IDENTIFICATION OF MUSCLE-SPECIFIC CALPAIN AND β-SARCOCYLAN GENES IN PROGRESSIVE AUTOSOMAL RECESSIVE MUSCULAR DYSTROPHIES

J S BECKMANN,1* I RICHARD,1 O BROUX,1 F FOUGEROUSSIE, V ALLAMAND,1 N CHIANNILKULCHAI,1 L E LIM,2 F DUCLOS,2 N BOURG,1 L BREGUER,1 P PASTURAUD,1 C QUETIER,1 C ROUDAUT,1 Y SUNADA,2 J MEYER,2 P DINCER,3 G LEFRANCE,4 L MERLINI,3 H TOPALOGLU,3 F M S TOME,4 D COHEN,4 C E JACKSON,8 K P CAMPBELL,2 and M FARDEAU1

1Genethon/CNRS URA 1922, Evry, France; 2Howard Hughes Medical Institute, Iowa City, IA, USA; 3Hacettepe University, Ankara, Turkey; 4UMR CNRS 9942, Montpellier, France; 5Istituto Ortopedico Rizzoli, Bologna, Italy; 6INSERM U153, Paris, France; 7CEPH — Foundation Jean Dausset, Paris, France and 8Henry Ford Hospital, Detroit, MI, USA

Abstract—The autosomal recessive forms of limb-girdle muscular dystrophies are encoded by at least five distinct genes. The work performed towards the identification of two of these is summarized in this report. This success illustrates the growing importance of genetics in modern nosology.

Key words: Limb girdle muscular dystrophy, LGMD, positional cloning.

INTRODUCTION

Limb-girdle muscular dystrophies (LGMD) are a group of inherited neuromuscular diseases presenting clinical heterogeneity. LGMDs are characterized by progressive weakness of the pelvic and shoulder girdle muscles, and their genetic aetiology has yet to be elucidated. These disorders may be inherited either as an autosomal dominant or recessive trait. The latter being more common with an estimated prevalence of 10−4 [1]. In view of the lack of a consensus on the specific nosological definition of each entity, it was proposed at the 30th–31st ENMC workshop to lump these autosomal recessive entities temporarily under the common denomination of LGMD2 [2]. The fact that this decision is far from having generated unanimous support among clinicians involved in this area, illustrates the complexity of this medical and scientific problem.

The first demonstration of a genetic basis for any form of LGMD was obtained in 1991 using bi-allelic RFLP markers [3]. This study was performed in a genetic isolate from the Isle of La Réunion, localizing by linkage analysis a recessive form (LGMD2A, MIM number 253600) to chromosome 15q. This demonstration provided the legitimacy for the existence of this clinical entity, because the latter had until recently been challenged.

Confirmation of this localization was subsequently reported in northern Indiana Amish [4] and Brazilian [5] pedigrees. The latter and subsequent work demonstrated that the autosomal recessive forms (LGMD2) constitute a genetically heterogeneous group involving five genes. These were shown to map to either chromosomes 2p13–p16 (LGMD2B) [6], 13q12 (LGMD2C) [7–9], 15q15.1 (LGMD2A) [3], 17q12–q21.33 (LGMD2D) [10, 11] and 4q12 [12, 13]. The LGMD2D gene was identified recently: it encodes the 50 kDa adhalin glycoprotein [10, 11, 14, 15], now known as α-sarcoglycan. The LGMD2C locus has also just been identified: it encodes for the 35 kDa α-sarcoglycan [9, 16]. Both sarcoglycans are members of the dystrophin glycoprotein complex.

Our efforts focussed mainly on the LGMD2A and LGMD2E loci. These genes

*Author to whom correspondence should be addressed at: Génethon, 1 rue de l'internationale, 91000 Evry, France.
authors suggest that the latter may regulate the activity of this protease.

CHARACTERIZATION OF THE MURINE CANP3 GENE

cDNA clones encoding mouse muscle-specific calpain were selected by hybridization and sequenced [31]. The mouse gene encodes for an mRNA of similar size to the human CANP3 mRNA, directing the synthesis of a protein 821 amino acids long. A partial characterization of the genomic organization of the CANP3 gene suggests that it follows the same overall structure as its human homologue. The results obtained on a somatic cell hybrid panel indicates either a localization on mouse chromosome 2 or on chromosome 4. Information of synteny conservation with human supports the chromosome 2 localization. Interspecies comparison of the CANP3 DNA sequences shows a high level of identity both at the DNA and amino acid levels (Table 1). The degree of similarity is indicative of the fundamental biological role of this protein. Together with the other known animal sequences, the mouse CANP3 sequence would be useful to assess missense mutations in LGMD2A patients. This sequence information can be used to study the spatio-temporal and tissue-specific expression as well as to develop knock-out mouse models. It thus represents a useful tool for gene-regulation, physiopathological studies of the disease as well as for therapeutic investigations.

GENETIC HETEROGENEITY OF AUTOSOMAL RECESSIVE LGMD IN A GENETIC ISOLATE (AMISH) AND EVIDENCE FOR A NEW LOCUS

Cases of autosomal recessive LGMD among members of the old order of Amish of northern and southern Indiana had already been described by Jackson and Carey [32] and Jackson and Strehler [33]. The families of these communities are interrelated by multiple consanguineous links and common ancestry which can be traced back to the 19th century in the Canton of Bern, Switzerland. It was by studying these families that Young et al. [4] confirmed the chromosome 15 localization of the LGMD2A locus. All Amish LGMD patients from northern Indiana were subsequently shown to be homozygous for the same Arg769Glu mis-sense CANP3 mutation [27]. Allamand et al. [34] reported the exclusion of the LGMD2A locus in six Amish kindreds from southern Indiana that are related by multiple consanguineous links to the same northern Indiana families in which the involvement of the chromosome 15 locus had previously been demonstrated. These findings disclosed the genetic heterogeneity of LGMD in this Indiana Amish isolate. In view of the high consanguinity and the similar clinical presentation of all Amish LGMD patients, this result was totally unexpected. Furthermore, genetic analyses also ruled out the possible involvement of the other known LGMD2 loci, thus demonstrating that a mutation within at least one additional locus leads to this condition. Several candidate genes putatively involved in neuromuscular disorders were also excluded. A systematic genome-wide genetic study undertaken with six of these families, eventually uncovered genetic linkage to the pericentromeric region of chromosome 4 [13], thereby providing legitimacy to the LGMD2E locus.

IDENTIFICATION OF THE LGMD2E LOCUS

Meanwhile, the human cDNA of β-sarcoglycan, a 43 kDa dystrophin-associated glycoprotein (also known as A3b), was cloned and mapped to the same chromosome 4 region [13]. Several chromosome 4 pericentromeric markers and a β-sarcoglycan intragenic polymorphic CA repeat cosegregated perfectly with the disease in Amish families from southern Indiana with autosomal recessive LGMD. In addition, reconstruction of haplotypes allowed the identification of chromosomes descending from a common ancestor. Furthermore, a Thr-to-Arg mis-sense mutation, present in a homozygous state in all patients, was identified.
Fig. 1. LGMD2A YAC contig. The map shows the contiguous array of YAC clones which are presented as lines (only clones with more than one characterized STS are drawn). The names of the STSs are given above the line representation of the contig. The position of polymorphic microsatellite markers and coding sequences are indicated by the presence of, respectively, a dot or a star above the marker name. The presence of the LGMD2A locus, encoding the calpain3 gene, is indicated by a large dotted red arrow. The STSs are drawn equidistant except when they are issued from the same gene (and hence are arbitrarily ordered), but for the 5' end of the fibrilin gene. Adjacent markers which cannot be oriented with respect to this map are placed within brackets. Inverted triangles: STSs having supplied YACs upon screening of the CEPH libraries. Upright triangles: STSs which failed to yield YACs in this screening. Bars on full lines indicate positive STSs. Dotted lines mean that a clone was found to be positive for a given internal STS. Arrows at the end of the clone indicate that a clone was found to be negative for the adjacent STSs. The parallel lines represent the minimal tiling path consisting of 10 YAC clones spanning the contig. The chromosome 15 YAC end-STSs (R or L) are shown in full squares, and YAC end constituted repeated sequences as dotted squares. The ends are arbitrarily drawn whenever both ends failed to map within the contig. One extremity of clone 911B6, a double insert and microtically unstable YAC, is shown as a hatched square, in the middle of the contig, as this clone was shown to be positive for external STSs. A drawing showing approximate G-banding pattern of the corresponding region of chromosome 15 is placed above the contig presentation. The thick green lines represent the YACs that were initially used to define the cytogenetic interval. The smallest recombinant interval defining the LGMD2A region is shown by the vertical arrows on the regional chromosome map, and the distance spanning the recombinant intervals is given above.
MUTATIONS IN THE PROTEOLYTIC ENZYME, CALPAIN 3, CAUSE LGMD2A

The CANP3 gene, encoding the large subunit of the muscle-specific calcium-activated neutral protease 3, belongs to the family of intracellular calpains, requiring calcium for their catalytic activities (for a review see references 25 and 26). Its genomic organization was determined, by sequencing cosmids containing this gene, which is composed of 24 exons spread over more than 40 kb [27]. Validation of a candidate gene requires the identification of the pathogenic mutations. A total of 16 non-sense, splice site, frameshift or mis-sense mutations in this muscle-specific gene were initially found to cosegregate with the disease in LGMD2A families, six of which were found within Réunion Island LGMD2A patients [27]. A digenic inheritance model was proposed to account for the unexpected presence of multiple independent mutations in this small inbred population. One of the mutations encountered in the Isle of La Réunion, was found to be a synonymous mis-sense mutation leading to a predicted Gly-to-Gly substitution. In this manner, we could demonstrate that this apparently neutral mutation creates a new splice site, and is thus effectively equivalent to a null mutant [28]. As a matter of fact, to reveal this mutation, we had to screen lymphocyte illegitimate transcription products [29], because the CANP3 locus shows a tissue-specific pattern of expression, and because of the lack of muscle tissues available from this as from most patients. Last but not least, the finding that calpain mutations cause LGMD2A is the first demonstration of an enzymatic rather than a structural protein defect causing a muscular dystrophy, a defect that may have regulatory consequences, perhaps in signal transduction [27].

Having identified the role of the calpain gene in LGMD2A, we next analysed the segregation of markers flanking the LGMD2A locus and carried out a search for CANP3 mutations in LGMD2 pedigrees from various geographical origins. Altogether, over 50 distinct (mis-sense, frameshift, non-sense, splice site, and small in-frame deletion type) mutations were uncovered. These are dispersed throughout the entire length of the CANP3 gene (unpublished data). Whereas approximately 50 per cent of the identified CANP3 mutations are mis-sense mutations (Fig. 2), the identification of carriers of two null type mutations demonstrate that the lack of calpain activities is pathogenic. These data indicate that muscular dystrophy caused by mutations in the muscle calpain gene are found in patients from all countries so far examined, with no predominant mutation, and further support the wide heterogeneity of molecular defects in this rare disease.

Subsequent to the incrimination of calpain in the aetiology of LGMD, another element of the puzzle fell in place when Sorimachi et al. [30] demonstrated that calpain 3 can bind to titin. This binding is through one of calpain's specific regions, IS2. Although the exact significance of this interaction is not understood as yet, these

---

**Fig. 2. Distribution of the CANP3 mutations.** The genomic structure, the distribution of the mis-sense and null-type (non-sense, frameshift, splice site and small deletions/additions) mutations, and the protein domains are represented.
were identified following either a classical positional cloning (for LGMD2A) or a combined positional and functional cloning (for LGMD2E) strategies. The various steps involved in this process are reviewed here.

**Mapping of a Chromosome 15 Region Involved in LGMD2A**

Four years ago, few highly informative genetic markers were available. Fortunately, the Généthon program had started and had already produced a score of new markers [17, 18]. In parallel, we also set about to developing in a targeted fashion, markers from the LGMD2A region [9, 20]. After genotyping the newly available chromosome 15 microsatellite markers, two were found to flank the disease locus within an interval that was assessed, based on segregation studies on the CEPH reference families, as spanning 7 cM (centiMorgans). The screening of the CEPH YAC (yeast artificial chromosome) libraries with the corresponding probes allowed the isolation of YACs which were used in fluorescence in situ hybridization to define the LGMD2A cytogenetic interval as 15q15.1–15q21.1 [19]. Different approaches were pursued for the establishment of the physical map of this area. This eventually allowed the assembly of an uninterrupted YAC contig spanning an estimated 10–12 megabases (Mb) (Fig. 1), with a current average STS (sequence tagged site) resolution of 50 kb, and for the 36 polymorphic microsatellites on this map, of 300 kb. Twelve genes and 25 genetic markers were initially positioned in this contig, which is constituted of a minimum of 10 clones.

**A Primary Expression Map of Chromosome 15q15 Region Containing the Recessive LGMD2A Gene.**

In order to progress towards the identification of the gene involved in LGMD2A, a primary transcription map of the LGMD2A genomic region was generated [23]. The direct cDNA selection strategy was used with three YACs covering the candidate region and two different muscle cDNA libraries. Seventeen different transcription units were identified among 171 cDNA fragments analysed. Five sequences corresponded to known genes, and twelve to new ones. They were characterized for their sequences, physical positions within the YAC and cosmide contigs, and expression patterns. There are now 29 coding sequences mapped onto the LGMD2A YAC contig (Fig. 1). Among those specifically transcribed in muscle, the calpain 3 gene — whose cDNA sequence has been known since 1989 [24] — appeared to be a good positional candidate for LGMD2A, although it was not considered, a priori, as a functional candidate.
within the β-sarcoglycan gene that leads to a dramatically reduced expression of β-sarcoglycan in the sarcolemma and a concomitant loss of α- and γ-sarcoglycan, thereby disrupting the integrity of the dystrophin glycoprotein complex [13]. Similar evidence was concurrently obtained by Bönner et al. [12] on a single severely affected Italian patient who was shown to have a compound heterozygote carrying null alleles. These data demonstrated that a defect in β-sarcoglycan is responsible for the recessive muscular dystrophy present in southern Indiana Amish families and in this Italian child. The β-sarcoglycan gene is thus the fifth locus (LGMD2E) involved in autosomal recessive LGMD, and the third sarcoglycan involved in a myopathy (Table 2).

CONCLUSION

The past year seems to have been a good vintage year for geneticists and clinicians interested in autosomal recessive progressive muscular dystrophies. Indeed, in a little over a year, four LGMD2 genes have been identified, three of which belong to the dystrophin-associated protein complex, the fourth one being an enzyme (Table 2). These discoveries raise more questions than they provide answers; but this is not a matter of complaint — they are also a new source of hope. The elucidation of the molecular defects underlying these diseases already offer new prospects in genetic counselling. Furthermore, this knowledge can, on the one hand, be used to learn more about the physio-pathological processes involved. On the other hand, with the identification of the pathological mutations one can now proceed and perform careful examinations of the phenotype–genotype relationships. Such studies already showed that LGMD2A patients carrying 2 null calpain mutations have in general a worse prognosis than carriers of mis-sense mutation(s) [35]. Furthermore, although one should be cautious not to draw general conclusions based on anecdotal observations, it is clear that this knowledge will eventually enable a clearer definition of the specific nosological boundaries of each one of these similar, yet different, clinical entities, a knowledge that could be of great immediate benefit to the patients themselves. Last but not least, it opens prospects for new therapeutic avenues.

Acknowledgments—We are particularly indebted to all patients, their families and clinicians. We thank Drs B. Bady, J.-M. Burgunder, M. Conneally, R. Chemaly, J. Feingold, N. Feingold, C. A. Garcia, G. Halaby, D. Hillaire, D. M. Kurnit, C. Legum, J. Loislet, C. Moomaw, A. Nivelon-Chevallier, E. Olagnon-Roman, G. Restagno, N. Romero, C. Slaughter, M. R. Passos-Bueno, J. Tischfeld and M. Zatz for their participation in parts of this work. We also acknowledge many colleagues for providing clones, chromosome 4 specific cosmid library; microsatellite repeat probes; for performing muscle biopsies; and finally, members from our respective laboratories for their help. This research was supported by grants from the Muscular Dystrophy Association and the Association Française contre les Myopathies (AFM), the European Commission within the context of the European Genome Mapping (EUROGEM) Project and the Groupement de Recherche Européen sur le Génome, V.A. is funded by a grant from AFM. K.P.C is an investigator of the Howard Hughes Medical Institute.
REFERENCES


limb-girdle muscular dystrophy (AR LGMD) maps a sixth form to 5q33-34 (LGMD2F) and indicates that there is at least one more subtype of AR LGMD. *Hum Mol Genet* 1996; 6: 815-820.
