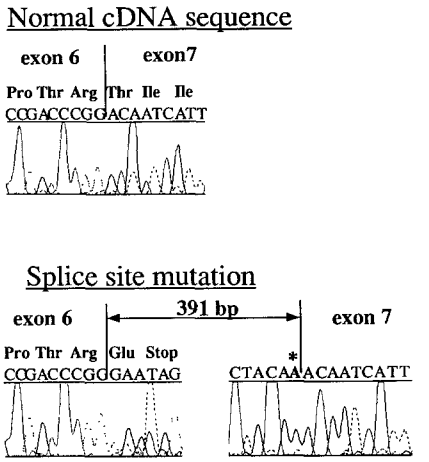

Figure 7. Sequence Analysis of the Junction between Exons 6 and 7 of *CANP3* of Illegitimately Transcribed mRNAs

The corresponding sequence from a normal muscular *CANP3* cDNA (above) or from a LGMD2A patient (below) are shown, together with the corresponding amino acids. The asterisk indicates the G→A transition responsible for the loss of the normal acceptor site leading to the use of a cryptic splice site within intron 6. This aberrant splicing results in the presence in the mRNA of an additional 391 bp intronic sequence, represented by an arrow, and of a stop codon as indicated above the sequence.

events. The presence of these mutations was confirmed on genomic DNA. The net result is presumably the loss of CANP3 activity since they lead to premature termination by a stop codon that is adjacent to the aberrant splice site (Figure 7).

Amish Families

As expected, owing to multiple consanguineous links, the examined northern Indiana Amish LGMD2A patients were homozygous for the carrier haplotype (Allamand et al., 1995b). A G→A missense mutation was identified at nucleotide 2306 within exon 22 (see Figure 6), transforming Arg-769 to glutamine. This residue, which is conserved throughout all members of the calpain family in all species, is located in domain IV of the protein within the third EF hand at the helix-loop junction (see Figure 4). This mutation was encountered in a homozygous state in all patients from 10 chromosome 15–linked Amish families. We also verified that this nucleotide change was not present in patients from the six southern Indiana Amish LGMD families for which the chromosome 15 locus was excluded by linkage analyses (Allamand et al., 1995a), thus confirming the genetic heterogeneity of this disease in this genetically related isolate.

Brazilian Families

As a result of consanguineous marriages, two Brazilian families (B501 and B519) are homozygous for extended *LGMD2A* carrier haplotypes (Allamand et al., 1995b). Affected individuals from family B501 were shown to have the same exon 22 mutation as the Amish LGMD2A patients (see Figure 6), but embedded in a completely different haplotype (Allamand et al., 1995b). In family B519, the patients carry a C→T transition in exon 2 in a homozygous

Table 5. *CANP3* Polymorphisms

Location	Families	Nucleotide Position	Nucleotide Change	Amino Acid Position	Restriction Site Change
Promoter	M42	−408	T→C	—	—
1	M31	96	ACT→ACC	32	DdeI→SecI
3	M37, M40	495	TTC→TTT	165	Without MboII
13	R14, R17	1668	ATC→ATT	556	—
Intron 22	R11, R19, R20, M35, M37	2380 + 12	Deletion A	—	—

See Table 4 legend for details.

state, replacing Arg-328 with a TGA stop codon (see Figure 6), thus presumably leading to a very truncated and inactive protein (see Figure 4).

Analysis of Other LGMD Families

Having validated the role of *CANP3* in the chromosome 15 ascertained families, we next examined by heteroduplex analysis LGMD families for which linkage data were not informative. These included one Brazilian (B505) and 10 metropolitan French pedigrees. Additional mutations were uncovered enabling the ascertainment of eight more LGMD2A families.

Heteroduplex bands were revealed for exons 1, 3, 4, 5, 6, 8, 11, 22, and the promoter region of one or more patients (see Figure 5). Of all sequence variants, 10 were identified as pathogenic mutations (five missense, one nonsense, and four frameshift mutations) and four as polymorphisms. Altogether, two morbid alleles were identified in five families, one in three, and none in three others (Table 4). Identical mutations were uncovered in apparently unrelated families. The mutations shared by families M35 and M37 or by M2888 and M1394, respectively, are likely to be the consequence of independent events since they are embedded in different marker haplotypes. In contrast, it is likely that the point mutation present in exon 22 of the Amish and in the M32 kindreds corresponds to the same mutational event, as both chromosomes share a common four-marker haplotype (*D15S779–D15S512–D15S782–D15S780*) around *CANP3* (unpublished data), possibly reflecting a common ancestor. The same holds true for the exon 22 substitution shared by families B505 and R14.

In addition to the polymorphisms present in exon 13 in families R14 and R17 (position 1668) and in the intragenic microsatellites, four additional neutral variations were detected (Table 5).

Discussion

Several lines of evidence implicate the *CANP3* gene in the etiology of LGMD2A. This gene is localized inside the 3 Mb *LGMD2A* interval, in the region suggested by linkage disequilibrium studies (Allamand et al., 1995b). Southern blot experiments (Ohno et al., 1989) and sequence-tagged site screening (data not shown) suggest that there is but one copy per genome of this member of the calpain family. Transcription studies suggested that it is an active gene rather than a pseudogene, and its muscle-specific pattern of expression is consistent with the phenotype of this disorder (Sorimachi et al., 1989; Chiannilkulchai et al., 1995).

A minimum of 17 independent mutational events were identified in families from different ethnic and geographic groups. These represent 15 different mutations, distributed throughout the gene (Figure 4), that cosegregate with the disease in LGMD2A families. The discovery of two nonsense, one splice site, and five frameshift mutations in *CANP3* supports the hypothesis that a deficiency of this gene product causes LGMD2A. All eight mutations result in a premature in-frame stop codon, leading to the production of truncated and presumably inactive proteins (Figure 4). Evidence for the morbidity of the missense mutations come from several sources: their relative high incidence among LGMD2A patients, the failure to observe these mutations in control chromosomes, and the occurrence of mutations at evolutionarily conserved residues, in regions of documented functional importance, or both (Sorimachi et al., 1989). Of seven mutations, four change an amino acid that is conserved in all known members of the calpain family in all species (Figure 3). Two of the remaining mutations affect less conserved amino acid residues, but are located in important functional domains: V354G is four residues before the asparagine at the active site; S744G within the second EF hand may impair the calcium-dependent regulation of calpain activity or the interaction with a small subunit (Figure 4). Several missense mutations change a hydrophobic residue to a polar one or vice versa (Table 4), possibly disrupting higher order structures.

The Réunion Paradox

Only four different mutations were identified by heteroduplex and sequence analyses in the families of La Réunion, despite the fact that all exons were scanned. Eventually, the use of illegitimate transcription allowed us to uncover aberrant splicing, leading to the recognition of the same splice site mutation in haplotypes I and II (Table 4). Thus, five different mutations have been identified so far in this population. But given that none of them was found on the remaining carrier haplotypes, there should be at least one other (as yet unidentified) mutation.

The presence of at least six different mutations among patients from the La Réunion island demonstrates the validity of the suggestion based on haplotype analysis of a multifounder effect (Allamand et al., 1995b). This is, how-

ever, an unexpected result given the multiple consanguineous links in these families. Indeed, the affected patients of La Réunion all belong to a small genetic isolate, presumed to derive from a single ancestor who immigrated to this island in the 1670s. These patients were all thus expected to carry the same *LGMD2A* mutation. The occurrence of multiple independent events in other small populations is not unprecedented (e.g., Bach et al., 1994; Rodius et al., 1994; Heinisch et al., 1995), although no satisfactory explanation has been forwarded so far.

Thus, the presence of multiple mutations in the La Réunion population needs to be reconciled with the reported low prevalence of this disease, a problem we refer to as the Réunion paradox.

To begin, the global prevalence of LGMD2A could just be much more common than initially presumed, but this seems highly unlikely. Another hypothesis assumes that heterozygous healthy individuals could benefit from a selective advantage (as in β-thalassemia), resulting in a high local prevalence. This hypothesis, however, seems also unlikely since one would not expect such selective pressures to lead to these results in such a limited timespan (i.e., 320 years at most).

We therefore speculate that this condition, which has this far been considered as a monogenic disorder, may reflect a more complex inheritance pattern, in which expression of the calpain mutations would be dependent on genetic background (nuclear or mitochondrial). Consider, for instance, a digenic model: only in the presence of specific alleles at a permissive second unlinked locus (e.g., a compensatory, partially redundant, regulatory, or modifier gene) will there be expression of calpain mutations. Since one would need mutations at both loci to be affected, the disease prevalence would remain low. Under this model, members of the La Réunion island community would, as a result of genetic drift, have a disease-associated allele at the hypothesized second locus at high frequency (or even fixed in this small population), conditions that would explain the apparent complete penetrance of the calpain mutations. The complete penetrance of this disease in the Amish and in the other described LGMD2A pedigrees would also be under control of the second locus.

If this model is true, there may be fewer selective pressures against the appearance of *CANP3* mutations, as a result of the conditional penetrance. In other words, the frequency of the calpain variants in the overall population can be much higher than initially deduced, based on the estimates of the prevalence of the disease under a simple monogenic model. Assume, for instance, recessive, independent, and fully penetrant expression at two autosomal loci, with similar frequencies for both deleterious alleles. Considering the reported prevalence (10^{-5}; Emery, 1991) and the estimated genetic heterogeneity, 1 in 25 persons would be a carrier for a deleterious calpain allele (and 1 in 625 would carry deleterious alleles at both loci). The frequency of calpain variants could thus approach that of the cystic fibrosis gene. If we assume the allele frequency at the second locus to be, respectively, 0.1 or 0.2, this would give us one *LGMD2A* carrier per 58 or 116 individuals. But, since expression of the trait would require the simultaneous presence of both sets of deleterious alleles, the overall prevalence remains, as expected, on the order of 10^{-5}.

Under this model, some of the families for which the role of the *LGMD2A* locus was previously excluded, based on linkage analyses assuming simple monogenic inheritance, might be authentic LGMD2A families, reflecting differential segregation of these two unlinked genes. The digenic inheritance model thus predicts that in a number of kindreds, there will be healthy individuals with two mutant calpain genes. To test the validity of this model, we are currently performing a systematic screening for the presence of calpain mutations in DNA from affected probands from all small nuclear families that cannot be identified as LGMD2A on the basis of linkage analyses. We will then assess the cosegregation of the mutations in these pedigrees, searching for asymptomatic carriers of two mutant calpain genes.

An alternative possibility assumes that the calpain mutations are pathogenic only in a specific mitochondrial context. It is worthwhile remembering that the mitochondria are central suppliers of energy and could therefore influence the fate of muscle cells. The mitochondrial model predicts a monogenic inheritance pattern within nuclear families (all sibs share the same mitochondrial genome). In other pedigrees, however, we should fail to get expression of the disease in individuals carrying calpain mutations in a nonpermissive mitochondrial genetic background.

It is important to stress that the observations in favor of the proposed departure from simple monogenic inheritance are not unprecedented. There has been the report of digenic inheritance of retinitis pigmentosa (Kajiwara et al., 1994). And there are a number of additional descriptions of traits for which expression is clearly influenced by the presence of unlinked mutations (e.g., Oppenheim et al., 1990).

The concomitant involvement of two or more loci in the *LGMD2A* phenotype may have some immediate practical consequences. Since the current estimates for genetic heterogeneity are based on a simple monogenic model, these would be truly inadequate under digenic or more complex inheritance models. In addition, if one of the latter turns out to be true, its impact on genetic counseling will also need to be carefully evaluated.

Clearly, the mechanism underlying the results reported here may have major bearing on our comprehension of simple genetic traits. They may imply that one may occasionally face similar situations in other inbred human populations, such as Finland. The proposed di- or oligogenic model could also explain some of the failures to reproduce known or expected human phenotypes in mouse gene knockout experiments, since the genetic backgrounds allowing for expression of the corresponding traits may not have been present. One particularly convincing example is the work of Rudnicki et al. (1993) on *MyoD* and *Myf-5* transgenic mice, where only double knockout animals express a pathological phenotype. Another example are the *Igf1* knockouts, which show variable survival rates depending on the genetic background (Liu et al., 1993). One

prediction of this model is that the mouse *CANP3* knockouts may also fail to have a pathological phenotype.

CANP3 and LGMD2A

Identification of *CANP3* as a defective gene in LGMD2A suggests a novel pathological mechanism leading to a muscular dystrophy in which this condition is caused by mutations affecting an enzyme and not a structural component of muscle tissue. This result is to be contrasted with all known other muscular dystrophies, such as Duchenne and Becker (Bonilla et al., 1988), severe childhood autosomal recessive (Matsumara et al., 1992), Fukuyama (Matsumara et al., 1993), merosin-deficient congenital muscular dystrophies (Tomé et al., 1994), and primary adhalin deficiencies (Roberds et al., 1994).

The understanding of the *LGMD2A* phenotype needs to take into account the fact that there is likely to be no active CANP3 protein in several patients, a loss compatible with the recessive manifestation of this disease. Simple models in which this protease would be involved in the degradation or destabilization of structural components of the cytoskeleton, extracellular matrix, or dystrophin complex must therefore be ruled out. Furthermore, there are no signs of such alterations by immunohistochemical studies on LGMD2 muscle biopsies (Matsumara et al., 1993; Tomé et al., 1994). Likewise, since LGMD2A myofibers are apparently not different from others that are dystrophic, it seems unlikely that this calpain plays a role in myoblast fusion, as proposed for ubiquitous calpains (Wang et al., 1989).

Alternative hypotheses must therefore be forwarded. One could imagine that the CANP3 participates in the activation of an enzyme or other protein involved in muscle metabolism, or it could be responsible for the catabolism of protein compounds in the muscle; the absence of such activity would result in a toxic accumulation of these compounds and eventually in the degeneration of muscle fibers. We favor another hypothesis wherein the CANP3 protein plays an active role in transduction of a signal, a hypothesis that takes the previously described properties of the ubiquitous calpain and of CANP3 into account.

Indeed, studies of CANP3 expression (Sorimachi et al., 1993b) reported that its mRNA is abundant (its expression being 10-fold higher than that of the ubiquitous calpains) and specific to fully differentiated myotubes and myofibers. Yet attempts to detect this protein failed (Sorimachi et al., 1993b). This is interpreted as a consequence of its extremely rapid turnover mediated by autocatalysis, possibly reflecting the need for precise regulation of its activity. When the CANP3 protein was expressed and measured in transfected COS and L8 myoblast cells, it was shown to have a nuclear localization (Sorimachi et al., 1993b), possibly mediated by the nuclear translocation signal in the IS2 region. This result suggests that this cellular compartment could be its natural site of action. The calpains have been proposed as having regulatory rather than degradative role, mediated by restricted proteolysis of specific proteins (Wang et al., 1989; Suzuki and Ohno, 1990; Croall and Demartino, 1991). In fact, ubiquitous calpains have been reported to regulate transcription by specific cleavage of c-Jun and c-Fos (Hirai et al., 1991) and by controlling NF-κB activities through proteolysis of its IκB-α inhibitor (Miyamoto et al., 1994). These reports suggest that CANP3 protein could be also involved in the control of gene expression by regulating the turnover or activity of transcription factors or of their inhibitors (e.g., members of the Id family; Benezra et al., 1990).

The finding that a defective calpain underlies the pathogenesis of LGMD2A may prove useful for the identification of other loci involved in neuromuscular diseases. Other forms of LGMD may indeed be caused by mutations in genes whose products are calpain substrates or in genes involved in the regulation of CANP3 expression. Techniques such as the two-hybrid selection system (Fields and Song, 1989) could lend themselves to the isolation of the natural protein substrate(s) of this calpain and thus potentially help to identify candidate loci. These approaches, inasmuch as they will lead to the identification of the other partners responsible for the LGMD phenotype, could eventually also clarify the molecular genetic basis of the inheritance pattern of LGMD2A.

With the caveat of the as-yet-unclarified inheritance mode, identification of mutations in the *CANP3* gene could provide means for direct prenatal or presymptomatic diagnosis and carrier detection in families in which both mutations have been identified. Gene-based classification of LGMD2A families should prove useful for the differential diagnosis of this disorder and for the establishment of phenotype–genotype relationships. The fact that morbidity results from the loss of an enzymatic activity raises hopes for novel pharmacotherapeutic prospects for this neuromuscular disorder or other myopathies. Finally, the evidence of pathological consequences of loss of a tissue-specific proteolytic activity, having a possible regulatory role, suggests that this may be a process of general significance.

Experimental Procedures

Description of the Patients

The LGMD2A families analyzed included three Brazilian families (Passos-Bueno et al., 1993), 10 interrelated nuclear families from La Réunion (Beckmann et al., 1991), 10 French metropolitan families, and 10 Amish families (Jackson and Carey, 1961). The majority of these families were previously ascertained by linkage analysis to belong to the chromosome 15 group (Beckmann et al., 1991; Young et al., 1992; Passos-Bueno et al., 1993). However, some families from metropolitan France, as well as one Brazilian family (B505), had insignificant lod scores (less than 3.0) for chromosome 15.

Sequencing of Cosmid c774G4-1F11 and EcoRI Restriction Map of Cosmids

Cosmid 1F11 (Figure 1C) was subcloned and sequenced following conventional procedures. The sequences were analyzed and alignments performed using the XBAP software of the Staden package, version 93.9 (Staden, 1982). Gaps between sequence contigs were filled by walking with internal primers. EcoRI restriction map of cosmids was performed essentially as described in Sambrook et al. (1989).

Analysis of *CANP3* in LGMD2A Patients

Human DNA (100 ng), obtained from peripheral blood lymphocytes, was used per PCR (Fougerousse et al., 1994; Table 3). Heteroduplex analyses were performed on 10 μl of PCR products essentially as in Keen et al. (1991). For sequence analysis, the PCR products were subjected to dye–dideoxy sequencing, after purification through Micro-

Table 6. PCR Primers for the Illegitimate Transcription of the *CANP3* Gene

Primer Name	Primer Sequence (5′–3′)	Position within the cDNA	PCR Product Size on cDNA (in Base Pairs)	Annealing Temperature (in Celsius)
CANP3-A1a	CAGCTCGGTTTTTAAGATGG	−188 to −169	674	57
CANP3-A1m	CGTCCACCCACTCTCCATAG	507–527		
CANP3-B1a	AAGTTCCCCATCCAGTTCGT	274–293	1006	57
CANP3-B1m	GAAGCTTGTCAGACTGCAGA	1260–1279		
CANP3-C1a	GTCACGCCTACTCTGTCACG	998–1017	912	57
CANP3-C1m	GCTTTTGCTTATCAGGGCTT	188–1903		
CANP3-D1a	TTGAAAATACCATCTCCGTG	1751–1770	869	57
CANP3-D1m	GGGGTAAAATGGAGGAGGAA	2590–2619		
CANP3-A2a	CTTTCCTTGAAGGTAGCTGTAT	−63 to −42	548	57
CANP3-A2m	AAGATCCCTGCGTAGTTTTCG	468–488		
CANP3-B2a	ATGGAGCCAACAGAACTGAC	341–360	914	57
CANP3-B2m	GGCCGTGAGGTTGCAGATCT	1235–1254		
CANP3-C2a	AGGTGGAGTGGAACGGTTCT	1085–1103	760	57
CANP3-C2m	GCTCCTTGTTGCTGTTTGCT	1824–1843		
CANP3-D2a	CCATCATCTTCGTTTCGGAC	1792–1811	721	57
CANP3-D2m	GGGTGAAACTGAAATCCTGA	2503–2522		

con devices (Amicon). When necessary, depending on the nature of the mutations (e.g., frameshift mutation or for some heterozygotes), the PCR products were cloned using the TA cloning kit from Invitrogen.

Total cellular RNA was extracted from lymphoblastoid cell lines by the RNAzol method (Bioprobe Systems). Reverse transcription was performed under standard conditions using 1 μg of RNA in a total volume of 30 μl containing 20 U of RNasin (Promega) and 200 U of Superscript reverse transcriptase (Bethesda Research Laboratories). We performed 20 cycles of PCR (40 s at 92°C, 30 s at 57°C, and 1 min at 72°C) on 1 μl of the cDNA sample using one of four sets of primers (A1, B1, C1, and D1; Table 6) to obtain four overlapping fragments spanning the whole *CANP3* coding sequence. The PCR product (1 μl) was then subjected to a second round of PCR amplification using nested primers (A2, B2, C2, and D2; Table 6). The PCR amplification products were analyzed by agarose gel electrophoresis and by direct double-stranded DNA sequencing.

Acknowledgments

Correspondence should be addressed to J. S. B. We gratefully acknowledge the help of Drs C. Auffray, P. M. Conneally, D. Fugman, D. Caterina, I. Cadjee, N. Fonknechten, J. Weissenbach, and P. Bucher. We thank Dr. H. Cann for critical reading of the manuscript and Dr. E. Bakker for suggesting the involvement of the mitochondrial genome. Above all, we are particularly indebted to all patients, to their families, to their clinicians, and to B. Barataud without whose participation and encouragement this work would never have seen the light. This research was supported by grants from the Association Française contre les Myopathies. We also acknowledge the Muscular Dystrophy Association, Fundacao de Amparo a Pesquisa do Estado de São Paulo (FAPESP), and Programa de Apoio de Desenvolvimento Cientifico e Tecnologico (PADCT) for their support for the collection of the American and Brazilian families.

Received November 10, 1994; revised February 1, 1995.

References

Allamand, V., Broux, O., Bourg, N., Richard, I., Tischfield, J., Hodes, M. E., Conneally, P. M., Fardeau, M., Jackson, C. E., and Beckmann, J. S. (1995a). Genetic heterogeneity of autosomal recessive limb–girdle muscular dystrophy in a genetic isolate (Amish) and evidence for a new locus. Hum. Mol. Genet. *4*, 459–464.

Allamand, V., Broux, O., Richard, I., Fougerousse, F., Chiannikulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Pereira de Souza, A., Roudaut, C., Tischfield, J. A., Conneally, P. M., Fardeau, M., Cohen, D., Jackson, C. E., and Beckmann, J. S. (1995b). Preferential localization of the limb–girdle muscular dystrophy type 2A gene in the proximal part of a 1 cM 15q15.1-q15.3 interval. Am. J. Hum. Genet., in press.

Bach, G., Moskowitz, S. M., Tieu, P. T., Matynia, A., and Neufeld, E. F. (1994). Molecular analysis of Hurler syndrome in Druze and Muslim Arab patients in Israel: multiple allelic mutations of the IDUA gene in a small geographic area. Am. J. Hum. Genet. *53*, 330–338.

Bashir, R., Strachan, T., Keers, S., Stephenson, A., Mahjneh, I., Marconi, G., Nashef, L., and Bushby, K. M. D. (1994). A gene for autosomal recessive limb–girdle muscular dystrophy maps to chromosome 2p. Hum. Mol. Genet. *3*, 455–457.

Beckmann, J. S., Richard, I., Hillaire, D., Broux, O., Antignac, C., Bois, E., Cann, H., Cottingham, R. W., Jr., Feingold, N., Feingold, J., Kalil, J., Lathrop, G. M., Marcadet, A., Masset, M., Mignard, C., Passos-Bueno, M. R., Pellerain, N., Zatz, M., Dausset, J., Fardeau, M., and Cohen, D. (1991). A gene for limb–girdle muscular dystrophy maps to chromosome 15 by linkage. CR Acad. Sci. (Paris) *312*, 141–148.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA-binding proteins. Cell *61*, 49–59.

Birnstiel, M. L., Busslinger, M., and Sturb, K. (1985). Transcription termination and 3′ processing: the end is in site! Cell *41*, 349–359.

Blackwell, T. K., and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science *250*, 1104–1110.

Bonilla, E., Samitt, C. E., Miranda, A. F., Hays, A. P., Salviati, G., DiMauro, S., Kunkel, L. M., Hoffman, E. P., and Rowland, L. P. (1988). Duchenne muscular dystrophy: deficiency of dystrophin at the muscle cell surface. Cell *54*, 447–452.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J. Mol. Biol. *212*, 563–578.

Bushby, K. M. D. (1994). Limb–girdle muscular dystrophy. In Diagnostic Criteria for Neuromuscular Disorders, A. E. H. Emery, ed. (Baarn, The Netherlands: ENMC), pp. 25–31.

Chelly, J., Concordet, J.-P., Kaplan, J.-C., and Kahn, A. (1989). Illegitimate transcription: transcription of any gene in any cell type. Proc. Natl. Acad. Sci. USA *86*, 2617–2621.

Chiannilkulchai, N., Pasturaud, P., Richard, I., Auffray, C., and Beckmann, J. S. (1995). A primary expression map of the chromosome 15q15 region containing the recessive form of limb–girdle muscular dystrophy (*LGMD2A*) gene. Hum. Mol. Genet. *4*, in press.

Croall, D. E., and Demartino, G. N. (1991). Calcium-activated neutral protease (calpain) system: stucture, function, and regulation. Physiol. Rev. *71*, 813–847.

Dynan, W. S., and Tjian, R. (1983). The promoter-specific transcription

factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79–87.

Emery, A. E. H. (1991). Population frequencies of inherited neuromuscular diseases: a world survey. Neuromusc. Disord. *1*, 19–29.

Emori, Y., Ohno, S., Tobita, M., and Suzuki, K. (1986). Gene structure of calcium-dependent protease retains the ancestral organization of the calcium-binding protein gene. FEBS Lett. *194*, 249–252.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein–protein interactions. Nature *340*, 245–246.

Fougerousse, F., Broux, O., Richard, I., Allamand, V., Pereira de Souza, A., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., Chiannilkulchai, N., Hillaire, D., Bui, H., Chumakov, I., Weissenbach, J., Cherif, D., Cohen, D., and Beckmann, J. S. (1994). Mapping of a chromosome 15 region involved in limb–girdle muscular dystrophy. Hum. Mol. Genet. *3*, 285–293.

Gosset, L. A., Kelvin, D. J., Sternberg, E. A., and Olson, E. (1989). A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. Mol. Cell. Biol. *9*, 5022–5033.

Heinisch, U., Zlotogora, J., Kafert, S., and Gieselmann, V. (1995). Multiple mutations are responsible for the high frequency of metachromatic leukodystrophy in a small geographic area. Am. J. Hum. Genet. *56*, 51–57.

Hirai, S., Kawasaki, H., Yaniv, M., and Suzuki, K. (1991). Degradation of transcription factors, c-Jun and c-Fos, by calpain. FEBS Lett. *1*, 57–61.

Imajoh, S., Kawasaki, H., and Suzuki, K. (1986). Limited autolysis of calcium-activated neutral protease (CANP): reduction of the Ca^{2+} requirement is due to the NH_2-terminal processing of the large subunit. J. Biochem. *100*, 633–642.

Jackson, C. E., and Carey, J. H. (1961). Progressive muscular dystrophy: autosomal recessive type. Pediatrics *28*, 77–84.

Kajiwara, K., Berson, E. L., and Dryja, T. P. (1994). Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science *264*, 1604–1608.

Keen, J., Lester D., Inglehearn, C., Curtis, A., and Bhattacharya, S. (1991). Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. Trends Genet. *7*, 5.

Kosak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acids Res. *12*, 857–872.

Liu, J.-P., Baker, J., Perkins, A. S., Robertson, E. J., and Efstratiadis, A. (1993). Mice carrying null mutations of the genes encoding insulin-like growth factors (*Igf1*) and type-1 IGF receptor (*Igf1r*). Cell *75*, 65–73.

Matsumara, K., Tomé, F. M. S., Collin, H., Azibi, K., Chaouch, M., Kaplan, J.-K., Fardeau, M., and Campbell, K. P. (1992). Deficiency of the 50K dystrophin-associated glycoprotein in severe childhood autosomal recessive muscular dystrophy. Nature *359*, 320–322.

Matsumara, K., Nonaka, I., and Campbell, K. P. (1993). Abnormal expression of dystrophin-associated proteins in Fukuyama-type congenital muscular dystrophy. Lancet *341*, 521–522.

Minty, A., and Kedes, L. (1986). Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol. Cell. Biol. *6*, 2125–2136.

Miyamoto, S., Maki, M., Schmitt, M. J., Hatanaka, M., and Verma, I. M. (1994). TNF-α-induced phosphorylation of IκB is a signal for its degradation but not dissociation from NF-κB. Proc. Natl. Acad. Sci. USA *94*, 12740–12744.

Murachi, T. (1989). Intracellular regulatory system involving calpain and calpastatin. Biochem. Int. *18*, 263–294.

Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M., and Suzuki, K. (1984). Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein? Nature *312*, 566–570.

Ohno, S., Minoshima, S., Kudoh, J., Fukuyama, R., Shimizu, Y., Ohmi-Imajoh, S., Shimizu, N., and Suzuki, K. (1989). Four genes for the calpain family locate on four different chromosomes. Cytogenet. Cell Genet. *51*, 1054.

Oppenheim, A., Yaari, A., Rund, D., Rachmilewitz, E. A., Nathan, D., Wong, C., Kazazian, H., Jr., and Miller, B. (1990). Intrinsic potential for high fetal hemoglobin production in a Druze family with β-thalassemia is due to an unlinked genetic determinant. Hum. Genet. *86*, 175–180.

Passos-Bueno, M.-R., Richard, I., Vainzof, M., Fougerousse, F., Weissenbach, J., Broux, O., Cohen, D., Akiyama, J., Marie, S. K. N., Carvalho, A. A., Guilherme, L., Kalil, J., Tsanaclis, A. M., Zatz, M., and Beckmann, J. S. (1993). Evidence of genetic heterogeneity in the autosomal recessive adult forms of limb–girdle muscular dystrophy following linkage analysis with 15q probes in Brazilian families. J. Med. Genet. *30*, 385–387.

Richard, I., Broux, O., Chiannilkulchai, N., Fougerousse, F., Allamand, V., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., Lorenzo, F., Sebastiani-Kabatchis, C., Schultz, R. A., Polymeropoulos, M. H., Gyapay, G., Auffray, C., and Beckmann, J. (1994). Regional localization of human chromosome 15 loci. Genomics *23*, 619–627.

Roberds, S. L., Leturcq, F., Allamand, V., Piccolo, F., Jeanpierre, M., Anderson, R. D., Lim, L. E., Lee, J. C., Tomé, F. M. S., Romero, N. B., Fardeau, M., Beckmann, J. S., Kaplan, J.-C., and Campbell, K. P. (1994). Missense mutations in the adhalin gene linked to autosomal recessive muscular dystrophy. Cell *78*, 625–633.

Rodius, F., Duclos, F., Wrogemann, K., Le Paslier, D., Ougen, P., Billault, A., Belal, S., Musenger, C., Brice, A., Dürr, A., Mignard, C., Sirugo, G., Weissenbach, J., Cohen, D., Hentati, F., Ben Hamida, M., Mandel, J.-L., and Koenig, M. (1994). Recombinations in individuals homozygous by descent localize the Friedrich ataxia locus in a cloned 450 kb interval. Am. J. Hum. Genet. *54*, 1050–1059.

Rudnicki, M. A., Schnegelsberg, P. N. J., Stead, R. H., Brau, T., Arnold, H. H., and Jaenisch, R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. Cell *75*, 1351–1359.

Sambrook, J., Fritsh, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, New York: Cold Spring Harvor Laboratory Press).

Shapiro, M., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucl. Acids Res. *15*, 7155–7174.

Sorimachi, H., Imajoh-Ohmi, S., Emori, Y., Kawasaki, H., Ohno, S., Minami, Y., and Suzuki, K. (1989). Molecular cloning of a novel mammalian calcium-dependant protease distinct from both m- and mu-type: specific expression of the mRNA in skeletal muscle. J. Biol. Chem. *264*, 20106–20111.

Sorimachi, H., Ishiura, S., and Suzuki, K. (1993a). A novel tissue-specific calpain species expressed predominantly in the stomach comprises two alternative splicing products with and without Ca^{2+}-binding domain. J. Biol. Chem. *268*, 19476–19482.

Sorimachi, H., Toyama-Sorimachi, N., Saido, T. C., Kawasaki, H., Sugita, H., Miyasaka, M., Arahata, K., Ishiura, S., and Suzuki, K. (1993b). Muscle-specific calpain, p94, is degraded by autolysis immediately after translation, resulting in disappearance from muscle. J. Biol. Chem. *268*, 10593–10605.

Staden, R. (1982). An interactive graphic program for comparing and aligning nucleic acid and amino acid sequences. Nucl. Acids Res. *10*, 2951–2961.

Suzuki, K., and Ohno, S. (1990). Calcium-activated neutral protease: structure–function relationship and functional implications. Cell Struct. Funct. *15*, 1–6.

Tagle, D. A., Swaroop, M., Lovett, M., and Collins, F. S. (1993). Magnetic bead capture of expressed sequences encoded within large genomic segments. Nature *361*, 751–753.

Tomé, F. M. S., Evangelista, T., Leclerc, A., Sunada, Y., Manole, E., Estournet, B., Barois, A., Campbell, K. P., and Fardeau, M. (1994). Congenital muscular dystrophy with merosin deficiency. CR Acad. Sci. (Paris) *317*, 351–357.

Uberbacher, E. C., and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc. Natl. Acad. Sci. USA *88*, 11261–11265.

Walton, J. N., and Nattrass, F. J. (1954). On the classification, natural history and treatment of the myopathies. Brain *77*, 169–231.

Wang, K. W., Villalobo, A., and Roufogalis, B. D. (1989). Calmodulin-binding proteins as calpain substrates. Biochem. J. *262*, 693–706.

Young, K., Foroud, T., Williams, P., Jackson, C. E., Beckmann, J. S., Cohen, D., Conneally, P. M., Tischfield, J., and Hodes, M. E. (1992). Confirmation of linkage of limb–girdle muscular dystrophy, type-2, to chromosome 15. Genomics *13*, 1370–1371.

GenBank Accession Numbers

The accession numbers for the sequences reported in this paper are X85032 (5′ sequence), X85031 (3′ sequence), and X85030 (cDNA sequence).